Dear Associate Editor Eppelbaum,

We would like to thank you for your time and careful consideration in handling our manuscript. It is our opinion that the revised manuscript has been significantly improved by the suggestions of the reviewers.

The appendix below details our response to the reviewer comments. We hope that this revised manuscript addresses their minor concerns completely and appropriately. Please let us know if you have any questions regarding this revision.

Best Regards,

--

Matthew G. Finley, Ph.D.
Postdoctoral Research Associate
Geospace Physics Laboratory, NASA Goddard Spaceflight Center
Department of Astronomy, University of Maryland, College Park
Email: matthew.g.finley@nasa.gov
Phone: (319) 899-7872
Appendix

Key:

Black Italics – Original Reviewer Comment

Black – Original Text, Included for Context

Blue – Author Response

Red – Changed/Added Text

[Red Brackets] – Description of Large Revisions

Reviewer Comments:

Reviewer One (Mark Moldwin)

This brief report describes a lab created data base of “simulated” noise sources (reaction wheels primarily) for use in testing and validating magnetic noise identification algorithms for spaceflight. The data set is useful for the space magnetometer community. My main comments include expanding description of the quality and characteristics of the data (e.g., resolution and amplitude of noise signals), rationale for the layout of the four magnetometers, and providing example time series as well as the Tables describing the signals. See attached pdf for other minor comments and elaboration of the above.

Thank you for your thorough review of this manuscript, we have responded to your comments, individually, below.

L27-28: Why did you avoid Earth in this list? Could provide MMS for HEO type orbit mission or ST-5 for LEO, etc.

This sentence has been updated to include Earth, with citations for both the MMS magnetometers and the ST-5 mission.

‘…enabling magnetometer suites capable of making measurements from Earth (Russell et al., 2016; Slavin et al., 2008), to Mars (Connerney et al., 2015) and the harsh radiation environment around Jupiter (Connerney et al., 2017), and even to the sun (Bale et al., 2016).’

L69-73: Suggest making present tense. [Some minor annotations in the PDF correspond with this comment.]

This paragraph was updated to use the present tense.

[Paragraph 3 in Sec. 1 (Introduction) was updated to use present tense.]
What is the impact if the assumption of being calibrated is not true? Do you rely on absolute field measurements or just variometer measurements? What is the resolution of the data at 200 Hz? Why such high data rate?

If the assumption that the Twinleaf VMRs were well-calibrated is not true some methods such as Ness-style gradiometry may suffer, but several of the more recent interference mitigation techniques may not be impacted. The authors have now included the original, unfiltered data captured by each magnetometer, such that users can perform calibration, as well as additional combination and processing steps as necessary to suit the algorithm they may wish to test. The revised text has been updated, in Sec. 3.2.1 (Level 1 – Filtering and Truncation) to indicate that the original data is output with this dataset, and can be used for calibration if necessary:

‘The Level 0 data files are first read in, parsed, and filtering steps appropriate for each type of data are applied. Note that the Level 0 data files have been converted to an easily read .csv format and saved as part of this dataset to enable further data calibration, combination, and processing by potential users.’

As for the absolute versus relative field measurements and resolution: the Twinleaf VMR datasheet (https://twinleaf.com/vector/VMR/) claims that the VMR is ‘particularly useful for measurements with nanotesla precision, and while it has excellent stability it does not feature high accuracy.’ They don’t provide much information about the resolution in the data sheet, and I assume they there is some built-in filtering of the output to enable higher apparent resolution.

Unless style guide has you spell out or abbreviate depending on if leading off a sentence or not, be consistent with either Figure or Fig. (see above that Figure 1 is spelled out).

[All instances of the abbreviations ‘Fig.’ and ‘Sec.’ have been removed for consistency.]

Similar to comment regarding format of Figure. Spelled out Section above and have it abbreviated here and unless style guide has different rules for leading off a sentence or not, suggest spelling out everywhere.

[All instances of the abbreviations ‘Fig.’ and ‘Sec.’ have been removed for consistency.]

This typo has been updated to read:

‘…meaning that there was no active compensation being applied…’

Are these “lab” noise fields identifiable? If so, are they random or periodic, large amplitude?

The local magnetic field in the University of Iowa Merritt coil facility contains a number of potential random and periodic sources, including the building’s elevators (which interestingly have some periodicity when measured by a magnetometer – one abrupt change in field when the elevator car moves past the floor where the magnetometers are being tested, followed shortly after by another change in field when the elevator’s counterweight moves past), cars and trucks driving nearby, and of course the folks working in the lab carrying their smartphones and other devices
with distinct magnetic signatures. Additional sources include the ever-present 60 Hz hum, as well as signatures from other electronic devices in the facility (the fume hood in that lab is notorious for an ~8 Hz hum, although it was switched off for the duration of these experiments). Many of these sources are unavoidable and unpredictable, and they can serve to enhance the difficulty of the interference mitigation problem. These sources are described both in the previous and revised manuscript, in Sec. 3.2.1 (Level 1 – Filtering and Truncation):

‘The simulated reaction wheel interference captured as described in Table 2 contains near-DC offsets from various static and time-varying local sources (including the building’s elevators, nearby cars, and individuals carrying ferromagnetic objects).’

L118: Typical gradiometer scheme is co-aligned, so are you focusing on the two co-aligned sensors in the middle? What is the rationale for the placement of the outer pair of magnetometers?

We agree that a traditional gradiometer scheme assumes co-linearity, and as such the inner magnetometers M1 and M2 would be desirable for algorithm testing. However, the desire to reduce boom length or eliminate the need for booms entirely would require bus-mounting the magnetometers. A configuration of M2-M4 is representative of this bus-mounted configuration.

The first paragraph in Sec. 2.4 has been updated, for clarity, to read:

‘Figure 4a shows the four Twinleaf VMR magnetometers, mounted in a configuration that enables analysis of multiple magnetometer configurations. For example, using only the co-linear combination of M1 and M2 is representative of a traditional gradiometer configuration, whereas the combination of M2 - M4 may be more similar to bus-mounted configurations. This enables users to test interference mitigation algorithms against various magnetometer suite topologies, which can be useful for application to missions that do not utilize a traditional co-linear gradiometer configuration (e.g., the upcoming HERMES NEMISIS magnetometers).’

Figure 4: What is the coordinates of your test set up? (x, y, z that go along with your data).

We have updated Figure 4, and its associated caption, to define the coordinate system for both the Merritt coil system and the magnetometers used in this effort:

[Figure modified to include coordinate system.]
Figure 4: Experimental setup used for data collection. (Left) Photograph of experimental setup. (a) Four Twinleaf VMR magnetometers used to capture magnetic field data; (b) DC motors with attached ferromagnetic plates, used to simulate spacecraft reaction wheels; (c) Merritt coil system used to simulate geophysical fields. Note that the black rectangular object between (a) and (b) is a reference magnetometer used in this effort to ensure proper coil system operation, not for data collection. (Right) Schematic of experimental setup with measurements, in meters, referenced to the Merritt coil. Merritt coil and magnetometer coordinate systems are also illustrated.

Table 1: Suggest adding a representative time series of the magnetic field in nT for each type of your noise to get a sense of the amplitude and wave form of the individual noise and signals as well as the mixed signal time series shown later.

Thank you for this suggestion. We have added five figures to a new Appendix to the revised manuscript, detailing the time series for each of the component signals (i.e., trend, interference, geophysical) and illustrating the spectra for the more complex signals (i.e., interference and geophysical). We have also updated some of the text in Sec. 3.1 (Data Acquisition) to point towards this new Appendix.

In Sec. 3.1:

‘Table 3 describes the stimuli, applied to the coil formers and within the coil system, used to create pseudo-geophysical wave packets and signals. The component signals corresponding to each stimulus in Tables 1 – 3 is shown, as measured by the center-most magnetometer, in detail in this manuscript’s Appendix.’

Following Sec. 5 (Conclusion):

‘Appendix

This Appendix provides a more detailed view of the component signals (i.e., each near-DC trend, each physically-synthesized interference interval, and each pseudo-geomagnetic phenomena) described in Tables 1-3, as measured by the central magnetometer (M2). Specifically, Figure 7 shows the time series associated with the near-DC trend signals described in Table 1. Note that the measurements shown are the axes in which the dominant signal was applied for the first three trend signals; for the last two trend signals, the Z-axis was chosen for consistency with the other plots. The spectral content associated with the near-DC trend signals is not shown, since the very low frequency of the stimuli and the substantial filtering applied to the trend signals results in uninformative spectra. Figure 8 and Figure 9, respectively, show the time series and spectral content associated with the physically-synthesized interference described in Table 2. The Z-axis is shown since it measured the highest magnitude of interference. Finally, Figure 10 and Figure 11, respectively, show the time series and spectral content associated with the pseudo-geophysical signals described in Table 3. The Z-axis is shown for these signals since it was the magnetometer axis on which the pseudo-geophysical signals were predominantly applied (i.e., for all stimuli except for GeoSignal5 in Table 3).

[Added 5 new figures, detailed in the above text, showing the component signals in Tabs. 1 – 3.]
Table 3: Why didn't you use an orbit of Swarm data (e.g., use real geomagnetic “clean” data to drive the coils to have the “real” signal).

The pseudo-geophysical data generated as part of this effort were intended in part to create some combinations of signals with significant simultaneous overlap in the time/frequency domain. This has been shown in several recent works on interference mitigation (i.e., the MSSA-based techniques, as well as UBSS and WAIC-UP) to be a challenging problem. Although using data from Swarm would have been more realistic, it may have been more difficult to provide these challenging benchmarks for existing and future interference mitigation algorithms.

Additionally, the Merritt coil system at the University of Iowa (as mentioned in Sec. 2.2) has difficulty driving low-frequency signals natively. Complex data (such as intervals from Swarm) could have been driven, via a DAC, through the coil formers (which were used to generate the pseudo-geophysical signals via a function generator) to create additional geophysical signals, but this additional development and instrumentation was outside the scope of this effort. The authors would certainly be open to collaborating in order to expand this dataset in the future.

Table 3: Manuscript would definitely benefit from showing an example of this time series data. This is a good suggestion and has been addressed per one of your previous comments. A new Appendix was added to the manuscript with several new figures illustrating the component signals more clearly.

L151: are [in regards to a typo using ‘is.’]

The first sentence in the first paragraph of Sec. 3.2 has been updated to read:

‘The data exported by the Twinleaf I/O software are considered the Level 0 data product.’

L153: (Deletion of ‘maximally’)

The third sentence in the first paragraph of Sec. 3.2 has been updated to read:

‘The remainder of this section will discuss the data product processing pipeline necessary to convert this into a maximally useful dataset for the validation of…’

L165: Can you provide evidence for the “hard working” claim?

Unfortunately, this claim is subjective and not easily verified. The text describing the time-varying sources captured by the magnetometers has been revised to read:

‘…including the building’s elevators, nearby cars, and individuals carrying ferromagnetic objects.’

L166: were [in regards to a typo using ‘was’]

The related text now reads:

‘The sinusoidal simulated interference data were brought down to a near-zero mean by…’

L179: with [in regards to an incomplete sentence.]
The text in the final paragraph of Sec. 3.2.2 now reads:

‘…whereas the interference measured by the sensors will fall off with distance from the body of the spacecraft.’

*Figure 5: The grey scale intensity of the higher frequency signals are really faint.*

This is a result of the background noise being extremely close to the harmonics, in terms of the PSD—they’re only a few decibels apart. We tried several different scaling schemes on the colorbar, and this one provides the most clarity. Bringing the harmonics up means that the background noise will start to obfuscate the other spectral features.

*L196: have [in regards to a typo using ‘has.’]*

The text in the second paragraph of Sec. 3.2.3 now reads:

‘Note that the combined data have been detrended with a twenty-second…’

*L202: Is the Bz component shown because it has the biggest amplitude of noise?*

The Bz axis is shown for most figures because the amplitude of the interference is the highest (it’s the direction pointing towards the oscillating ‘reaction wheels.’) However, upon reading this question and one of your previous questions regarding coordinate systems, the authors did find an error regarding the stimulus for the pseudo-geophysical signals described in Table 3. The revised manuscript has been updated to correctly indicate that the geophysical signals were applied to the Z-axis coil former, NOT the X-axis coil former:

<table>
<thead>
<tr>
<th>Geophysical Signals</th>
<th>Stimulus Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeoSignal1</td>
<td>0.75 Hz sinusoid applied to Z-axis coil former with amplitude swept from 0 V to 0.6 V_{rms} to 0 V. Stimulus applied at ~10-min and ~20-min.</td>
</tr>
<tr>
<td>GeoSignal2</td>
<td>0.2 V_{rms} sinusoid applied to Z-axis coil former with frequency swept from 0.75 Hz to 20 Hz. Stimulus applied at ~10-min and ~20-min.</td>
</tr>
<tr>
<td>GeoSignal3</td>
<td>5.0 Hz sinusoid applied to Z-axis coil former with amplitude swept from 0 V to 0.6 V_{rms} to 0 V. Stimulus applied at ~10-min and ~20-min.</td>
</tr>
<tr>
<td>GeoSignal4</td>
<td>10.0 Hz sinusoid applied to Z-axis coil former with amplitude swept from 0 V to 0.6 V_{rms} to 0 V. Stimulus applied at ~10-min and ~20-min.</td>
</tr>
<tr>
<td>GeoSignal5</td>
<td>Vigorous ferromagnetic wrench-waving inside coil system performed by exemplary postdoc. Stimulus applied at ~5, 10, 15, 20, and 25-min.</td>
</tr>
<tr>
<td>GeoSignal6</td>
<td>0.1 V_{rms} sinusoid applied to Z-axis coil former with frequency swept from 0.75 Hz to 20 Hz. Stimulus applied at ~8, 16, and 24-min.</td>
</tr>
</tbody>
</table>

Note that the revised manuscript now mentions the reasoning behind displaying certain components in the plots. This new text can be found in the Appendix, and has been copied into our response to one of your previous comments.

**Reviewer Two (Anonymous)**
The magnetic cleanliness program is significant to achieve scientific objectives related to magnetic field measurements. It should be noted that magnetic tests are required to suppress stray magnetic fields. Furthermore, any gradiometric approach and magnetometers on long booms would guarantee magnetic field measurement without magnetic interference. However, the cost involved in performing such magnetic tests and adopting a long boom is likely to be high. To reduce the cost, therefore, recent missions adopt shorter booms. Hence, the authors generated an open-source dataset of stray magnetic field data, which is to be used to develop and validate magnetic interference mitigation algorithm. The dataset contains temporal variations of magnetic field corresponding to near-DC trends, synthesized interference, and pseudo-geophysical signals.

The significance of the open-source dataset of stray magnetic field data is highly evaluated.

Thank you for your review of this manuscript, we have responded to your specific comments, individually, below.

Table 2: Motors 1 and 2 seem to correspond to reaction wheels whose directions are different. However, spacecraft may have more reaction wheels. Furthermore, their rotation rates and directions may change depending on attitude of spacecraft. That is, these interference patterns may be too simple, and Inter1 to Inter8 may not be sufficient.

The original manuscript mentions, in Sec. 4 (Discussion and Future Work), some of the limitations of the dataset proposed in this work. Specifically, we describe several methods to generate more challenging problems for interference mitigation: combinations of multiple interference or pseudo-geophysical signals into a single data product; time-shifting the data to provide more severe spectral overlap between interference and pseudo-geophysical signals; and, additional proxies for interference commonly seen in spaceflight magnetometer data (e.g., spacecraft heaters, magnetorquers, and solar panel/battery currents). We have updated the revised text to more specifically call out additional magnetometers as another avenue for future improvements to this dataset, in order to better serve the community. The second paragraph of Sec. 4 (Discussion and Future Work) now reads:

‘Firstly, the dataset can be extended into more complex and challenging scenarios. Although the data presented in this manuscript has only combined a single interference signal with a single pseudo-geophysical signal and near-DC trend, combinations of multiple interference signals or multiple pseudo-geophysical signals can be generated that may be more difficult for some interference mitigation schemes. Additionally, the data can be time-shifted to provide more substantial spectral overlap between the pseudo-geophysical signal and interference, which is a common and challenging problem when attempting to mitigate local magnetic interference. Finally, this dataset has only provided physically synthesized proxies for a single type of local magnetic interference (i.e., reaction wheels). Although this is often a dominant source of interference on many spaceflight missions, it is not the only source of interference seen in in-situ magnetic field measurements, and spacecraft will often have more than two reaction wheels. Additional reaction wheels, spacecraft heaters, magnetic torque rods, and electrical currents from subsystems such as solar panels are commonly present in magnetometer data (Angelini et al., 2022; Stolle et al., 2021). Physical proxies for these interference sources can also be generated and added.
to the dataset, providing a more thorough set of example data for use in the validation of interference mitigation techniques.

A new example script (‘example.py’) has been added to the repository where the dataset is stored, illustrating how the data can be loaded for utilization by users. The detailed comments in this script also describe how additional permutations can be generated from our output L2 CDF files, generating more complex scenarios with signals from more reaction wheels.

A new Appendix has been added to this manuscript showing time series and spectra for the component signals (i.e., each interference signal, each pseudo-geophysical signal, and each near-DC trend). These figures more clearly illustrate the similarity between our physically-synthesized reaction wheels and the behavior of reaction wheels seen on-orbit. Specifically, many of our interference signals change rate significantly over the course of each 30-min interval (see Interference Signal 4, 5, 6, and 8) or have different rates during the entire duration of the interval (see Interference Signal 3, 7).

[Appendix added to more clearly illustrate the similarity between our physically-synthesized reaction wheels and the behavior of reaction wheels on-orbit.]

Figure 5: It seems that the figure caption for Figure 5 is insufficient. For example (a)-(l) are not mentioned at all. Temporal variations of spectra are shown, but a period for which frequency analyses are carried out is not defined.

The caption associated with Fig. 5 has been updated to provide additional clarity to readers in terms of the component signals (from Tab. 1-3) used for the three data samples shown:

Figure 5: Sample intervals captured by M2 in laboratory-generated dataset. (Column 1-2) Detrended time series and spectra associated with the combined data product; (Columns 3-4) Time series and spectra corresponding to the ground truth interference; (Row 1) Data associated with the combination of Trend2, Inter4, and GeoSignal2; (Row 2) Data associated with the combination of Trend4, Inter1, and GeoSignal4; (Row 3) Data associated with the combination of Trend3, Inter6, and GeoSignal6.

Additionally, an Appendix has been added to the manuscript showing both the time-series behavior and dynamic frequency-domain content of each of the component signals in each category (i.e., interference, geophysical, trend).

[Appendix added to show more clearly the dynamic time-series and frequency-domain content of each component signal.]

Reviewer Three (Anonymous)

The manuscript describes the creation of an open-source dataset of magnetometer interference signals. The dataset of stray magnetic field signals is exhaustive and enables many possible
combinations of different interference and pseudo-geomagnetic field signals to test magnetometer interference removal algorithms. This work enables magnetometer scientists to test and develop their magnetometer payloads for a broad range of purposes from UAVs to spacecraft of many different volumes.

My main recommendation is to include an example script / tutorial in the dataset that shows how to load the data and create different signal combinations. This script would improve the readiness and accessibility of the dataset to the scientific community. Please see the attached minor comments that ask for more descriptive captions, several citations, etc.

Thank you for your review of this manuscript, we have responded to your general and specific comments, individually, below.

The accessibility of this work would greatly benefit with the inclusion of an example/tutorial script to show loading the data and creating custom signal permutations.

The lack of example code was an oversight on the part of the authors. We have updated the temporary repository to contain an example Python script, titled `example.py`, which illustrates one method of loading an L2 CDF, separate the different source signals, and load the CDF metadata explaining which stimuli was applied to generate each source signal (i.e., which row of Tab. 1-3 was used in the combination). The comments within this script also highlight several potential methods for the generation of new data combinations. Should this manuscript be accepted, this code will be transitioned into an Institutional Repository for long-term storage and reuse.

[Generated new Python tutorial script for the loading and parsing of the L2 CDF files, added to repository.]


The text in paragraph 3 of the Introduction has been updated to include this reference. The revised manuscript now reads as:

‘To reduce the cost and complexity associated with magnetic field measurements, many modern missions now utilize shorter booms. For example: Parker Solar Probe has fluxgate magnetometers deployed up to 2.7 meters from the host spacecraft (Bale et al., 2016) and CASSIOPE/Swarm-Echo operates with a 0.9-meter boom (Wallis et al., 2015). The increasingly popular CubeSat form factor may require even shorter booms, due to their significant constraints on mass, volume, and cost (Clagett et al., 2017; Miles et al., 2016). Even though these short booms reduce mission cost and technical complexity, they diminish the effectiveness of the standard gradiometric interference mitigation approach, especially in the case where the dominating interference is caused by time-varying sources such as reaction wheels.’

The list of References has also been updated to include:

Figure 4: Please describe in the caption the black item located between the magnetometers and the reaction wheel apparatus. Is this another magnetometer?

The figure caption has been updated, for clarity, to read:

‘Figure 4: Experimental setup used for data collection. (Left) Photograph of experimental setup. (a) Four Twinleaf VMR magnetometers used to capture magnetic field data; (b) DC motors with attached ferromagnetic plates, used to simulate spacecraft reaction wheels; (c) Merritt coil system used to simulate geophysical fields. Note that the black rectangular object between (a) and (b) is a reference magnetometer used in this effort to ensure proper coil system operation, not for data collection. (Right) Schematic of experimental setup with measurements, in meters, referenced to the Merritt coil. Merritt coil and magnetometer coordinate systems are also illustrated.’

Note that this reference magnetometer has already been mentioned in the text describing Fig. 4, and as a result the in-body text has not been updated:

‘Figure 4b shows the two simulated reaction wheels, seen in greater detail in Figure 3. The Merritt coil used to generate large magnetic trends is partially shown in Figure 4c. The large rectangular object between the Twinleaf VMRs and simulated reaction wheels is a reference magnetometer used during initial testing and setup of the coil system and is not relevant to the output dataset.’

L161: Suggest using the same wording as L91 when discussing amplitude discontinuities below 1 Hz to improve readability.

The wording describing the amplitude steps produced by the Merritt coil system when driven at frequencies less than ~1 Hz has been updated to read:

‘As discussed in Section 2.2, the Merritt coil system produces amplitude discontinuities when driven at frequencies less than ~1 Hz.’

L165: The comment regarding “hard working postdocs” should be omitted to maintain a professional tone.

The authors agree that this claim is subjective and not easily verified. The text describing the time-varying sources captured by the magnetometers has been revised to read:

‘…including the building’s elevators, nearby cars, and individuals carrying ferromagnetic objects.’

Figure 5: The caption requires additional details regarding the rows depicted.

The caption associated with Fig. 5 has been updated to provide additional clarity to readers in terms of the component signals (from Tab. 1-3) used for the three data samples shown:

‘Figure 5: Sample intervals captured by M_2 in laboratory-generated dataset. (Column 1-2) Detrended time series and spectra associated with the combined data product; (Columns 3-4) Time
series and spectra corresponding to the ground truth interference; (Row 1) Data associated with the combination of Trend2, Inter4, and GeoSignal2; (Row 2) Data associated with the combination of Trend4, Inter1, and GeoSignal4; (Row 3) Data associated with the combination of Trend3, Inter6, and GeoSignal6.’

L199-200: The phrasing “their associated interference” is unclear. Also, is the whistler mode wave the interference or the pseudo-geomagnetic signal in this case?

Thank you for catching this typo. The revised manuscript has been updated for clarity, and to correctly call out the wave packets as pseudo-geomagnetic signal as:

‘The pseudo-geomagnetic signals associated with the combinations shown in rows one and three are frequency-swept wave packets simulating…’


The text has been updated to include this reference:

‘…simulating geophysical phenomena such as whistler-mode waves (Teng et al., 2019).’

L245: Recommend changing “easily serve” to “… well-suited…”

The revised text has been updated to read:

‘…this work can be well-suited to validate interference mitigation efforts across a wide range of fields.’