Response to Reviewers' Comments

We thank the reviewers for their detailed reading of the paper, for catching some errors, and for the helpful suggestions for improvement. Please see the following pages for a detailed response.

Sincerely,

Snizhana Ross, Arttu Arjas, Ilkka I. Virtanen, Mikko J. Sillanpää, Lassi Roininen, and Andreas Hauptmann
22 December 2021

Response to Reviewer 1

I found the description of this hierarchical convolution technique to be clear and well-organized, and I have high confidence that I could implement the technique based on reading the paper. I think this is an exciting area of development for processing radar data and, in particular, incoherent scatter radar data, and I look forward to future developments. I have some specific comments that follow, but they mainly touch on areas where I think additional information or clarification would improve the paper.

Response: We thank the reviewer for the encouraging evaluation and for excellent suggestions to improve the clarity of the paper. In the following we outline our specific responses to the raised points.

1. *In the paragraph containing Equation (7), it is introduced as, "In order to reach resolutions better than the elementary pulse length". I found this slightly confusing on the first read-through because I initially failed to recognize that Equation (7) is a discretization of Equation (5) since my attention had been directed to the resolution issue. The quoted clause implied to me that the form of the following Equation (7) is specialized in order to achieve increased resolution, but in truth the equation would look similar in all cases and is necessary just for discretization. I suggest removing the quoted clause and placing discussion of how to achieve resolutions better than the elementary pulse length to after the description of Equation (7).*

   Response: We agree with the reviewer and will change the text before Equation (7) as follows: "In order to reach resolutions better than the elementary pulse length, → In order to move from the continuous time signals to discrete samples,"

   We will then add the sentence after Equation (7): "In order to reach resolutions better than the elementary pulse length, we oversample the signal, i.e. use sampling steps shorter than the elementary pulse length."

2. *Using a mean of 0 for the Gaussian Process prior for P is described as a "convenience", and I appreciate from my own experience with GPs that it is indeed such. Are there other justifications you can provide for why that is an appropriate assumption in this case?*

Response: Our target is in estimating the variability, and especially the high frequency parts (non-stationarities) of P. Our approach is in modelling this variability via the non-stationary covariance function with zero-mean GPs. Alternatively, we could choose a non-zero mean parameter or a continuous-parameter profile. These we could also estimate within our model. However, if we have a continuous-parameter profile, this could lead to overparameterisation and unidentifiability of the unknown objects, as the high-frequency parts would be both in the non-stationary covariance, as well as in the estimated continuous-parameter mean. A straightforward alternative would be to use some other measurements for mean estimation (this has been done e.g. for ionospheric tomography by using ionosonde measurements for mean estimation). As ISR can be considered as the baseline measurement for ionosphere, this is pretty much impossible to achieve with standard experiments (perhaps excluding rocket experiments).

In summary, the zero-mean choice is a rough simplification, but as our target is to detect the non-stationarities and providing general purpose tools (rather than tuning the model for specific cases), we feel that making a more complex choice is not needed for the purpose in this paper. We will add a short comment accordingly after introducing the 0 mean in Section 3.1 and remove the wording "convenience".

3. *Similarly, can you provide additional justification for why a Matérn covariance with* $\nu = 1.5$ *is chosen? Including a quick statement in the text will help readers who are less familiar with Gaussian Processes so they don't have to turn to one of the references to find the answer.*

Response: We agree that this needs more explanation. The parameter ν determines the smoothness of the underlying process. In the case $\nu = 1.5$, the process is once mean square differentiable. We choose $\nu = p+0.5$, $p \in \mathbb{N}$, as this provides a Markov approximation for the model, and thus there exists a simple form for the covariance function via stochastic differential equations. Hence by constructing and by choosing $\nu = 1.5$, the square-root of the inverted covariance matrix has a tridiagonal structure – which is numerically convenient.

We will add an explanation accordingly to the text.

4. *It is noted that* L_l *is a tridiagonal matrix with reference to Roininen et al. 2014. I suggest adding a quick statement saying why this is the case (e.g. finite*

differences approximating the derivative) and why it is useful (e.g. efficient computation especially as the problem size scales up). Providing an explicit expression of L_l as a function of l_i here would also be good for clarity, although I do note that it appears in-text later in lines 204 and 205.

Response: This is correct, and as pointed above, the Markov approximation leads to numerically and computationally useful presentation. Moreover, this also allows us to model ℓ_i via increments, thus simplifying both the model and computations.

We will add a short motivation where it is first mentioned and point the reader to the following section for the explicit representation.

5. *The Figure 3 labels and text discussing the figure refer to u as the "length-scale function". I think it would be clearer to note that this is the log of the underlying length scale, so that statements like "by factor 3-5 large in smooth parts of the profile" can more easily be associated with the log scale under discussion. Better yet would be to reference the physical units associated with the underlying length scale values.*

Response: We agree and will change the figure captions. We also point to our response to Reviewer 2 in Technical corrections 4, that we will add [arb.units] as this depends on the sampling accuracy of the profiles to be recovered.

6. *The alpha tuning parameters were optimized to minimize the mean squared error between P and $\hat{P}$, and the resulting estimates all underestimate the peak power of the sporadic E layer. Presumably this is because the length scale would need to reach a smaller value at those altitudes in order to permit the large gradient that exists there. Did you test higher values for the alpha parameters, and does that end up fitting the sporadic E peak better? What does that do for the quality of the estimates at other altitudes for the background ionosphere? In other words, if one was more interested in the highest quality estimates of either a narrow feature or the background ionosphere at the expense of the other, how does that affect the decision for setting the alpha parameters?*

Response: This is a very good point, that we missed to discuss. We have conducted some more experiments to test if a higher peak could be reached with more parameter tuning. In fact, the particular values for α_C and α_{TV} are already (very) close to optimal for the sporadic E peak. Tuning the α parameters for the estimation of the narrow feature of the peak will primarily affect the outer layers and does not improve the reconstruction of the highest peak power. This nicely underlines the benefit of the length-scale function in the estimation procedure, as it is robust for the narrow layers and the tuning parameters affect primarily the desired smoothness of the outer layers.

We will add a comment to Section 5.1. as follows: "The presented modelling proves to be robust in recovering the peak power. Specifically, choosing different tuning parameters to estimate the narrow feature of the highest peak power not increasing significantly the quality of the reconstruction in comparison to the optimal values of α_{C} and α_{TV}. This underlines the need for an adaptive length-scale function in the estimation procedure."

7. *Following on from the previous comment: did you test any other prior distributions (i.e. not Cauchy or TV/Laplace) for the length scale difference that might be better suited to really sharp gradients? If not, can you point to directions for future work in this area?*

Response: This is a good point for discussion, in our study we have not chosen any other priors. Naturally, there is a large selection of different priors one can use for specific applications: Gaussian priors (easy to use for continuous models, but not good for rough features), Besov priors (good for rough features, but have dyadic structures due to wavelets), geometric priors (requires model-specific constructions), and data-driven priors via neural networks, especially GANs (requires training data).

As mentioned, for this particular study, we concentrated on the Cauchy and TV priors. This is because we want to avoid the dyadic structures, specific geometries and dependence of data, which could be well suited for future studies, but out-of-scope here. For instance, one could use rougher features by using general alpha-stable processes, but losing the analytical properties of Cauchy distributions, which would further complicate the process due the needed computations of approximations of probability density functions of alpha-stable densities. Moreover, as the Cauchy probability density function is already an infinite-variance model, we suspect that the result of going to rougher models would have a marginal effect to the first-layer GP non-stationary model.

We will add a comment accordingly to the Discussion in Section 6 and point to possible improvements in future studies.

8. *How specifically did you choose the tuning parameter values for the PMWE and PMSE results? (i.e. What "performance" [line 296] is being optimized?)*

Response: Naturally, the ground truth is not available for the PMWE and PMSE measured signals. As such, the tuning parameters in this case were chosen empirically and were validated visually by professional judgment to improve reconstruction characteristics. Specifically, concentrating on the resolution of the narrow layer and continuity over time (in Figure 7), while maintaining smooth characteristics of the outer layers.

We will add a comment in Sections 5.2 and 5.3 accordingly.

Response to Reviewer 1. Technical corrections

1. *(line 313) "from in TV prior"* $\rightarrow$ *"from the TV prior"?*
(line 334) "Cauchy difference TV" $\rightarrow$ *"Cauchy and difference TV"?*

Response: Thank you for carefully reading the manuscript, we will correct the errors.