Response to the referee comment #2 for manuscript EGUSPHERE-2023-588 with title “Verification and Calibration of a Commercial Anisotropic Magnetoresistive Magnetometer by Multivariate Non-linear Regression” by Belsten et al.

We thank the reviewer for their insightful comments and questions on the manuscript. Thanks to the reviewer, the new version of the manuscript is much improved, particularly in references to previous works and in discussion of the limitations of our available calibration data.

In our references to work by Archer et al. and the other attitude independent calibration papers we more carefully alert the reader to the important differences between the attitude independent methods and our method which requires external attitude knowledge. We have provided a new paragraph explaining the AERO-VISTA magnetic sensor concept of operations to make this distinction clearer. We have significantly increased the explanation of our proposed calibration model to aid readers.

We agree with the reviewer’s comments which discuss the various limitations of our data collection methods. In the new version of the manuscript we more clearly explain that the intended scope of the work is a verification of hardware and calibration methodology and not the derivation of calibration parameters which will be used on orbit. The new version of the manuscript discusses the intended scope earlier in the motivation section and collects the limitations together into a new subsection (3.2 in the new version). In the discussion section we describe why, despite these limitations, the successful fitting of the x-axis of the magnetometer indicates that the pairing of instrument and model meets RMS error requirements and why the x-axis results are generalizable to the y- and z-axes.

Below is our response (in blue) to the issues raised in the review (printed in *italics*).

Specific Comments

p. 2, Subsection 1.3

*The concept of magnetometer calibration, the conditions of its operation on the satellites AERO and VISTA, and the requirements for measurement accuracy are not entirely clear. For example, from paragraph 1.3 a reader may conclude that a magnetometer calibrated on the ground should provide a measurement accuracy of 100 nT. On the other hand, from the information in sections 3.2, 3.3 and 4 (“*In orbit, AERO and VISTA will gather calibration data at low latitudes using a global magnetic map as a reference source. The regression parameters will be used to achieve the desired accuracy in the science gathering region near Earth's aurora.*”) it is unequivocally stated that the instrument is supposed to be calibrated in orbit.*

The AERO and VISTA magnetometers will be calibrated on orbit. It is not required that any calibration parameters derived on the ground still be applicable on orbit. However, it is important to perform ground-based performance verification prior to launch. Our manuscript addresses performance verification on the ground by applying our proposed calibration method to data collected on the ground.

We thank the reviewer for pointing out the confusion introduced by delaying discussion of AERO-VISTA operations until later in the manuscript and we have added a new paragraph to section 1.3 to further explain the use of magnetometers during the AERO-VISTA mission. This new paragraph also explains that the intended scope of this manuscript is to verify the performance of the AERO-VISTA magnetometer hardware together with calibration method, not a derivation of calibration parameters to be used on orbit.

First, under what conditions (temperature range and field measurement range) is it necessary to ensure a measurement accuracy of 100 nT? The value of 100 nT itself refers to the modulus of the field or to each component, 100 nT - is this the maximum or root mean square value? Secondly, if in-orbit calibration by the proposed method is supposed, how will this be done? The method compares the readings of the components of the tested and the reference magnetometer, and not the field module, as is done, for example, in the attitude-independent calibration method in the 1 paper by Archer et al. In addition to the position in orbit and, accordingly, the values of the field components according to IGRF or other global magnetic model, it is necessary to know the orientation of the spacecraft in order to recalculate the magnetic field components in the satellite coordinate system. After all, there will not be a reference magnetometer there, right?

The requirement is for 100 nT root mean square accuracy over the range of conditions encountered during science operations on orbit. Thermal analysis indicates this will range from about 0 C to 40 C, but these bounds could change under different modes of operation. For this work we demonstrated fit RMS errors within requirements when perturbed in temperature by ~35 Celsius, therefore demonstrating calibration under temperature. As pointed out by Reviewer 1, it would be ideal to perform additional calibration testing at cold temperatures (and under vacuum) but it is not possible to simultaneously perform magnetic characterization testing and cold + vacuum testing with the resources available to our CubeSat project.

The range of fields during measurement is set by Earth's magnetic field—assumed to range ± 50 uT over all axes. The magnetometer will experience 150 uT fields during spacecraft magnetorquer operation (between scientific observations), which determined the measurement range for the hysteresis testing in section 2.3.4.

In addressing the earlier question about collection of calibration data on orbit we have expanded Section 1.3 to include the concept of operations (ConOps) of our magnetic sensor calibration. In this new paragraph we clearly state that the spacecraft star tracker will be used as a reference source and therefore our magnetometer calibration method is not attitude independent like the work in Archer et al. We thank the author for noting this point of confusion and believe that the explicit categorization of our calibration as not-attitude-independent addresses these questions.

p. 3, Subsection 1.4

*"*The work by Archer et al. fits calibration coefficients for gain, offset, and angular position using on-orbit magnetometer data and the IGRF as a reference.*" In fact, Archer et al. estimated also temperature coefficients of gain and offset, at least for the outboard magnetometer.*

We thank the reviewer for catching this oversight. Our measurement equation only differs from the measurement equation of Archer et al. in the characterization of temperature dependence of the cross-axis terms, although the fit methods differ due to our access to independent attitude knowledge, and therefore a reference magnetic vector. This has been updated in text.

p. 3, Subsection 2.1

"A 3D printed mechanical mount for the EDU constrains the DUT in space at about 1 cm distance to the reference magnetometer (a Meda FVM400)." *Is there no mutual interference of the sensors with such a small distance between them?*

Magnetic screening was performed on all elements of the test and reference magnetometer to estimate the necessary distance to avoid coupling observable within our measurement sensitivity. Additionally, any coupling which is constant with time will be calibrated out as an offset so the verification of calibration performance is still valid; this is discussed in the new Section 1.5. The relatively short distance of 1 cm was chosen to retain good field uniformity within the Helmholtz pair used for ground-based testing. We have included this concern in the new "limitations" section of the results.

p. 4, Subsection 2.3

"Cross-axis coupling" The cross-axis effect in magnetic sensors is "a change in sensitivity based on the applied in the transverse, or cross-axis, direction". (https://aerospace.honeywell.com/content/dam/aerobt/en/documents/learn/products/sensors/application-notes/AN205_Magnetic_Sensor_Cross-Axis_Effect.pdf). The response of the sensor is non-linear in respect to the applied transverse signal. The calibration model (Eq. 1) does not take into account this non-linear effect, in my opinion. The parameter "cross-axis coupling" may be misunderstood by a reader as the cross-axis effect. It would be useful to clarify the meaning of this parameter.

We thank the reviewer for identifying this potential point of confusion. The calibration model (Eq. 1) does not take into account the gain variability caused by the "cross-axis effect" identified in Honeywell application note AN205 and is instead intended to address both non-orthogonality and instrument misalignment with the reference axis. These terms also account for soft iron errors. We have renamed this parameter 'off-axis coupling' instead of 'cross-axis coupling' throughout the paper to reduce confusion. Thanks to the reviewer's suggestion we also have added the cross-axis effect as a potential calibration extension in the future work section with a citation to the Honeywell application note.

p. 8, Subsection 2.3.4

"Given that both magnetometers reported similar hysteresis effects, the source of the hysteresis is likely magnetization of material near both magnetometers and not an effect inherent to either magnetometer alone." *This is an important issue, if there is an object near magnetometers, which disturbs the calibration field. Would not such an object introduce distortions into the results of other tests?*

This would be problematic if we were trying to derive calibration coefficients on the ground to be used on orbit because the change in magnetization environment would change the readings from the magnetometer. As we have now clarified, the point of this work is to evaluate the accuracy of the magnetic sensing instrument with an associated calibration method, not to derive these parameters on the ground for use in orbit. That the hysteresis is nearly equally present in both magnetometers indicates that this is a magnetic cleanliness issue that will need to be addressed in implementation of the flight models. The hysteresis that is attributable to the test magnetometer itself is the difference between the test and reference magnetometers, and this magnitude is within our accuracy requirement.

In response to a comment from Reviewer 1 we have added a comment about the possible sources of the larger hysteresis effect seen in both magnetometers (namely Raspberry Pi connector housings). We additionally have added the following explanation to the end of subsection 2.3.4: “The 50 nT difference in hysteresis between test and reference magnetometer is attributable either to a gradient in the magnetic field caused by the source of interference, or is inherent to our magnetometer instrument, or some combination of both. Given that our requirement is 100 nT, we consider this to be acceptably low sensor hysteresis even if all is attributable to instrument effects.”

The implementation of magnetic cleanliness for AERO-VISTA is outside the scope of this work focusing on the sensor, but we have updated the manuscript to refer readers to prior work on the project that discusses magnetic cleanliness efforts.

p. 10, Section 3

*“*This work extends the calibration equation reported in work by Archer et al. (2015) (Eq. 1) by including parameters for linear drift of all gain and offset parameters with temperature.*” The calibration equation presented in Archer et al. (2015) differs significantly from the one proposed in this paper. First, Archer et al. used the reference signal as an independent variables (or predictors) and the magnetometer data as a dependent variables (or response variables). The present paper uses the instrument magnetic and temperature data as predictors and the reference magnetometer data as response variables. In some cases, errors in the independent variables (predictors) can lead to biased parameter estimates.*

Second, Archer et al. evaluated the set of calibration parameters that minimizes the square difference of the field magnitude. This makes their calibration method attitude-independent – the knowledge of the satellite orientation is not necessary. In the present work, three subsets of calibration parameters are evaluated, each minimizing the difference of one of the three magnetic field components.

*And finally, Archer et al. also included in their model temperature dependence of the magnetometer gains and zero offsets: “*Therefore, we subsequently applied a temperature-dependent calibration to the science mode data to account for the large temperature drift during this interval. This was achieved by modifying the attitude-independent procedure, requiring a linear relationship of the offsets and gains with the temperature measured by the thermistor at each time, e.g. Ox(t) = cx T(t) + dx , where Ox(t) is now a time-varying magnetometer offset, T(t) is the temperature measured by the thermistor and cx and dx are the constants estimated through the iterative calibration procedure.*”*

We have used the work by Archer et al. as a initial guide for the fit measurement equation, namely incorporating the use of a sensitivity matrix and offset vector. We thank the reviewer for pointing out the important differences between our calibration method and that of Archer et al. and we have significantly edited and expanded our explanation of the magnetometer calibration equation to help clarify this for the reader. In the new version of the manuscript we begin with the basic matrix sensitivity and offset vector with the actual field as the predictor, as is done in Archer et al. and several other references. In this work, we also include a discussion of other effects such as nonorthogonality, misalignment, and soft iron effects which are parameterized by the matrix equation. We then rearrange the equation such that the predictors are the measured magnetic field and temperature and the independent variables are the reference magnetic field.

We have experimented with using the reference as a predictor and the measured field as a response, as well as evaluated the RMS fit error when using the measured as a predictor and reference as the response. When we rearrange the parameterized model fit with the reference as the predictor, we achieve much worse prediction of the reference field (biased estimates on the order of 2 uT). The reference field is what we care about for this work so we have chosen to use the reference field as the response and the measured as the predictor.

p. 11, Subsection 3.1

*"*The K Syy term is also anomalously large at -0.052 as compared to less than magnitude 0.01 for all other sensitivity terms. … This shows a that lack of characteristic calibration data can cause overfitting due to degeneracy of the fit to the available data.*" Does it mean that the estimated value (-0.052) of the K Syy term is wrong? If yes, why the data set was not expanded till the level sufficient to properly estimate all calibration parameters? For example, temperature experiments can be done so that the Y probe and then the Z probe are parallel to the axis of the calibration coils. If the proposed calibration method is so sensitive to the nature of the data, it is necessary to perform some preliminary operations in order to assess the applicability of the method. By the way, Archer et al. estimated how uniform the data covers of the attitude sphere in order to reliably extract calibration parameters. If a magnetometer actually has a scale factor temperature coefficient of 5%/°C, then it is unlikely that such an instrument can measure a 50,000 nT magnetic field with an accuracy of 100 nT (0.2% of full scale in another formulation).*

Yes, we agree with the reviewer that this does mean that there was insufficient data to properly fit the KSyy. In this paper, we aim to utilize data collected on the ground to verify the performance of our magnetic sensing hardware and proposed calibration method. To achieve this verification we exposed the magnetometer to a range of magnetic fields and temperatures, but did not include every possible combination of magnetic field and temperature. As the reviewer points out, temperature experiments with the magnetic coil along the y probe and z probe would likely yield the necessary data to better fit the KSyy (and KSzz) terms. Infrastructure for magnetic testing of satellite hardware is expensive and difficult to access for resource-constrained missions such as the AERO-VISTA CubeSats, so during experimentation we focused on validating the method by applying temperature perturbations with the applied field along only one axis. We are making an assumption that the successful calibration of one axis during temperature deviations indicates that all three axes can be similarly calibrated.

To address this comment, we have first explained in more detail in the motivation that the intent of this work is validation of performance of a combination of hardware and calibration methods, not the derivation of final parameters to be used on orbit. Additionally, we have more thoroughly explained the need for uniform data coverage on orbit in the discussion with an expanded explanation of the work by Archer et al. Finally, we have created a new subsection in the conclusion entitled "Limitations" which addresses these issues.

p. 13, Section 4

*"*This experiment has simultaneously validated the magnetometer design and calibration method for use on the AERO-VISTA mission*" In my opinion, the results presented in the manuscript are not sufficient to draw the above conclusions. 3 The magnetometer calibration parameters had not been validated by applying them to other datasets. There are doubts about the accuracy of estimating the calibration parameters KSyy, Oz , and KOz. The applicability of the calibration method to processing on-orbit data is not clear without understanding how the reference magnetic field values derived from global magnetic models will be transformed to the satellite coordinate frame.*

We believe that the low residual fit error in the x-axis in particular is a positive result that verifies that the magnetometer can be calibrated to the level of RMS error performance required for the mission. We agree with the reviewer that the parameters estimated from our limited dataset would not perform well on a uniform dataset like that acquired on orbit. Given the limitations in our dataset, we have clarified in the conclusion that the positive result is only a verification of calibration capability over varying environmental conditions, not a successful calibration of the entire magnetometer. In particular we acknowledge that that y-axis and z-axis terms are not well fit due to a lack of characteristic data, but the calibration of the X-axis, which did undergo temperature excursion, is proof that the physical processes that lead to the errors discussed can be calibrated with the proposed method. We have strengthened this argument by pointing out that each predicted value is dependent on its own subset of parameters, independent of the other parameters. Therefore, we do believe that we have achieved characteristic calibration of the x-axis to within the fidelity of our data gathering methods. We include the limitations imposed by our dataset and measurement methods in the new "Limitations" subsection.

Technical Corrections

p. 1, Abstract "...multivariate non-linear regression using a 27 parameter measurement equation"

p. 10, Section 3 "...9-element model..." *I counted only 24 calibration parameters estimated by the method of multivariate non-linear regression. They are 9 elements of the matrix S, 9 elements of the matrix KS, 3 elements of the vector O and 3 elements of the vector KO. 9+9+3+3=24. A subset of 8 calibration parameters is evaluated at each of the three executions of the MATLAB fitnlm function.*

We thank the reviewer for this correction. It is indeed an 8-element model for each axis for 24 parameters in total. This has been corrected throughout the manuscript.

p. 7, Subsection 2.3.3

"The measured fields over temperature are reported in Figure 3. The linear fit to the X-axis data pictured derives a linear temperature coefficient of 4.37 nT per degree C." *In my perception of Figure 3, X component was drifting at -4600 nT over a temperature range of 34 Celsius, so the linear temperature coefficient is -135 nT/°C.*

We thank the reviewer for catching this error; the correct linear fit slope is -140. nT/°C.

p. 8, Figure 3

The temperature varies in the range 33 – 73 K (Kelvin) , whereas in the subsection 2.3.3 we read "*the DUT was heated to about 65 Celsius and allowed to cool to steady state—approximately a 30 degree Celsius temperature range...*".

This figure mistakenly labeled the x-axis in Kelvin instead of degrees °C and has been corrected. Additionally, while the unit was heated to 73 °C as indicated in the figure, we only apply the liner fit starting at 65 °C to allow several seconds for thermal gradients from the heating process to even out. We have adjusted the range of Figure 3 (Figure 4 in the new manuscript version) to be consistent with the text discussion.

p. 11, Table 7

"Derived regression coefficients. Units of °C and µT" *In my opinion, if units of Bact and Bmeas is µT, then S has to be dimensionless, KS – 1/°C, O – µT, KO – µT/°C, RMSE – µT. Thus, only terms O and RMSE have unit of µT, units of all other coefficients are neither °C, nor µT. If units of Bact is µT and Bmeas is dimensionless, then S has to be µT, KS – µT/°C, O – µT, KO – µT/°C, RMSE – µT. Thus, none of the coefficients has a unit of °C.*

We agree with the reviewer that it is much clearer to indicate the actual units of the regression coefficients and not the units of the predictor variables. We have added a column to Table 6 with the units of the fit coefficients and removed the confusing units from the Table 7 caption.