The paper Mathematical foundation of Capon's method for planetary magnetic field analysis provides the underlying formalism for applying Capon's method to planetary magnetic fields and illustrates it with simulated data relevant for BepiColombo mission. While this is a valuable contribution to the field, a number of points need to be addressed before publication.

1. Extension of Capon's method to planetary magnetic fields. As indicated in the first and last paras of Section 3, L73–75 and L202–206, the paper generalizes the Capon method, previously used for the analysis of wave data. This is a major result that could be emphasized better, perhaps in a separate discussion section. This section could include a closer analysis of the case in the paper as compared to the wave case, by referring to Motschmann et al. (1996).
The discussion section could also detail the key principle(s) underlying the method, like maximum likelihood / minimum variance, along the line of Narita (2019). The divergence free feature of the magnetic field could be discussed on top, as in Motschmann et al. (1996).

Reply: Agreed. We will add a section dedicated to the connection of the Capon method to other inversion methods (e.g., maximum likelihood, least square fit). We will highlight the difference from the wave analysis method (Motschmann et al., 1996) in section 6. It is also a good idea to add a discussion about the divergence-free nature of the magnetic field in the introduction.

Related:
• L166-168: I am not sure I understand the text here, even though I could essentially follow Eqs. 30 to 41. Eventually, the underlying model makes the main contribution to the data (e.g., in the test application of Section 5). I guess this 'minimal contribution' has rather the meaning of Motschmann et al. (1996), where the filter w absorbs all the energy not associated with k (here not associated with the parametrized field) and leaves the part related to k undistorted (here the part related to the parametrized field). Same issue at L256.

Reply: Surely it is possible that the underlying model makes the main contribution to the data. But the distribution of the parameterized and the non-parameterized parts is unknown. Therefore, we stay conservative and assume safely that a large part of data is influenced by the noise and non-parametrized signals. (Explanation will be added on page 7)

• L200-201: Is this expression derived by Narita (2019)? Or could be derived by further processing of the maximum likelihood estimator? (e.g., in the suggested discussion section?)

Reply: Yes, Narita (2019) derived the expression for Capon's estimator by regarding the likehood function as nearly Gaussian (particularly around the peak of the likelhood function).

2. Illustration & Validation. In view of upcoming BepiColombo data, the authors chose to illustrate the method with simulated observations of Mercury magnetic field. While this is certainly helpful to prepare BepiColombo, I wonder if it is also the best test bed for the method. Earth magnetic field is known much better, at various altitudes – such that the weight of the external field and its influence on the results could be analyzed too. Including an example at the Earth, or at least a brief discussion of this validation possibility, would be more than welcomed.

Reply: The application of Capon‘s method to the analysis of Mercury‘s magnetic field will been emphasized as an example. Surely, the test against the Earth' magnetic field would be a good alternative. But such an analysis would have the size of a separate paper and it could not be pressed in a paragraph. The simulated data have the advantage that the ideal solution is known and thus a doubtless evaluation is possible. (Discussion will be added on page 14)

Regarding the test exercise of Section 5, Table 1 shows that the largest errors are associated with g21 and, to some extent, with h21 . Is this by chance, or related to some systematics?

Reply: The deviation of these coefficients is related to the underlying model and therefore it is systematic noise. (Discusson will be added on page 14.

3. Technicalities. Considering the target audience of the journal, different to a good extent from the signal processing community, the mathematical language of the paper may prevent the optimal transmission of the message. Additional explanations may help, inserted in the text or collected in an Appendix – when detailing the math would perturb the flow too much:

• *L69-71: Please clarify this sentence, possibly including an example.*
Reply: For example, when the magnetic field data are known on a dense grid in the vicinity of the planet, the Gauss coefficients can be estimated via integration. But in the case of a limited data set those integrals cannot be evaluated. (Sentence to be added on page 3.)

• *L93-97: The intuitive introduction of the filter matrix w via Eq. 8 s a bit confusing, since eventually the non-parametrized (external) part of B does not show up in the g C formula, Eq. 41.*
Reply: Eq. 8 has been written down, since it is the first intuitive idea to solve the inverse problem. An explanation that the non-parameterized parts <v> are unknown and have to be truncated by w will be added on page 4.

• *Eqs. 9 and 10 fall pretty much out of the blue. The use of M and P becomes clear later, but some clarification would be good already at this point.*
Reply: Agreed. We will add motivation on page 4.

• *L106-110: Please detail why the determinant vanishes (even though it may look straight), how does statistical average prevent this, how is statistical average achieved.*

Reply: It is a mathematical nature. The calculation of vanishing determinant of the outer product will be shown in Appendix.

• L127: Please indicate also the second order moments.
Reply: Maybe the structure of the sentence was confusing, but <g o g> are the second order moments.

• Eq. 21: Please explain why 2<Hg> o <v> and not <Hg> o <v> + <v> o <Hg> (given that, in general, the external product does not commute).
Reply: The outer product commutates in this special case because <Hg> and <v> have the same dimension (to be added on page 6)

• Eq. 27 and L154: Please explain why this is not enough to uniquely determine w.
Reply: It is because the parts that have to be truncated are unknown. (will be added on page 7)

• L155-158: Feels confuse. As long as the filter matrix truncates the non-parametrized part, it is not clear why its contribution to the data matters, neither how 'this yields the following procedure'.
Reply: The contribution matters because the parts that have to be truncated are unknown. The sentence 'this yields the following procedure' will be deleted.

• L191: Why is tr P a convex function?
Reply: trP is a convex function since it is the sum of the quadratic averaged expansion coefficients and thus it follows the same structure as for example the function f(x)=x^2 (explanation to be added on p. 8)

•Eqs. 42 and 43: Please detail what is meant by 'input' and 'output'. Regarding Eq. 43, is there an equation analogous to Eq. 23, to clarify the meaning of 'signal' and 'noise' also for output?
Reply: By input we mean the measured data. By output we mean the filtered data (will be added on p. 9);
Yes, one can construct an equation analogous to Eq. 23, which results, when w^T(...)w is applied to this equation. This can be seen within SNR_o.

•Eq. 44: Please explain why this ratio is dominated by 1/trace.
Reply: The array gain is dominated by 1/trace, since P, H and v are given by the model and the data and do not depend on the method (w does). (explanation will be added on page 9)

•L266-267: Please provide a brief demonstration.
Reply: The eigenvalues of M will be calculated in the Appendix.

•L267-269: Please explain briefly what is this about.
Reply: Explanation will be added on page 11.

•L278: How is the ‘compromise’ quantified?
Reply: The compromise can be understood in the sense that trace reaches its minimal value under the condition that P_d is maximal. (to be added on p. 11)

•L278-282: This is quite opaque for those not familiar with signal processing and in particular with these techniques.
Reply: The basic idea of the Tikhonov regularisation will briefly be explained on pages 11–12.

4. Others
• L17-18: What is non-ideal orbits?
Reply: The sentence was adversely formulated. We change it to „data sampled on single orbits“.

• L18: simulated Mercury magnetic field data
Reply: Formulation will be changed in the manuscript.

• L54: ‘closing the void’ => ‘covering the range’ ?
Reply: Formulation will be changed in the manuscript.

• Eq. 54 is identical to Eq. 41.
Reply: Reference will be added in the manuscript.

• L325: in => at
Reply: Formulation will be changed in the manuscript.

• L352: In principle, one could analyze also the external field, if some model is adopted.
Reply: That‘s very true! We will modify the sentence accordingly.