Response to Reviewer 1 Comments on "Hybrid Instruments Network Optimization for Air Quality Monitoring**" (Manuscript** AMT-2023-173**)** submitted to ***Atmospheric Measurement Techniques***

The original review comments are included in normal black text, and our responses to each comment follow in *italicized text*.

Reviewer 1

Overall the paper is well written and clearly presents its methodology and results.

I find that there is a conceptual problem with requiring emissions information for PM2.5 as an input to your optimization framework, when, in the absence of a monitoring network, this may not be well quantified (as is the case for the Mumbai example you present). You may want to discuss this as a potential limitation of your work and present some possible solutions, e.g., using existing model or satellite-derived information as a proxy for local concentrations during the network design phase, and/or iteratively chancing the network as newly collected data update the prior estimates of concentrations in the different grid cells. Furthermore, there is a potential disconnect between PM2.5 emissions and PM2.5 concentrations (which are to be measured by the network), with the possible impacts of secondary aerosol formation and pollution transport not being accounted for by using emissions information alone; maybe emissions are being used as a proxy for concentrations, but that was not clarified in the text.

Response: *We appreciate the reviewer's feedback. We acknowledge the challenges associated with quantifying PM2.5 emissions in areas lacking an established monitoring network, as evident in the Mumbai case. We have now included these solutions such as considering existing models or satellite-derived data as proxies for local PM2.5 concentrations during the network design phase. Also, after the placement of instruments using some existing models or proxy data, one could iteratively update the placement of the network as newly collected data update the prior estimates of concentrations in the different grid cells. We have now added this in the starting of the second paragraph in the Conclusions section (see page 19).*

We agree with the distinction between PM2.5 emissions and PM2.5 concentrations. In our approach, we initially prioritize PM2.5 emissions as the foundational data for network design. Our primary objective is to strategically deploy sensors and monitors to identify and analyze

pollution sources in a given area. While we acknowledge the non-linear relationship between concentrations and emissions, the absence of concentration data at the grid level led us to utilize emissions as a proxy. However, the placement of the instruments can be updated as better estimates of PM2.5 concentrations become available after the initial placement of sensors. The above discussion has been added as a footnote on page 4.

Reviewer 1

Similarly, while you clearly state that you are aiming to optimize public satisfaction through your sensing network design, there are many other potential objectives which might be the goal of a monitoring network, e.g., improving estimates of population exposure, monitoring the largest known sources, etc. I would suggest adding some commentary to your conclusions discussing how your approach might be modified to achieve these other objectives.

Response: *Thank you for your valuable feedback. We acknowledge that, in addition to optimizing public satisfaction, monitoring networks can serve various other important objectives. These objectives may include improving estimates of population exposure, monitoring the largest known sources of concern, addressing specific environmental or health concerns, or even optimizing resource allocation. To address these alternative objectives, we could make the following modifications to our approach. First, the objective function could be defined appropriately whether it is optimizing public satisfaction, estimating exposure, or addressing specific environmental issues. Also, depending on the chosen objective, we may need to adapt the data collection methods used. For example, if the goal is to estimate population exposure, we may need to tailor the data collection frequency accordingly.*

The analysis methods and models used for decision-making can be customized based on the objective. For instance, if the goal is to address specific environmental concerns, sophisticated modeling techniques may be employed to assess pollutant dispersion.

When other objective functions are used then the fitness function in the genetic algorithm will get modified. The selection, crossover and mutation operators will not change if the constraints remain the same and there would only be change in the objective function. Similarly, the greedy algorithm will have a modified gain function s^* *and the rest of the algorithm will remain the same provided the constraints in the problem remain the same. Thus, our approach can be*

flexibly adapted to address a range of objectives. The above stated discussion has been included in revised manuscript in lines 436-447 on pages 19 and 20.

Reviewer 1: General Comments:

While you note that sensors and monitors have different capabilities, in your formulation they are treated equally in terms of your utility function (i.e., people will be equally satisfied to be located near a monitor or near a sensor). Could you justify this further, or discuss how your results might differ if monitors were given a higher weight?

Response: *We thank the reviewer for this comment. We accept that utility should not be the same for sensors and monitors if the accuracy of these instruments is known to the users and our optimization formulation can be modified to consider that. However, in many practical air quality monitoring scenarios, users may not be either interested or be able distinguish between data collected from monitors and sensors (if the information related to the type of instrument is not openly available), particularly in a hybrid network. From the user's perspective, the primary concern may be just to obtain reasonable air quality information, rather than worry about the specific source of the data. We have now added this lines 98-102 on page 4.*

Reviewer 1: Specific Comments

1. Line 10: Use of "reasonable" here is a bit unspecific; I suggest "less accurate" or a similar description instead, to contrast them with the reference stations.

Response: *We agree with the reviewer's comment and have replaced the term "reasonable" with "less accurate" in line 11 of page 1.*

2. Line 11: Remove "as"

Response: *We agree with the reviewer's comment and have removed the term "as" on line 11 of page 1.*

3. Line 17: "selects locally best choice" should be "selects the locally best choice"

Response: *We agree with the reviewer's comment and have replaced the term "selects locally best choice" with "selects the locally best choice" in line 18 of page 1.*

4. Line 27: Remove "the"

Response: *We agree with the reviewer's comment and have removed the term "the" in line 27 of page 1.*

5. Line 54: "limitations that" should be "limitations in that"

Response: *We agree with the reviewer's comment and have replaced the term "limitations that" with "limitations in that" in line 55 of page 2.*

6. Line 56: Suggest replacing "in the previous-to-previous paragraph" with "previously"

Response: *We agree with the reviewer's comment and have replaced the term "*in the previous-to-previous paragraph*" with "*previously*" in line 57 of page 2.*

7. Lines 60-61: The distinction between sensors and monitors has already been defined earlier in the paper

Response: *We agree with reviewer's comment and has removed this sentence "In this paper, we propose deploying a combination of low-cost sensors (referred to as sensors) and reference stations (referred to as monitors), termed hybrid instruments, in a specific region."*

8. Line 64: This definition for hybrid instruments has already been stated

Response: *We agree with reviewer's comment and have removed this sentence "We refer to the combination of sensors and monitors as hybrid instruments".*

9. Line 67: "noble" should be "notable"

Response: *We agree with the reviewer's comment and have replaced the term "noble" with "notable" in line 65 of page 3.*

10. Line 70: "Therefore, following" should be "Therefore, the following"

Response: *We agree with the reviewer's comment and have replaced the term "Therefore, following" with "Therefore, the following" in line 68 of page 3.*

11. Line 77: "Next section" should be "The next section"

Response: *We agree with the reviewer's comment and have replaced the term "Next section" with "The next section" in line 75 of page 3.*

12. Line 80: "of hybrid" should be "of a hybrid"

Response: *We agree with the reviewer's comment and have replaced the term "of hybrid" with "of a hybrid" in line 78 of page 3.*

13. Line 84: Consider restating the objective to better explain "people satisfaction", e.g., "Our approach focuses on placing sensors in order to maximize a utility function quantifying popular satisfaction with the sensor placements".

Response: *We agree with the reviewer's comment and have replaced the term "The approach focuses on the utility gain of placement of sensors as per people satisfaction." with "Our approach focuses on placing sensors and monitors in order to maximize a utility function quantifying popular satisfaction with the instrument placements." in lines 82-83 of page 3.*

14. Line 87: "g(d) be" should be "g(d) must be"

Response: *We agree with the reviewer's comment and have replaced the term "g(d) be" with "g(d) must be" in line 88 of page 3.*

15. Equation 1: describe how the parameter theta is set

Response: *We thank the reviewer for his/her valuable feedback. By introducing theta as a decaying parameter of distance, we effectively control the rate at which g(d) decreases as the distance increases. Depending on the largest distances that are considered in a grid network and the precision that is being considered, θ should be appropriately decided. For instance, if the computation precision being used is say about 10^{-5} and the largest distance is say 10 units then $\theta = 1$ might reasonable since $e^{-\frac{10}{1}} = 4.5 * 10^{-5}$. This has now been added as a footnote on page 4.*

16. Line 101: While emissions have an influence on local PM2.5 concentrations, secondary aerosol production and pollution transport also play a role.

Response: *We agree with reviewer's comment. In our formulation, we focus on PM2.5 emissions as a starting point for network design. The goal is to strategically deploy sensors and monitors in an area to identify and analyze sources of pollution. We have now the changed the sentence "PM2.5 emission indicates the level of fine particulate matter in the air in that grid" with "PM2.5 emissions are an indicator of the level of fine particulate matter in the air within that grid (secondary aerosol production and pollution transport also play a role in the concentrations but they are not considered here due to lack of data)." on line 111-113 on page 4.*

17. Lines 106-107: Move this sentence right after the first one in this paragraph.

Response: *We thank the reviewer for his/her valuable insight and hence shifted the sentence "The notations are summarized in Table 1 of appendix." after the first sentence in this paragraph in lines 118-119 on page 4.*

18. Line 110: "where monitor" should be "where a monitor"

Response: *We agree with the reviewer's comment and have replaced the term "where monitor" with "where a monitor" in line 124 of page 5.*

19. Equation 4: It is not clear why sensors deployments should be required, but monitor deployments should not be.

Response: *We thank the reviewer for inquiring regarding the Equation (4) in Problem Statement. The locations for sensors could be hospitals, nursery, malls, market or crowded area etc. where authorities want to know the air quality of that place. However, monitor deployment has not been considered that flexibly because monitors cannot be place anywhere, they need to be set a specific place where electricity is available, they are big and heavy as compared to sensors, skilled engineers would be required for their maintenance. Also, monitors are much costlier and so we cannot put them everywhere. We have added this in lines 135-137 of page 5.*

20. Equation 5: Similarly, it is not. clear why monitor deployments are restricted, but sensor deployments are not.

Response: *We thank the reviewer for this comment. The potential locations for placing instruments can be places with sparse population, water bodies, etc. However, it may not be cost-effective or practical to deploy expensive monitors in such areas and thus monitor deployments are restricted, but sensor deployments are not. The above stated sentence have been added to the manuscript in lines 138-140 of page 5.*

21. Line 115: Please define d(a,b).

Response: *Thank you. We define d(a, b) as the distance between grid a and grid b. We have now provided a clearer definition of d(a, b) in lines 131-132 of page 5.*

22. Line 129: "or" should be "of"

Response: *We agree with the reviewer's comment and have replaced the term "or" with "of" in line 148 of page 6.*

23. Line 156: "carried" should be "carried out"

Response: *We agree with the reviewer's comment and have replaced the term "carried" with "carried out" in line 174 of page 7.*

24. Line 164: describe how the parameter alpha was chosen

Response: *We thank the reviewer for this comment. The choice of the value alpha, set to 10^(-5), was made based on extensive experimentation and consideration of the algorithm's behavior in our optimization framework. Alpha serves as a stopping criterion in our genetic algorithm, influencing its convergence behavior. This specific value was determined through a systematic tuning process. We initially experimented with a range of alpha values, including both higher and lower values, to observe their impact on convergence and the quality of solutions generated. After careful assessment and analysis of multiple runs, we found that alpha = 10^(-5) consistently produced satisfactory convergence behavior while maintaining computational efficiency. The above stated explanation has been added in lines 268-270 of page 11.*

25. Line 171: I believe that a maximization expression is missing in the equation here.

Response: *Thank you. We have now written* $s^* = argmax_s \ \sum_{a=1}^{n} m_a \left(g\left(d'(a, K \cup s)\right) - g\left(d'(a, K)\right)\right)$ *in line 189 of page 7.*

26. Line 175: "reduce" should be "subtract"

Response: *We agree with the reviewer's comment and have replaced the term "reduce" with "subtract" in line 193 of page 7.*

27. Line 179: Stopping criteria are not described for the greedy algorithm.

Response: *Thank you. The greedy algorithm stops whenever P' ≈ 0 or there is not enough amount of budget left for the placement of hybrid equipment that is considered as stopping criteria. A sentence has been added in the manuscript in lines 197-198 of page 7.*

28. Line 190: Please provide citations or links to the World Bank and TERI datasets used here.

Response: *Thank you. Our population data was obtained from an open source site called WorldPop and we have now provided the link for that. We have now provided this as a footnote in page 9. However, the emission data from TERI was obtained after our request and is not available as an open source. However, we have now also provided both the population densities and emissions used for different grids in Surat (please see Figures 2, 3 and 4 on page 10) and the population densities used in Mumbai (please see the footnote on page 18).*

29. Line 220: "of the Mumbai" should be "of Mumbai"

Response: *We agree with the reviewer's comment and have replaced the term "of the Mumbai" with "of Mumbai" in line 357 of page 16.*

30. Line 224: "maintained consistently as above in" should be "the same as in the example for"

Response: *We agree with the reviewer's comment and have replaced the term "maintained consistently as above in" with "the same as in the example for" in lines 361-362 of page 16.*

31. Line 225: "we have larger number" should be "we have a larger number"

Response: *We agree with the reviewer's comment and have replaced the term "we have larger number" with "we have a larger number" in line 363 of page 16.*

32. Line 244: "solution is" should be "solution that is"

Response: *We agree with the reviewer's comment and have made the change in line 383 of page 17.*

33. Figure 5: It is unclear how the size of the grids is being varied; Is this the same example for Mumbai? Are the sizes of grids being reduced, or is the area of coverage being increased?

Response: *We thank the reviewer for the comment. In Figure 12 we illustrate the impact of changing the area of coverage rather than altering the sizes of the grids. Specifically, the size of the grids has been held constant at 1 km x 1 km throughout the example for Mumbai. The variation in Figure 12 pertains to the expansion of the area covered by our monitoring network, which involves increasing the number of grids. This expansion allows us to investigate how our optimization framework performs as the coverage area grows.*

34. Table 1: Describing g(d) as a function of d is not very informative; consider expanding the description and referring back to Equation 1 for the definition.

Response: *We thank the reviewer for pointing out this. Now we have defined the g(d) function in Table 2 as "an individual's satisfaction as a function of his or her distance d to the closest sensor or monitor".*

35. References: It appears that a citation is missing for Lerner et al. 2019

Response: *We thank the reviewer for this comment. In response to your feedback, we have added the reference to Lerner et al. (2019) in the reference section (see page 22).*