Author Response for Paper 6449 We thank the reviewers for their time and helpful feedback! Below we summarize and respond to as many major reviewer comments as space permits. Please note that any in-text references use the same numbering as the manuscript.

1.The neural-adjoint appears similar to some recently-proposed methods. The authors should clarify the degree of its novelty, and reconsider its branding: Reviewer #1 shared several recent publications and, after reviewing these publicataions (and related works that we found), we agree that several of these papers do propose methods that are fundamentally similar to the Neural Adjoint (NA). We thank the reviewer for bringing these relevant publications to our attention.

Given this new literature however, we find that our original claims of novelty are still valid. In our manuscript (Section II) we explained that the Neural-Adjoint (NA) is based directly upon the approach proposed in [21]. The approach in [21] and the NA share the following methodology: a deep neural network is used as a surrogate for the forward model, and gradient descent is used to optimize the forward model output with respect to its input. In Section II we claimed two main novelties with respect to [21]: (i) we propose the boundary loss and Γ design that result in the NA method, and (ii) we provide thorough empirical evidence that the NA is competitive or superior across many inverse tasks. We find that the new references, and their proposed methodologies, share (*roughly*) the same similarities and differences with respect to the NA as [21]. And therefore our original claimed contributions are still valid. As reviewer 1 noted though, the NA does represent a specific variation on a broader class of recently-proposed methods - more than just reference [21].

Given the expanded related work, however, we do agree we should revise our NA branding. We still believe it is reasonable to employ a name, in this case "Neural Adjoint", so that our method can be easily referenced both within and outside of our paper. However, we will revise our manuscript to expand the related work and clarify that the NA method has major methodological similarities with many recently proposed methods, explain the similarities/differences, and explain that NA is the name for our particular variation on this general class of recent methods.

2.The authors should report the posterior distribution of solutions (e.g., Maximum Mean Discrepency measurement (MMD): We did not include MMD largely to control the scope/length of our paper. However, we agree with the reviewer's that MMD results would be valuable, and we provide them in Table 1. Note that D4: Meta-material is not included due to the intractability of rejection sampling. We find that cINN always has the best (lowest) MMD, closely followed by NA. Due to space limitations, we will need to add Table 1 and accompanying explanations to the appendix. We will briefly summarize these results in the manuscript.

Table 1: Posterior matching MMD score

Data	NA	TD	cVAE	INN	cINN
D1	0.07	2.62	0.07	2.03	**0.04**
D2	0.04	2.84	0.03	1.07	**0.03**
D3	0.06	2.70	1.62	0.11	**0.04**

3.Assumptions and constraints imposed by the boundary loss, $\mathcal{L}_{bnd}$. We found that our trained neural networks produced poor forward predictions outside of the training data domain. Therefore $\mathcal{L}_{bnd}$ was designed to encourage the model to seek solutions within the training data domain (i.e., where the forward model is accurate), at the cost of limiting the solution space. As shown in the appendix, this tradeoff consistently yielded substantial performance improvements. In all of our benchmark problems, the training data sampling space is well approximated by a hyper-cube of the form $|\hat{x} - \mu_x| - 2\sigma_x$, motivating our $\mathcal{L}_{bnd}$ design. However, we agree with reviewers that this imposes a specific prior on the solutions e.g., a high uniform prior within the hypercube. We also agree that this approximation will *not* work for more complex distributions (e.g., non-convex). We do believe $\mathcal{L}_{bnd}$ can be extended to such cases e.g., via kernel density estimates, and other methods. Based upon reviewer feedback, we will revise our manuscript to clarify these assumptions of our methodology for readers so they understand its limitations.

4.Contribution beyond prior meta-material work in [24] In [24] the authors considered one inverse approach, and did not quantify its performance. Here we quantitatively compare five methods. In [24] the single inverse model predictions were simulated with the real electromagnetic simulator, which is very slow. This inspired us to propose to train a high accuracy neural simulator (we achieve MSE of 6e-5 here versus 1e-3 in [24]) and treat it as the real simulator (e.g., sampling training data from it, and evaluating re-simulation errors using it), which was not done in [24]. This allows us to easily share our neural simulator with other groups (also not done in [24]) and for others to easily replicate and build upon our study of this modern problem.

5.Mixture density network (MDN) not included for comparison We did not include the MDN because we think performance comparisons with the cINN and INN are sufficient to support our conclusions. However, we do agree with the reviewer that inclusion of MDN would strengthen our conclusions. However, we did not have time to code, quality assure, and test the MDN on all datasets within the rebuttal period. Therefore we will be unable to address this limitation.