1 Comments from Reviewers

1.1 Reviewer 1

General Comments

This work discusses the operating principal and measured accuracy performance of the scalar reference magnetometer currently flying on ESA's JUICE mission. The paper begins with helpful background on the use of the scalar magnetometer MAGSCA as a calibration source for the fluxgate magnetometers of J-MAG, the larger magnetic sensing suite. A major strength of this work is the bridging of theoretical analysis of CDSM physics with the engineering implementation, operation, and verification of magnetometer hardware. However, the terms precision, accuracy, and requirements are used without necessary context such that it is difficult to compare the observed and expected performance to previous works. The hardware presented is apparently an iteration on the magnetometer previously flown on the CSES mission, which the authors cite, but it is not clear how the ground verification and environmental calibration presented here built on this previous effort, or on any other work. The methods would be better justified with further comparisons to previous ground verification and calibration efforts, either of the CSES mission or those utilizing other magnetometer technologies. The authors' conclusion that their ground testing indicates that the requirement of 0.2 nT accuracy can be met is justified by the presented results. The conclusions in their current form read like an internal engineering report and no effort has been made to connect these results to a broader audience. Discussion of the source of the 0.2 nT requirement, and how the verification affects JUICE mission design and operation would make the results more broadly applicable to other space missions and magnetic sensing efforts.

Specific Comments

1) The motivation for the work should be strengthened by justifying the required accuracy of 0.2 nT within the context of the JUICE science mission. The motivation for this experiment appears to be a combination of instrument verification and determination of the ideal operating parameters and necessary post-processing. This motivation is discussed briefly in Section 4, but the manuscript would be improved by clearly stating this motivation and using it to justify the test configuration (Section 2). 2) Throughout the work, the treatment of precision and accuracy is not clear. In some places the precision is defined as the standard deviation over measurements made at different angular orientations (e.g. line 114). In other places the precision seems to be defined as repeatability at a given sensor angle (e.g. line 89). In general, the conditions under which the expected or observed accuracy and precision are reported should be clearly stated. What (and how many) samples are combined to yield standard deviations? Also, are you assuming that a given level of repeatability during ground testing should provide the same level of repeatability in the spacecraft environment? 3) Sensor noise and sample rate are parameters of great interest to most magnetic sensing missions, but these considerations are not discussed

at all in this work. Comparisons between MAGSCA and previous work would be more useful with additional information on sample rate or spectral noise. 4) The observed performance in this experiment is demonstrated to be sufficient for the JUICE mission requirement, but without discussion of the source of this requirement the significant of this experiment to the broader JUICE mission and scientific community is not clear.

1.2 Reviewer 2

The authors present the great performance of an innovative scalar magnetometer which meets the challenging 0.2nT accuracy requirement. The paper is worth to publish.

Below please find a few minor comments:

1. The Heading characteristic seems to be an offset error – independent of ambient field magnitude. It is mentioned in chapter 1.1/1.2 as absolute error, which depends on laser bias current and rubidium vapour temperature. Please add here the independence on field strength.

2. In the error discussion two error quantities are introduced, angular averaged precision (standard deviation to nominal field value) and angular averaged accuracy (averaged displacement of measurement to nominal field). The second error quantity is probably introduced because the heading error leads generally to an underestimate of the field magnitude. Is there any physical reason, that the measured value is in average lower than the real field magnitude? If yes, please indicate.

3. In chapter 2 the measurement setup in the COBS facilities is describe. A coil system is used for compensating the Earth field and for setting artificial test fields. Since the homogeneity inside the test volume is limited, additional errors have to be taken into account (AEcoil and Pcoil). This could simply be avoided by investigating the heading error in the very homogeneous and known field in the observatory - without any coil system and similar to tests performed with proton magnetometers. Please add a sentence why the coil system facility has been used.

2 Answers to Reviewers

Thank you very much for taking a lot of time and effort to comment on our paper. The comments help us to increase the quality of our work.

2.1 RC1

A detailed response to the many comments are found in the attached file.

2.1.1 Specific Comments

1. The motivation for the work should be strengthened by justifying the required accuracy of 0.2 nT within the context of the JUICE science mission. The motivation for this experiment appears to be a combination of instrument verification and determination of the ideal operating parameters and necessary post-processing. This motivation is discussed briefly in Section 4, but the manuscript would be improved by clearly stating this motivation and using it to justify the test configuration (Section 2).

   Reply: Currently, the J-MAG team is working on the instrument paper which will cover not only MAGSCA but also both fluxgate magnetometers. This paper will list all the relevant details concerning the JUICE missions for a broader scientific community and will explain how the requirements are defined.
   Our submitted manuscript aims to show that the required accuracy is achieved by MAGSCA and how it was measured since, the measurements and their evaluation are not straight forward. Nevertheless, we have added several annotations to the new version of our manuscript to explain and justify the 0.2 nT requirement for MAGSCA. Please check the detailed response to the many comments of reviewer 1 in the attached file.
   We have clarified the motivation at at beginning of section 2 of the new manuscript.

2. Throughout the work, the treatment of precision and accuracy is not clear. In some places the precision is defined as the standard deviation over measurements made at different angular orientations (e.g. line 114). In other places the precision seems to be defined as repeatability at a given sensor angle (e.g. line 89). In general, the conditions under which the expected or observed accuracy and precision are reported should be clearly stated. What (and how many) samples are combined to yield standard deviations? Also, are you assuming that a given level of repeatability during ground testing should provide the same level of repeatability in the spacecraft environment?

   Reply: In the new version of the manuscript we have tried to clarify the definitions of the used accuracy and precision. The heading characteristics defines the accuracy of the scalar magnetometer. The heading characteristics is an absolute error of MAGSCA depending on the orientation of the magnetic field vector in relation to the sensor axis.
   We distinguish between the accuracy of a measurement performed at a constant sensor angle and at a constant magnetic field strength and the "angular averaged accuracy" which is the mean value of all measurements performed for a heading characteristic (measured at equidistant sensor angles and constant magnetic field strength). The standard deviation of all measurements for a heading characteristics we define as "angular averaged

precision".
The sampling rates are added to the new manuscript.
We are pretty sure that if the operational parameters (as used for the presented ground verification: such as vapour temperature and laser bias current) are re-established the performance of MAGSCA will be the same during the space mission.

3. Sensor noise and sample rate are parameters of great interest to most magnetic sensing missions, but these considerations are not discussed at all in this work. Comparisons between MAGSCA and previous work would be more useful with additional information on sample rate or spectral noise.

   Reply: Please check the detailed response to your comments in the attached file. We added the sampling rate and noise level. I want to emphasize that the instrument paper will cover this topic in greater detail since also the performance of both fluxgate magnetometers will be of great significance for the scientific community.

4. The observed performance in this experiment is demonstrated to be sufficient for the JUICE mission requirement, but without discussion of the source of this requirement the significant of this experiment to the broader JUICE mission and scientific community is not clear.

   Reply: Please see response to "specific comment 1" and the detailed response in the attached file.

2.2 RC2

1. The Heading characteristic seems to be an offset error – independent of ambient field magnitude. It is mentioned in chapter 1.1/1.2 as absolute error, which depends on laser bias current and rubidium vapour temperature. Please add here the independence on field strength.

   Reply: The manuscript shows that the independence from the magnetic field strength is achieved in the magnetic field strength range (300 to 1500 nT). The manuscript also discusses the highly linear relation ship between the detection frequency and the magnetic field strength. However, for larger magnetic field strength (in Earth's magnetic field range (Pollinger et. al. 2020)) second order effects impact the heading characteristics (see response to comment 3 of RC2) where the offset error can change. Thus, it is - in general - not true that the offset error is independent of the ambient field magnitude.

2. In the error discussion two error quantities are introduced, angular averaged precision (standard deviation to nominal field value) and angular averaged accuracy (averaged displacement of measurement to nominal field).

The second error quantity is probably introduced because the heading error leads generally to an underestimate of the field magnitude. Is there any physical reason, that the measured value is in average lower than the real field magnitude? If yes, please indicate.

Reply: This offset is assumed to be a result of light shift effects. The light shifts are atomic level shifts caused by the interaction of the laser light with the atomic states. The mentioned direct laser current modulation creates a complex laser light spectrum for the excitation of the CPT resonances. The precise impact of each of these light fields (at different frequencies) on each atomic level (participating in the formation of the CPT resonances) cannot be modelled easily. Under certain conditions, light shifts are indistinguishable from Zeeman shifts (= atomic level shifts created by the ambient magnetic field). However, the measurement principle of MAGSCA is designed to minimise these light shifts. The residual light shifts appear as heading characteristics of the magnetometer. It is assumed that the imbalance of the individual components of the laser light spectrum together with their frequency detuning to their respective atomic transitions cause the offsets of the magnetic field strength measurement. The discussion of light shift effects is a very complex one, thus it was found to be not the scope of this work. A detailed discussion of the heading characteristics is found in (Amtmann 2022).

3. In chapter 2 the measurement setup in the COBS facilities is described. A coil system is used for compensating the Earth field and for setting artificial test fields. Since the homogeneity inside the test volume is limited, additional errors have to be taken into account (AEcoil and Pcoil). This could simply be avoided by investigating the heading error in the very homogeneous and known field in the observatory - without any coil system and similar to tests performed with proton magnetometers. Please add a sentence why the coil system facility has been used.

   Reply: The Merritt coil system was used since it allows to verify MAGSCA at the magnetic field strengths of 300 to 1500 nT required for the Ganymede phase (compared to the Earth's magnetic field strength 48000 nT). For larger magnetic field strengths ($>$ 10000 nT) the MAGSCA measurement principle requires an internal reference frequency correction (as it was used for (Pollinger et. al. 2020)) which is not implemented within MAGSCA. The higher magnetic field strength can have an impact on the heading characteristics and its offset (see response to comment 1). Additionally, the Merritt coil system allows (compared to the manual rotation of the sensor unit in the Earth's magnetic field) for much faster measurements which increased the parameter space which can be sampled in a feasible time frame. We have added an annotation (line 167 to 168 of the new manuscript) to clarify that we need the Merritt coil system to apply magnetic fields in the required range for the JUICE mission.

3 Detailed Reply to the Comments from Reviewer 1

The mentioned Lines refer to the lines of the manuscript with the comments.

1. Line 7: In a scalar magnetometer, all inaccuracies are deviations of the magnetic field strength measurement from the ambient field strength. Clarify that you are focusing on those effects of intrinsic angular variation due to the physics of your magnetometer operation. As oppose to, for example, interference effects.
   Reply: Clarification added to the abstract.

2. Line 10: is the performance requirement precision or accuracy? What sample rate? Does this include improvements by any calibration or post processing?
   Reply: The performance requirement is the accuracy of 0.2 nT (1-σ) with a sample rate of 1 Hz. No further calibration but post processing in case ratings B, C or D.

3. Line 15: later this is called a requirement. Verifying a requirement vs designing to a target are different motivations and signal different methodologies. To me this paper reads more like requirements verification
   Reply: Sentence removed of the updated manuscript for more clarity.

4. Line 26: It is worth mentioning that thermally induced offset drifts can also be modeled and either compensated or calibrated for. This is mentioned in the Merayo et al. 2000 paper.
   Reply: Annotation added to the new manuscript at line 28.

5. Line 27: This is an important point. Include a citation justifying this point or explain the relative dynamics of the Ganymede magnetic field as it relates to satellite operations.
   Reply: References added to the new manuscript at lines 29-31.

6. Line 34: Explain how
   Reply: Comment added to the new manuscript at line 40 and 41 including reference: A MAGSCA accuracy (0.2 nT) will permit calibration of the magnetic field vector data from the fluxgates to an accuracy sufficient to resolve the higher order moments of Ganymede's dynamo field as well as the 10.5 h, 171.7 h and 27 day induction signals from Ganymede's ocean.

7. Line 35: add citation
   Reply: This information is based on the J-MAG team's internal simulations. There is no citation available.

8. Line 36: Include some justification for why this requirement was set, or at least provide a reference where it is discussed
   Reply: See comment for Line 34.

9. Line 37: Explain the context of this requirement here. Does averaging over many measurements improve error? Does that still meet the 1-sigma requirements? Is there calibration or post-processing applied?
   Reply: The context of the requirement is already explained in the reply to comment for Line 34.

10. Line 57: A diagram of how the instrument is constructed around the gas cell would be helpful (or a citation to one)
    Reply: Reference (Ellmeier et al., 2023) shows a schematic diagram of the sensor unit. This is added in line 94 of the new manuscript.

11. Line 61: is this the difference of the frequencies of the two systems? or is "difference frequency" a term.
    Reply: Reworded to: "The magnetic field strength is derived from the difference of the resonance frequencies of a set of coupled CPT resonances ..." at the line 65 of the new manuscript.

12. Line 63: This is confusing. Do you mean that the consequence is a linear relationship between the measured frequency and magnetic field strength?
    Reply: Yes. This section between the lines 65 and 74 of the new manuscript were adapted for more clarity. Also, Fig. 2 was updated. Now the resonance frequencies of the CPT resonances are added.

13. Line 64: if derived, explain how. If taken from another source provide clear citation
    Reply: See previous reply.

14. Line 70: Is this effect included in your proposed accuracy requirement?
    Reply: If there would be one – luckily there is none – it would have to be included. Sentence removed in new manuscript.

15. Figure 2: You've explained why different resonances are required for different sensor angles, but why do the boxes also imply a range of heading characteristics? By my reading of this figure, heading characteristic is the independent variable of the line plots across sensor angle.
    Reply: The boxes do not imply a range of heading characteristics in the "y" axis of the plot. The boxes are indented to show that in the angular region where two boxes overlap, a magnetometer operation at one or the other resonance is possible. MAGSCA is just configured to switch at the sensor angles of 60°, 120°, 240° and 300°.

16. Line 80: Does this difference in design result in a different calibration procedure? You state in the abstract that your post processing is novel but you haven't made it clear why this post processing wasn't developed for the earlier iterations of this magnetometer technology.
    Reply: This coil system was specifically installed for the investigations of the MAGSCA performance and thus, it was not available for the CSES mission. In lines 80 to 86 of the new manuscript it is now mentioned that

the magnetometer for the CSES-1 mission was tested at another facility where it was compared to a Caesium magnetometer.

17. Line 87: is this referring to the measured magnetic field strength?
Reply: Yes. Sentence reworded in the new manuscript at line 112.

18. Line 89: standard deviation over what? Measurement to measurement? At a characteristic time scale?
Reply: Lines between 113 and 118 reworded for clarity.

19. Line 91: but given the deviations observed in Figure 2, this is not necessary for your required accuracy of 0.2 nT as long as you select the correct resonance
Reply: For the heading characteristic of figure 2, a correction of the heading characteristics is not necessary to achieve an accuracy below 0.2 nT. However, to get the most accurate measurement - even below 0.2 nT - the correction would be necessary. This statement was removed from the new manuscript for clarification.

20. Line 92: And in the conclusion you report that you do not need to correct for heading characteristics
Reply: Statement was removed from the new manuscript for clarification.

21. Line 99: which optical components? and what material are they made of?
Reply: Especially the optical fibres, reference added at line 126 of the new manuscript.

22. Line 100: what level of attenuation is expected. Is the laser bias operation sufficient to overcome this attenuation?
Reply: Based on radiation analysis and test, the - in the worst case - transmitted optical power at the end of the mission lifetime will be reduced to about 1/4 of the initially transmitted optical power. With a higher laser bias current and a sensor temperature, the operation of MAGSCA is possible when an attenuation to 1/4 of the initial optical power (before entering the vapour cell) is introduced. Annotation added to the new manuscript at lines 127 to 130.

23. Line 115: this is going to depend on sampling method. Are you assuming even sampling of angular space?
Reply: Yes, for the presented measurements of the heading characteristics the sensor angles are evenly sampled with 15° (MAGSCA stand alone setup) or 30° steps (J-MAG integrated setup). This information is added to the lines 244 and 251 of the new document.

24. Line 116: where did the error bars on the points in Figure 2 come from, if not from the angular averaged precision? (aka standard deviation)
Reply: Section 4 discusses that for the heading characteristics of Fig. 2, 3 and 4, each data point is the mean value of four separate measurements of the heading characteristics at four different sensor orientations, but with

the same operational parameters and the error bars are the corresponding standard deviations. This information is clarified at the lines 102 to 107 of the new manuscript.
The angular averaged precision is the standard deviation of all data points (magnetic field strength measurements at different, equally spaced sensor angles) of a heading characteristic and thus, it is not depicted in Fig. 2, 3 and 4.

25. Line 119: This is a big assumption. Spacecraft do not tend to evenly sample angular space unless specifically operated to do so. Is this the case? If not, state this clearly as an assumption
Reply: Added to new manuscript at the lines 150 and 151 for clarification: ... achievable during the Ganymede phase of the mission. The primary operational phase of MAGSCA is during the Ganymede phase of the mission. The stated assumption relates to this Ganymede phase.

26. Line 124: This rating of A,B,C,D could be removed. The critical point is that different combinations of angular averaged AE and P require different kinds of post processing. This would be more clearly indicated with a new "necessary post processing" column in Table 1
Reply: Column added to Table 1.

27. Line 124: I see now that you also use this rating system in Table 5 and the conclusion. If you choose to keep the A,B,C,D rating structure, you should at least motivate it by saying that a contribution of this work is to determine what type of a signal post processing may be needed for different operating modes. Than A,B,C,D becomes "post processing types" not just an arbitrary letter grade.
Reply: Comment added to new manuscript at line 154 and 155.

28. Figure 4: As in earlier figure, it's not clear what these error bars represent
Reply: See reply to Line 116.

29. Line 135: to what level? Provide a citation or measurement justifying this claim
Reply: More information concerning the observatory added within the manuscript including a reference at lines 171 to 174 of the manuscript.

30. Line 150: Merritt coil systems also do have a small residual inhomogeneity, is this what you mean?
Reply: Yes. Clarification added to the new manuscript at line 188.

31. Table 2: what sensor? Do you expect the fluxgate sensor repeatability to be at the level of 0.01 nT like you indicate in this table?
Reply: The engineering model of the DFG-magnetometer developed for the Magnetospheric Multiscale mission was used for the measurements. Reference Russell et al. (2014) added to the manuscript. The 0.01 nT is a result of the mathematical calculation of the inhomogeneity (measured in a +5cm, 0cm -5cm grid) per centimeter.

32. Line 165: is this LSB the dominant source of current uncertainty?
Reply: Yes.

33. Line 169: Why is this? It seems like an important part of accuracy verification.
Reply: The accuracy verification was carried out in the x-y-plane of the Merritt coil system to ensure easy rotation and alignment of the sensor. Therefore, no additional test fields were applied in the Z-direction. Annotation added to the new manuscript at the lines 204 to 206.

34. Line 178: what sample rate? or how many total samples?
Reply: The sampling rate was 128 Hz for all measurements presented in the manuscript. Sampling rate was added at line 102 of the new manuscript.

35. Line 198: Include photos or diagrams of both setups
Reply: For the J-MAG integrated setup, the fluxgate sensors were placed outside of the Merritt coil system. Thus, for both setups, only the MAGSCA sensor was inside of the Merritt coil system. The front-end electronics box (for both setups) was located in a distance of more than 7m to the Merritt coil system to prevent any interference. Comments added at lines 252 and 253 of the new manuscript. Change of word "within" to "with" in manuscript at line 235 to avoid confusion.

36. Figure 6. I think this is the MAGSCA stand-alone setup, but this should be clarified
Reply: See previous reply.

37. Line 206: How far away is the MAGSCA front-end electronics in this test? Is there any possibility of interference from the control electronics?
Reply: See reply for Line 198.

38. Equation 7: This equation is confusing and/or the parameters need to be more clearly defined. If a_X is the field along the coil x axis caused by the x coil, and b_Y is the field along the coil x axis caused by the y coil, then both a_X and b_Y in $B_{app}(X)$ should be multiplied by $cos(\beta)$
Reply: For $\beta = 0$ a magnetic field is only applied in the x-axis. No field is applied in the y-axis. Thus, the impact of the coil pointing in the y-axis is 0 in the x-direction, which is satisfied by $\sin(\beta)$. b_Y describes the impact of a magnetic field applied with the Y-axis coils in the Y-direction.

39. Equation 8: you previously defined this residual vector as B_res, why not just make this a vector equation of B_res? (Otherwise at least make it clear that B_x B_y B_z are components of B_{res})
Reply: Equation 8 is rearranged in the manuscript. $B_{res} = (B_x, B_y, B_z)$ added at line 285 for clarity.

40. Line 248: A general weakness of non-linear fitting to scalar measurements is the risk of overfitting if fit functions contain linearly dependent unknown

parameters. You manage to avoid this with the orthogonal functions of β multiplied by k1 through k5 in equation 15. This merits some discussion. Similarly, eqs 10-14 contain 4 unknowns by count, so you can likely solve given the k_i for $i = 1 : 5$. this should also be mentioned. Presumably you are aware of this and this is why you took the steps to reduce the number of fit parameters; you can motivate this section with this point.
Reply: Yes, you are right. We are aware of that and therefore, we reduced the number of fit parameters. To point this out, annotations were added to the new manuscript at the lines 304 to 306.

41. Line 249: how do you know this?
Reply: This is known from measurements with the fluxgate used for measurements of the inhomogeneity. See reply for Table 2.

42. Line 269: Again these parameters need to be more clearly explained. By my reading, b_X and b_Y are the cross-coupling of fields from coils into the orthogonal sensor direction, and therefore are not necessarily equal
Reply: See reply to Equation 7. Both, b_X and b_Y describe the effect of the non-orthogonality between the X- and Y- coils and must therefore be identical.

43. Line 279: How did you achieve exactly 90° rotation? Do you know the expected misalignment of your sensor positioning? Will this contribute to any errors in your fitting?
Reply: This is addressed in the latest version of the manuscript at lines 326 to 332.

44. Line 315: Any magnetically soft material in the sensor can be remagnetized by the applied magnetic field and therefore may depend on applied magnetic field strength. Additionally, magnetic material properties in general do change with temperature. This assumption needs further justification. The addition of an earlier discussion of the materials in the sensor unit may suffice
Reply: The sensor materials are addressed in the new manuscript at line 95.

45. Line 324: Needs further explanation. What uncertainty are you referring to? Which parameter in particular is uncertain?
Reply: The uncertainty introduced by the fitting of the data. Clarification added at line 374 of the new manuscript.

46. Line 333: What might explain this deviation?
Reply: The temperature within in tunnel is about 6°C all year. When a magnetic field is applied to the Merritt coil system for the compensation of the Earth's magnetic field (electric currents in the range of about 3 A for the vertical - Z-axis - component) it takes some time (usually 6 h) until the coil system is in thermal equilibrium. The reported deviation

was (most likely) caused by a too early start of the measurements before the thermal equilibrium of the Merritt coil system was reached.

47. Line 340: Also mention that this figure confirms your point that errors are dominated by heading characteristics
    Reply: In line 77 of the new manuscript we mention that the heading characteristics define the absolute error of the instrument.

48. Line 343: do you have the ability to control the vapour cell temperature in flight?
    Reply: Yes.

49. Line 344: in addition to what? It doesn't seem like any correction at all is needed
    Reply: "Additional" removed

50. Line 349: Is there a proposed explanation for this?
    Reply: A higher laser bias current has a negative effect on the accuracy and precision which can be explained by larger light shift effects due to the higher light intensity within the vapour cell. Reference added.

51. Line 359: Word removed
    Reply: Word removed

52. Line 363: Tie this to the motivation for this work. What does this mean for the JUICE mission, how it will be operated, and how the data will be processed?
    Reply: Annotation added at line 421 of the new manuscript, that the operational parameters (vapour temperature of 25°C and laser bias current of 2.14 mA) were defined as the default configuration of MAGSCA.

53. Line 366: You observed variation in instrument performance at different laser bias currents, but is this difference due to the laser power output or due to the illumination of the gas cell? i.e. if you get radiation darkening of the fiber and increase the laser current to compensate, how do you know you won't see performance similar to the lower current mode pre-darkening?
    Reply: The comment is correct. The heading characteristics are mainly caused by the light intensity within the vapour cell. The coupled CPT resonance amplitude is an indicator for the effective light intensity within the cell. Currently, we plan to utilise the resonance amplitude (already characterised right after launch during the JUICE commissioning) for tracking the radiation induced darkening.

54. Line 369: at least in the environmental and operational conditions you tested. This does not necessarily guarantee that this is true in all cases.
    Reply: Annotation added to line 425.

55. Data availability: Will the fitting software be made similarly available?
    Reply: Remark added to manuscript.

56. References: doi text is correct but hyperlink is broken
    Reply: Should be working now.