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 and a summary of our changes.

Sincerely,

Snizhana Ross, Arttu Arjas, Ilkka I. Virtanen, Mikko J. Sillanpää, Lassi Roininen, and Andreas Hauptmann 19 January 2021

Response to Reviewer 1

I found the description of this hierarchical convolution technique to be clear and wellorganized, 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, \rightarrow 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."

Changes: We have changed the wording as follows.

In line 127: In order to move from the continuous time signals to discrete samples

In line 136, 137: 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".

- **Changes:** Starting from line 165, we have removed "Convenience" and added the short reasoning: "which is a simplified assumption, but sufficient to detect the desired non-stationary features in this study."
- 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.

- **Changes:** We have added the text in lines 172-174: "We choose $\nu = p + 0.5$, $p \in \mathbb{N}$, as this provides a Markov approximation for the model. By the construction, the square-root of the inverted covariance matrix has a tridiagonal structure which is numerically convenient".
- 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.

- **Changes:** We added text in line 182-184: "[...], motivated by the Markov approximation that leads to a computationally useful presentation. Moreover, this also allows us to model ℓ via increments, thus simplifying both the model and computations. The explicit representation of \mathbf{L}_{ℓ} is given in the Section 4."
- 5. The Figure 3 labels and text discussing the figure refer to u as the "lengthscale 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.
- **Changes:** Accordingly, the labels to the Figures 3, 4, 6, 8 were changed from "Length-scale function" to "Log of length-scale function [arb. units]"
- 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 $\alpha_{\rm C}$ and $\alpha_{\rm TV}$. This underlines the need for an adaptive length-scale function in the estimation procedure."

Changes: We have added the following to line 260-263: "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 $\alpha_{\rm C}$ and $\alpha_{\rm TV}$. 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.

Changes: We have added a short discussion of possible other priors to the Conclusions as a possible direction for future studies (Lines 413-417): "In this study we have not investigated other priors than the Cauchy difference and TV prior. Naturally, there is a large selection of priors one can use for specific applications (Gelman et al., 2013): Gaussian priors, geometric priors, and even data-driven priors via neural networks (Adler and Öktem, 2018). The reason why we have concentrated on the Cauchy difference and TV priors, is because we want to avoid dyadic structures, specific geometries and dependence of data, which could be well suited for future studies, but out-of-scope here."

- 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.

Changes: We added the corresponding text to lines 315-318: "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."

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.

Changes: All the errors have been corrected.

Response to Reviewer 2

1. While the model and the algorithm are quite complicated, the authors made a large effort to make them as clear as possible in the text. In addition, no serious errors and flaws were found. On the other hand, it is very hard to overview the whole structure of the model at a glance. For the sake of readers, I give some minor comments in what follows.

The current version of Fig. 2 simply shows the relationship between the parameters, and the model structure is described in detail part by part throughout the sections 2-4. However, the current structure requires readers to go back and forth in the text until the model is understood and this is rather painful. In my opinion, Fig. 2, or perhaps better to add another figure, should also include the model structure itself to grasp the whole structure at a glance. More specifically, it should illustrate relationship of the Gaussian Process and Matern covariance, the additive epsilon and Gaussian pdf and so on in the diagram, as well as MCMC and MAP.

- **Response**: Thank you for the encouraging evaluation and the critical comments. We agree that the structure might be a bit hard to follow at first and welcome the suggestions how to make it more accessible. We have discussed how to best address the reviewers comments to improve the presentation and came the the conclusion that we will include an additional subsection 4.1 on "Model overview", that aids to summarise the whole process. In particular, to address the suggestion of adding another Figure, we will include a flowchart that shows which steps have to be taken and how these relate to each other.
- **Changes:** We have now added a new section 4.1 on "Model overview". We have decided not to include a flow chart, in contrast to what we wrote in the previous response as we felt it will be redundant. Nevertheless, we have now formulated the full hierarchical model in equation (16) followed by a pseudo-code for both, the MAP estimate and MCMC inference. We believe this should help the reader to follow the necessary steps, while all information is now collected in one concise subsection.
- 2. In addition to the logical relationship of the model parts mentioned above, it is recommended if possible schematically to show the sequence (in time) of the procedures to show which part of the model and how to start the calculation from.

- **Response**: Following the previous point, we will not only add the flowchart as visual illustration, but also include a summarising pseudo-code for the MAP and MCMC estimation. We believe that this should provide the necessary overview of the model in a concise manner.
- **Changes:** As previously outline, we have now included a pseudo-code in section 4.1.
- 3. In L.152, which is "Here $p(P_m|P, L)$ is the likelihood..." (L is intentionally capitalized for readability purpose in this communication), can $p(P_m|P, L)$ be $p(P_m|P)$? It is because P_m is presumably conditionally independent from L given P.
 - **Response:** This is correct, we will add a short comment that it can be also simply $p(P_m|P)$ due to conditional independence.
 - **Changes:** Text added in lines 153, 154: "and due to conditional independence is just $p(P_m|P)$ "
- 4. Equation (12) indicates the name of prior PDFs (Cauchy & Laplace) but does not show their mathematical forms. While this is accepted in case actual expressions are not concerned, it is recommended to show them in this paper because the definitions of $\alpha_{C/TV}$ are needed in the following discussions.
 - **Response:** We fully agree on this point and the corresponding mathematical expressions will be added add the corresponding location in Section 3.2. after Equation (12).
 - **Changes:** Lines 188-190: Text added "The probability density functions of Cauchy and Laplace distributions are given as $p(x) \propto ((x x_0)^2 + s^2)^{-1}$ and $p(x) \propto \exp(|x x_0|s^{-1})$, respectively, where x_0 is the center and s is the scaling of the distribution."
- 5. In Figure 5, what is the reason by which the sidelobe of the left plot (PMWE) is wavy while the other (PMSE) is quite smooth?
 - **Response:** We agree, that it is beneficial to include more detail about the sidelobe behaviour. We will add some more explanation in Section 5.2.

Specifically, for the wavy sidelobes we will add: "Considerable side lobes are produced in matched filtering of the shortcodes that are not Barker codes".

Regarding the PMSE, we will add that "the range side lobes are smooth because the long codes behave reasonably well in matched filtering, and sidelobe patterns of each code in the long code sequence are different."

Finally, the corresponding caption to Figure 5 will be changed to: "The pair of 10-bit codes used in the PMWE observation produces significant side lobes in matched filter decoding, while the sequence of 128 61-bit codes used in the PMSE observation leads to a smooth pattern of smaller side lobes"

Changes: We have made the following changes.

Line 311, 312: Text added "Considerable side lobes are produced in matched filtering of the shortcodes that are not Barker codes"

Lines 344, 345 : Text added "the range side lobes are smooth because the long codes behave reasonably well in matched filtering, and sidelobe patterns of each code in the long code sequence are different."

Caption to Figure 5, text added "The pair of 10-bit codes used in the PMWE observation produces significant side lobes in matched filter decoding, while the sequence of 128 61-bit codes used in the PMSE observation leads to a smooth pattern of smaller side lobes"

- 6. In Figures 6 and 8, what is the reason by which u_C and $u_T V$ are quite different where they are higher than 4.0?
 - **Response**: We acknowledge that this needed more explanation. The fundamental reason is that the Cauchy and TV process priors are different models, and have different parameterisations, thus leading to different estimators. The crucial point is that we want both models to detect two distinct parts, the smooth part and the high-frequency part, that is, even though the estimators look different, both of them lead to estimators which clearly model the targets as wanted. Even though the differences between the models may seem to be significant, the resulting backscattered power profiles are similar, thus one can claim that the model is robust against parameter tuning within this range.

We will add corresponding comments to the Discussion in Section 6. Following Reviewer 1, comment 7, as well as the next point.

- **Changes:** We have added a short comment in the discussion starting line 373: "In this study we have chosen two process priors, Cauchy difference and TV, which have different parameterisations, thus leading to different estimators. The primary goal herein is that we want both models to detect two distinct parts, the smooth part and the high-frequency part." Additionally, we have added a new paragraph to the discussion, starting in line 386: "Indeed, Cauchy difference priors lead to marginal distributions which are either unimodal or bimodal (See Markkanen et al. 2019). Bimodality is the key ingredient in designing models with rough features, that is, the edges are modelled with bimodal probability densities. For the TV prior, the edges are modelled via the product of two exponential functions, by which edges have uniform density. This means that the Cauchy difference prior promotes rougher features than the TV prior, and this is, in our understanding, the reason for differences in the reconstructions. Nevertheless, even though the differences between the models may seem to be significant, the resulting backscattered power profiles are of similar appearance, thus we believe that the proposed models are robust against parameter tuning within the presented range."
- 7. On p. 17, the authors discuss the difference between the results from Cauchy and Laplace priors, but its underlying reason is not mentioned at all. Since the difference is very curious and interesting, it is preferable to mention some of your ideas about it if you have any.
 - **Response**: This nicely complements the previously raised points. Indeed, our main idea is that the Cauchy process priors lead to marginal distributions which are either unimodal or bimodal (See Markkanen et al. 2019). Bimodality is the key ingredient in building models with rough features, that is, the edges are modelled with bimodal probability densities. For the TV prior, the edges are de facto modelled via the product of two exponential functions, which means, that at the edges there is "uniform" density. This means that the Cauchy process prior promotes rougher features than the TV, and this is, in our understanding, the reason for differences in the reconstructions.

As previously mentioned, we will add more details on priors and their differences to the Discussion in Section 6.

Changes: Please see response to the last point 6.

Response to Reviewer 2. Technical corrections

- 1. L169 and L177: Roininen et al. (2014) corresponds to two papers in the reference list. Please identify which one it is.
 - **Response:** Thank you, the appropriate publications will be identified correctly in the text.
 - **Changes:** Lines 171, 181. The appropriate publication Roininen et al. (2014b) has been identified
- 2. L169: Is "partial differential equation" correct? (10) and (11) look like ordinary differential equations.
 - **Response**: As it depends on dimensionality, we will remove "partial" to avoid any ambiguity.

Changes: Line 171, "partial" was removed

3. L313: out from $in \rightarrow out$ in L334: difference $TV \rightarrow difference$ and TVL395: $STEL \rightarrow ISEE$

Response: Thank you for your careful reading, the errors will be corrected.

Changes: All the errors have been fixed

- 4. Figures 3, 4, 6, & 8: Is the "unit" of length-scale function [km] or [log km]?
 - **Response:** We agree with the reviewer that on the Figures 3, 4, 6 and 8 units of the log length-scale functions needs to be clarified. Following Reviewer 1 comment 5, we will change the caption to "logarithm of length-scale function". The units are more difficult, as they are non-physical and depend on the sampling resolution of the underlying profile and are assumed to be universally 1. Thus, we have decided to add [arb.units] here as well.
 - **Changes:** Labels to the Figures 3, 4, 6, 8 were changed from "Length-scale function" to "Log of length-scale function [arb. units]"