

Author's response to the editor and reviewers on May 2 regarding GI-2021-28

Knut Ola Dølven¹, Juha Vierinen², Roberto Grilli³, Jack Triest⁴, and Bénédicte Ferré¹

¹Centre for Arctic Gas Hydrate, Environment, and Climate, UiT The Arctic University of Norway, Tromsø, Norway

²Institute for Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway

³CNRS, University of Grenoble Alpes, IRD, Grenoble INP, 38000 Grenoble, France

⁴4H-JENA engineering GmbH Wischhofstrasse 1-3, 24148 Kiel, Germany

Correspondence: Knut Ola Dølven (knut.o.dolven@uit.no)

The author's response presented here are based on the author replies posted in the discussion on the manuscript (<https://doi.org/10.5194/gi-2021-28-RC1> and <https://doi.org/10.5194/gi-2021-28-RC2>), however there are some minor additions to these and changes which will be presented after the individual author comments. Section 1 and 2 are the author comments to the reviewers from the discussion forum and additional comments are found in Section 3.

5 1 Author reply to "Comment on gi-2021-28" by Anonymous Referee 1 submitted on on 13 Dec 2021

The authors would like to express their thanks for this very thorough, positive and constructive review. Here follows our response to the review report. Text from "Comment on gi-2021-28" on GI-2021-28 by Anonymous Referee 1 are in grey italic font and our responses are in black normal font. Text added to the manuscript is written in emboldened font and are in quotation marks.

10 Comments

1. In your simulated data and toy model, you use a measurement error proportional to um plus a constant noise floor term. After reading the full paper, I see where this comes from (your particular application/example), but do not think that it serves your objectives here as outlined in section 2.1 (l. 134ff.). While the lagged inflation/deflation of the estimate's \hat{u}_a uncertainty is well explained and understandable in the field example (Figure 6d and corresponding text), it raises question marks and

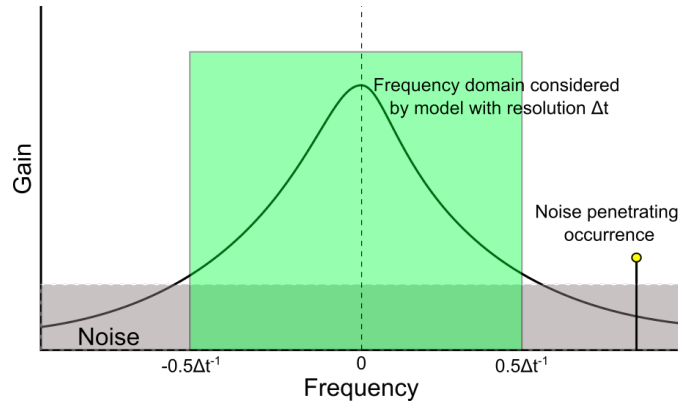
15 creates wrong conclusions here, early on in your derivation, e.g., "So it seems that the method accumulates/amplifies noise terms with time, as the \hat{u}_a uncertainty envelope grows and grows with time (Figure 3a and c). What would happen if there were a longer 0-period of say time 50 instead of time 5, would the error envelope be huge already at the step change? Can the method only be applied to short pieces of a longer time series??" Instead, I'd suggest to use only a constant measurement error term in the toy model for simplicity, to illustrate that your method does not suffer from accumulation of noise and (in

20 such a case) can keep a nearly constant \hat{u}_a uncertainty. However, ...

Section "2.1 Simulation and Δt determination" has a threefold aim: A) Validate the method in a numerical experiment B) explain how the method is implemented in practice. C) Describe how regularization optimization is implemented (L-curve). An important property of this method is its ability to weigh the least squares solution to Eq. 6 by the measurement uncertainty of each measurement (uncertain measurements are weighted less and vice versa). In our aim to validate the method in a numerical experiment (A) we therefore think that the simulated data should contain varying uncertainty. However, the point made by the referee is valid considering that this section also aims to explain the practical implementation of the method (e.g. considering Figure 3a). We therefore added a parenthesis in the sentence prior to Eq. 14 such that it now reads: **"We assume that measurement errors are proportional to $u_m(t)$ (measurement uncertainty increase with increasing concentration) in addition to a constant noise floor term, providing a standard deviation for each measurement given by:"** and also rewrote the sentence found in 1.178/1.179 (in GI-2021-28) such that it now reads: **"The error estimate, given as 95% confidence intervals (2σ uncertainty), is obtained by Eq. 11 and the simulated measurement uncertainty (which is proportional to $u_m(t)$, Eq. 14)."**

2. I find it hard to understand and believe that (e.g., with a constant σ_j) the uncertainty of the reconstruction \hat{u}_a does not have/show a dependence on the magnitude of correction (e.g., difference between \hat{u}_a and \hat{u}_m). I.e., that the reconstructed signal of a sensor pretty much in equilibrium (in no-gradient regions) should be as certain or uncertain as the reconstructed signal where the sensor experiences a strong gradient (and needs a strong correction). Intuitively, with small corrections required (no-gradient case), sensor noise should not be of paramount importance (just slight amplification, reflected in smaller estimate's uncertainty envelope), whereas with large corrections (strong gradient case), the same level of sensor noise should have a stronger impact on the reconstructed signal (more strongly amplified, so larger estimate's uncertainty envelope), shouldn't it? The figures 2 and 3 tend to support this intuition: "Overshoots" in the reconstructed signal \hat{u}_a tend to be a more prominent issue in strong gradients (here: step change, which violates the smoothness assumption as noted in the Appendix A, agreed. But the same is true if \hat{u}_a has a "smooth" strong gradient within Δt_j resolution). Similarly, fit residuals (Figure 2d) show a different character at the step change/strong gradient, where they are more coherent (and more coherently wrong) across multiple t_j (suggesting $\hat{u}_a(t_i)$ uncertainty should be larger), whereas they show only high frequency fluctuations around a zero-mean in the later no-/small gradient (suggesting validity of a smaller $\hat{u}_a(t_i)$ uncertainty). This is probably the most important point that needs to be addressed by the authors (and testifies to the otherwise really excellent work): Should the uncertainty on \hat{u}_a not be larger when there is a large correction compared to a small one?

As observed in the model fit residuals, there are indeed artifacts in the reconstructed signal where the step change occurs - even with our example of a "good" Δt choice ($\Delta t = 1.35$, Figure 2d). Model fit residuals should ideally be roughly normally distributed and without strong systematic patterns, which can indicate errors in the model assumptions or model itself. The model fit residuals can for instance reveal if our assumptions about the growth law equation is correct or tell if the complexity of the model is sufficient to resolve the property of interest. In the step-function case, this is indeed what is happening: The



AR 1. Drawn conceptual figure showing the frequency response of a growth law equation and domain considered by a model (as outlined in GI-2021-28) with a time-step assumption of Δt which is indicated with the green shaded box. When considering the noise and noise penetrating occurrence one can picture that the underlying signal is flat in frequency domain with an event indicated by the yellow circle.

extra wiggle at the step change indicates that in this part of the time-series, the model complexity is insufficient to properly model the signal. Figure AR 1. illustrates this by the growth law frequency response, which essentially is a low-pass filter, and indicated frequency region considered by a model with time-step Δt : The convolution dampens high frequencies in the input signal, however, events with considerable high frequency components (compared to the rest of the time-series - like the step-change) can still make its mark on the measured data. In our model, we assume that the signal we try to reconstruct only changes linearly with time-steps Δt or larger (or that there is no information we consider useful between these points). The result is that the measurements m'_j will deviate from this assumed linearity in $u_m(t'_j)$ (our model assumption) between the modelled points $u_m(t_i)$, thereby resulting in increased model fit residuals here. In Figure 1. of the manuscript, this effect is clearly visible as the discrepancy between the red line (modeled measurements $u_m(t'_j)$) and the red dots (real measurements m'_j).

In practice, it will naturally be more difficult to reconstruct a rapidly varying signal, since more correction needs to be applied to resolve this, which in turn puts higher demand on the measuring equipment. However, for the algorithm itself it is not the case that the method is generally less reliable/has a higher uncertainty for rapidly changing signal, e.g. where the most likely solution suggests that the property of interest is constant, a possibility still exists that the property is actually varying within modelled uncertainty and that the instrument is incapable of resolving this due to its accuracy limitations. That being said, artifacts due to poor complexity is more of an issue where variability is high and these regions can be identified in the model fit residual (essentially showing that there is more information to be found there).

In practice, there are probably many situations where it is difficult to completely avoid local deviations in the model fit residuals, since time-series often have varying local variance (e.g. parts slowly varying, parts fast varying signal). In the L-curve optimization, Δt is chosen as a compromise between model fit residuals and noise amplification for the *whole* time series which can result in insufficient model complexity for high variance regions of the time-series (the degree of this problem of

course depends on the capabilities of the sensor, i.e. accuracy and response time and the time-series). Fortunately, such regions are identifiable in the model fit residuals, providing information that can be used in data evaluation to either point out regions with slightly increased uncertainty (if the deviation is relatively small) or to re-evaluate the Δt assumption and for instance slightly increase the model complexity. It is also possible to split the time-series and apply different Δt s to different sections and thereby retain high accuracy for regions with smaller high-frequency components. This of course implies that the end result is a more complicated data set (with different time-steps).

To make this aspect clearer we

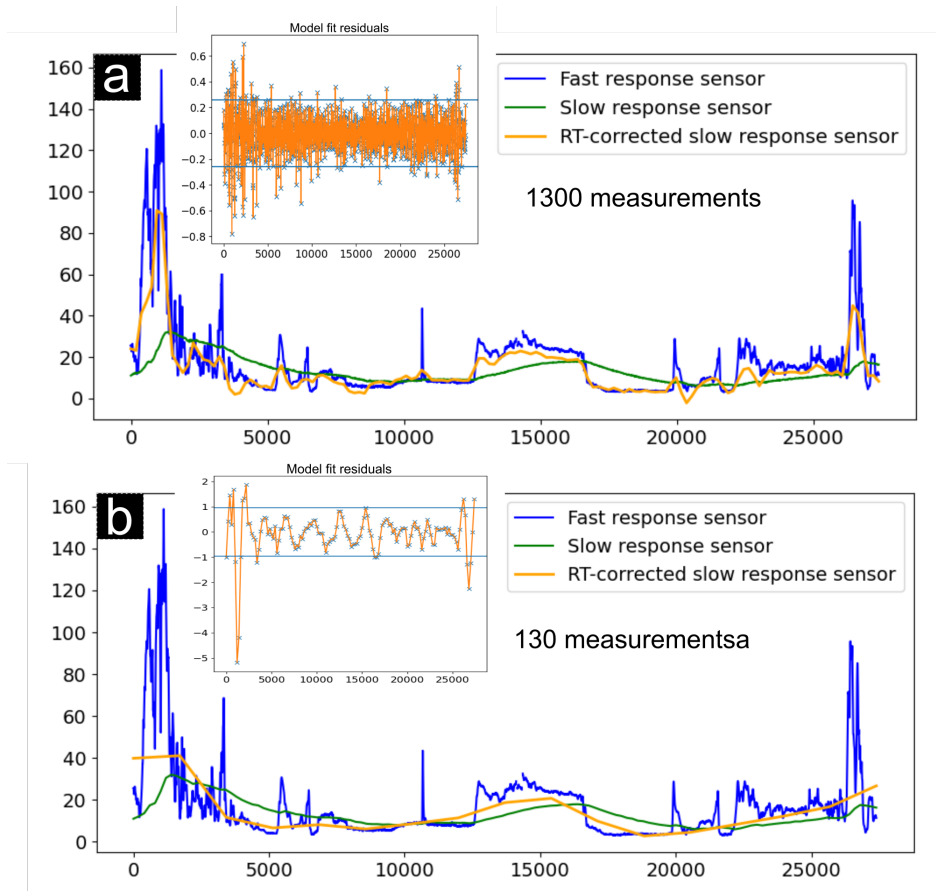
- expanded and rewrote the end of the paragraph starting from l.177 (in GI-2021-28) to: **"Since we assume that the property of interest changes linearly between our model points, a small irregularity in the model fit residuals remains at the step (Figure 2d) due to the models inability to capture the high frequency components of the step function. In practice, the fit residual irregularity arise due to the discrepancy between $u_m(t'_j)$ and m_j at the step change (the effect is clearly visible in the schematic, Figure 1). Nonetheless, the step-change is well represented within the limits of the resolution provided by the model assumption without very eye-catching fit residuals (Figure 2d)."**
- We also added a paragraph at the end of this section 2.1 (which also touches upon several of the comments addressed here): **"Even though the step function is an unrealistic scenario in a practical application, it is likely that the variability of a measured property can change considerably within a single time-series, for instance in profiling applications. Since the L-curve criterion provides a Δt which is a compromise between error amplification and model fit residuals, this can result in a model complexity which is too crude to resolve important high variability sections of the data. Generally, the model fit residuals should be roughly normally distributed, and should not have strong irregularities or systematic patterns. Inspection of model fit residuals can identify sections of the data set with too low complexity (resulting in residual spikes, as shown in the simulation experiment) and considerations can be made such as increased caution in data interpretation. Alternatively, it is also possible to manually increase model complexity for the whole or certain parts of the time-series to reduce the fit residuals in these regions."**
- Additionally, we made some small changes to Figure 1 and a couple of other minor edits throughout (see the tracked changes document).

3. The authors present examples where there are highly-resolved observations together with a very long time constant. This is arguably a very favourable case for response time corrections, where the high resolution allows for a lot of averaging, thus keeping noise amplification low. The authors should add a comment in their manuscript on more poorly resolved scenarios, where τ_{63} is closer to Δt_j and not orders of magnitude between them ($\tau_{63} \gg \Delta t_j$). Is there a limit (e.g., to the applicability, utility, ...) and if so, what to watch out for?

It is indeed an advantage when the sensor is able to collect large amounts of data, since this method is based on providing a meaningful separation between noise and useful information in any data point and harness this information to provide a good estimate of the ambient concentration - i.e. more data gives better estimates. There is no fixed limit to the applicability, although it of course becomes meaningless to apply this method when τ_{63} approaches the characteristic variability in the property of interest. Since this method models error amplification and can be evaluated for general consistency (by model fit residuals) the limit to its applicability will become apparent in each individual case. In other words, performance will degrade with decreasing amount of information and/or measurement accuracy. For instance when τ_{63} approaches Δt_j , L-curve optimization will most likely give a solution showing that there is not much more information to retrieve from the measurements. In this case, if τ_{63} is also very large and the property of interest is in reality varying considerably, the instrument is probably incapable of providing meaningful information for the property of interest (the RT cannot retrieve information which is not there). Figure AR 2. shows the application on the field data with 10^{-1} , and 100^{-1} of the data ($\Delta t_j = 20$ s/1300 data points and $\Delta t_j = 200$ s/approximately 130 data points). The figure shows that the amount of variability which is possible to resolve is reduced as the measuring frequency and correspondingly amount of information is reduced (as expected), however, there is no critical point where the performance suddenly drops. If the instrument had higher accuracy or faster response time the resolution of the RT-corrected signal would be of correspondingly higher resolution/quality (and the L-curve criterion would yield a different result). The general advice would be the same for any application of this method: An L-curve resembling an L and roughly normally distributed model fit residuals (and if deviations inspect and/or comment) are good indications, however, limitations on what variability can be resolved is of course dependent on the resolution and quality of the data as well as response time of the instrument. In 1.131 we added **"The quality of the solution relies on an appropriate choice of regularization parameter Δt and noise/uncertainty in the measurements, but also on the ratio between the RT of the sensor and variance in the property of interest."** In addition, the additions under Comment 2 are also relevant here.

4. As outlined in the introduction, the method may find application in a wide range of fields and settings (e.g. in profiling, on moving platforms, ...). Among them are scenarios, where there is a difference in scale across the time series, both in expected dynamics (e.g., deep ocean with less dynamic vs. surface layer with more dynamics) or in measurement resolution (e.g., a lower and a higher resolution Δt_j part). Please add some comment on how to deal with such irregularly spaced scenarios: Would you recommend to split such a time series and use a specific model Δt_i for different parts for better resolution, or rather keep it as one (with uniform scale Δt_i) for better L-curve analysis? Or instead of an evenly sampled grid t_i (l. 71) stitch a few pieces with different Δt_i sample spacing together?

I understand that generic recommendations are hard to do, so if you want, imagine a profiling scenario in the open ocean where response time is on the order of 100 s and measurement resolution varies from 500 s from 2000 m to 1000 m depth (ca. 20 samples), to 100 s from 1000 m to 300 m depth (ca. 80 samples), and finally to 20 s from 300 m depth to the surface ca. 150 samples.



AR 2. a) Result of the response time algorithm with automatic Δt selection for field data with simulated 20 s and b) 200 s measuring interval (rather than ~ 2 seconds) to illustrate hypothetical cases where the instrument records at lower frequency.

We have touched on this topic under comment 2 but will further address this in comment 6 since they are related.

5. Side question: Can a Δt_i smaller than the max. Δt_j be selected by the regularization? From your example in Figure 3d it seems like it.

You can choose (manually) to have a Δt_i which is smaller than Δt_j , but this would not be meaningful since the region
 140 between Δt_{j+1} and Δt_j is by our (discrete) definition linear (i.e. there is no information here). This is also the case for our model assumption, i.e. the property of interest can only change linearly between any Δt_{i+1} and Δt_i . A solution where $\Delta t_i < \Delta t_j$ would not be chosen in the L-curve criterion since such a solution is 1. of higher complexity and 2. would have higher error amplification but similar model fit residuals than a solution where $\Delta t_i = \Delta t_j$. Both these aspects would make

our implementation of the L-curve criterion (our measurement and fit residual norm definitions, Eq. 15 and 16 in manuscript)
145 prefer the sparser solution.

6. With your approach in general, is there a difference between full series or subseries application (given same parameters and neglecting edge effects)? I.e., in case of very long time series (and thus a large matrix G), is there a trade-off of splitting the time series to avoid memory problems?

It could indeed be beneficial to split and stitch data sets to reduce memory issues and solve high/low variability sections
150 separately. This is entirely possible and there are numerous ways to split and stitch data set to limit edge effects. There are no particular issues when doing this when using the proposed method other than being aware of the increased uncertainty towards the edges of the reconstructed time-series (thus it is advised to have overlapping seams). In general, such an approach becomes relevant when

- The number of data-points in the model (t_i) approach $\sim 10^5$. In practice, this depends on the total size of the data set
155 and the difference between Δt_i and Δt_j . In the field data set presented in the manuscript, m' was of length $M = 13700$ while the length of u (modeled measurements) was $N = 498$ for the chosen solution.
- When there are sections in the model fit residuals which show a strong effect of too poor complexity in the model. See also comment 2 here.
- This is also the case for the reviewers example with varying time-steps in the measurements: If there is a need for fast
160 response in parts of the time-series, but not in others, splitting the time-series can be a good option. In general, for a deep ocean profile where there is for instance two different regimes with regards to variability and where it is crucial to resolve the high variability regime it could be advisable to split and stitch the time-series, or manually select a model with enough complexity to resolve the high variability regime and accept unnecessary high uncertainty in the low-variability sections.

165 The added text referred to under comment 2 should address these concerns in the text. Regarding memory usage, there is also a warning trigger implemented in the code.

7. Code and data availability: Providing the code of this manuscript will be of tremendous help to anyone trying to apply this excellent work. For code repository, I would however discourage a “static” manuscript supplement, which does not allow for bug fixes or feature updates, and instead use a code repository where this is possible (e.g., github/zenodo or similar).

170 We completely agree and there is a github repository for code and data here: https://github.com/KnutOlaD/Deconv_code_data. This is also linked in the manuscript under code and data availability.

Minor comments and typos

- l.69: *Please add a reference (e.g., to a textbook?)* Reference added.
- l.106: *"The constant $\gamma \gg \sigma_j^{-1}$ is a numerically large weighting constant" Like 10x numerically large or like 1000x numerically large? (Does it matter?) Please add some guidance.* We expanded this sentence such that it now reads:
175 **"The constant $\gamma \gg \sigma_j^{-1}$ is a numerically large weighting constant (we used $\gamma = 2\Delta t \cdot 10^5 \sigma^{-1}$), which ensures that when solving this linear equation, the solution of the growth-law equation will have more weight than the measurements."**
- Eq. 8: $k_i + (\Delta t)^{-1}$ when $i = 1$ and $j = i$ Typo? There was indeed a typo, but I believe the typo was it in the line
180 below where in OS-2021-28 said " $-(\Delta t)^{-1}$ when $i = 1$ and $j = i + 1$ " which is now corrected to " **$(\Delta t)^{-1}$ when $i = 1$ and $j = i + 1$** ", this should now be consistent with Eq. 2.
- Eq. 15: *This is not normalized by N (as N changes with different Δt_i), correct?* Yes.
- l.175: *remove "is"* We removed the "is".
- l.211: *Worth adding an equation for this (like Eq. 15)* Sure, we added an equation here.
- l.242(ff.): *Please check the authors guidelines on how to format numbers in exponential notation (I'd have expected $5.747 \cdot 10^{-4} s^{-1}$ or similar).* We changed this accordingly.
- l.248: *Correct equation referenced?* Indeed not. We corrected this and now the correct equation should be referenced.

2 Author reply to "Comment on gi-2021-28" by Anonymous Referee 2 submitted on on 03 Feb 2022

We thank the referee for a thoughtful review, which has helped us improve the quality of the paper. Text from "Comment on
190 gi-2021-28" on GI-2021-28 by Anonymous Referee 2 are in grey italic font and our responses are in black normal font. Text we added to the manuscript are in emboldened font in quotation marks.

Reply to the views proposed by reviewer 2

1. *There are too many curves in Figure 3. it is difficult to see clearly;*

The curves represent the amplified noise and uncertainty in the reconstructed signal and it is certainly difficult to see what
195 this signal is supposed to show due to the noise. The very noisy graph in Figure 3a looks noisy with intent, since illustrates a case where there is too much noise amplification, and it is important that this figure remains.

2. *The experimental results of the paper show that the real response signal can be extracted from the measurement signal of slow response sensor to eliminate the influence of transmembrane effect, which is in good agreement with the measurement results of DTB sensor. However, the experimental results of the algorithm are introduced in the summary. It is not understood that the correlation R has increased from 0.18 to 0.91. Because the slow response curve is very different from the fast response curve, the correlation between the two must be very low. The correlation between the fast response signal extracted from the slow response signal and the fast response signal measured directly must be very high. It doesn't feel that it can be said to be "improved", nor can it reflect the advantage of this algorithm to obtain the fast response signal;*

The reviewer brings up a good point about the applicability of the Pearson correlation coefficient (or coefficient of determination - the Pearson correlation coefficient squared, i.e. the R^2). The metric of comparison we used (R^2) has, as any such metric, drawbacks (see e.g. Barrett, 1974). One of the drawbacks are its limitations in inferring causality. We have made changes to the manuscript to clarify that the correlation coefficient only indicates that the two time series are more similar (see for instance 1.284). The key requirement for using the R^2 in such an application is that the model is validated beforehand. We validated that the technique work (in principle) via simulations and a controlled laboratory settings. These tests are presented prior to the field application which makes the R^2 applicable in our case. The simulation and controlled laboratory test are documented in sections 2 and 3. We also used the Mean Absolute Error to supplement R^2 as a metric for comparison.

3. *This paper mainly analyzes the influence of time step on the stability of the algorithm. Are there other factors?*

We will address this view under view 4., since we believe these two points are related in both scope and content.

4. *Based on the relevant knowledge of slow response and fast response sensors, is it a good way to directly measure fast response signals? Or is it better to extract from slow response?*

These are indeed very relevant points to address (viewpoint 3 and 4). There are several factors that affect how much temporal resolution can be recovered. The two factors are the decay time, and the measurement noise variance. The slower the sensor, the more difficult it is to recover high temporal resolution fluctuations. The measurement noise also affects how well the deconvolution can be done. The purpose of the L-curve procedure is to determine how good of a time resolution can be obtained. This procedure finds a balance between how noisy the reconstruction is, and how well the reconstruction agrees with the measurements. We have made changes to the manuscript to hopefully explain these factors more clearly for instance in 1.131 where we added in: **"The quality of the solution relies on an appropriate choice of regularization parameter Δt and**

noise/uncertainty in the measurements, but also on the ratio between the RT of the sensor and variance in the property of interest."

225 3 Additional comments to reviewers and the editor

We did a few minor changes and correction of typos we found to increase readability. These are shown in the track changes document. We also did a few minor changes to Figure 1 to clarify and make it fit better with the text.

Additions to author's comment to reviewer 1

230 In our reply to reviewer 1 regarding her/his comment on Eq. 15, we might have made a misinterpretation. The equation is indeed correct, but the comment from the reviewer is relevant since our formulation favors less complex solutions in the L-curve analysis when E_s and E_m is not normalized. We added a sentence about this stating that **"Note that our definition of E_s is not normalized with the number of model points (N) implying a slight favoring of less complex solutions (it is possible to normalize with $N - 1$ and M in E_s and E_m resulting in slightly more complex solutions in the bend of the L"**. We also made a couple of minor changes in a couple of sentences to make it clear that E_s and E_m can be defined in slightly different
235 ways if this is justified (see the marked up document).

References

Barrett, J. P.: The Coefficient of Determination—Some Limitations, *The American Statistician*, 28, 19–20, <https://doi.org/10.1080/00031305.1974.10479056>, 1974.