I thank the reviewer for their comments and corrections. Below I have responded to the comments and have indicated where changes or corrections in the manuscript have been made.

One of the questions that I asked and documented in this paper was if Method 2 could provide reasonable or superior out-of-sample estimates of the geoelectric field with respect to Method 3 for the purposes of prediction. To my knowledge, this is a question that has not been addressed in the MT literature. This was noted in the last sentence of the introduction. That is, in the MT literature, the end product is a transfer function with small error bars and papers like Jones et al. 1989 compare the computed transfer functions and their error bars but do not compare the performance of the models in predicting the geoelectric field. This is a very important point that I revisit several times in my reply.

The background and motivation for the results presented in this work stem from analysis that was performed in attempting to predict GIC measurements with locally measured magnetic fields. I found that the prediction performance using Method 1 was very poor and that using robust methods was better, but still poor. So I asked if one could do better and started with the basic approach of Method 2. For the data sets that I was considering, I found the simple and straightforward Method 2 to be much better. This was somewhat unexpected given the MT literature that I was familiar with. (Because of constraints on the publication of the GIC data, I was not able to present the results from my initial analysis that motivated this work.)

Please show plots for all three transfer functions (obtained with the three methods); these are traditionally plotted in log scale multiplied by square root of period in seconds. Alternatively, please show the apparent resistivities and phases for the three methods. Note that transfer functions need to be smooth to reflect the underlying physics.

I intentionally did not display the transfer functions in this work. The main objective of this work was to determine the out-of-sample prediction performance of different methods for the geoelectric field. It is expected that Method 2 will produce inferior and flawed estimates of the transfer function for purposes of estimating the ground conductivity (especially at short and long periods). As noted by the reviewer below, this issue was addressed over the past 30 years. A second reason for excluding the transfer function plots is that we agree that the generation of transfer functions (for the use of conductivity estimation) is best left to MT professionals, as noted by the reviewer below. The final reason is that this manuscript is about prediction of the geoelectric field and not on transfer function estimation.

Egbert & Booker (1986; Geophys. J. R. astr Soc. 87, 173-194) provide a thorough discussion on the advantages of robust estimation vs LS estimates. Indeed, under certain simple circumstances (such as the relatively very clean time series segments selected for analysis in this manuscript) the LS estimate can work well. In other cases, especially when noise is not obvious in the time series, they can provide entirely misleading results. It should be noted that clear and unbiased electric time series is an extremely rare occurrence and it often takes an expert to identify noise and bias, including cultural, tidal and temperature effects.

I agree with this statement (with the caveat that they provide a thorough discussion of the advantages of robust estimation *for the purpose of conductivity estimates*) and am familiar with the literature on the problems with estimating robust and bias corrected transfer functions using LS methods *for the purposes of estimating ground conductivity*. From a GIC perspective, what is of interest is the geoelectric field, independent of if it contains contributions that are not explained by ground conductivity structures. The question that was addressed in this manuscript was not whether LS methods give reasonable estimates of transfer functions associated with ground conductivity structures. I asked if LS methods, which are optimized to produce low LS errors in the prediction of the geoelectric field, provided significant differences in the estimate of the geoelectric field to robust methods, which are not fully optimized to produce low LS errors. As noted in the last paragraph of the introduction, the motivation for this question was that GIC practitioners generally assess the quality of a prediction using a LS metric. Another motivation was based on experience - I found that codes that computed robust and bias corrected transfer functions tended to perform worse in predicting the geoelectric field than transfer functions computed using Method 2. This is consistent with the results presented in this manuscript for somewhat "clean" time series. I agree that if one was not dealing with “clean” time series in the derivation of the transfer function using Method 2, one may get different results. What I have shown is that Method 3, which used clean and un-clean parts of a time series to derive a transfer function, produced slightly inferior predictions on clean parts of a time series than Method 2, which used clean parts to derive a transfer function.

To address the concern, I have modified the abstract so the original sentence

“We compare three methods for estimating the geoelectric field given the measured geomagnetic field at four locations in the U.S.”

to

“We compare three methods for estimating the geoelectric field given the measured geomagnetic field at four locations in the U.S. during time intervals where the measurements had few data spikes and no base-line jumps.”

Moreover, it is not appropriate to ignore the transfer function error bars in such a comparison as they contain vital additional information about the TF estimates. Since this is precisely what the techniques discussed in this manuscript do, the discussion should reflect this clearly, and the conclusions should be much more modest. The only conclusion that can be made from this is that for clean data sets, and only for the purpose of local E-field estimation, brute force LS estimate may be preferred to using the robust estimate at face value, without accounting for the TF error estimates in the analysis.

Consistent with our response to the first comment, we are not interested in whether LS methods can produce transfer functions with small error bars and thus have omitted a comparison plot of the transfer function and their error bars. Based on the extensive literature on LS methods for estimating transfer functions that can be used to infer ground conductivity structure, I am convinced that LS transfer function derivation methods are inferior in this regard. This manuscript asks if LS methods, although flawed in making TF estimates for purposes of inferring ground conductivity structures, can they provide reasonable or superior out-of-sample predictions of the geoelectric field for the type of data segments considered. In an attempt to make this point more clear, I have added the sentence

> "Although the conventional least squares method has been shown to be flawed with respect to transfer function estimation for the purpose of ground conductivity estimation (Egbert and Booker 1986), we have shown here that it can produce improved out-of-sample predictions of the electric field on data segments without many defects."

to the third paragraph of the Summary and Conclusions section.

Given the rich body of rigorous published literature on TF estimation over the past 30 years, it is strongly suggested that TF estimation is left to MT research professionals in order to avoid potential significant biases in the predicted electric fields. Also, if the utility of LS estimates (obtained from CLEAN time series!) for E-field estimation can at least be argued for, the suggestion that "it is an open question whether Method 2 produces reliable estimates" of the TFs for conductivity inversion (p. 10 line 216) is entirely inappropriate. Indeed, this was an open question... 30 years ago. Robust estimates provide a transfer function that is, well, robust! - in the sense that whichever segment of the time series is used for the estimate, the TF is the same within the error bars. This is a necessary feature for conductivity inversion, which in vast majority of applications should not depend on time. It is extremely ill-advised to publish a statement such as lines 215-218 on p 10.

I have removed this line. The intended meaning was that if we don't know the ground truth (that is, that we have an independent direct measurement of the ground conductivity), we can't say which model is better. In retrospect this seems somewhat obvious and does not need mentioning. I agree that there is a vast literature on why the model associated with Method 3 is *expected* to be closer to the ground truth.

Part of the motivation for omitting a plot of the transfer function is explained by the statement that transfer functions (for the use of conductivity estimation) are so difficult that only a MT research professional can do it. This is also justification for using Method 2 - GIC practitioners can implement Method 2 easily, but must become a MT professional in order to use Method 3 properly.

On a different topic, p. 10 lines 221-223 plainly suggests that all 1D estimates are inappropriate. This is mostly true; however there is an additional complication that needs to be taken into account. Estimated TFs are often affected by very local features (known as "surface distortion") in which case they might provide a less accurate regional estimate than a 1D inversion/ approximation at the same location. Of course one would actually need to invert the TF rather than use a large-scale regional 1D conductivity model to provide such a comparison. However, this option is missing from the manuscript so again, conclusions should be milder.

I respectfully disagree that it was plainly suggested that all 1D estimates are inappropriate. In that sentence I note that Methods 2 and 3 are viable options based on the results presented but this does not imply that Method 1 is not appropriate. Moreover, in the next sentence I note note that it is an open question as to how much revised estimates of historical geoelectric field estimates made with Method 1 will change. The reader should conclude from this paragraph that we have shown Methods 2 and 3 to be viable for the sites and data segments considered and that the question of the viability of Method 1 was not commented on directly and more work is needed to determine where it is viable.

Finally, p 6 lines 119-123 and p 4 line 156 make a strong statement that periods outside of the TF frequency range are predictable. As far as I am aware, this is mathematically impossible in general. Much more evidence is needed to substantiate this claim.

Periods outside of this range are expected to be predictable based on the MT literature. In the MT literature, the coherence is often found to be very high in the period range of 10-20,000 seconds with a tapering off outside of this range. This non-zero coherence also indicates that there is a predictable component outside of this range, but the MT algorithms use a cut-off value of coherence to determine the range in which the transfer function is estimated. In principle, the transfer function could be estimated outside of this range, but the error bars will be higher and there are physical arguments as to why this part of the transfer function does not represent ground conductivity structures.

The mathematical limitations on the periods for the transfer function are determined by the cadence of the measurements (1-second, corresponding to a 2-second Nyquist frequency) and the length of the data set (two days); with this one can estimate periods from 2 seconds to 86400*2 seconds in a spectra or a transfer function.

Minor comments:

p 2 line 4: misprint (as a either -> as either)

Fixed. Thanks.

p 2 magnetotelluric misspelt throughout

Fixed. Thanks.

p 2 line 28 A data citation for USArray is now available:

Schultz, A., G. D. Egbert, A. Kelbert, T. Peery, V. Clote, B. Fry, S. Erofeeva and staff of the National Geoelectromagnetic Facility and their contractors (2006-2018). "USArray TA Magnetotelluric Transfer Functions". doi:10.17611/DP/EMTF/USARRAY/TA. Retrieved from the IRIS database on Month Date, Year

The reference has been modified to include "USArray TA Magnetotelluric Transfer Functions" and "Retrieved from the IRIS database on Month Date, Year", which did not appear because of an error in the BibTeX file.

This reference is on line 30. The reference on line 28 was for the over-arching project.

p 2 line 30 While I appreciate the attempt to cite this work, the complete citation for the online database is:

Kelbert, A., G.D. Egbert and A. Schultz (2011), IRIS DMC Data Services Products: EMTF, The Magnetotelluric Transfer Functions, doi:10.17611/DP/EMTF.1.

(see http://ds.iris.edu/ds/products/emtf/#citation)

The reference has been modified to include "IRIS DMC Data Services Products: EMTF, The Magnetotelluric Transfer Functions", which did not appear because of an error in the BibTeX file.

p 5 line 103 The correct citation for the robust method that was used to obtain USArray MT TFs is Egbert & Booker (1986); Eisel & Egbert (2001).

The reference has been modified.

p 6 line 125 Note that a two-day interval is not nearly sufficiently long to extract adequate estimates up to 18,000 secs (5 hrs). At this period, you get a sample of less than 10. While it happened to work (I guess) for the four examples in this paper, it should not be expected to work in general.

I agree that a two-day interval may not be sufficient to estimate the transfer function at periods of 5 hours (with small error bars). However, the objective of this paper was not to obtain reliable estimates of a transfer function. It was to determine how well one can predict the out-of-sample geoelectric field using different transfer function estimation methods. We showed that when the transfer function derived using one two-day interval was used to predict another two-day interval, the prediction efficiencies were similar.

(Note that the uncertainty in the spectra at a given frequency is dependent on the number of samples in the interval and the uncertainty of the signal at that frequency and not the number of full waveforms at that frequency in the interval considered as suggested. The spectrum at a given frequency is not computed using only one value per wave - all values of the waveform are used. In the case of a white noise signal, the uncertainty in the spectra is independent of frequency (and the error bars are determined from a chi-squared distribution with two degrees of freedom). We agree that the uncertainty will be higher at longer periods for the type of signal considered, which has larger amplitudes at long periods, but the uncertainty is not directly dependent on the number of full waveforms in the time span considered as suggested.)

p 7 line 157 What exactly are the "smoothed error spectra"? Smoothed in what way?

The smoothing method was described in the paragraph that followed this sentence.   I modified the sentence to now say "The smoothed error spectra (described below)". In addition, I have modified the first sentence in the paragraph after this sentence to refer to Figure 1 to make this point more clear.

p 8 line 170 misprint (the similar -> similar)

Fixed. Thanks.

p 9 line 200 1D assumption also implies that Zxx = Zyy = 0.

I have added this statement.