Dear Reviewer#1,

Thank you for your second assessment of our manuscript, and for your positive feedback. Your two pending suggestions have been considered, and the following changes have been implemented to address them.

In the following point-by-point responses your comments are in normal font and our responses are in *italic*. The line and figure numbers refer to the revised version of the manuscript. In the marked revised manuscript the sections with changes are highlighted in blue.

Hoping that these improvements will fulfill your expectations,
Best regards,

Lionel Benoit, on behalf of the authors.

I am satisfied with the authors response to my comments during the first round of revision. However, I still have two comments:

I/- I was not clear enough in my comment IV/- Before assessing the spatial pattern, is it possible to look at the seasonal and inter-annual variability for a sub-sample of stations (section 3.3)?

This comment was made for the evaluation of the simulations. In particular, the reproduction of seasonal and inter-annual variability is essential for hydrological simulations. I would like to see these are well simulated for the different climatic divisions. Besides doing that for the conditional and unconditional simulations, would nicely illustrate the need for actually doing conditional simulations.

Thank you for clarifying your request. The assessment of the simulation of seasonality and inter-annual variability at a sub-sample of stations is now presented in the Supplementary Material 1, which is introduced in the main text at line 350-353. Results show that both seasonality and inter-annual variability are properly reproduced in simulations, with a lower uncertainty in conditional simulations.

II/- My second comment is about the double sequential conditioning, what is the gain in performance obtained from conditioning daily rainfall maps to monthly totals

Thank you for this comment that gave us the opportunity to better demonstrate the performance of our model. To do so we added the new figure 8 and its description in the main text (l 546-573). The results presented in this figure quantify the gain in performance obtained from conditioning daily rainfall maps to monthly totals in a sparsely gauged area, and show that the improvement reaches 90% in terms of mean bias, 46 % in terms of mean absolute error and 44% in terms of mean CRPS.

Dear Reviewer#2,

Thank you for your second assessment of our manuscript, and for your positive feedback about our revisions. Your pending suggestions have been considered, and the following changes have been implemented to address them.

In the following point-by-point responses your comments are in normal font and our responses are in *italic*. The line and figure numbers refer to the revised version of the manuscript. In the marked revised manuscript the sections with changes are highlighted in blue.

Hoping that these improvements will fulfill your expectations,
Best regards,

Lionel Benoit, on behalf of the authors.

The authors received three detailed and demanding reviews. They have done an excellent job of taking their comments into account. The article is now much clearer. In particular, the abstract has been completely rewritten and now presents the novelties and specifics of the article much more clearly. The method section has also been completely rewritten and reorganized, with more detail on certain points. I enjoyed reading this new version and the efforts made by the authors to improve their manuscript. I still have a few comments that should be easy to deal with:

- I still had trouble understanding the covariance modeling. In fact, I only understood it when I read the conclusion (l529-532). “Similarly, the covariance function of the latent field is first assumed to be stationary within relatively small climate divisions during the estimation of the covariance parameters, then the covariance parameters are assigned to the barycenter of each climate division, and finally these parameters are interpolated through space and incorporated in a model of non-stationary covariance.” This is very clear, and these three steps (stationary adjustment, assignment to the barycenter, interpolation) should be presented in this way in section 3, which remains unclear on this subject. For example, barycenter assignment only appears in brackets on line 201 (“preliminarily assigned to the barycenter of each climate division”).

Thank you for this comment that helped us improving the description of the covariance modeling, and in particular how the non-stationary model of covariance is built. To improve this description we decided to split the section 3.3 Model calibration in two sub-sections: 3.3.1 Estimation of model parameters assuming local stationarity and 3.3.2 Spatial interpolation of model parameters. We did this split to emphasize that model calibration is a two-steps process.
We also re-wrote the content of the section 3.3.2 (cf. l 210-219) to present parameter estimation in the way proposed by Reviewer#2.

- I appreciate that the authors have provided more details on conditioning to monthly totals, as requested by one of the reviewers. However, the MCMC procedure is quite technical and I think all the text from lines 261 to 285 could be moved to the appendix.

We agree that the description of the MCMC procedure is quite technical, but we decided to keep it in the main text to not cancel a change that seemed to satisfy another reviewer, and because we believe that uninterested readers can easily skip it.

- I also appreciate the efforts made to justify the Gamma/Matern/geometric anisotropy choices, as requested by one reviewer. However, it should be noted that other choices could have been made without calling into question the main contribution of the article, which is the use of a fully non-stationary trans-Gaussian model (as indicated in line 59).

This is a good point. We added a small paragraph in Sect. 3.1 (l 127-130) to clearly state that the parameterizations we choose throughout the paper are related to the Hawai'i dataset, but that they could be changed without calling into question the use of the fully non-stationary trans-Gaussian model, which is the main focus of the study.

Dear Reviewer#3,

Thank you for your second assessment of our manuscript. All your suggestions have been considered, and the following changes have been implemented to address the questions you raised in your review.

In the following point-by-point responses your comments are in normal font and our responses are in *italic*. The line and figure numbers refer to the revised version of the manuscript. In the marked revised manuscript the sections with changes are highlighted in blue.

Hoping that these improvements will fulfill your expectations,
Best regards,

Lionel Benoit, on behalf of the authors.

I would like to thank the authors for their careful revision of the manuscript and their responses to my comments. The new version is clearer and the added explanations help the reader better understand the modeling choices and the novelties of the approach.

Having now better grasped the strengths of the paper, I believe that since the work combines existing ideas (non-stationary covariance and non-stationary marginal distributions), the validation methodology could be strengthened to more precisely demonstrate the benefits of this combination.

Minor comments

1. Notation and definitions (l102, p5)
- The notation $\mathcal{D}$ is introduced to denote the set of rain gauges but is not consistently used later in the manuscript, where it would be valuable to distinguish between the discrete set of rain gauges and the continuous spatial domain.

This is a good point. In the revised manuscript we use the notation $\mathcal{D}$ *more consistently (l 102, 121, 174, 216, 223) to refer to the study domain. We also introduced new notations for the set of rain gauge locations (l 173-174, 193-194) and for the set of target locations where simulation in performed (l 223) in order to avoid confusion between these different sets of locations.*

- On a related note, it is unclear in Equations 6 and 8 how Ψ is known for arbitrary locations s when the model is only trained on discrete stations. Please clarify whether this interpolation is addressed later; if so, it would be helpful to announce this when first introducing these equations.

We agree that this is important to clarify this aspect. The interpolation of the model parameters (i.e. the parameters of Ψ and C_Y) is explained in Section 3.3 and is performed before model parameters are used in Equations 6 and 8. To make it more clear we decided to split section 3.3 in two sub-sections whose titles reflect how model parameters are first estimated at the local scale (3.3.1. Estimation of model parameters assuming local stationarity) and then interpolated throughout the domain $\mathcal{D}$ of interest (3.3.2. Spatial interpolation of model parameters). In addition, we added a short paragraph at the end of section 3.3 to emphasis on the interpolation of model parameters (l 216-218).

- A notation for the continuous spatial domain would also be useful to avoid confusion between fitting at stations versus generating over the fine spatial domain. For instance, in Equations 3, 6, and 8, it is not entirely clear whether the variables are defined on the station set or on the spatial domain.

We agree with this comment and added two new notations to avoid confusion: one for the set of rain gauges (l 174-175, 194-195) and one for the set of target locations where simulation in performed (l 224).
In addition gauge locations are now systematically referred to with the letter ***s*** *and target locations with the letter* ***u****. We also added one sentence l 223-225 to make it explicit.*

2. Section structure and terminology (Section 4)
- The subsections in Section 4 should be renamed more descriptively. It appears that Section 4.2 addresses multisite validation against BSNLG2022, while Section 4.3 focuses on spatial validation against CAI. More explicit titles would help readers navigate the comparisons.

To better distinguish between multi-site and fully spatial simulation, the section 4.2 has been split into two sub-sections, with explicit titles: 4.2.1 Multi-site rainfall simulation and 4.2.2 Rainfall fields simulation. In addition, the first paragraph of section 4.2.2 has been re-written to allow for a better transition between multi-site and fully spatial simulation (l 383-387).

- The terms "conditional" and "unconditional" are somewhat confusing in this context. It seems these refer to multisite versus spatial generation, respectively. Please make this distinction more explicit in the text.

Actually the term "unconditional" is used in this paper to refer to the generation of synthetic rainfall (i.e., without anchor to observations), while "conditional" refer to simulations that are forced to honor a set of observations (which allows for interpolation and therefore mapping). This is mentioned in the introduction (l 35-39), and explained in details in Sect. 3.5 (l 243-263). Nevertheless, we agree that this explanation occurs early in the paper, and readers may have forgotten about it when reaching the case study (Sect. 4). To avoid ambiguity, we removed all mentions to "unconditional simulation" in this section 4, and made sure that when the terms "conditional" or "conditioning" are used we explicitly mention with respect to what the conditioning is performed; this way, we hope to better guide the reader and avoid misunderstanding.

- Throughout the plots, text, and tables, consider using abbreviated model names rather than generic phrases like "non-stationary trans-Gaussian model" and "the benchmark model." This would make it easier to distinguish when BSNLG2022 is used versus when CAI is the comparison model.

Thank you for this advice, this is an excellent idea. We implemented this change, and therefore replaced "non-stationary trans-Gaussian model" and "our model" by "NSTG model", and replaced "benchmark model" by either "BSNLG2022 benchmark model" or "CAI benchmark model" throughout Sect. 4.

- Why is CAI not included in the model comparison table? Including it would provide a more comprehensive overview.

This is a good point. We included CAI in the model comparison table (Appendix C, l 676) in the revised manuscript.

3. Visualization choices
- Make sure your color map are colorblind friendly. The "turbo"-like color map can distort the perception of color differences.

We used the website https://www.color-blindness.com/coblis-color-blindness-simulator/ to check how the figures would appear to colorblind persons, and after modification of the figure 6 (cf. response to the next comment) all figures appear to be colorblind friendly.

- Similarly, in Figure 6 what is the point of showing positive and negative errors in the color bar but having the same dark color at the end of both sides?

This is a good observation. We originally made this change to answer the comment of a reviewer about using the same colormap for signed and unsigned quantities, but your comment made us realize that this change degraded the readability of the figure. We therefore decided to change it again, and to use the same colormap for all plot of Figure 6 but use triangles pointing up and down to denote positive (resp. negative) values. This results in a better readability, and after check with the color-blindness-simulator website the final figure appears to be colorblind friendly.

Unclear points regarding validation

Several aspects of the validation procedure require clarification:

1. Simulation strategy
- In the multisite validation, are the metrics computed using a single simulation, or are they averaged over multiple simulations? This should be stated explicitly.

The metrics are computed using a single simulation. We slightly modified the description of the simulation process to make it explicit. The revised sentence reads as (l 347-350): "For comparison purposes the simulation is repeated the same number of times as the number of days within each rain type, and the 7305 daily simulations are ordered and concatenated to create a single 20-year time series with the same rain type timeline as the observations".

2. Cross-validation procedure
- The training/validation split using cross-validation is not detailed enough. Are you applying this procedure to all stations? Is it a leave-one-out approach (removing one station at a time)? Are the validation metrics averaged over all folds? Additionally, is the benchmark model trained using the same cross-validation scheme to ensure a fair comparison?

The cross-validation strategy we use is a basically a leave-one-out approach, but we remove data from gauges located in a 5km radius around the target gauge to avoid artificially good results in case of almost co-located gauges. This procedure is detailed lines 444-450, where we added the explicit mention to the leave-one-out strategy and the fact that the same procedure is applied to all gauges (l 447-448): "The same procedure is applied to all gauges sequentially, and ultimately the cross-validation implemented here is equivalent to a leave-one-out approach".
Regarding the benchmark model it is indeed trained using the same scheme to ensure a fair comparison, as stated l 478-480.

3. Benchmark model selection
- BSNLG2022 is described as the benchmark model, but then CAI is used for spatial comparisons

because BSNLG2022 is only multisite. Please clarify the rationale for this choice and how it affects the interpretation of results.

The BSNLG2022 model is multi-site, and therefore cannot be used for spatial interpolation. That is it why the CAI model is used instead of BSNLG2022 as a benchmark for rainfall mapping. We use the best benchmark model for each task at hand in order to demonstrate that the proposed NSTG model outperform state-of-the-art benchmarks for both stochastic rainfall generation and rainfall mapping (through stochastic interpolation).
The reason why BSNLG2022 cannot be used to benchmark rainfall mapping is now stated twice: l 383-387 where we explain that it is not suited for the simulation of spatially continuous rainfall fields, and l 475-477 where we explain why CAI is preferred to BSNLG2022 to benchmark rainfall mapping.

4. Choice of metrics
- The use of MB and MAE for validation is not well motivated. These metrics increase the number of quantities to examine, yet their specific value is not discussed. Moreover, they compare observed values with simulated values directly, whereas weather generators are typically evaluated on their ability to capture the statistics of observed values. Please discuss these choices in more detail or consider alternative metrics that better capture the statistical properties of the data.

Here MB and MAE (along with MCRPS) are used to evaluate the performance of rainfall mapping, i.e., the spatial prediction of rainfall for a given day at an ungauged location conditionally to neighboring observations. In our opinion these metrics are standard to evaluate rainfall mapping in a cross-validation experiment, and within the references cited in the present paper these metrics are used for instance in Fouedjio et al. (2016), Frazier et al. (2016), and Lucas et al. (2022). We therefore decided to not justify this choice further.

We agree that in addition to a good point prediction (evaluated through MB, MEA and MCRPS) we also want out model to capture and reproduce the statistics of the observed values when the model is used as a stochastic rainfall generator. We would like to emphasize that this aspect is already evaluated in great details in Sect. 4.2, and in particular in figures 4 and 5. Hence we decided to not develop this aspect further in the revised manuscript.

Suggestions for strengthening the validation

- The contribution of full non-stationarity versus the imposed weather regimes is not entirely clear. It is possible that much of the spatial variability is already captured by the weather regime structure. A benchmark comparison against the same model with stationary covariance (and possibly stationary marginals, but the interest of the latter is more intuitive) would more clearly demonstrate the benefit of incorporating non-stationary covariance beyond what is learned through the weather regimes alone.

We do not fully understand this comment. In our framework the rain types (i.e., weather regimes) are used to capture the temporal variability of daily rainfall, but they are not used to capture the spatial variability, in particular because for a given day the rain type is the same for the whole domain of interest (here the all island of Hawai'i).
However this is true that different rain types are associated with different set of model parameters, and the model must be able to properly capture and reproduce the spatial statistics and spatial variability of rainfall for each type separately. This is the case, and this is now shown in the new supplementary material 2 (introduced in the main text l 371).

- Additionally, it is unclear whether CAI and BSNLG2022 include non-stationarity (are empirical copula non stationary?); this information is not evident in Table C1 and should be clarified.

Thank you for this comment. We forgot to mention this important information in Table C1, and we added it in the revised version of the paper.
We also re-wrote the description of the CAI model to explain which part of the model is non-stationary (the marginal distribution) and which part is stationary (the covariance function). The updated description of the CAI model is found lines 477-488.

- The CAI model is not validated using BS and MCRPS. Including as a benchmark the stationnary version of your model could allow to easily compare these metrics against your full model.

The CAI model is not validated using BS and MCRPS because these metrics are used to assess the performance of ensemble forecasts, but the CAI model provides only a deterministic interpolation (i.e., a single value is estimated for each day and location). This feature of the CAI model is now explained l485-488.