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.

**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.

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.

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.

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?

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.

**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.

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.

**INTRODUCTION**

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

Magnetic field (cf. the following sentence)
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?

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)

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.

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

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.

L.72: for t users

Users or the 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)

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?

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?

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

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?

THEORY

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

For (almost) all subsurface media

L.119: low induction number

Explain (and reference)

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.

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)

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?

METHODOLODY

L.150: using EMagPy

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

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

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

Just state ‘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’.

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

Rephrase

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

Rephrase

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

How?

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.

**RESULTS & DISCUSSION**

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

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

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'?

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

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

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?

L.315: *balances performance with reduced field effort*
What do you mean by this? You should clarify this aim in your introduction

L.319: circle

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

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 )

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 where you expecting the VCP/HCP to outperform PRP? Provide the full argumentation.

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?

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.

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

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

**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*)

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.

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.

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’.

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.

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

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)

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?

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.

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).