Response to Anonymous referee 2

We would like to thank Anonymous referee #2 for his/her insightful review and for bringing up several interesting topics for discussion. Most of the comments are related to the model components and/or about motivating model assumptions. We will clarify this part in our revised MS. In addition we will edit/rewrite the awkward sentences that are mentioned in the review.

See our detailed answers to referee #2 below.

Kind regards,
Thea Roksvåg and co-authors.

R2: *"A potential weakness of the method, which has been mentioned by the authors, is that the model does not prevent negative run-off predictions in some (unlikely situations). This is due to the Gaussian likelihood and Gaussian GRF. The authors mention that log-runoff could be used instead, but then linearity of Eq. (6) is lost, which is an impediment. Another way of preventing negative predictions would be to use log-Gaussian likelihood and log-Gaussian random fielsd for x(u) and α(u) in (4). This would be a marginal change, since INLA/SPDE allows for log-gaussian likelihood and LG random fields at almost no cost. As a result, predictions for x and α would always be positive. I wonder how this would work. Ideally, I'd like the authors to try this option, but I'd be happy if they only discuss this possibility."*

Response:
It should be possible to make a model like this in inla, to avoid negative predictions. The drawback is that a log GRF will make it more difficult to interpret the two spatial fields. As the paper already is quite long, we will add this to the discussion, as possible further work.

R2: *"The GRFs x(u) and α(u) are independent. This assumption is never clearly stated and it is not discussed. Is this a reasonable assumption? Is this an assumption you could check or validate? How useful/difficult would it be to relax this assumption?"*

Response:
We will clearly state that *x(u) and α(u)* are assumed to be independent, and discuss whether this is reasonable. We can compare *x(u) and α(u) to investigate whether they are independent.*

To relax the independence assumption we could include a third spatial field that is both included in the factor multiplied with h(x) and added to the model (multiplied with a scaling coefficient). This will result in a model that is harder to interpret, but might give better results. A drawback of this option, is that an additional spatial field will significantly increase the computational complexity of the model, and the computational complexity is already quite high. The increased computing time is probably not worth it, as the current model gives quite good results as it is.

R2: "The GRFs x(u) and α(u) are assumed to be stationary. Are you able to check that this assumption is supported by the data?"

Response: The GRFs x(u) and/or α(u) might be non-stationary. The spatial dependency structure of precipitation and runoff can change with for example elevation. We would e.g. expect the spatial range of x(u) to decrease with elevation. Other non-stationary effects could also exist.

We could investigate the non-stationarity of the spatial fields by fitting non-stationary models. From this we can see whether the non-stationary effects are significant. We could also compute empirical ranges and variances from the data, for different areas. However, modelling x(u) and/or α(u) as stationary is a choice we have made based for the following reasons:
* Modelling x(u) and/or α(u) as non-stationary, would introduce additional parameters to the model which will represent an increase in computational complexity.
* The type of non-stationarity can be difficult to identify from the data.
* According to Fuglstad et al (2015), non-stationary processes can in many cases be modelled by stationary models.
* Even if the underlying process is non-stationary, we think that our model should work well because we have two spatial fields. Together, x(u) and α(u) give a flexible model that is able to capture different dependency structures in the data.

R2: *"To my knowledge, the product of an exponential variogram with a fractal variogram is not a valid variogram. However, the product of an exponential covariance function with a fractal variogram might be a valid variogram. Please double-check and provide references if necessary. "*

Response: We fitted the default variogram type in the rtop package. According to the package documentation (https://cran.r-project.org/web/packages/rtop/rtop.pdf) this is a *"multiplication of a modified exponential and fractal model"* (model="Ex1").

R2: *"Regarding the results: is it really desirable to get a correlation of 1 between measures and predictions? I would relate this to the fact that the coverage is 83%, which shows that the SVC is over-confident in the UG setting. Please comment."*

Response:
It is not necessarily desirable to get a correlation of 1 (as for the orange points in Fig 7), and the coverage of 83 % indeed shows that the SVC is slightly over-confident in the UG setting. By using a model with two spatial fields, we get a model that is quite flexible. This can explain why we sometimes get correlations close to 1. Furthermore, the HBV model aims at getting perfect predictions, which would imply a correlation of 1.

Whether a correlation of 1 is desirable or not, depends on what we want from our runoff map. If it is important that the map is correct in areas where we have data, a correlation of 1 between the map and the data is good. Alternatively, we could accept a lower correlation, and get a model that might perform better in ungauged areas (more spatial smoothing).

After the MS was prepared, we did some more experiments with the SVC model, where we used a larger dataset with 180 gauged catchments and 450 partially gauged catchments. We also used a different hydrological model (WASMOD instead of the HBV model). In these experiments, the correlation between the observations and predictions were lower than 1. Hence, whether we get a correlation of 1 depends on the dataset. In general, we would expect a lower correlation if we have more data as it then becomes more difficult for the model to fulfil all data “constraints”. In addition, the correlation depends on the hydrological product used and how it is calibrated.

In the discussion, on line 720, we briefly mention that we cannot expect a perfect fit (correlation 1) between the observations and the runoff map. In a revised version of the manuscript, we can rewrite/extend this part, and/or rewrite the results section (around line 595 and 600).

R2: Comments about awkward sentences and spelling mistakes.
Response: We will rewrite/edit these sentences.

References:
Geir-Arne Fuglstad, Daniel Simpson, Finn Lindgren, Håvard Rue,
Does non-stationary spatial data always require non-stationary random fields?,
Spatial Statistics,
Volume 14, Part C, 2015,
Pages 505-531,
ISSN 2211-6753,
https://doi.org/10.1016/j.spasta.2015.10.001.
(https://www.sciencedirect.com/science/article/pii/S2211675315000780)