Dear Jack Muir:

We have reached a decision regarding your submission to Seismica, "Bayesian eikonal tomography using Gaussian processes". We decline to publish the manuscript at this time and are attching the reviews that we received.

Thank you for submitting a novel method to address the inference of spatial distributions of seismic phase delays (as proxies for propagation velocity), along with often-missing uncertainty estimates, from spatially-localized and heterogeneously accurate observations without the need for choosing a model parameterization.

The reviewers, who were selected on account of their expertise in quantitative seismology in general and ambient seismic noise analysis in particular, value your contribution and would have looked forward to evaluating your method's potential in their own research. However, reviewer 2's articulation of the relative inaccessibility of the prose and the equations to a general audience of geophysicists prompts me to request one of several possible different approaches towards disseminating your innovative work. Examples of alternate approaches include

1. Follow reviewer 2's suggestion and submit to a more mathematically or computationally oriented geophysics journal.

2. Contribute your work in the form of an educational lecture & lab in the format of ROSES 2020/2021 (Remote On-line Sessions for Emerging Seismologists).

3. Re-cast your paper for a more general (and less specialist) audience of seismologists or geophysicists. Although seismologists might not be generally familiar with GPs and the corresponding statistics and equations, geophysicists are often acquainted with the method of kriging, which is essentially a GP with an outdated name yet might be worth a mention in a re-cast manuscript. Adapting notation to common geophysical variable names, and defining them profusely including their typical dimensionality, would be helpful to capable practitioners who are not familiar with GPs. In addition, I strongly recommend you also take the suggestion of both reviewers into account on selecting a more realistic, physically reasonable wavefield than the synthetic "toy example" used in the paper. In fact, two different examples with realistic wavefields, including realistic or real data, would go a long way towards illustrating your method's potential. Reviewer 1 further adds that "a rigorous comparison between GP and standard interpolation derived phase maps, and traditional tomography would be needed".

If you decide to go with option 3, which would be welcomed by Seismica, there are many other reviewer comments to take into account, including whether "tomography" is apt, extension to amplitude fields, separation of structural effects and data noise and their influence on tomographic results (e.g. phase velocity maps), clarification of terms and symbols used in equations of what is shown in figures. Please see the reviews for pertinent details.

If you decide to resubmit to Seismica, please explain in detail how a new version of your manuscript has taken the reviewers' comments into account.

I agree that your paper is well written, and constitutes a novel approach that may well improve the way our community conducts eikonal tomography in the future. However, the main impediments to publishing, currently, are the target audience (which appears to be data scientist specialists rather than seismology researchers), the lack of realistic examples, and the lack of suggestions on how the estimated uncertainties in the inferred phase fields can be used to improve

eikonal tomography.

Thank you for considering Seismica for dissemination of your work.

Kind regards,

Suzan van der Lee

Dear Suzan,

Thank you for collecting the reviewers comments and for your own editorial input. In light of the comments, I have made several substantial edits to the paper, most noteably including an entirely new example for a realistic velocity field, and changing the notation to hopefully be clearer. I have responded to the reviewer comments in blue below.

Best wishes,

Jack B. Muir

1 Reviewer A

In this manuscript, Muir targets to improve eikonal tomography using Gaussian processes. Instead of utilizing spline to create phase delay maps, the author describes a process of creating phase delay maps using Gaussian processes which more appropriately account for data error (if I understand correctly). While I believe the idea is worth pursuing and can improve the existing eikonal tomography methodology, the manuscript is written in a way that is extremely hard to read by a typical geophysicist (which I consider myself is). I tried hard twice to read the manuscript and gave up both times around halfway through (around page 7). If the intended audiences are mathematicians, I think the author can ignore my comments below and resubmit the manuscript to a more math-oriented journal. If the goal, however, is for general geophysicists to adopt the proposed processing scheme, I think the manuscript needs to be rewritten in a way that is more approachable for geophysicists. I only include some main comments below, which hopefully can help the author to improve the manuscript.

Thank you for raising these concerns — it is my hope that a relatively wide cohort in geophysics could read this paper and find it to be useful. As such, I have expanded and altered the notation to try to improve legibility as much as possible. I have also directly responded to the concerns below.

The mathematic notations used in the manuscript and the associated references are often unfamiliar to geophysicists. More descriptions are needed to provide some intuition needed to understand what is going on. Here are a few examples that I think require clarification.

"Gaussian noise η with the distribution $N(0; \sigma)$". It will be great to explain what is the N function.

There is now a longer definition of what a GP is in terms of defining how it operates on finite collections of points; as part of that the notation for N is defined (a multivariate Gaussian).

In eq(3), "$y = \tau(X) + \eta$", why y is capital but η is not. Isn't the dimension of the variables should be consistent on two sides?

Thank you for pointing out this, as I agree it was stated a bit confusingly; I've now explicitly mentioned that boldfont corresponds to observed data; as you cannot observe the error directly it is not in bold.

For eq(4), "$\tau(x) = \tau_0(x) + f(x)$", it says $\tau_0(x)$ is a reference phase delay field, for example for a laterally homogeneous medium. But it only says f is a zero-mean GP. What's the physical meaning of the f term? Does it represent the phase delay perturbation due to structural heterogeneity? But that explanation doesn't seem to agree with the expression of eq. (5).

The model is essentially "homogenous travel time field" + "heterogeneity & errors"; i.e. f accounts for lateral heterogeneity and also data errors. Note that f is a function distribution, not a function, which is now mentioned in the text.

In eq. (6), I don't think ρ is ever defined.

ρ and l_i are now explicitly defined immediately after the equation for the squared exponential k

On line 107, what is the physical meaning of $K_{\boldsymbol{X}\boldsymbol{X}'}$?

This matrix gives the covariance of the observed data in respect to the points that you want to estimate the travel time field at. By definition, it is the transpose of $K_{\boldsymbol{X}'\boldsymbol{X}}$ which has the easier interpretation of the covariance of the desired points in respect to the observed data.

What's the definition and physical meaning of joint multivariate Gaussian distribution N mentioned in eq. (7)?

The definition is now added earlier in the manuscript as stated above; the physical meaning is that all of the data and predicted values are correlated to each other via the covariance matrix + the reference model, which comes directly from the definition.

Related to the item above, I do not understand eq. (8) as well.

This equation is just a standard result in GPs / kriging; the derivation is quite long and so is not included. I've included a fuller explanation of what conditioning is just before the equation.

On line 113, what is CD?

This is now defined earlier — C_D is the data covariance matrix for some collection of data with correlated Gaussian errors (as opposed to the simpler assumption actually used in the paper and most other geophysical inversions of $C_D = \sigma^2 I$ where all of the data is assumed to be I.I.D. Gaussian with error scale σ.)

I am not familiar with the expectation operator E discussed in eq. (9).

The expectation operator gives the weighted average of some random variable (weighted by a probability distribution). The operator $\mathbb{E}$ is standard mathematical notation.

There are more but the list above is probably sufficient to demonstrate the challenge for readers.

I think the author tries hard to generalize everything which makes it extremely hard to follow. To me, the only practical application of eikonal tomography is 2D surface wave tomography. In that case, it is better (in my opinion) to explicitly express the terms for that and reduce the number of variables. For example, instead of saying "d coordinates", it would be more straightforward to get rid of "d" and say x represents the 2D location. Also, I will suggest reserving y as the variable for coordinate instead of data. One line 142, I am curious about what higher dimension applications the author has in mind when noting that "dimensions immediately generalize". If practically there is no such

application, I do not see the point to emphasize it. In eq(6), in what physical situation will the independent length scales (li?) be different? To me, it is more straightforward to replace the summation in eq (6) with the square of 2D distance normalized by the length scale instead of keeping the formula general.

Thank you for this discussion, which I have considered seriously. Ultimately, I decided that I disagree with the premise of specializing the paper to 2D in all cases, firstly because I feel that there are use cases in 1D and 3D for eikonal tomography as I now mention in the introduction, secondly because some of the results in the paper are generally applicable beyond eikonal tomography, and thirdly because making everything 2D does not actually substantively simplify the notation. That being said, I have made major changes to the notation in response to this comment, for example replacing y for data with d, using x and y as the 2D coordinate variables for 2D examples, replacing x with u in the saddlepoint expansion PDF to avoid confusion, etc. Hopefully the paper is now more consistent and easier to read. In regards to the independent length scales for the covariance function; if either the distribution of data (i.e. station distribution) or the underlying travel time field has anisotropic structures, then the performance of the regression is substantially improved by allowing independent length scales, hence motivating the definition used in the paper.

Unfortunately, all the math jargon has prevented me from understanding the part that I am actually interested in. Are there preset parameters being used to construct the phase delay map? That is, do the data uncertainty and/or the independent length being assumed in the process? For the spline approach, it basically assumes that there is no data uncertainty and that all measurements are independent of each other. This is obviously wrong, but it is unclear to me how the GP process resolves the trade-off between data uncertainty and variation due to real small-scale structural heterogeneity. I guess without fulling understanding the math expressions, I do not understand exact how the GP process constructs a map from input data.

The hyperparameters (length scales, data uncertainty, model variation, underlying mean velocity) are set by minimising a function called the negative log marginal likelihood; essentially this gives the hyperparameters that would be most likely to generate the observed data. This was discussed in the original draft but I have now signposted it earlier in the manuscript.

In Figure 1, more description is needed to understand what is shown in each subplot. I assume the Data Points plot shows the travel time of an example wavefield excited by either a real earthquake or a virtual ambient noise source. Is it physical to have a travel time maximum near the center of the map? While this might not be relevant to what the author wants to demonstrate, it is unclear why not use an example closer to a realistic wavefield.

I've expanded the caption to describe the experiment more fully; the data is the travel time assuming a source at (0,0), with added Gaussian noise. I've now more fully described in the text why I use a "nonphysical" example (noting that it does satisfy the eikonal equation) — in the synthetic example, I wanted to fully highlight the performance difference between simple spline models at the GP model. I've now added an entirely new section with a realistic velocity model (Ridgecrest, CA regional model of White et al. 2021) in addition.

Also in Figure 1, is GP Mean basically the input travel time map for the test? Is the Spline Difference the difference between GP Mean and Smoothing Spline maps? Is the GP Mean Difference the difference between GP Mean and

GP reconstruction in Figure 2? And is GP Std. Dev. the error estimated based on the difference between Data Points and GP Mean? Clearly, I am confused here as there is simply not enough description provided. Also, not sure why GP reconstruction in Figure 2 is plotted as a different aspect ratio compared to plots in Figure 1. Also in Figure 2, not sure why the X2 and X1 for the cross-sections were selected with such complicated numbers.

The GP results are the mean and standard deviation of the estimate of the phase delay given the observed data (this is now specified in the caption). The difference maps are of the spline / GP mean relative to the true data. The GP Std. Dev. is the posterior standard deviation of the GP (noting that the GP is a function distribution, not a single function). In general, the caption has been expanded. The numbers for the example cross sections have been adjusted slightly to become whole numbers. The aspect ratio is a tradeoff between accuracy and legibility / aesthetics; since this example is an end to itself and not important for future interpretative work, I've chosen to maintain the given aspect ratios for aforesaid aesthetic reasons.

Recommendation: Resubmit Elsewhere

2 Reviewer B

The study by Muir addresses the derivation and definition of uncertainties in phase gradient measurements applied in eikonal tomography. The author presents a theoretical framework that applies Gaussian processes (GP) to interpolate the phase (delay) field between individual data points and derive uncertainties of it. The manuscript is well organized and written, and a welcomed contribution of a barely addressed aspect of eikonal tomography.

The interpolation of single phase delay maps from a finite number of irregularly spaced data points, which this study addresses, deserves further attention and I think that this study can contribute to the discussion. I have a few, mostly general questions and remarks from a practitioners perspective, which can be likely dealt with in a minor revision.

Thank you for the very helpful review comments and the positive engagement with the study; my responses to them are below.

2.1 Review

Very generally, I am wondering whether "eikonal tomography" is the right framing for the study. Eikonal tomography refers to the generation of phase velocity maps that are obtained by averaging over several phase fields which propagate across the structure of interest in various directions (e.g. Lin et al., 2009). This study targets the interpolation of a single phase map which is just one, albeit important, element of eikonal tomography - the second element is averaging over multiple fields to obtain a phase velocity map.

I feel that, while in practice every eikonal tomography result will include an average over multiple sources, the essence of the technique is the differentiation of the phase delay field. Personally, I would call a single phase velocity estimate using the eikonal equation an "eikonal tomography", although I wouldn't particularly trust the results! Note that for example, the recent paper of Chevrot

and Lehujeur spends only two paragraphs talking about averaging in an otherwise longish paper. As such, I don't think it is unusual to not talk about the averaging process except in passing whilst still saying the paper is about eikonal tomography (Leaving aside that one could argue that "eikonal tomography" in and of itself is not a "real" tomography but rather a gradiometry!). Further to this, the section on the implications for sample statistics is a somewhat oblique investigation of this; one of the motivations for this study was to show that if the number of estimates taking part in the average are small then the PDF for the phase velocity can be biased and very heavy tailed, which is important in the case of anisotropy studies, especially near the edges of the domain. I've tried to make this case more explicit in the text.

GP are presented as an alternative interpolation scheme for single phase delay maps. Maybe I am missing something substantial, but if GPs are applicable to regression problems in general, I wonder why they shouldn‘t be similarly applicable to amplitude fields? Moreover, if differentiation of GPs is a linear process, couldn‘t one similarly fancy a Helmholtz tomography using GP (i.e. combining several interpolated phase and amplitude fields)?

Indeed, I think that GP regressions could be a useful way of performing Helmholtz or wavefield gradiometry studies, as mentioned in the discussion — unfortunately, I will have to leave this investigation for future work. One potential issue is that the saddlepoint approximation may be harder to derive for the Helmholtz tomography case.

As outlined in the introduction, eikonal tomography refers to ignoring amplitude variations in the wave propagation and derive phase velocities from the gradient of the phase (or phase delay) alone. The interesting aspect in the GP formulation is clearly that uncertainties in the determined phase field are provided. This has been barely addressed previously, and if so, rather empirically.In principle, deviation of the shape of the phase field from a radially symmetric (near-source) or planar (teleseismic) field has two reasons, Earth structure and data noise. As I understand it, the "GP mean" of a single phase field is the equivalent to the result of other (e.g. spline) interpolation schemes. The advantage of GPs is to provide an associated uncertainty. This uncertainty covers, however, both structural effects and data noise combined. It is not clear to me that a better representation of a single phase field (GP mean)will lead to a better final tomographic result.

I think that, in the limit of huge numbers of sources, such as isotropic inversions of ambient noise data in a big array, the outcome will not be that different between the GP method and others; the real benefit is in situations where there is less data, such as in anisotropic inversions where azimuth bins need to be treated seperately. In this case, explicit handling of the uncertainties is important to understand what the errors actually are in this low-data case; I've tried to make this more explicit in the discussion.

In eikonal tomography, multiple phase fields are combined/averaged to compute an (azimuthally an-)isotropic phase velocity map. This averaging can be understood to separate structural information from data (or interpolation) uncertainties. The most relevant question is therefore, how can GP derived uncertainties be propagated to the averaging over multiple wavefields in a useful manner? How can the GP uncertainties be exploited to provide more useful (anisotropic?) phase velocity maps? A general outline of an implementation strategy that accounts for these uncertainties and describes a way to improved

phase velocity maps would be needed.

The example that illustrates the implementation is generally instructive. However, the phase field in Fig. 1 seems to be non-physical if one thinks of a radially propagating seismic wavefield. The phase (delay) should increase radially away from the source whereas in Figs. 1,2 the phase delay decreases at larger distances from the origin. The motivation for the exemplary phase field should be clarified, maybe the reference model could be included.

I've now further mentioned why this model was used in the text (to better highlight difference), and also metioned that there is no reference model, in that it the example is constructed using the method of manufactured solutions.

Beyond the purely synthetic example it would be interesting to see an (additional?) example with a more realistic phase field, e.g. based on a simulation through a realistic seismic anomaly. To convince practitioners that GPs should be preferable over standard interpolation methods, a rigorous comparison between GP and standard interpolation derived phase maps, and traditional tomography would be needed. I'd be curious to see an evaluation of GP against e.g. spline interpolation across a synthetic structure, or even real data example, both for single and averaged wavefields. This would provide improved insight in the usefulness of deploying GP interpolation rather than routine (e.g. spline) interpolation algorithms. I am aware that this requires significant extra work and may justify a separate study, so I refrain from calling for it as part of this review. Nevertheless, for this study, a clearer formulation and discussion of how eikonal tomography (averaging multiple phase fields) could benefit in practice from a better representation of the (single) phase field interpolations with GP should be included.

I have now included a new example using a realistic synthetic velocity map, showing that similarly improved results are obtained. I've kept the focus on single non-averaged inversions as mentioned above.

Fig. 3: are the two point locations from the example in Figs. 1, 2? If so, they could be marked in Fig. 1. It should be clarified why these two points were chosen, what makes them representative? The red crosses and lines are barely visible, maybe they could be drawn by a thin line atop the orange? The points are now marked in figure 1, and the GP-FD markers are now smaller and laid on top as suggested. The points were chosen as being near the edge, and near the center with many points, which is specified in the text

Recommendation: Revisions Required

Dear Jack Muir:

I hope this email finds you well. Thank you for thoroughly considering the reviewers's input in revising your manuscript.

Reviewer 1 is happy with the revisions and finds the revised manuscript more accessible. Nevertheless reviewer 1 requests additional clarifications, which I expect will be quite doable, and I look forward to seeing them addressed. In addition, please also respond to my own inquiries, detailed at the very end of this message.

Based on reviewer 1's and my reviews, your manuscript is likely to be suitable for publication after additional revisions. When you are ready to resubmit the revised version of your manuscript, please upload:

A 'cleaned' version of the revised manuscript, without any markup/changes highlighted. A pdf version of the revised manuscript clearly highlighting changes/markup/edits. A 'response-to-reviewers' letter that shows your response to each of the reviewers' points, together with a summary of the resulting changes made to the manuscript. Once I have read your revised manuscript and rebuttal, I will then decide whether the manuscript requires further minor changes, or can be accepted. I look forward to seeing the requested revisions. Please don't hesitate to contact me with any questions or comments about your submission, or if you have any feedback about your experience with Seismica.

Thank you for an innovative contribution to Seismica and seismology.

Kind regards,

Suzan van der Lee

Dear Suzan,

Thank you for running the second round of reviews and for your own editorial comments. I have responded to the issues raised in blue below. I've also lightly edited the abstract so that I don't use the word "result" three times in two sentences like in the previous version!

Best wishes,

Jack B. Muir

1 Reviewer A

First of all, I apologize that it took longer than expected for me to find time to review this manuscript.

Overall, I find the revised manuscript much easier to comprehend compared to the previous version. While it is still quite mathematically focused, I feel I can understand most of the concepts and the proposed process. I can see the advantage of the Gaussian process approach proposed in the manuscript, which can put velocity measurements derived from eikonal tomography on a more solid statistical ground. I do have some lingering comments that I hope the author can address, which hopefully will further improve the manuscript. I should admit that some of the math derivations are still elusive to me, particularly the section describing the saddle point method, which I decided not to comment on.

1.1 Major Comments

- The manuscript suggests that the uncertainties reported for eikonal tomography are often underestimated. It will be helpful to clarify this statement and make sure the assessment is fair. Does this statement apply to only regions where station coverage is poor (e.g. edge of the map)? In Figures 8 and 9, the example is taken from a location close to the edge of the map, where the result is expected not to be reliable in the traditional eikonal tomography approach. It will be nice to show an example where station coverage is good. Based on Figures 5 and 7, it seems the spline approach works pretty well for regions with good station coverage. The discussion around Figures 8 and 9 principally relates to areas where the coverage is poor. I've included the example of point 2 in Figure 8, which shows that the interior points produce quite nice posterior distributions. This paper is not meant to totally replace spline workflows (indeed, the two workflows are closely related, as I discuss under the next point), but rather show how the machinery of GPs allows us to automatically tune a non-parametric model, obtain the full posterior and approximate it efficiently using the saddlepoint method.

- It is unclear if resolving the GP hyperparameters is part of the proposed process. Assuming it is, there is no discussion on how the resolved hyperparameters agree or disagree with the input value and whether there is a tradeoff between each parameter. Furthermore, it is unclear if the Gaussian distribution and the proposed squared-exponential kernel (eq. 7) truly reflect the expected structure heterogeneity and measurement errors. If either the heterogeneity or error can't be described by GP, will the proposed process still outperform the traditional spline approach? To me, it makes sense the proposed Gaussian process works when the underlying GP assumptions are all satisfied but it seems the spline approach relies on fewer assumptions. Section 2.2 discusses the methodology for determining the GP hyperparameters; recovering them is not meaningful for either of the two synthetic experiments shown as the travel time field is not generated using a GP for either of them (nor would we expect travel time fields in real life to be!). Ultimately, splines and GPs are quite similar; they are both non-parametric regression methods that can either be smoothing or constrained to fit the data exactly, and have hyperparameters that describe their properties such as characteristic length scales. In fact, Kimeldorf and Wahba (1970) show that smoothing splines are in fact a subset of certain GPs, although the correspondance is very technical and the squared exponential kernel I use in this paper does not reproduce splines. The advantage of the GP approach is partly flexibility, in that we can choose a desired covariance kernal that is appropriate to our problem, and partly the nice mathematics that result, which makes finding posterior distributions easy. See Kimeldorf and Wahba, The Annals of Mathematical Statistics, 1970, Vol. 41, No. 2, 495–502.

1.2 Minor Comments

- On lines 71-72, I am not sure if the statement "over-smoothing the map will result in a measurement of Cp that is too large; conversely, maps that

are too rough will result in too small Cp" is accurate. The Cp is related to the gradient of the travel time map, but the roughness/smoothness of the map describes the second derivative of the map. If the travel time map is rougher than it should be, there should be areas that become faster than expected but there should also be areas that become slower than expected. If there is any sum effect, I would expect the rougher surface will lengthen the total ray path, which will effectively make the Cp slightly faster.

- It will be good to avoid using the same notation for different things to avoid confusion. For example, the f(x) in equation (3) and equation (5) , the rho symbol in equation (7) and line 184, and the two Tx in eq (15) which only encompass Ty. I have tried to further avoid any notational clashes; both the f are replaced (equation 3 with m to suggest mean, and equation 5 with r to suggest residual, so that neither clashes with the f in the saddlepoint discussion), the first ρ is replaced by a for amplitude, and the first Tx is replaced by grad T as should have been the case in the first place. I haven't changed one of the K's (either the covariance matrix or the cumulant generating function) as they are both nearly totally standard within their own field — I have to hope that the two sections are sufficiently distinct to make which one I am using clear by context.

- It is unclear how the smoothing spline travel time plot was made in Figure 1. Is the method used the same as the minimum curvature surface fitting method used by Lin et al. (2009)? Looking at the map, it is unclear if the map fits all the input data points and it is surprising to see the striping pattern in the Spline Difference plot. If it is not the minimum curvature surface fitting, it will be good to have that result presented as well. The spline method used here is the default smoothing spline of FitPack (i.e. what a user would get if they fit a spline using scipy without changing any defaults). This spline normally performs well for generic regression tasks. In this case, enforcing that the spline fits the data exactly is a bad idea because the data has added independent Gaussian random noise — I'm inclined against adding another spline comparision because the figure is already quite large. I've chosen to implement this spline (along the lines of Chevrot and Lehujeur, 2022, rather than Lin et al. (2009)) because I think it is a fairer representation of what a "modern" spline based eikonal equation workflow would look like.

- There is a missing y-label in the upper left corner of Figure 3. Thank you for raising this; the label is not missing, but the lack of the label is something of a deficiency of corner plots as a genre despite their general usefulness for exploring joint probability distributions; for the 1D histograms, the y-axis is traditionally not labelled (it is value of the PDF for that variable). I've now stated that in the caption.

- It would be nice if some intuitive explanation could be given for the distribution shown in Figure 8. It will be good to show another example with a different source location. The plots now include point 2 (the point closer to the interior, which has a more well-behaved posterior distribution).

- Lines 323-324 and Figure 9, it is unclear to me what represents the 4^n draws and what represents the 100,000 repeats. Is one of these reflecting

the different source locations? If yes, how does this after the spline result? In this experiment, I draw a batch of 4^n samples from the GP posterior, and calculate the mean or median. I then do this 100,000 times; this gives the distribution of the sample statistics (i.e. the width of the distribution of the sample means is the standard error in the mean if all of the distributions are normal.) The point of this is that the sample distribution of means is not particularly normal, even for a decent number of draws (which could be repeated individual eikonal tomograms from similar source locations), when the underlying distribution for phase velocity is heavy tailed like it is for point 1 in figure 8. This is the sort of setup that often occurs on the edges of arrays, and especially for attempts to infer azimuthal anisotropy. I have added to the discussion to make this more explicit.

Recommendation: Revisions Required

1.3 Editor Review

- I feel "These" in "These hyperparameters are" (l.122) refer to the l_i mentioned in line 120 (and rho). The sentence in between is useful but disconnects "these" from what it is referring to. Perhaps replace with "The hyperparameters l_i are" I have made this more explicit as suggested. Unfortunately at this point I can't write out all of the hyperparameters completely as I haven't yet introduced s0 and sigma.

- l.162 I feel you mean equation 10 (not 11) in "mean given by Equation 10 and covariance given by Equation 12:" Thank you, this has been corrected.

- l.249: I feel you can say that u could represent the slowness squared, or phase velocity, for example. This is a good suggestion, and has been implemented.

- l.253: it is not clear what the role of s is in eqn 17, what it represents, and what its dimensionality is (and why). Without the latter the concept of saddle point is hard to get. This is even more confusing when, at the end of the section, you say that higher-than-one-dimensional application is left for future work. Unfortunately, s doesn't really have a particularly clear interpretation — the cumulant generating function is a useful calculation tool, but you rarely use it directly; normally there is some operation that you evaluate with a particular value of s that gives you something useful; for instance, the saddlepoint approximation here, or e.g. the cumulants of the distribution by evaluating the derivatives of K at s=0. In regards to the higher-than-one dimensional case, I've tried to restate this to be clearer — what I meant was, determining the joint distribution of two random variables; the saddlepoint aproximation in this paper is for a single random variable (e.g. Cp at a particular point) that is a function of a random field (i.e. the GP) in any dimension, although I only present it for the 2D GP case for ease of exposition. To make this clear, I have changed the reference to "root finding in two variables" rather than "two dimensions".

- l.254: remove one "then" Thank you for pointing this out, the extraneous then has been removed.

- l.261 here K is a CGF rather than a covariance matrix, so its relation to X seems to fall out of thin air, given the abstract introduction of the math in this section (u is a "random variable" and s remains undefined). I've added a note that the two K's are different; unfortunately, as I mention also to reviewer 1, K is standard notation in the two different fields discussed in this paper.

- l.294 I feel you can say explicitly that g is the phase velocity and u is the slowness squared. I've tried to keep the distinction between functions and variables; to do this better I have made this sentence more explicit to define g as a transformation.

- l.302 missing article in "joint distribution function two points"? Thank you for pointing this out; corrected

- Overall, section 3 (Approximation of the posterior using the saddlepoint method) remains confusing (I am in agreement with reviewer 1). It is clear this discussion underlies Figure 8, and might therefore be necessary, but I doubt that readers will find enough concrete information to apply this method themselves. This might lead readers to resort to more Monte Carlo computations, which, promisingly, seems to be what you are trying to avoid. I agree that this section is difficult; in particular many of the underlying proofs in the statistics literature require a lot of contour integration, which is something I'm certainly happy to have left behind in undergrad! I think it is worth the inclusion however, as it is really a lot faster and can potentially be more accurate than Monte Carlo sampling when we are looking at the tails of the distribution. I've rewritten the introduction of the saddlepoint approximation to make it a direct quotation rather than implying that going from equation 17 to equation 18 should somehow be obvious. In order to apply the method, users can therefore just directly calculate through the equations until they reach equation 26. The attached Pluto notebook contains a full implementation as well. In regards to the success of the saddlepoint method itself — I was being honest in the manuscript when I called it serendipitous. I haven't read a truly convincing explanation as to why it works as well as it does (compared to e.g. just approximating the posterior directly using Laplace's approximation) in many cases, beyond what I have already stated in the paper. However, in this application it does work extremely well, and has the distinct advantage of dealing with the CGF which is much more convenient for sums of random variables than the PDFs themselves. Failing that, and on another note, one of the nice features of GPs is that even if users want to perform Monte Carlo sampling, they can do so very efficiently as we have access to samples directly from the posterior; this is distinct from methods such as Markov Chain Monte Carlo in which samples from the posterior can only be accessed indirectly by proposing and rejecting samples.

- l.373 You are the first person I encounter in geophysics who uses the Julia programming language in Pluto notebooks. Perhaps this reflects my limited perspective of the geophysics world, but I do know I share this with many readers. Therefore, please elaborate a little bit on how users can run

your notebook on their computer, including environmental requirements and what they might need to install first. I have now added instructions to the Data and code availability statement. One of the great advantages of Pluto is that they contain an internal package manifest, so all that the user needs to do is install Pluto itself and then run the notebook; all further dependancies will be automatically installed at the correct version! This is a massive improvement over the dependency nightmare induced by trying to line up Python versions in Jupyter, and a major part of my advocacy for Julia.