Response from the authors to the comments by editor Gerrit de Rooij

We would like to thank the editor for taking the time to provide this feedback.

Dear authors,

I sent your revision back to the original reviewers and received one report.

Significant issues remain, but I concur with the reviewer that these are differences of opnion that can be addressed in scientific debate and should not disqualify the paper from publication.

But I do think you should be able to carry this debate over from the discussion (which will be published alongside the paper) to the paper itself. That is, I believe it is worthwhile to address the remaining issues in the paper.

The reviewer indicates some shortcomings in your approach that you can acknowledge without taking away from the main thrust of the paper (if you agree with them). There may be other cases where you disagree where it may well be possible to develop an counterargument in the discussion section or present and assess alternatives to your approach in the Introduction.

From the discussion so far it transpires that you seem to agree with many points raised in the review process, but did not use the revision to make modifications that reflect this. In effect, you are relying on the discussion to address points that can also be addressed directly in the main text.

I therefore ask you to go over the paper once more with this feedback in mind. Ideally, the text of the paper will reflect the outcome of the discussion that will accompany it. Since there were no new issues raised, I believe this should be achievable with minor revisions.

We thank the reviewers and editor for their constructive criticism. We regret that it seemed that we were not open to revising the paper. In fact, as detailed below, this version represents a significant change in the approach. Namely, we are using the more complete forward model rather than the McNeil approximations. As we had expected, the final results are essentially unchanged. But, we agree with the author that using the more complete solution removes any concern about deficiencies of the widely used McNeil equations. To highlight changes made in the text, we have included quotes from the revised paper in this response letter.

The argumentation in the discussion for using the simple model (McNeil) will not be carried over to the manuscript, because we have changed the forward geophysical model to a Maxwell-based full solution

Following added to section 2.1

“This study uses a complete forward model when estimating EC_a to capture the changes in spatial sensitivity introduced by variation in subsurface EC. However, there is no hindrance to use a simpler geophysical model or a model describing a different process.”

Comments from previous review that were not used in the revision to make a direct modification. Some of which are also addressed by changes from this round of replies.

Moreover the authors choose a rather arbitrary selection covering a very broad range of subsurface properties for the forward models. The chosen ECa range is rather high and from the practical point of view many field sites vary by a delta ECa not more than 20 mS/m which would cover only two classes (e.g., van Hebel 2018, McLachlan 2017, Robinet 2018, Reyes 2018).

“The ranges of EC used in the forward model were chosen to represent a wide spectrum of soil types and water contents. This is to capture different scenarios of EMI use e.g., a survey of a large heterogenous area. The lowest EC represents a dry sandy soil and the highest EC represent an agricultural soil with a combination of high clay or water content (Triantafilis and Lesch, 2005; Robinson et al., 2008; Harvey and Morgan, 2009). ”

Given the option of EMagPy it seems to me more convenient, even for an unexperienced user, to run a forward model with several instrument configurations (HCP, VCP, PRP and coil distances) for the specific application with some prior knowledge of texture, salinity etc..

Added to section 3.1:

“Each of the three coil orientations was modelled for three different coil separations and three different instrument heights, the 27 instrument configurations cover both the more typical configurations for field applications of EMI and some more uncommon configuration.”

Added to section 3.3

“This study uses layer EC and thickness as prior knowledge, but any information can be considered to constrain the range of cases.”

Added to section 4.5

“The ML provides a quantitative measure of the shared information among model parameters (Table 2 and Fig. 7) to compare the likely success of each configuration.”

Secondly, you start by stating that using modelling to predict the response of multiple soil models is computationally too challenging (I think that is what you mean in L.80). I don’t think this is the case, particularly not for 1D modelling, as you perform yourself. So, either this point is incompletely made in the manuscript, or it may be (partially) incorrect. For one, simply presenting the sensitivities of the considered coil configurations would already elucidate much of their application potential.

Following

” In this study 27 instrument configurations in combination with 100000 subsurface models is considered the full ensemble. Using the presented ML approach to assess data value for our full ensemble is more efficient than an inverse (Furman et al., 2007; Khodja et al., 2010; Song et al., 2016) or sensitivity (Hanssens et al., 2019) approach. In some cases, evaluating all instrument configuration will not be necessary, which means the inverse or sensitivity approaches become more efficient. The ML approach requires a certain size of model ensemble to yield stable results therefore model run time will reduce efficiency, but this affects the inverse analysis more because it generally requires more model runs. Ultimately the efficiency of the ML, inverse and sensitivity approaches depends, in the EMI case, on model run time, number of layers, parameter boundaries and the number of considered configuration and the combination of these in the applied case will determine which method is more efficient."

You equally do not consider other factors such as (instrumental) noise.

We now consider the effect of 5% gaussian measurement noise and added it as appendix B.

Appendix B The effect of noise on Inferring subsurface parameters and feature importance

To assess the impact of noise 100 realizations of heteroscedastic gaussian noise with a standard deviation of 0.05. The ensemble from the full solution was multiplied by the random noise prior to ML application to the full ensemble (no restrictions). This was repeated for each realization of noise and the average fit and their standard deviation are shown in table B1.

Table B1 The root mean square error (RMSE) between the prediction from the gradient boosted (GB) model and the testing data. The machine learning procedure was repeated with each of the five subsurface parameters as targets, thus creating five models.

Target	ECA	ThickA	ECB	ThickB	ECC
Unit	mS/m	m	mS/m	m	mS/m
RMSE	7.09	0.29	18.8	0.51	2.98
RMSE (noise)	12.4	0.41	23.3	0.58	8.00
Std (noise)	0.05	0.0008	0.05	0.0008	0.02

The average feature importance over the 100 realizations (Fig. B1) affects ThickA, ECB and ThickB the most. Here the feature importance is distributed more evenly among the configurations compared to without noise.

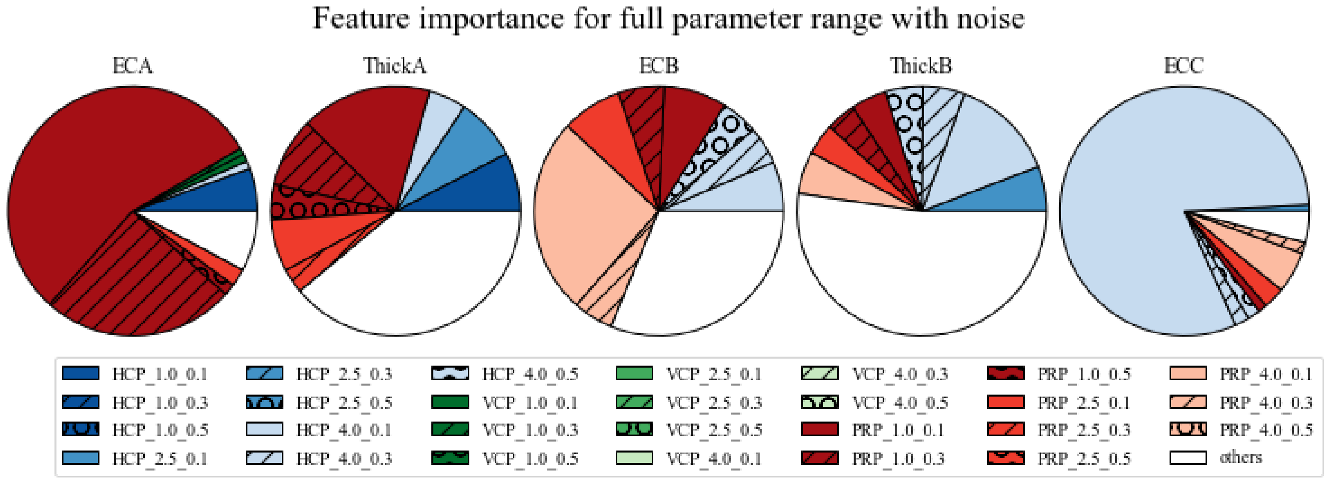

Figure B1 Feature importance for inferring each of the five parameters from a decision tree analysis of the full parameter range. The feature importance from all 27 configurations sum to 1. The eight most important configurations for inferring each

of the five parameters are shown with a unique color and pattern combination. The remaining 19 configurations are aggregated into the "others" category and displayed with white.

"

L.229: *Here, we examine how reducing the uncertainty of one soil EC parameter improves the EMI-based inference of other parameter values and whether this additional information changes the composition of the optimal EMI configurations to include in a survey*

Essentially a sensitivity analysis of your model/EM configuration to the EC and thickness of the respective soil layers you consider, which will be strongly related to the spatial sensitivity of the considered coil geometry.

Added to section 3.3

"Additionally, to the sensitivity of the configurations this analysis provides the parameter values that results in significantly lowered identifiability of any one of the five subsurface parameters"

L.465: *taken together Fig. 7 and 8 provide a direct guide to an EMI user when designing a survey with a specific target*

Again, I think this is a very complicated guide. What you do in the section above is describe the observations you make in your analysis, based on the importance of features in your ML approach. You hereby circumvent discussing the physical basis for this, which lies in the spatial sensitivities of the EM configurations. Your discussion now is very descriptive and data-driven. While there is nothing wrong with this, essentially, I really think you cannot aim to provide practical insight into EM survey strategizing without laying out these fundamental theoretical concepts. This is, for instance, done very clearly by Tabbagh 1986 (see ref. above).

Added to section 4.5

"Designing a combination of optimal configurations based on a conceptual understanding of the spatial sensitivities (rule of thumb) is not a reasonable task. Furthermore, measurement optimization requires a quantitative measure of the information content. The ML provides a quantitative measure of the shared information among model parameters (Table 2 and Fig. 7) to compare the likely success of each configuration."

Response from the authors to the comments by anonymous referee

We would like to thank the referee for taking the time to provide this constructive review.

Dear authors,
many thanks for the detailed responses. I gather from your responses that we disagree on a couple of aspects regarding your methodology and its meaning, but that of course is not a problem. It makes things interesting.

Dear Referee, thank you for finding the discussion interesting.

My main general concerns remain similar to those in the previous review round and are related primarily to [1] the use of an incomplete forward model, and [2] the usefulness of possible alternative approaches to yours by conducting a more basic sensitivity analysis. I stand by my previous review comments that these factors should be addressed. The limitations of your study (which does lean strongly on aspects of EM theory) should be properly stated.

For [1] I do think you should mention (or, ideally do) the following:
- assessment of the impact of an incomplete forward model on a selection of results. Ideally, you would use a more complete forward model to evaluate the limitations of this aspect of your methodology. This is raised by other reviewers, but you ignore this. Either state this very clearly or - ideally - provide a fuller solution approach as a comparison. The fact that 'McNeill is still the most widely used ...' is not a valid argument. It is incomplete and it would serve the use of EM methods better if alternative open-source models (again, as presented in the previous review round) are available.
I strongly disagree with: ' Furthermore, the connection between results from ML analysis and theory becomes simpler and the discussion is accessible to a wider audience.'. How does something become simpler when it has fundamental limitations that strongly impact the interpretative potential of your results, and you do not mention these impacts? You also state that there is no hindrance to use a more complete forward model. If that is the case, why have you not done this?

We agree that it serves EM usage and the community to use more complete forward models if open-source options are available. Both to achieve more realistic modelling results and spread the knowledge of the software that offers the complete solutions.

We have therefore repeated all our modelling (EM and ML) and results with a Maxwell-based full solution that EMagPy offers. These results (tables and figures) replace the previous ones where we used the simple CS model (rather than accompany), because we want the focus to be on the suggested approach itself. For our cases the outcome did change slightly, but the overall picture remained.

Therefore, the theory section is now:

“We apply the Maxwell-based full solutions (eq. 1, 2 and 3) from Wait (1982) to calculate the relationship Q between the secondary field (H_s) and the primary field (H_p). The solution works for a one-dimensional subsurface and it is valid for low frequencies because it assumes that the electromagnetic fields spread due to conduction currents:

$$Q_{VCP} = Im\left(\frac{H_S}{H_P}\right)_{VCP} = Im\left(-s^2 \int_0^\infty R_0 J_1(s\lambda)\lambda d\lambda\right), \tag{1}$$

$$Q_{HCP} = Im\left(\frac{H_S}{H_P}\right)_{HCP} = Im\left(-s^3 \int_0^\infty R_0 J_0(s\lambda)\lambda^2 d\lambda\right), \tag{2}$$

$$Q_{PRP} = Im\left(\frac{H_S}{H_P}\right)_{PRP} = Im\left(-s^3 \int_0^\infty R_0 J_1(s\lambda)\lambda^2 d\lambda\right), \tag{3}$$

Where Im means that only the imaginary component is considered, R_0 is an interlayer reflection factor, J_0 and J_1 are Bessel functions of respectively zeroth and first orders and λ is the radial wave number. The integrals of eq. 1, 2 and 3 represent Hankel transform and in the EMagPy software (McLachlan et al., 2020) these are calculated with linear filtering (Anderson, 1979; Guptasarma & Singh, 1997). The LIN approximation proposed by McNeil (1980) assumes that depth of investigation does not depend on the EC of the subsurface. Therefore, a method similar to that of von Hebel et al. (2019) is used through EMagPy (McLachlan et al., 2020) to estimate EC_a from Q. The EC_a is estimated by minimizing the differences between a predicted or measured Q_{pred} and a Q value calculated for an equal homogenous half-space, Q_{homo}. The minimized difference approach is valid for a broader range of EC_a compared to the LIN approximation (Von Hebel et al., 2019; McLachlan et al., 2020). We refer to Von Hebel et al. (2019) for a more detailed description of this method."

To the left are the previous McNeil based figures and tables and to the right is the full solution based figures. The differences are mainly revealed through the feature importance when inferring the ECA, ECB and ECC (fig 5 and 8).

Table 2: The root mean square error (RMSE) between the prediction from the gradient boosted (GB) model and the testing data. The machine learning procedure was repeated with each of the five subsurface parameters as targets, thus creating five models. The RMSE is normalized by the mean value of the target to get the normalized root mean square error (NRMSE).

McNeil						Full solution					
Target	ECA	ThickA	ECB	ThickB	ECC	Target	ECA	ThickA	ECB	ThickB	ECC
Unit	mS/m	m	mS/m	m	mS/m	Unit	mS/m	m	mS/m	m	mS/m
RMSE	7.34	0.29	18.7	0.49	1.51	RMSE	7.09	0.29	18.8	0.51	2.98
NRMSE	0.07	0.20	0.19	0.26	0.02	NRMSE	0.07	0.20	0.19	0.27	0.03

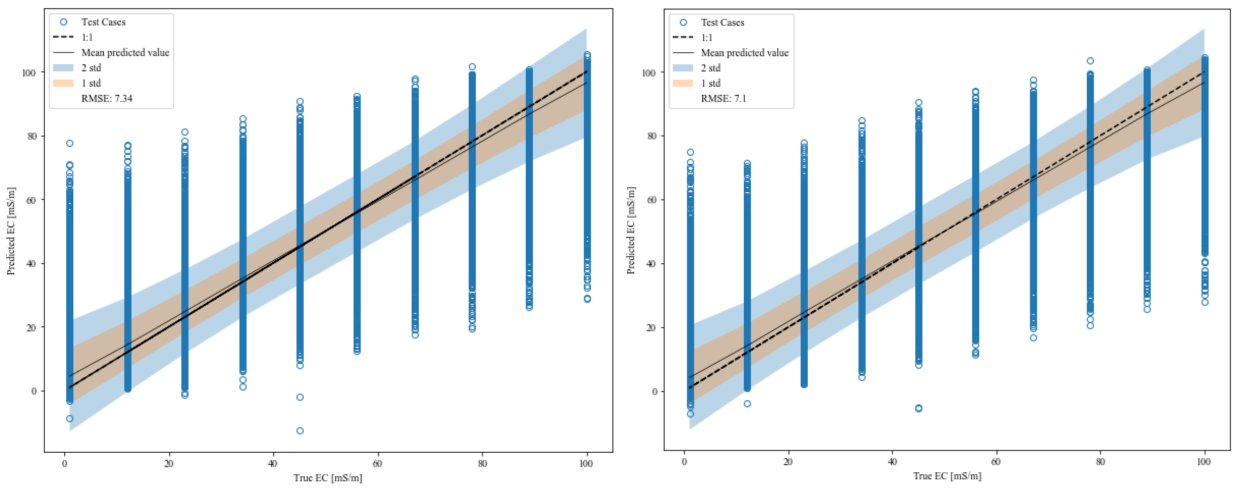

Figure 1: The result from running the DT with GB on the entire 100000 soil types and all 27 instrument configurations five times. The EC of the A-layer (ECA) is the parameter that is being predicted. The X-axis is the true value of the ECA, and the Y-axis is the predicted values for ECA.

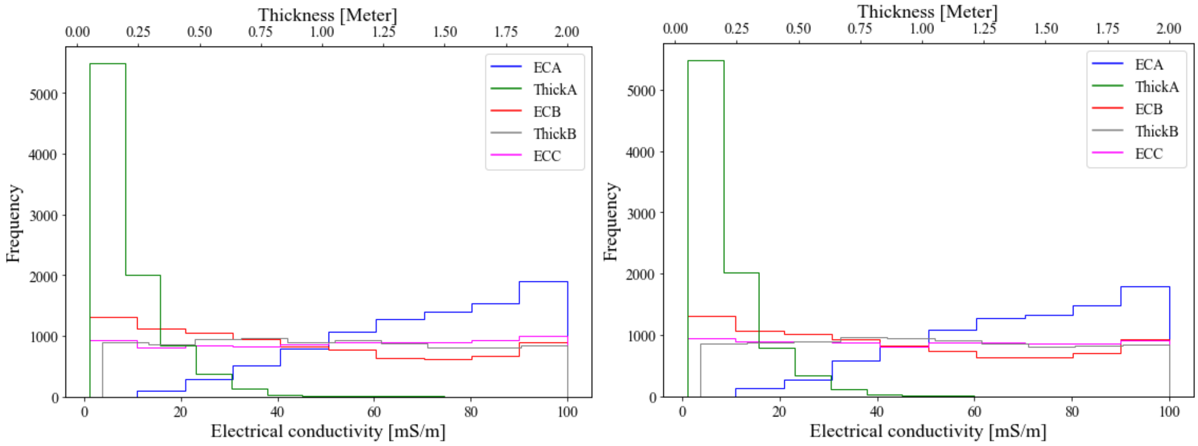

Figure 2: The ECA was inferred for 150,000 test cases. In 8894/8816 of the 150,000 cases the inference was more than one standard deviation away from the true value. The figure shows the distribution of five subsurface parameter values within the 8894 conditions. The top X-axis is the layer thickness, the bottom X-axis is the layer EC and the Y-axis is the frequency.

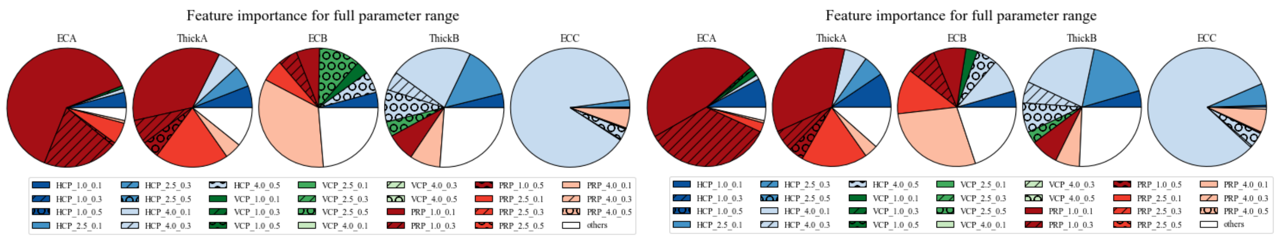

Figure 3: Feature importance for inferring each of the five parameters from a decision tree analysis of the full parameter range. The feature importance from all 27 configurations sum to 1. The eight most important configurations for inferring each of the five parameters are shown with a unique color and pattern combination. The remaining 19 configurations are aggregated into the “others” category and displayed with white.

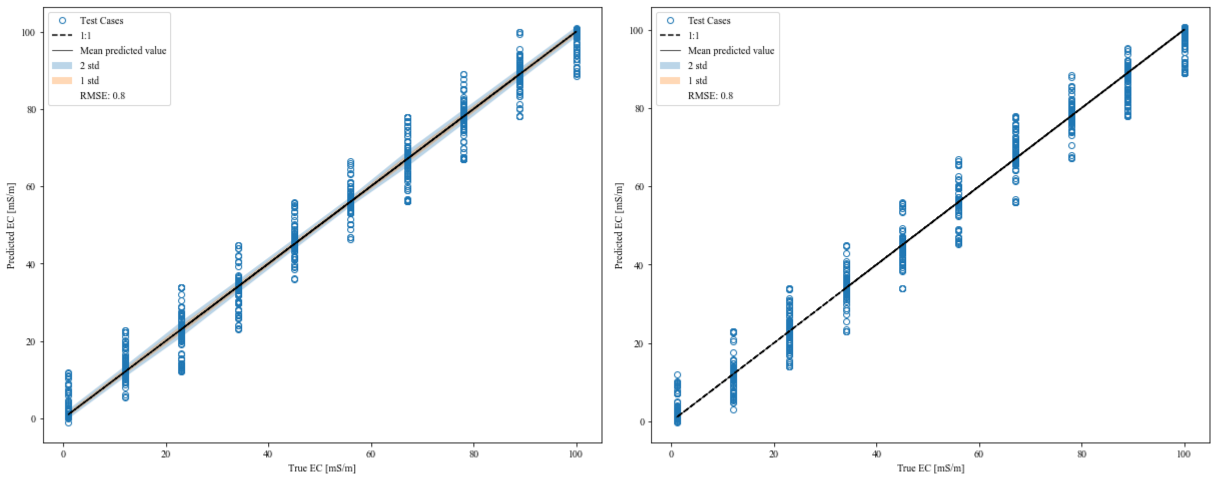

Figure 4: The result from running the machine learning algorithm on a subset of the ensemble where the thickness of the A-layer have been restricted. Only 20,000 soil types and all 27 instrument configurations remain in this restricted subset. The EC of the A-layer (ECA) is the parameter that is being predicted.

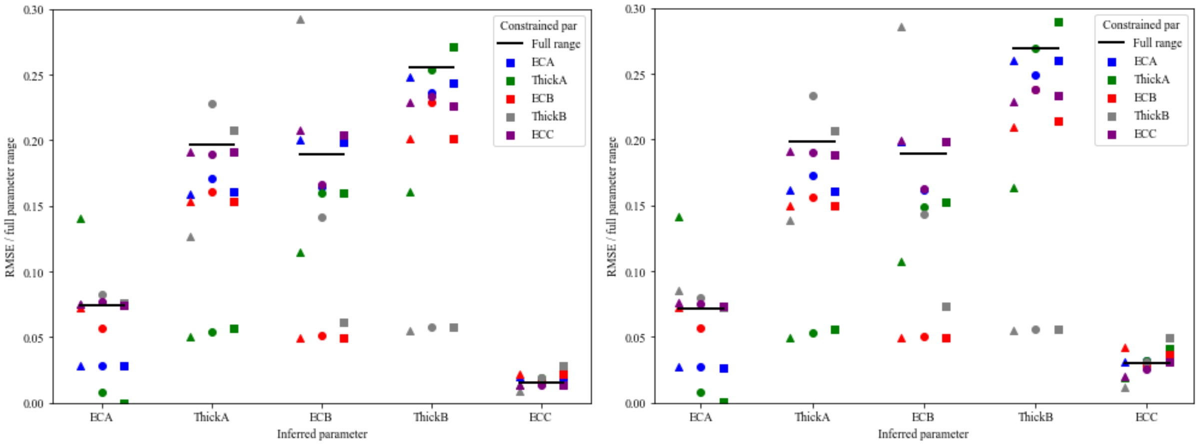

Figure 5: The changes in inference of the five subsurface parameters (X-axis) are based on a comparison between the RMSE from restricted case divided by the range of the parameter (Y-axis). The lines show how well the parameters are predicted when all parameters are full range. The color shows which parameter that is being represented and the location and symbol represents the three restriction patterns skewed low (left nudged triangle), centered (centered dot), skewed high (right nudged square).

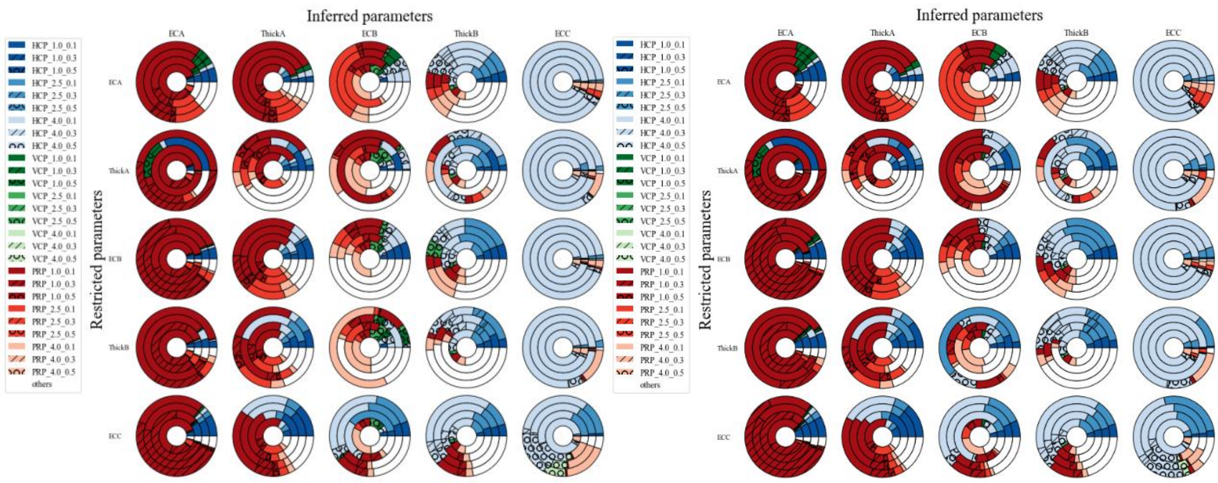

Figure 6: Feature importance for the 8 most important EMI configurations for every combination of the five inferred/restricted parameters and the three patterns. Each circle is subdivided into four rings that shows, from inside out, the feature importance for full range, centered, skew low, and skew high. Each column/row represents the each of the five inferred/restricted parameters. The coil orientations are colored so that Horizontal (HCP) is blue, Vertical (VCP) is green, and Perpendicular (PRP) is red. A dark and light hue represents respectively a short and long coil distance.

For [2] i do think that in many cases specific knowledge on the survey area etc. is available or can be gathered. Uncertainties (which is, I assume, what you in your responses dub as bias) can be integrated in such procedures to a certain extent (and are always inherent to geophysical applications/modelling). The argument that conducting an analysis that includes all possible EM instruments is only partly valid. Usually (always?) users evaluate a limited set of available instruments/configurations so, while absolutely useful and key in your manuscript, the evaluating the full suite of instruments is – in my opinion – not a standard requirement.

We have added the following to the discussion (section 4.5)

” In this study 27 instrument configurations in combination with 100000 subsurface models is considered the full ensemble. Using the presented ML approach to assess data value for our full ensemble is more efficient than an inverse (Furman et al., 2007; Khodja et al., 2010; Song et al., 2016) or sensitivity (Hanssens et al., 2019) approach. In some cases, evaluating all instrument configuration will not be necessary, which means the inverse or sensitivity approaches become more efficient. The ML approach requires a certain size of model ensemble to yield stable results therefore model run time will reduce efficiency, but this affects the inverse analysis more because it generally requires more model runs. Ultimately the efficiency of the ML, inverse and sensitivity approaches depends, in the EMI case, on model run time, number of layers, parameter boundaries and the number of considered configuration and the combination of these in the applied case will determine which method is more efficient. Designing a combination of optimal configurations based on a conceptual understanding of the spatial sensitivities (rule of thumb) is not a reasonable task. Furthermore, measurement optimization requires a quantitative measure of the information content. The ML provides a quantitative measure of the shared information among model parameters (Table 2 and Fig. 7) to compare the likely success of each configuration”

Furthermore (dotting the I’s), you do not fully integrate the instrument specifications (e.g. frequency and, more importantly, noise level). As an argument you state that you present a proof of concept of a novel use of ML analysis for measurement network optimization. That is only partly true, for your main manuscript, the balance between focus on EM and ML is about 50/50, so I believe you cannot use this argument as a rationale for using simplifications and omitting factors such as noise. At the very least, you should make this rationale – and the associated limitations – explicit in your manuscript.

We assessed the effect of noise. These results are added to appendix B in the manuscript. With a reference in section 4.2

“a table of how gaussian noise affect the RMSE is shown in appendix B.”

And section 4.3

“The influence of simulated noise on the results in Fig. 5 are shown in appendix B.”

“Appendix B The effect of noise on Inferring subsurface parameters and feature importance

To assess the impact of noise 100 realizations of heteroscedastic gaussian noise with a standard deviation of 0.05. The ensemble from the full solution was multiplied by the random noise prior to ML application to the full ensemble (no restrictions). This was repeated for each realization of noise and the average fit and their standard deviation are shown in table B1.

Table B1 The root mean square error (RMSE) between the prediction from the gradient boosted (GB) model and the testing data. The machine learning procedure was repeated with each of the five subsurface parameters as targets, thus creating five models.

Target	ECA	ThickA	ECB	ThickB	ECC
Unit	mS/m	m	mS/m	m	mS/m
RMSE	7.09	0.29	18.8	0.51	2.98
RMSE (noise)	12.4	0.41	23.3	0.58	8.00
Std (noise)	0.05	0.0008	0.05	0.0008	0.02

The average feature importance over the 100 realizations (Fig. B1) affects ThickA, ECB and ThickB the most. Here the feature importance is distributed more evenly among the configurations compared to without noise.

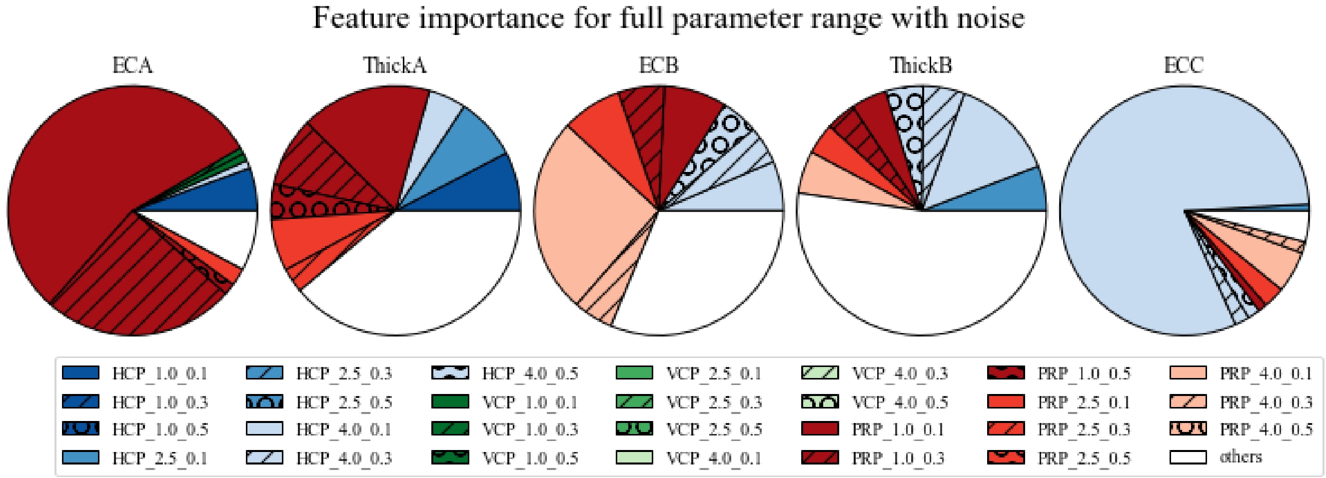

Figure B1 Feature importance for inferring each of the five parameters from a decision tree analysis of the full parameter range. The feature importance from all 27 configurations sum to 1. The eight most important configurations for inferring each of the five parameters are shown with a unique color and pattern combination. The remaining 19 configurations are aggregated into the "others" category and displayed with white.

"

I hope this helps finalize the manuscript.

It did indeed, thanks again for the feedback.