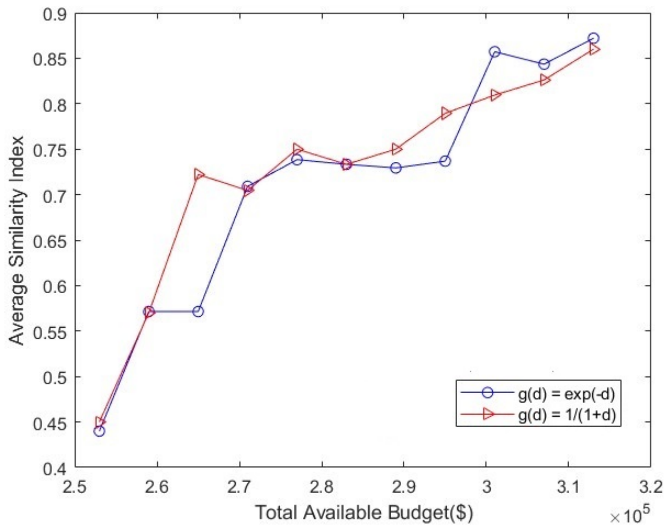
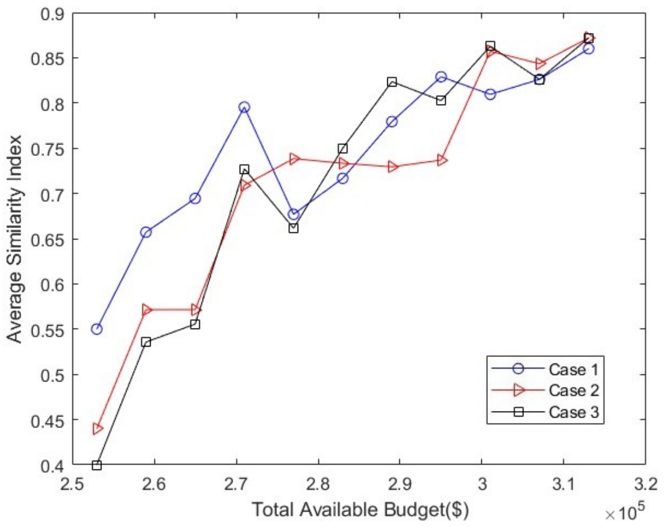
Response to Reviewer comments on "Hybrid Instruments Network Optimization for Air Quality Monitoring" (Manuscript AMT-2023-173) submitted to *Atmospheric Measurement Techniques*

Reviewer

In the new sensitivity studies presented in 3.1.1.1 and 3.1.1.2, while you compare the objective values and trends using the different definitions in the objective, the key question to investigate is how the placement of sensors varies (or does not vary) between these different approaches. To do this, you should define similarity metric (e.g., fraction of the grid cells with instruments which are the same in both approaches). You may need to run multiple trials with the genetic algorithm to get robust results, as the selected locations will be somewhat randomized in that approach.

Response: *We appreciate the referee's feedback. We have now defined a similarity index that quantifies the difference in the placement of hybrid instruments as obtained by different algorithms. Suppose the number of grids where the placement of hybrid instruments is identical is given by k (a grid is said to have identical placement by the two algorithms if the grid contains a sensor as determined by both the algorithms or a monitor as determined by both the algorithms). Also, let the maximum number of hybrid instruments that can be placed in the given constraints be equal to p. Then, similarity index is given by k/p. Since the solution obtained by the genetic algorithm is probabilistic, we tested five runs of genetic algorithm (for a given budget value) and compared the solution obtained by each run to the solution obtained by the greedy algorithm to determine the similarity indices and finally obtained the average similarity index by taking the mean of five similarity indices. The following figure shows the average similarity index for different budget values and for different* $g(d)$ *functions (while keeping equal weights for the percentages of population density and emissions). Note that similarity index is upper bounded by one. Also, we see that as budget values increase the average similarity index for both* $g(d)$ *functions increase. That is because as the budget increases the number of grids at which instruments can be placed increases and both the algorithms usually place sensors at most grids except at a few grids where monitors are placed to meet the requirement of minimum monitors. Note that the average similarity index is around 0.5 for low budget values due to the existence of solutions with that have varying placement but have close objective function values (but as the budget increases the variation in the* *placement reduces as explained before). Also, the average similarity indices obtained by the two $g(d)$ functions are close for most of the budget values. The figure after that shows the average similarity index for different budget values and different cases corresponding to the weights of percentages of population density and PM2.5 emissions (while keeping $g(d)$ as exponential function). Table 2 in the paper shows the weights corresponding to different cases. It can be seen that the average similarity index increases with budget values for the same reason mentioned before. Also, the values of similarity indices are close for most of the budget values. These two figures are in the revised manuscript as Figures 5 and 7 and this discussion is on pages 12 and 14.*





Reviewer

Lines 100-104: While I agree that the public might not distinguish between monitors and sensors in terms of data quality, the designer of the network should. This is a reason why additional objective might be included in the optimization related to the distribution of the monitors, e.g. so that they provide a robust baseline value for the rest of the network. Further, if there is no distinction between these from the point of view of the optimization algorithm, it will always choose the cheaper sensors, once the minimum number of monitors has been placed.

Response: *Thank you. We agree that the designer of the network may want to distinguish between sensors and monitors although the public may not distinguish. Therefore, we now provide an alternate optimization formulation whose objective is to maximize the weighted sum of satisfaction functions from monitors and sensors. Let w_s be the weight corresponding to the satisfaction from sensors and w_m be the weight corresponding to the satisfaction from monitors. Let $d(a)$ be the minimum distance between grid a and any grid containing sensors and $d'(a)$ be the minimum distance between grid a and any grid containing monitors. The remaining parameters and variables mean the same as before. Then, the formulation is as follows:*

$$\max w_s \sum_{a=1}^{n} m_a . g\big(d(a)\big) + \quad w_m \sum_{a=1}^{n} m_a . g\big(d'(a)\big) \qquad \text{(B1)}$$

$$\text{s.t. } \sum_{a=1}^{n} (cx_a + c'y_a) \le P \qquad \text{(B2)}$$

$$\sum_{a \in B} x_a \ge 1 \qquad \text{(B3)}$$

$$\sum_{a \in C} y_a = 0 \qquad \text{(B4)}$$

$$\sum_{a=1}^{n} y_a \ge h \qquad \text{(B5)}$$

where $d(a) = \min_{b \in V} \{x_b . d(a,b) + \overline{d}(a).(1 - x_b)\}$,

$\overline{d}(a) = \max_{b \in V} d(a,b)$ *and* $d'(a) = \min_{b \in V} \{y_b . d(a,b) + \overline{d}(a).(1 - y_b)\}$.

Thus, the relative values of the weights w_s and w_m decide the relative importance being given to monitors and sensors. Typically, w_m should be chosen larger than w_s as monitors are more accurate than sensors. One could solve the above formulation with minimal changes to the proposed genetic and greedy algorithms. This formulation is now provided in Appendix B of the revised manuscript.

Also, we agree that in the original formulation since there is no distinction between sensors and monitors, the optimization algorithms will usually select sensors after fulfilling the requirements of minimum number of monitors since sensors are cheaper. That is why in the greedy algorithm that we proposed, we select sensors (until the budget allows) once the constraints (4), (5) and (6) are met. Please check lines 199-202 on pages 7 and 8.

Reviewer

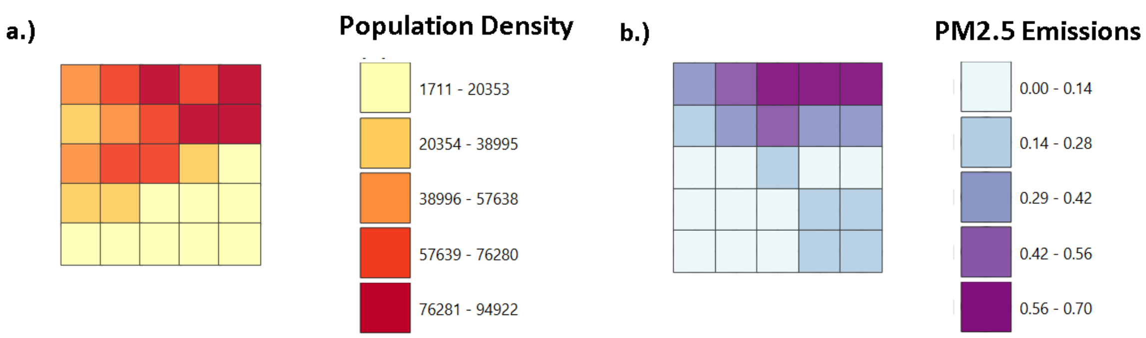
Lines 205-233: Suggest that this example be relocated to the appendix.

Response: *We thank the reviewer's feedback and have now shifted this example to Appendix C of the revised manuscript.*

Reviewer

Figures 3 and 4: Suggest these be presented as shaded grids where the shading intensity is proportional to the value. Ideally, these could also be overlaid on the map of Surat City.

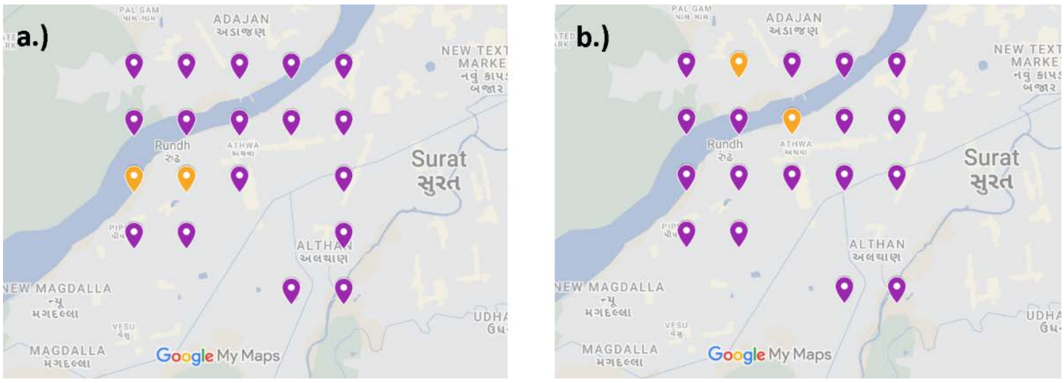
Response: *We thank the reviewer's feedback. We have now presented these grids with shades that reflect the respective intensities. Please check Figure 1 in the revised manuscript.*



Reviewer

Lines 264-267: The differences in placement are immaterial; as the sensors and monitors provide the same value, any placement covering all 25 grid cells which matches the constraints will have the same value. In the case of the greedy algorithm, at least, since the monitors are placed first, they represent the highest-valued pair of locations. The Genetic algorithm will have no such prioritization. It might be more illustrative to present a result where the budget only allows about half of the cells to contain instruments; comparing these might reveal more about the different placement approaches of the algorithms. Still, any difference in the placement of monitors v. sensors in the GA approach will be mostly random. This should be noted in the discussion of results.

Response: *We thank the reviewer's feedback. We have now replaced Figure 2 to show the results corresponding to the budget value equal to $295000 (instead for the budget of $313000 that was shown earlier). Now all the 25 girds are not covered with instruments (which is the case when the budget is $313000). Also, the solutions obtained by both the genetic and greedy algorithms have two monitors and seventeen sensors although the spatial placement of these instruments is not identical (it is possible that two solutions with very different looking spatial distribution can have the same objective function value). Note that there is no scope to further add any instrument in the solution of any of the algorithms as 2*122000 + 17*3000 = 295000. However, if the budget is sufficiently large and the optimal solution involves covering all the grids then genetic algorithm can provide solutions where at some places interchanging sensors with monitors will not change the value of the solution (because the objective function does not differentiate between monitors and sensors). We have now mentioned this point in lines 229-234 of page 9 and in lines 250-253 of page 10.*



Reviewer

Lines 300-324: The motivation for presenting this is not clear to me. You seem to be illustrating that the sub-optimal solutions are not optimal, which is intuitive. This may also be better suited for the appendix.

Response: *Thank you. We have now shifted this example to Appendix D of the revised manuscript.*

Reviewer

Line 328: I suggest not putting the equation in the section heading.

Response: *Thank you. We have now replaced '3.1.1.1 Results for g(d)=1/(d+1)' to '3.1.1.1 Sensitivity analysis with another g(d) function' at line 269 on page 11 of the revised manuscript.*