

Response from the authors to the comments by anonymous referee

We would like to thank the referee for providing constructive review and commentary.

Dear Authors,

Many thanks for your original contribution. Below you find my main comments and suggestions, which I hope will help you finalise your paper.

I have structured my comments per manuscript section, after my general remarks. References to specific parts of the text are made with line numbers (L.XX). In essence, I think the paper presents an interesting approach, but in certain aspects argumentation for specific hypotheses, decisions and conclusions is lacking or incomplete. This lack of (complete) argumentation for these aspects throughout the manuscript is why I have indicated the scientific significance of the manuscript in its current form as 'poor'. As such, I believe major revisions are required. In my opinion, these should address the general comments presented below.

Thank you for finding the presented approach interesting.

GENERAL COMMENTS

You propose a method to enable more efficient EM survey strategizing, mainly aimed at non-expert users.

You hereby start with the premise that the current means to determine the optimal EM instrument configuration, defined in your paper as a combination of coil configuration (geometry and spacing) and instrument height, are insufficient. However, it is unclear what you see as those current means (see for instance L.65-66 of the introduction)? It is therefore difficult to evaluate to which types of approaches you (want to) compare your approach. You equally do not define or specify the 'rules of thumb' for the application of EM instrumentation, again making it impossible to fully understand what you mean by this.

We now explicitly state the commonly applied rules of thumb within the introduction (see detailed comments).

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.

During optimization of measurement campaign, it is required to consider the range of possible designs and the degree of uncertainty in the conditions being surveyed.

In the simplest case, a researcher may consider all but one property to be well defined and consequently only having one adjustable parameter. Furthermore, they may be choosing among several single instrument configurations. For these cases, a researcher can conduct a single-parameter sensitivity analysis and the coil configuration with the highest sensitivity can be selected.

Optimization becomes more difficult if the campaign includes the combination of multiple coil configurations. For these conditions, the shared information of all sensors must be considered. A researcher can still use a sensitivity analysis, but each combination of sensors must be considered, and it is not clear how the sensitivities of multiple coil configurations should be combined.

It is more appropriate to conduct an inverse analysis with each combination of sensors and to use this to infer the combined information in each measurement set. Further complication arises when more than one parameter value is unknown. For these conditions, it is no longer appropriate to conduct a single parameter sensitivity analysis assuming all but the parameter of interest are known. The number of sensitivity analyses (with an associated inverse model for each) increases geometrically.

Consider the following conditions. We can use up to 27 different coil configurations. There are five parameters that characterize the system to be surveyed. Each of these parameters are represented by a range of 10 different possible values. The forward model is extremely fast (0.01 seconds) and the inverse model is fast (1 second).

Considering a single sensor and a single parameter, with all others known, only 270 forward models need to be run for a total run time of 2.7 seconds.

If we consider a pair of coils, there are 351 combinations of coil configurations, each of which requires 10 forward models and 1 inverse model. This is already a total of 386 seconds.

In a study such as that shown here, we consider 27 sensors which may be taken 1, 2, 3, 4, or 5 at a time and all five parameters can vary. To determine the optimal set over this range of conditions would require consideration of 10^5 possible parameter combinations and $(27 + 351 + 2925 + 17550 + 80730 = 101583)$ possible sensor sets.

Each parameter set requires one forward model (1000 seconds) and each sensor set requires one inverse model (10158 seconds). Therefore, taking a typical sensitivity analysis approach would require approximately 3 hours of simulation time.

In contrast, the approach described here requires the same number of forward models, the ML replaces the individual inverse models and only requires 60 min, and the underlying structure of an ML ensures that the solution balances goodness of fit with generalizability.

Then, if a user wants to know how information about one parameter might influence the survey design, they would down sample the forward models and, following a traditional approach as suggested, they would have to repeat all of the inverse models. That is, each examination of any set of existing information on the survey design would require hours of simulation time. In contrast, the ML only requires retraining on the reduced set of forward models which requires a few minutes.

It should be emphasized that the availability of EMagPy, and the nature of the EM problem, lead to very fast forward and inverse analyses. Many design problems require far more effort – especially if a more

complex forward model is chosen. The run time of the forward model is particularly limiting on traditional inverse analyses, which requires many forward models to be run during the inverse process. In contrast, the ML only requires that each forward model be run once, up front, and then the ML can be trained with a training time that is independent of the forward model run time.

You deploy machine learning to predict the optimal combination of coil configurations for targeting one of five subsurface parameters (EC + thickness of two layers, plus EC of a third layer with thickness set to infinity). Essentially, what you are doing is evaluating how sensitive the evaluated (27) instrument configurations are to each of these parameters. Or, more correctly, how sensitive the deployed forward model of those configurations is to these. Here, I do not fully see the difference between the machine learning approach you take towards this issue, and a simpler sensitivity analysis (e.g., Monte-Carlo based)? The latter, in my opinion, has at least two advantages: it is simpler (i.e., it is a straightforward, robust way to evaluate the influence of parameters on a model outcome), and it would be more straightforward to visualize.

As discussed above, the ML approach is fundamentally different – and we claim, more efficient – than a traditional inverse modeling approach to assessing data worth. First, an ML is trained specifically to balance goodness of fit with generalizability – this is not a common feature of inverse analyses. Second, the tree-based MLs used here conduct a data-worth analysis at each step as part of the training. This is imperfect, of course, but it naturally identifies data that do not contribute to the training and quantifies the contribution of the important inputs. In contrast, a sensitivity analysis (or, more accurately, and inverse analysis) would have to be repeated multiple times for subsets of the data to determine which combinations of observations are most informative. The key element of this work is the recognition that the feature importance provides a rank order of data worth that is produced without extra effort as the training seeks to improve the goodness of fit while avoiding overfitting. The result is that we can provide a very efficient calculation of the information content of different coil configurations for a user-defined range of site conditions. In this context, we think that a case can be made that this has clear advantages over a much more computationally expensive inverse analysis of data worth to support survey design.

Next, you deploy a forward modelling procedure, which you do not describe in detail. I think you use the (so-called) McNeil approximation, but you do not state this explicitly? You implement the modelling through EMagPy, but, again, without providing details on the model you use. This makes it difficult to evaluate the outcomes of your procedure (though I think you use McNeil, and have evaluated the following as such). If you use an approximation that is only valid within specific conditions (low induction condition), you essentially use a simplified (albeit elaborate) rule of thumb?

We appreciate the reviewer's point – we relied too heavily on the publication describing EMagPy. In the revised version we explicitly state that we use the McNeil approximation and provide a description so that the reader is not required to read the EMagPy paper.

“Then, the EC_a was calculated for many EMI instrument configurations through EMagPy (Mclachlan et al., 2020) version 1.1.0. EMagPy deployed the CS response functions from eq. 1, 2 (McNeil, 1980), and 3 (Wait,

1962) in combination with the summation of eq. 4, which assumes that the LIN approximations (McNeil, 1980) are valid.”

In your modelling procedure: you only consider quite a narrow range of EC variations (0-100 to meet – generally speaking – the LIN condition). This effectively limits the application potential of your approach (but would only imply deploying a forward model integrating the full solution – e.g. Hanssens et al. 2019 <https://doi.org/10.1109/MGRS.2018.2881767>). You equally do not consider other factors such as (instrumental) noise.

The EC range was chosen to both approximate the LIN condition and still cover the EC range of a large portion of agricultural fields. (Interestingly, the other reviewer commented that the EC range was too wide!) However, it should be noted that our approach is not reliant on using the forward model selected. A more complete forward model, such as that available in EMagPy, could be used or a user could even link to their own even more complete forward model, if it was necessary for their application. There would be no change to our approach other than the increased forward model run time. We chose the McNeil approximation because the model is widely used in the interpretation of EMI data and we believe that data collection and analysis should be linked. (That is, the data collection should be designed, to the degree possible, with consideration of the analyses that will be applied to the data.) Furthermore, we wanted to connect the results from ML with underlying concepts that would be accessible to a wide audience. We felt that this discussion would be easier to follow if we used the simpler McNeil solution rather than add considerations of changes in spatial sensitivity as a function of the EC structure.

The reviewer is correct that our analyses did not consider measurement noise. This would be an important extension of the work presented here, which is intended to be a proof of concept of a novel use of ML analysis for measurement network optimization.

DETAILED COMMENTS PER SECTION

ABSTRACT

L.14: *There are general, rule-of-thumb guides to choose an optimal instrument configuration for a specific survey*

While I understand this is not elaborated on in the abstract, you should explain which ones you mean and what the possible advantages/shortcomings are.

It is now explicitly stated which rules of thumb we refer to in the introduction (see detailed comments below).

L.15: *The goal of this study was to use machine learning (ML) to improve this design optimization task*

I assume the goal is to provide a robust, efficient way to strategize EM surveys. ML is not a goal, it is a tool.

The line is changed to:

“The goal of this study is to provide a robust and efficient way to design this optimization task.”

INTRODUCTION

L.47: *combined current is measured with a receiver coil*

Magnetic field (cf. the following sentence)

The line now reads:

“combined magnetic field is measured with a receiver coil”

L.55: *Finally, in some cases, the spatial sensitivity may depend on the absolute value and spatial distribution of the EC (Callegary et al., 2012).*

What do you mean, in some cases?

The dependency of spatial sensitivity on the absolute value and spatial distribution of the EC are small when working at low frequency and low EC (LIN condition). The cases we refer to are the ones outside of the LIN conditions. Changed the wording to:

“Finally, in some cases, the spatial sensitivity may have a higher dependency on the absolute value and spatial distribution of the EC (Callegary et al., 2012).”

L.57: *In this investigation, we make the common assumption that the spatial sensitivity only depends on the instrument configuration, but this dependence could be considered using more complete forward models of EMI response.*

Why not use a fwd model that integrates the other relevant aspects (see general comments)

We kept this first situation simple to be able to connect the result of the machine learning analysis to the physical concepts and be able to present to the discussion to a wide audience (see general reply).

L.64: *Developers of EMI instruments have long recommended using different configurations to measure layered ECa values, leading to simple rules of thumb such as*

using shorter coil separations for shallow mapping and larger separations for deeper investigations.

What are these rules of thumb you refer to? Make these explicit.

This is a good point. The rule of thumb we refer to is derived from the 70% cumulative response in LIN condition. Where 70% of the total response for VCP, HCP and PRP coils are accumulated from respectively 0.75, 1.5 and 0.5 coil separations of depths. In practice this leads to general rule of thumbs such as use a of short coil separation or VCP/HCP for shallow survey and larger separation or HCP for deeper survey. This is

We have made this more explicit in the introduction with the following:

“The depth of investigation (DOI) of EMI instruments is both in the scientific literature (Saey et al., 2009a; Saey et al., 2009b; Saey et al., 2012; De Smedt et al., 2014; Doolittle & Brevik, 2014; Adamchuk et al., 2015) and by the manufacturers (Duaem Inc., Canada n.d.) often estimated to be at the depth the has 70% of the cumulative response. There is a relationship between depth sensitivity of the instrument response and coil spacing and position. Therefore 70% cumulative response rule is in practice frequently converted to a rule of thumb that states larger coil spacings and HCP should be used for deeper investigations while short spacing and VCP/PRP should be used for shallow investigation (Acworth, 1999; Beamish, 2011; Cockx et al., 2009; K Heil & Schmidhalter, 2015; Kurt Heil & Schmidhalter, 2019). While this rule of thumb is not wrong the terms shallow and deep are subjective and will have different meaning depending on whether it is a hydrogeologist, archeologist, agronomist or a geophysicist who applies the terms. It also fails to make any distinction to the differences between using the VCP or PRP coil orientation.”

L.65: *But little specific guidance is offered.*

What do you mean? I see:

- a/2 rule
- 70% cumulative response in LIN conditions (McNeill)
- Forward modelling

It is true that the 70% cumulative response rule provides general user guidance. It fails to provide specific information about what subsurface conditions would cause a short VCP coil to be better than a slightly longer PRP coil or an even shorter HCP coil.

Forward modelling will be able to calculate the instrument responses from different instrument configurations and subsurface layering it needs to be combined with another tool to rank the many combinations of configurations for different subsurface layering. Such as what we do with machine learning.

L.66: *Furthermore, there is no way for a user to consider the possible impact of prior knowledge (e.g. bounds on the expected depth of the topmost layer) in the survey design.*

Unless I am missing something here, this is not true. Forward modelling can easily provide this information.

Some kind of tool need to join the forward modelling procedure to analyze the vast number of possible outcomes. E.g. Monte Carlo (As you suggested) or ML that we deploy.

L.72: *for t users*

Users or the users.

Corrected to “for users”

L.72: *This makes it difficult for users without theoretical background in geophysics to make an informed choice regarding the preferred instrument and configuration*

This is a subjective statement: what do you mean by the theoretical background in geophysics required to deploy EM instruments? One could say that without the necessary understanding of basic theoretical concepts, you can never critically deploy this instrumentation. (I am aware that in practice this is not necessarily the case for all users)

It is our impression (based on personal experience) that a lot of users are guided by the rule of thumb that we describe, and it is very few that are aware of the underlying concepts, (apart from geophysicist).

We have changed the line to be:

“This makes it difficult to make an informed choice regarding the preferred instrument and configuration”

L.75: *Each survey design includes multiple measurements at each location, each with a different configuration, that jointly provide the most useful information for inferring specific, user-identified subsurface properties.*

survey or survey design?

Survey design

L.76: *That is, a user is faced with the question of which combination of configurations is optimal given their measurement priorities and, ideally, incorporating any applicable constraints that they may have regarding the subsurface conditions. Any method that requires formal inversion of each proposed combination of configurations is computationally intractable for most users.*

What do you mean by this? Why inversion with survey design?

[See general reply](#)

L.82 onwards: this is methodology, isn't it?

[Yes, indeed, thanks for pointing this out. Majority of this section is moved to method section 3.2 that describes machine learning.](#)

[We kept some of the section within the introduction because we believe it is an important concept for this study.](#)

[“Machine Learning \(ML\) describes a wide range of regression algorithms used for pattern recognition. ML has grown in popularity and is now used regularly within and beyond science. The simplest ML tools are based on Decision Trees \(DT\), which are supervised ML techniques that perform classification or regression by sequential categorization based on observations. DTs are computationally inexpensive, but they can have limited predictive skill \(Hastie et al., 2001\). To improve their performance, DTs are often augmented by ensemble learning methods such as bagging \(Breiman, 1996\) and boosting \(Friedman, 2001\).”](#)

L.77: Feature importance key ability of DTs (with and without GB), which is a functions that quantify the importance of each feature for making the predictions of interest.

L.98: *without having to do multiple inverse models*

Forward models?

[Changed the sentence to: “This allows us to find the optimal instrument configurations for each subsurface parameter without having to do inverse modelling.”](#)

THEORY

L.113: *It is more common, especially on agricultural soils*

For (almost) all subsurface media

[Changed line to “In almost all subsurface media the EC varies with depth due to soil layering.”](#)

L.119: *low induction number*

Explain (and reference)

[The section has been changed to:](#)

“The model only strictly applies under low induction number (LIN) conditions. The LIN approximation proposed by McNeil (1980) and assumes that changes in the measuring frequency has no effect on the response and that the depth of investigation does not depend on the EC of the subsurface. Assuming LIN conditions therefore means the response depends only on the depth, coil separation length, and coil configuration with no regard for the subsurface EC distribution.”

L.120: *with no regard for the subsurface EC distribution.*

It would be good to mention that, generally, the output of commercially available EM instruments makes use of this approximation. The ‘no regard for the subsurface EC distribution’ is inherent to the ECa value, as you present yourself in the preceding paragraphs.

Added the line

“It is a common assumption for commercially available EMI instruments to operate under LIN conditions, despite being a simplification.”

L.125: eq. 3: if I’m not mistaken, equation for PRP is based on Wait 1962, not McNeill (who only presents response functions for coplanar configurations)

Modified the beginning of the paragraph to include Wait 1962:

“The simplest, most widely used depth sensitivity model is the Cumulative Sensitivity (CS) model of McNeill (1980) (eq. 1 and 2) and Wait (1962) (eq. 3).”

L.136: *EMagPy (McLachlan et al., 2020) offers the user the opportunity to use several models and makes them readily available to a wide audience, even users with no background in EMI modelling.*

What do you mean by this? The fact that it incorporates a GUI?

Geophysical software licenses can easily be priced at several thousand euros. The fact that EMagPy is free makes it more available and appealing to someone who is not a specialist. Effectively increasing the number of people who can be participants in the geophysical community.

METHODOLOGY

L.150: using EMagPy

Ok, you mention the python package you use, but you should elucidate (and reference) the deployed forward model

Added a reference to equations 1, 2, 3 and 4 to the sentence:

“Then, the EC_a was calculated for many EMI instrument configurations through EMagPy (McLachlan et al., 2020) version 1.1.0, using the CS response functions from eq. 1, 2, and 3 in combination with the summation of eq. 4”

L.158: *The lowest EC represents a dry sandy soil and the highest EC represent an agricultural soil with a combination of high clay, salinity, or water content*

For a max. EC of 100, you cannot say that you evaluate the influence of salinity

Thank you for noting this we removed reference to salinity in the paragraph:

“The ranges of EC used in the forward model were chosen to represent a wide spectrum of soil types and water contents. 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).”

L.169: *from thin (0.05 m) to relatively thick (2.0 m) ...*

Just state ‘from 0.05 m to 2.0 m thickness’.

Changed the sentence to:

“The ranges of soil layer thicknesses span from 0.05 m to 2.0 m thickness.”

L.164: *Note that all analyses were repeated for the Andrade (2016) EMI model.*

What do you mean by this? Explain the ‘Andrade model’.

We removed the sentence at line 164:

“Note that all analyses were repeated for the Andrade (2016) EMI model “

L.165: *The findings were not significantly different, so the results are presented for the simpler, more widely used McNeil model.*

Because you stay within LIN conditions. I expect the difference to be most important (within LIN) for the PRP configurations?

Also, these are results, not methods

We removed the sentence at line 165:

“The findings were not significantly different, so the results are presented for the simpler, more widely used McNeil model.”

L.179: *x is the inputs (features) and y is the response*

Rephrase

We modified the sentence to

"A training data set consists of n samples $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, where $x_{1:n}$ are the inputs (features) and $y_{1:n}$ are the corresponding outputs (targets)."

L.198: *gradient from which the algorithm named*

Rephrase

We modified the sentence to:

"Right side of the minus sign in equation 6. is the gradient from which the algorithm is named, and the residual r_{im} are named pseudo-residuals."

L.221: *optimal values for these parameters were found to be 0.1, 10, and 2, respectively*

How?

We changed the sentence to be

"The learning rate, maximum tree depth, and minimum samples per leaf were tuned by manual trial and error and the optimal values for these parameters were found to be 0.1, 10, and 2, respectively."

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.

Yes, this is a form of sensitivity analysis that looks at the boundaries for each parameter with the added bonus that it gives an estimate of the identifiability of the parameters given a specific soil.

RESULTS & DISCUSSION

L.255: *The variations are less pronounced for larger coil separations.*

as you would expect cf. spatial sensitivity of these geometries.

We agree that this is an expected outcome based on the spatial sensitivities.

L.256: *differences in the smoothness of the distributions*

I assume these are related to the EC of the upper soil layers? It is difficult to evaluate your results, as it is unclear which forward model you deploy. Is this just the 'McNeil-approximation'?

We added additional reference in the beginning of section 4 to make it more clear which forward models we deploy:

"In this section, we present the outcome from the forward modelling with the CS models for VCP (eq. 1), HCP (eq. 2) and PRP (eq. 3) and the summation from eq. 4 (section 4.1)."

L.301: *The finding is opposite for ECA*

if could it be the deployed forward model (approximation) strongly influences this as well? Furthermore, as this is (I think) still based on all 27 instrument configurations, this will have a significant influence as well. One would assume the poorly inferred cases are more likely related to configurations with a larger coil spacing?

These are the distribution of the parameter values from the outlier cases. Therefore, each case will indeed contain a response from each of the 27 configurations.

L.303: *this suggests that the method would be more likely to be successful*

Which method? Your approach?

Modified the sentence to:

"Practically, this suggests that identifying layer with an EMI instrument would be more likely to be successful"

L.305: *A more successful survey, based on the ability to infer ECA, would occur if the ECA values tend to be lower. That is, a center or low skewed restriction should show better performance*

Again: influence of the forward model?

The nature of the forward model will of course influence the outcome of the approach. But we want to verify the approach on a simple model before extending it to more complex forward models (see earlier comments).

L.315: *balances performance with reduced field effort*

What do you mean by this? You should clarify this aim in your introduction

Improving the quality of data or performance of the models that require said data while increasing the efficiency of field/lab work is an intrinsic part of optimizing experimental designs.

We have made this clearer in the introduction:

“One of the challenges of environmental investigations is to determine the optimal data to acquire. Data which is often used to provide structural information to a model or constrain model parameterization. Measurement optimization is an attempt to balance data quality and the work expended in the field and laboratory. The ultimate goal of was to develop an approach to measurement optimization that would be accessible to a wide range of users, with the hope that a similar approach could be developed for other measurement network design problems. The specific objective of this investigation was to present an approach to select sets of EMI configurations that are optimal given the specific survey goals and any independent knowledge of the subsurface electrical properties.”

L.319: circle

Corrected the typo

L.323: *However, he did not consider the PRP orientations*

Tabbagh (1986 – doi: <https://doi.org/10.1111/j.1475-4754.1986.tb00386.x>) did.

We are sorry we did not include that reference, very relevant (thanks). Reference to this is now included in section 4.3.

L.325: *To our knowledge, no other method, short of exhaustive comparisons of many synthetic inverse analyses, would have been able to show that a single configuration was so clearly dominant for inferring ECC.*

I disagree. Evaluating the QP sensitivity of a specific coil configuration to perturbing EC can be evaluated in a quite straightforward manner (see, for instance Hanssens et al. 2019 – doi: <https://doi.org/10.1109/MGRS.2018.2881767>)

We were not aware of the work by Hanssens et al. 2019. Thank you for bringing this interesting study to our attention. They too use a “*brute force method*” of calculating the sensitivities with each forward model being resolved multiple times based on the number of layers. Conducting a global sensitivity analysis (using all the soil) would be exhaustive.

L.325: *The small coil separation and low instrument height fit with general expectations, but the PRP orientation was not expected before conducting this analysis*

Why not? And, conversely, why were you expecting the VCP/HCP to outperform PRP? Provide the full argumentation.

We did not expect VCP/HCP to necessarily outperform PRP but expected a more equal performance between the VCP and PRP sensor.

L.335: *Perhaps more controversially, in the context of EMI instrument design and use, only 26% of the most informative configurations used the VCP orientation ...*

Why is this controversial?

The EM38 sensors make use of the VCP and HCP configurations and has a very widespread use. It is the most widely used EMI instrument in agriculture according to (Heil & Schmidhalter, 2017). It is counterintuitive that the most widely used instrument uses the least sensitive coil (VCP) rather than PRP. This also suggests that there is a gap between the community of EMI specialists and a large portion of end users.

L.339: *This may be partially explained by the spatial sensitivities of the orientations*

Why only partially? What you are doing is essentially evaluating the applicability of geometries/configurations with specific spatial sensitivities.

Changed the sentence to:

“This may be explained by the spatial sensitivities of the orientations”

L.341: *high spatial sensitivity redundancy for the HCP and VCP*

Why redundancy? You mean that these are not very complementary?

Based on the analysis we believe that the HCP/PRP pairing are more complementary relatively to the HCP/VCP pairing.

section 4.4 Parameter restriction analyses .

This may be a consequence of an incomplete understanding I may have on specific aspects of your ML (and your overall study aim), but I don't understand the point of this aspect. What will happen is that the uncertainty of the outcome will be reduced based on how sensitive your EM configuration (FWD model is) to a specific parameter. So, based on the previous section, you would expect that fixing the properties (EC and thickness) of the

first model layer (the most shallow layer) will have the strongest influence for most coil configurations.

This is essentially what you present in 4.4.2 (and emphasise in L.389: *The only clear exception was inferring ECA, which showed a greater improvement by restricting ThickA with a central or right skew*)

What we find is that reducing the range of thickness of A reduces the uncertainty of inferring ECA more than reducing the range of ECA itself. But fixing the properties of layer A is not necessarily the best option for inferring the remaining 4 parameters.

L.396 and beyond/ explanation for Fig. 7 'as a guide for planning an EMI survey':

I find this an overly complicated way to address the sensitivity of specific coil configurations to specific (combinations of) subsurface perturbations. I still do not see the advantage of your approach to a simpler sensitivity analysis.

We agree that Fig. 7 can be a bit of a handful. However, in practice it would not be needed to visualize all the combinations from Fig. 7. But rather put in the assumed ranges of each parameter from a targeted field and then compare it to the full range.

L.410: *From the perspective of an experienced user of EMI surveys, most of these general conclusions will be obvious, which helps to confirm the validity of the proposed approach*

This is an odd statement when put in the perspective of your study aims and introduction.

We want to ensure that the ML comes to reasonable conclusions. This is also why chose a simple forward model so we can verify the findings. If the approach was used on a model that describes a complex nonlinear system e.g. a groundwater aquifer then it would be difficult to confirm the result.

You mention there is '*no way for a user to consider the possible impact of prior knowledge*'. I think this is not true: you can use open-source forward models to do this. And I think you refer to this by stating that 'most of these general conclusions will be obvious'.

The '*no way for a user to consider the possible impact of prior knowledge*' is in the introduction and is referencing to the rule of thumb (that we now explicitly define). The sentence has been altered and now makes a direct reference.

L.412: *We see the value of this analysis as providing general guidance to less experienced users and to provide more fine-tuned guidance for site-specific conditions for those with more experience using EMI.*

Essentially, you provide a means to evaluate different realisations of a forward model. This is indeed useful.

Thank you for finding this part of our study to be useful.

L.413: *Furthermore, the guidance provided is quantifiable rather than based on general rules-of-thumb.*

You do not specify what you mean by ‘rules of thumb’? You also do not compare the outcomes of your analysis to the assessments provided by these rules of thumb.

One could also consider using the ‘McNeill-approximations’ (i.e. approximations under the LIN condition) as a rule of thumb.

We now specify what we mean by rule of thumb in the introduction.

Our approach is not tied to any specific model or even domain of models. One of the advantages is that it is very general and can be extended to deal with any type of input and output. We just chose to showcase this with a simple geophysical model of EMI instrument response.

L.434: *Figure 8 is somewhat information dense*

Very true. Cf. my previous comments, a simpler sensitivity analysis would offer more clarity (and, I think, perhaps partially make section 4.4 redundant)

Each individual ring can be represented by a more traditional tornado diagram, but it would require a lot of diagrams to show all the combinations. This figure is designed to display all the states within the boundaries we defined. In practice a user could constrain the ranges and would display a pie for each parameter like in Fig 5.

L.444: *This result could not be anticipated based on McNeil’s solutions*

What do you mean by this? Essentially, you are using the McNeil approximations, so I don’t understand this statement?

Changed the sentence to:

“This result could not be anticipated based on the rule of thumb”

L.450: *It is surprising, however, that one of the four observations place the instrument higher above ground. We suggest that this is a good example of a result that has both*

immediate practical value for survey design and could point researchers to ask follow-on questions about why this combination of observations is identified as optimal.

This is essentially a result of the spatial sensitivity (as captured in the deployed FWD model) of the evaluated configurations.

We believe the approach can be extended to other more complex forward models or combinations of forward models where the conclusions are less straightforward.

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).

Our ultimate aim is to provide a robust method for design optimization partially by showing that ML is a reasonable approach, which is also one of the reasons we choose a simple model. So here, we show that it works for a simple geophysical model and we can make direct connections between the result of ML and expected outcome. If a user was in a situation where they could decide between multiple geophysical methods for their problem. They could essentially all be incorporated into the approach by extending the ensemble with appropriate forward models of higher complexity.

References

Heil, K., & Schmidhalter, U. (2017). The application of EM38: Determination of soil parameters, selection of soil sampling points and use in agriculture and archaeology. *Sensors (Switzerland)*, 17(11). <https://doi.org/10.3390/s17112540>