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.

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.

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.

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).

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"

6. *In Figures 6 and 8, what is the reason by which u_C and u_TV 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.

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.

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.

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.

3. *L313: out from in → out in*
   L334: difference TV → difference and TV
   L395: STEL → ISEE

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

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.