Response to Reviewers

[Cover Letter]

Dear Editor;

Thank you and the reviewers for your valuable time reviewing our paper and offering insightful comments. Your feedback has been instrumental in enhancing the current version. I have meticulously reviewed and addressed each of the comments. I trust that the revised manuscript meets your expectations and welcome any additional constructive feedback you may have. Below, you will find my detailed response to the reviews, which includes a list of all pertinent changes made in the manuscript. I have carefully incorporated the revisions with all modifications in the manuscript based on the referees' and reviewers' feedback, as outlined in the authors' response. Also, you can follow all the relevant changes highlighted in the Author's track-changes file.

Sincerely,

Utkan Mustafa Durdag, Ph.D.

umdurdag@artvin.edu.tr

Asst. Prof., Faculty of Engineering, Department of Geomatics Engineering,

Artvin Coruh University, Türkiye

Response to Referee #1

RC1: 'Comment on gmd-2023-210', Anonymous Referee #1, 15 Dec 2023:

The comments about the manuscript "**Minimal variance-based outlier detection method using forward search model error in a leveling network**"

Outlier detection problem is one of important issues in geodesy discipline. Conventional and Robust methods have been used for the outlier detection problem. Especially the studies about the efficiencies of the methods are worthy for the interpretations of the results of the analysis. In this manuscript, Author has applied a new method for the detection of the outliers in geodetic networks. A leveling network has been simulated and different types of outliers' detection problems have been examined. In my opinion, the paper has been accepted after the comments given below are considered:

Author Response (AR) to anonymous referee #1

AR. The author thanks the referee for our article's detailed and comprehensive feedback. We appreciate the reviewer's time and effort in reviewing our work. The author's responses to each referee's comment are in the below:

1) The title should be changed. In the paper, the proposed method has been applied to the only leveling network; but the functional and stochastic models can be applied to all type of geodetic networks. Also, this method has been applied to the regression problem before. That's why, the title can be changed as "Minimum variance based outlier detection method using forward search model error".

AR1. The author agrees with the referee's suggestion to modify the title for clarity and inclusivity. The potential revised manuscript title is now "Minimum Variance-Based Outlier Detection Method Using Forward Search Model Error in Geodetic Networks."

2) In the Abstract, at the line 10, "…removes dependency" is not clear. Is the dependency between observations? Maybe it can be "…removes dependency between observations".

AR2. In lines 38 to 43, It is emphasized that If the rate of successful detection of an outlier using conventional and robust methods is 50%, and one outlier is determined incorrectly, the probability of correctly determining two outliers remains below 50%. This condition is based on the interdependence of each iteration. Incorrect determination at each step also reduces the possibility of identifying more than one outlier in the next step. In other words, it has been stated that the decision taken in one iteration stage directly affects the reliability of the results in the following decision stage. Eliminating this dependency will mean that outliers can be detected more reliably when investigating observation groups based on the minimum variance principle. So the sentence has been clarified as follows: "…variance and eliminates the interdependence of decisions made in iterations." in lines 14-15.

3) In the Abstract, at the line 10, "…to seek the novel outlier detection approach efficiency in…" should be "to seek the efficient outlier detection approach in…".

AR3. In line 10, the expression has been corrected to "to seek an efficient outlier detection approach in …" in line 15.

4) In the Abstract, at the line 14, "(i.e., 1<m<4)" should be deleted.

AR4. In line 18, we have removed "(i.e., 1<m<4)" as per the referee's suggestion.

5) In the Abstract, at the line 14, "Besides, the Forward Search of Model Error (FSME) is…" should be changed as "Besides, proposed model is…".

AR5. In line 18, we have adjusted the wording to "Besides, the proposed model is..."

6) Author has used "model" or "approach", in my opinion only the "model" can be used. Please, check the manuscript.

AR6. We have revisited the manuscript to ensure consistency in using either "model" or "approach." Following the referee's recommendation, we have standardized the usage of "model" throughout the manuscript.

7) In the introduction, at the line 19, "gross errors" should be "outliers". Not only the gross errors, but also outliers have contaminated effects on the results of LSE.

AR7. "Gross errors" has been changed to "outliers" in line 23.

8) In the introduction, at the line 20, "estimation" should be "estimated".

AR8. "Estimation" has been changed to "estimated" in line 24.

9) In the introduction, at the line 25, "Batilovic et al., 2020 or 2021?

AR9. It has been corrected as "Batilovic et al., 2021" in line 29.

10) In the introduction, at the line 26, "low efficiency" can be used instead of "unreliability".

AR10. "Unreliability" has been changed to "low efficiency" in line 30.

11) In the introduction, at the line 28, "…these novel methods…" should be "these mothods".

AR11. "These novel methods" has been changed to "these methods" in line 32.

12) In the introduction, at the line 29, "the reliability…" should be "the reliabilities…".

AR12. "the reliability" has been changed to "the reliabilities" in line 33.

13) In the introduction, at the line 32, "IF" should be "influence function (IF)".

AR13. "IF" has been expanded to "Influence Function (IF)" in line 25, where it is mentioned first.

14) In the introduction, at the line 32, "Maronna et al., 2006 or 2019?

AR14. It has been corrected to "Maronna et al. 2019" in line 36.

15) In the introduction, at the line 35, the sentence "Multiple outliers can be identified at most the number of possible outliers by repetitive test procedures" is not clear. Please, rewrite the sentence clearly.

AR15. The sentence has been rewritten: “Multiple outliers can be identified at most the number of possible outliers ($m_{max} \leq \frac{n-u}{2}$) by repetitive test procedures.” in line 39-40.

The number of maximum possible outliers $m_{max} \leq \frac{n-u}{2}$ given by Hekimoglu (2005).

The reference (Hekimoglu, S. (2005). Increasing reliability of the test for outliers whose magnitude is small. Survey Review, 38(298), 274-285.) has been added in line 367-368.

16) In the introduction, at the line 44, “conventional method” should be “robust methods”. Please check the reference.

AR16. "Conventional method" has been changed to "robust methods" in line 48.

17) In the introduction, at the line 45, “…observation(s) is included as an additional unknown parameter in the…” should be ““…observation(s) is(are) included as an additional unknown parameter(s) in the…”

AR17. The phrase has been corrected to "observation(s) is(are) included as an additional unknown parameter(s) in the…" in lines 49-50.

18) In the introduction, at the line 53, “…value were flagged…” should be “…value are flagged…”.

AR18. The phrase has been changed to "…value are flagged…" in line 57.

19) In the introduction, at the line 54, “combination pace…” should be “combination step…”.

AR19. The phrase has been revised as suggested in line 58.

20) In the introduction, at the line 56, the sentence “The primary purpose of this study is to apply seek the proposed outlier detection method efficiency in geodetic networks.” should be changed as sentence “The primary purpose of this study is to apply the proposed outlier detection method to geodetic networks and to seek its efficiency.”

AR20. The sentence in lines 60-61 has been revised as suggested.

21) In the section 2, at the line 63, “…**P** a positive definite weight…” should be “…**P** be weight…”.

AR21. The phrase has been corrected to "…P be a weight…" in line 67.

22) In the section 2, at the line 64 “…$\mathbf{x}_{ux1}$ a vector…” should be “…$\mathbf{x}_{ux1}$ be a vector…”; “…$\mathbf{l}_{ux1}$ an observation…” should be “…$\mathbf{l}_{ux1}$ be an observation…”; “…$\boldsymbol{C}_{ll_{nxn}}$ an a priori…” should be “…$\boldsymbol{C}_{ll_{nxn}}$ be a priori…”.

AR22. Corrections have been made in lines 67-70 to ensure proper grammar and syntax.

23) In the section 2, at the line 65, “…$\mathbf{Q}_{\mathrm{ll}_{nxn}}$ a weighted…” should be “…$\mathbf{Q}_{\mathrm{ll}_{nxn}}$ be a weighted…”; “…σ_0^2 an a priori…” should be “…σ_0^2 be a priori…”; “…where n and u a number…” should be “…where n and u are the number…”.

AR23. Lines 68-69 has been revised to ensure accurate grammar and syntax.

24) At the line 76, "…the following hypothesis" should be "…the null hypothesis."

AR24. The phrase has been corrected to "…the null hypothesis" in line 80.

25) At the line 79, in the Eq.(6) and at the line 80, "τ" should be "w"; also, τ_i and w_i should be explained after related equations.

AR25. The symbol "τ" has been changed to "w" in Eq.(6). Explanations for τ_i and w_i have been added after the related equations in lines 84 and 87.

26) At the line 84, "are used iteratively…" should be "is used iteratively…".

AR26. The phrase has been corrected to "is used iteratively" in line 88.

27) At the line 90, the reference "Huber 1964" should be added to the reference list.

AR27. The reference "Huber 1964" has been added to the reference list in line 377.

28) At the line 100, "$\hat{\mathbf{x}}^{\mathbf{k}}$" should be "$\hat{\mathbf{x}}^{\mathbf{r}}$".

AR28. The symbol "$\hat{\mathbf{x}}^{\mathbf{k}}$" has been corrected to "$\hat{\mathbf{x}}^{\mathbf{r}}$" in line 104.

29) At the line 103, is "3σ" a priori or a posteriori?

AR29. The phrase "3σ" is corrected as "$3\sigma_0$" in lines 110 and 228. Also, expressions have been added to lines 236 and 242.

30) At the line 128, in the Eq.(22), is it (+) inverse or (-1) inverse?

AR30. It is pseudo (+) inverse.

31) At the line 155, the expression square root should be removed. Because, variance is calculated with the Eq.(30), not standard deviation.

AR31. The equation has been corrected in line 157.

32) In the section 3, Author uses "σ" and "s" for variance. Please, select one of them and use in the text.

AR32. The consistency has been provided using "s" for variance in section 3.

33) I think the sentence at the line 178 "whether the model……or not." should be move to the line 175 (after α = 0.05; before the sentence "if both…".

AR33. The sentence has been moved to line 178.

34) At the line 189, "Hekimoglu and Koch (2000) have…" should be "Hekimoglu and Koch (2000) has…".

AR34. The sentence has been revised to "Hekimoglu and Koch (2000) showed that a finite-sample breakdown point determined the global reliability of an estimator and a test procedure." in lines 192-193.

35) At the line 196, "Erdogan et al. (2019) have…" should be "Erdogan et al. (2019) has…"

AR35. The phrase has been corrected in line 198.

36) At the line 199, "…biased…" should be changed as "…contaminated…".

AR36. The word "biased" has been changed to "contaminated" in line 202.

37) At the line 205,"…novel methods…" should be "proposed methods…".

AR37. The phrase has been changed to "…proposed methods…" in line 207.

38) At the line 221, "…for Robust methods…" should be "…for robust methods…".

AR38. The phrase has been changed to "…for robust methods…" in line 226.

39) At the line 223, "…the A priori variance…" should be "…the a priori variance…".

AR39. The phrase "…the A priori variance…" has been changed to "…the a priori variance…" in lines 227 and 232.

40) At the line 225, "Pope's test had a lower MSR than Baarda's did. However, the MSRs of the FSME (Forward Search of Model Error) are…" should be "Pope's test had a lower MSR than Baarda's test. However, the MSRs of the FSME are…"

AR40. The phrase has been corrected to "Pope's test had a lower MSR than Baarda's test. However, the MSRs of the FSME are…" in line 230.

41) At the line 231, maybe "affect" can be used instead of "impact".

AR41. The word "impact" has been changed to "affect" in line 236.

42) At the line 241, the reference "Durdag 2020" should be added to the reference list.

AR42. It has been corrected as "Durdag 2022" in line 246.

43) At the line 259, the reference "Durdag 2021" should be added to the reference list.

AR43. It has been corrected as "Durdag 2022" in line 264.

44) In the text, Author uses different types of outliers. How did Author define influential outliers? Please add some explanations.

AR44. The sentence "An influential outlier is a situation that, independently or when combined with other biased observations, adversely affects the outcomes of an analysis. Even a single influential outlier may ruin the estimation parameters." has been added in lines 209-211.

45) At the line 292, the sentence "When the redundancy of the observations decreases…" should be "When the redundancies of the observations decrease…".

AR45. The sentence has been corrected to "When the redundancies of the observations decrease…" in line 296.

46) In the conclusion, the first and second sentences may be changed as “This study was designed to determine the usability of the FSME method in geodetic networks. For this aim, FSME methods have been applied to the leveling network. The design of the FSME method is based on identifying the minimum variance from all possible combinations that assume observations as model errors in the Gauss-Markov model. Although, only leveling network has been simulated, the functional and stochastic models of FSME methods can be applied to all type of geodetic networks. This method gives….”.

AR46. The sentences have been revised in lines 316-319.

47) Although Hekimoglu and Erdogan (2013) has been added to the reference list, it has not been cited in the text.

AR47. The reference has been removed from the manuscript.

Response to Reviewer Xinyue Yang

CC1: 'Comment on gmd-2023-210', Xinyue Yang, 19 Dec 2023

This paper presents a minimal variance-based outlier detection method, using forward search model error in leveling networks. It differs from traditional additive bias models by considering potential outlier combinations in observations to enhance efficiency in geodetic networks. The manuscript would benefit from revisions supporting the underlying principles of the research design and situating the results within prior work. Specific enhancements needed include:

Author Response to Reviewer Xinyue Yang #

> **AR.** Thank you for your thorough and constructive feedback on our manuscript. The author appreciates the insightful suggestions to improve the clarity and depth of our work.

1. The abstract should more specifically state the research's unique contributions and main findings, clarifying the method's innovative aspects.

> **AR1.** Abstract Clarity: The abstract has been revised in lines 8-19.

2. A clearer description of the proposed method's computational complexity and details of the data or networks used for testing is suggested. Consider adding step-by-step explanations or visual aids.

> **AR2.** Computational Complexity and Method Details: If unclear computational steps or equations are specified to eliminate computational complexity, actions can be taken for more concrete correction. The calculation section in the manuscript presents the step-by-step method and a flow chart. If the parts that are not understood at these stages or need to be detailed are specified, it will be possible to add the necessary explanations. Please access the input data of the leveling networks worked on in the manuscript, such as coefficients, measurement, weight matrices, and so on, from the assets and dataset section under the preprint. References are given about how the data is generated. Any detailed questions regarding this section will be answered, and necessary corrections will be made to the requests with the editor's approval.

3. More comprehensive and detailed experimental setup is advised, particularly regarding geographic network usage, error scenario testing, and comparative analysis.

> **AR3.** Experimental Setup: Different types of errors have been considered for the geodetic networks presented in this manuscript. Calculations were made by examining the occurrence of these outliers in different scenarios (influential outlier, gross error, small and large outliers) and their behavior in 3 different networks. If there is a missing error type or scenario example, please specify more clearly. Missing parts can be added to the revision with the editor's approval.

4. Ensure simulations involving predefined error scenarios encompass a broad range of real conditions and backgrounds. This includes expanding the variety of tested outlier scenarios, integrating more real-world data examples, and testing a wider range of outlier magnitudes and frequencies.

> **AR4.** Simulation Scenarios: Simulation studies were carried out similarly to actual measurements. The random errors added to the measurements are similar to those occurring if the observations were made on the land. Since the study focuses on the concept of reliability, simulation is a necessity. The approaches in the study can also be applied to actual values.

5. In the discussion, enhance the study's contributions and limitations, especially its statistical significance and impact on the field.

> **AR5.** Discussion: Conclusion has been revised in lines 312-330.

Response to Referee #2

RC2: 'Comment on gmd-2023-210', Anonymous Referee #2, 23 Dec 2023

This paper proposed a new method for the detection of outliers in geodetic networks. Detection of outliers is an important issue within the field of geodesy. The paper is well-written, clear, and concise. The subject of the problem, the idea, the theoretical background, and the experimental setup are clearly set and explained. The paper has been accepted after the comments given in the pdf file are considered.

Response to anonymous referee #2

> **AR.** The author greatly appreciates the Referee's thorough review and constructive comments on our paper. We recognize the importance of addressing this challenge to enhance the practicality of our approach. Thank Referee for drawing attention to the strategies proposed by Neitzel (2004) and Ebeling (2014) in the context of the Maximum Subsample Method (MSS). Indeed, the MSS presents a promising avenue for reducing the number of combinations in outlier detection procedures. As Ebeling demonstrated for deformation monitoring, the MSS could be a valuable tool to streamline our method's applicability, especially in more extensive geodetic networks.

The paper (gmd-2023-210) titled “Minimal variance-based outlier detection method using forward search model error in a leveling network” proposed a new method for the detection of outliers in geodetic networks. Detection of outliers is an important issue within the field of geodesy. The paper is well-written, clear, and concise. The subject of the problem, the idea, the theoretical background, and the experimental setup are clearly set and explained. However, there is one important issue in the paper I must point out.

- It is obvious that this approach quite easily leads to a large number of combinations depending on how many observations and outliers there are. For example, for a geodetic network with 20 observations, of which 3 observations are contaminated with outliers, the number of combinations is

$$\binom{20}{1} + \binom{20}{2} + \binom{20}{3} = 20 + 190 + 1140 = 1350.$$

Since we often encounter geodetic networks of several hundreds of observations in practice, it is necessary to consider and propose a strategy for reducing the number of combinations. For example, Neitzel (2004) and Ebeling (2014) proposed a few strategies for reducing the number of combinations in the procedure identification of the largest congruent group of points (MSS - maximum subsample method).

Ebeling, A. (2014). Ground-Based Deformation Monitoring. PhD Thesis. Calgary: University of Calgary, Department of Geomatics Engineering.

Neitzel, F. (2004). Identifizierung konsistenter Datengruppen am Beispiel der Kongruenzuntersuchung geodätischer Netze. PhD thesis. München: Deutsche Geodätische Kommission, Reihe C, Nr. 565.

> **AR.** In light of the Referee's suggestion, the author has mentioned the potential benefits of incorporating the MSS to mitigate the computational challenges of many combinations in the potential revised manuscript in lines 325-330. This acknowledgment adds a valuable dimension to our research, and we appreciate the Referee's insightful input.

Response to Chief Editor

CEC1: 'Comment on gmd-2023-210', Juan Antonio Añel, 20 Dec 2023

Dear authors,

Unfortunately, after checking your manuscript, it has come to our attention that it does not comply with our "Code and Data Policy".

https://www.geoscientific-model-development.net/policies/code_and_data_policy.html

You have archived your code on GitHub. However, GitHub is not a suitable repository for scientific publication. GitHub itself instructs authors to use other alternatives for long-term archival and publishing, such as Zenodo. Therefore, please, publish your code in one of the appropriate repositories, and reply to this comment with the relevant information (link and DOI) as soon as possible, as it should be available before the Discussions stage. Also, please, include the relevant primary input/output data.

In this way, if you do not fix this problem, we will have to reject your manuscript for publication in our journal. I should note that, actually, your manuscript should not have been accepted in Discussions, given this lack of compliance with our policy. Therefore, the current situation with your manuscript is irregular.

Also, you must include in a potentially reviewed version of your manuscript the modified 'Code and Data Availability' section, the DOI of the code (and another DOI for the dataset if necessary). Moreover, in the GitHub repository there is no license listed. If you do not include a license, the code continues to be your property and nobody can use it. Therefore, when uploading the model's code to the new repository, you could want to choose a free software/open-source (FLOSS) license. We recommend the GPLv3. You only need to include the file 'https://www.gnu.org/licenses/gpl-3.0.txt' as LICENSE.txt with your code. Also, you can choose other options that Zenodo provides: GPLv2, Apache License, MIT License, etc.

Juan A. Añel

Geosci. Model Dev. Executive Editor

Response to Chief Editor:

AR. Dear Editor; the Code and Data availability section has been corrected as below in lines 338-342:

The current code version is available from the project website: "https://doi.org/10.5281/zenodo.10417506" under the "MIT license." The exact version of the model used to produce the results used in this paper is archived on Zenodo (Utkan Mustafa, D. (2023). Godesist/OutlierDetectionForGeodeticLevelingNetwork: Initial Release (0.1.0). Zenodo. https://doi.org /10.5281/zenodo.10417506), as are input data and scripts to run the model and produce the plots for all the simulations presented in this paper.