# CAOR-D-22-01397
# "The Hampered K-Median Problem with Neighbourhoods"

## Answer to Reviewers' Comments

### December 27, 2023

We wish to thank the editors and reviewers for their valuable comments and advice which allowed us to further improve the quality and presentation of our article through this revision.

We revised the manuscript taking into account all the suggestions of Reviewers 1 and 2. We highlight all the changes in the revised manuscript in blue. In the following, we report our changes inside the coloured textboxes.

---

### Reviewer 1

*This paper considers the hampered k-median problem with neighbourhoods. This is a continuous location problem on a 2-dimensional space in which there are both neighbours and barriers present. The authors claim that the problem has applications in the delivery industry and in the areas of inspection and surveillance activities.*

*They develop valid mixed-integer formulations for two versions of the problem, first assuming that neighbourhoods are not visible from one another, and later for the case where this assumption is dropped. They develop valid inequalities and present computational experiments that both aim to find the optimal solution and also quick solutions (of good quality) using a metaheuristic.*

### Answer R1.1

We appreciate your comments that summarise our work.

---

*From an applications perspective, the contribution is minimal as the applications have not been explored. There is no case study presented, and no effort made to speak to data requirements and availability.*

### Answer R1.2

To deal with the Reviewer's requirements, along the manuscript we present different real-life scenarios where these models can be adopted. In the following, we remark on different paragraphs that have been included to argue the combination of neighbourhoods and barriers in the $k$-median:

- The use of neighbourhoods has interest on drone delivery and inspection problems. Neighbourhoods can represent regions that the drone must reach and where customers are willing to pick up the orders (they can be seen as uniform probability densities) in the delivery industry. Moreover, they can also be used for modelling some areas that must be inspected by the drone (whenever visiting a point of these areas is enough to consider them as inspected). **Section 1. Introduction: Paragraph 3.**

- The use of barriers in the $k$-median, that is known as $k$-median problem with barriers are deeply studied in the literature (see Klamroth (2002)). In this case, there exist some areas that cannot be traversed, and it is necessary to compute the minimum distance between each pair of points in the graph induced by the facility locations and demand points. In the context of planar location modelling, this problem represents restrictions that appear in a real-life context. One of them can be the case in which

there are regions (called forbidden regions) in which the placement of a facility is forbidden but transportation is still possible. They can model state parks or regions where geographic characteristics forbid the construction of the facility. For a survey of location problems with forbidden regions, see Hamacher and Nickel (1995) or Nickel (1995). The case in which transport is possible, but only at higher cost (called congested regions) is studied in Butt and Cavalier (1996) or Mitchell and Papadimitriou (1991). In this case, in these regions, different travel speeds or travel costs are considered. Finally, the existence of military areas, buildings, mountain ranges, lakes, large rivers, and highways cannot allow transportation. These examples of barrier regions involve circuit board design (LaPaugh, 1980), pipe network design for ships (Wangdahl et al., 1974), or location and routing with robots (Lozano-Pérez and Wesley, 1979; Blanco et al., 2022). **Section 1. Introduction: Paragraph 4.**

- From a drone routing perspective, these barriers can simulate buildings in urban areas that drones cannot cross. Another application can be seen as obstacles that pedestrians must avoid in urban or rural areas. **Section 1. Introduction: Paragraph 5.**

We describe a realistic application of the problem studied in this document to perform the last mile delivery. We focus on the problem of allocating $k$ vans from where some drones depart to deliver some products to the represented population. In particular, we consider a neighbourhood in the city of Córdoba called Fátima, which is located in the northeast of the city. We express sources that are represented by green circles that simulate zones in which vans can park; and targets, also modelled by circles, that define the zones in which customers are willing to pick up the delivery. Barriers represent buildings that drones cannot cross. To simplify the problem, we assume the drones have enough endurance to go from any source to any target and they cannot ascend to avoid the buildings, and once the number of vans is fixed, we can know the number of required drones assigned with each van. **Section 6. Case study.**

We also discuss data availability by exposing that the advance of technology permits obtaining maps with accurate representations of two-dimensional obstacles by means of various software tools designed for geographic information systems (GIS). An open source such as Quantum GIS (QGIS Development Team, 2009) provides robust capabilities to create, edit and export maps with detailed two-dimensional building footprints and attributes. These tools enable users to incorporate building information into spatial analyses, urban planning, and environmental assessments. We include some examples where QGIS has been successfully used. The work of Arsanjani et al. (2013) explores the use of open-source GIS software to map and analyse urban structures, highlighting the importance of accurate two-dimensional building data. In drone routing, the work of Mangiameli et al. (2013) proposes an innovative approach to the construction of a map of obstacles based on Geographic Information Systems (GIS) technologies. **Section 1. Introduction: Paragraph 6.**

Finally, we have also restructured our computational results section including a full comparison between SCF and MCF models. In addition, we have included the analysis of the influence of the link distance in the difficulty of problem solving. Due to the large amount of tables to compare all the factors in the problem, namely different formulations, weight $\omega_L$ for the link distance, number of neighbours and barriers, use of initial solution, we have organised the results in tables comparing statistical significance ($p$-values). We believe that this presentation is much more informative and compact. **Section 5. Computational Experiments.**

*From a theory perspective, the contribution is better but in my opinion is not significant enough to justify publication in C&OR. It is in my view a good incremental paper in this area of research, but I don't think that is good enough for a journal like C&OR.*

**Answer R1.3**

This paper analyzes the $k$-Median problem with neighbours and polygonal barriers. This problem is new and has important actual applications in different fields, as mentioned in the paper. Its analysis and formulation cannot be seen as a simple incremental contribution, since the tools required to analyse and solve it go beyond any standard overlapping of techniques used in the separated problems. We present a valid formulation that relies, apart from some already known blocks of constraints taken from $k$-median and network design, on new tools borrowed from computational geometry. Specifically, one of our important contributions is presented in subsection 3.4. There, we derive a new test, representable by linear constraints, to check whether given two arbitrary variable vertices $P, P' \in V_X$, the edge $(P, P')$ belongs to $E_X$, with $X \in \{\text{KMPHN}, \text{KMPN}\}$, i.e., whether the line segment $\overline{PP'}$ does not intersect with any barrier of $\mathcal{B}$. The referee may note that this is a new methodological contribution to the field and that it will be used in the future in different problems involving barriers. Apart from the exact method, we also propose a new matheuristic based on the knowledge obtained from the exact analysis. All together, this brings a new perspective to the problem considered and contributes to a better understanding of the logistic problems with neighbours and barriers.

*The authors propose a new variant of the k-median problem in which potential facility and client locations are given as neighborhoods (compact regions that can be described using second order cone constraints). In addition, there are line segments representing barriers, and allocations between clients and open facilities should not cross a barrier - instead, they can "touch" the end-points of these line segments, forming barrier-free paths between each allocated client-facility pair. The authors consider two versions of the problem - with and without "hidden neighborhoods".*

*The authors propose a MINLP problem formulation with bilinear constraints using binary variables. Then, they use standard linearization techniques, to obtain a mixed integer SOCP (for one problem variant). They use the concept of single-commodity flows to model the client-facility paths. They propose a simple initialization heuristic as well and then they evaluate the new model with and without this initialization heuristic on a set of instances with up to 80 nodes.*

*The content is interesting and relevant for the readership of COR, however, there are several major points that need to be improved before the article can be considered for publication. These points are listed below.*

> **Answer R2.1**
>
> We thank the referee for his/her detailed evaluation which we appreciate very much.

## Clarity of problem definition

*On page 3, Section 3.1: please provide major assumptions regarding the neighborhoods. Are they supposed to be compact, closed, convex sets? Maybe already here you should emphasize that these sets should be such that one can describe them using SOCP constraints?*

> **Answer R2.2**
>
> In *Section 2.1. Parameters and Assumptions of the Problem*, we have included the following paragraph:
>
> In this work, the set of neighbourhoods $\mathcal{N} = \mathcal{S} \cup \mathcal{T}$ are assumed to be compact, closed, and convex sets. These sets can be represented by second-order cone constraints, as explained in Section 3.3.

*Moreover, condition $A_4$ is not clear to me - it is not clear to me why it is important for the input of the H-KMPN that there is no straight-line connecting a client-facility pair without crossing any barrier? And why introducing such a possibility causes later non-convexities in the linearized MISOCP formulation. These points have to be better explained on page 10 (where the authors just state that product of continuous variables in alpha-constraints introduces non-convexities).*

> **Answer R2.3**
>
> In the part of the manuscript in which $\alpha$ constraints are introduced, we have introduced the following paragraph:
>
> The fact of considering the possibility of going directly from one neighbourhood to another leads to include product of continuous variables, presented in the determinants of the $\alpha$ constraints of the model. Specifically, for the tuples $(P, Q, Q') \in V_{\mathcal{B}} \times E_{\mathcal{ST}}$, the determinant $\det(P|QQ')$ produces bilinear terms. These products cause the general H-KMPN to become non-convex. However, for the hidden version, since two of the three arguments of the determinant are fixed because $E_{\mathcal{ST}}$ is not considered, the constraints $\alpha$ become linear. This difference justifies the comparison between the two formulations in terms of the computational cost studied in Section 5.

*In Section 2.2. in the definition of the objective function, it should be mentioned that one*

*minimizes the two sums, one of the weighted lengths and one of weigthed link distances. More-over, more motivation for a choice of such objective function is needed.*

---

**Answer R2.4**

We have described with more precision the objective function represented in our problem after the description of the first formulation. We have also inserted a reference that motivates the use of the link distance in this kind of problem:

The objective function takes into account both the Euclidean and link weighted distances to join the selected sources with their assigned targets. The term related to the Euclidean distance is divided into two parts. The first accounts for for the variable distances, while the distances in the second are known beforehand. Inspired by a drone routing application, the link distance refers to the vehicle's rotation cost (see Maheshwari et al. (2000) for more details).

---

*When introducing the graph $G_X$, it is not clear whether points $P_S$, $P_T$ are given in advance, or whether we are searching for their coordinates (the latter is the case, but this becomes clear much later in the paper after introducing the full MINLP model).*

---

**Answer R2.5**

We have given details in the description of the problem by including the notation for the points selected in $\mathcal{S}$ and $\mathcal{T}$. Additionally, we have better explained how the visibility graph is built by arguing that it is a "variable" graph that depends on the location of its vertices to determine which are its edges. **Section 2. Description of the problem. Paragraph prior to Proposition 1.**

---

*For the set $E_{\mathcal{S}\mathcal{T}}$, shouldn't we also add in the definition that these line segments should not cross any other barrier from $B$?*

---

**Answer R2.6**

Yes. We have inserted the correct description of the set of edges $E_{\mathcal{S}\mathcal{T}}$.

---

*When introducing constraints ($\mathcal{N}$-C), please mention that $P_N$ are variables.*

---

**Answer R2.7**

Thank you. We have explained that $P_N$ represents the coordinates of the variable point selected in $N$.

---

## MINLP formulation

*When introducing model variables, what is the domain of variables - for example, for $\alpha(P|QQ')$, where does $P$, $Q$, and $Q'$ come from? The same has to be clarified for all variables used in the model. I recommend to use notation $f^{ST}_{PQ}$ instead of $f(PQ|ST)$, first, because it is a common way to state flow along a segment $PQ$ for the commodity $ST$ in that way, and second, to avoid confusion with the determinants notation (see minor remarks below). For all introduced constraints, please introduce quantifiers and variable domains for which these constraints are imposed.*

---

**Answer R2.8**

In Section 3, we have performed a deep revision to unify the notation to be more similar to the common way of representing the variables. These changes are described in the following:

- $y(S) \rightarrow y^S$.
- $x(ST) \rightarrow x^{ST}$.
- $f(PP') \rightarrow f_{PP'}$.

---

*The authors state that they use single-commodity flow formulation to model client-facility allocation paths, however, their model is a mix between single- and multi-commodity formulation (SCF and MCF, respectively). This has to be corrected, and, in addition, there are indeed two ways to model it, as described below.*

- *In the MCF approach, we need binary variables $f_{PQ}^{ST}$ to be set to one for each ST pair, if line segment connecting P and Q is used. For constraints (f-C), what is the definition of $f(PQ)$ and what is the link between $f_{PQ}^{ST}$ and $f(PQ)$? I believe constraints (f-C) should be replaced by stronger ones: $f_{PQ}^{ST} \leq \varepsilon(PQ)$ for all P, Q (please add the variable domain) and for all S, T (please add variable domain). Finally, there is a quantifier missing in the flow-preservation constraints (flow-in equals flow-out): for all P is missing.*

- *In the SCF approach, we would need flow variable indexed by source nodes S, modeling a tree routed at S and connecting all targets T such that $x(ST) = 1$. These variables can then be denoted as $f_{PQ}^{S}$ and their value is integer, counting the number of targets T allocated to S, that use segment PQ. In that case, the constraints (f-C) would read $f_{PQ}^{S} <= |T|\varepsilon(PQ)$, for all P,Q,S and in the flow preservation constraints, we will have that the RHS for the flow-preservation constraints for $P = S$ is equal to $\sum_{T \in \mathcal{T}} x(ST)$.*

**Answer R2.9**

Thanks for your contribution. We have taken into account your comments to differentiate between two formulations that can model the problem we study. In addition, we have structured Section 3 by splitting the formulation in different blocks of constraints that appear in the problem:

- Section 3.1 describes the classical $k$-median constraints.

- Section 3.2 represents the path source-target by means of single- and multi-commodity approaches as suggested by the referee.

- Section 3.3 states the second-order cone constraints defining neighbourhoods and Euclidean distances.

- Section 3.4 makes use of the results of the computational geometry to model the visibility graph.

- Section 3.5 presents the single- and multi-commodity formulations for the problem and discusses the linearisation of the objective function using McCormick's envelopes.

*The authors state on page 9 "The fourth ones are the flow conservation constraints, where the units of commodity launched from the source must be the number of targets that are assigned*

6

*to that source. The fifth group of inequalities ensures that the trip can reach, in endurance terms, the target T when it starts from the source S." From the first sentence, it seems to me the authors had in mind a SCF formulation, but this is not what the current model is providing. The second sentence mentions the fifth group of constraints which is actually missing in the model. Besides, what is meant with "endurance terms"? In the next MINLP model on page 10, there is a constraint that uses "endurance" but I do not think the endurance is defined earlier in the paper.*

### Answer R2.10

We agree with the referee. In a previous version of the paper, we had considered an endurance term that bounds the total distance between each pair of source-targets. However, although it is possible to model this restriction, we have decided to remove it to simplify the problem.

*Proposition 1 should be mentioned right after the problem definition at the end of Section 2. Besides, please give a formal definition of the k-median with geodesic distances and and a reference where it is shown that the latter problem is NP-hard.*

### Answer R2.11

We agree with the referee. We have stated this result right after the problem definition. We have also included this paragraph that argues for this result:

Note that once a point is fixed in each neighbourhood, the problem that results in the induced graph $G_X$ is the $k$-median with geodesic distances. This problem is a version of the discrete $k$-median in which the distances are computed beforehand by solving a shortest-path problem for each pair source-target in $G_X$. Hence, it is NP-complete (Kariv and Hakimi, 1979).

## Computational results

*It is important to make publicly available instances used in this paper.*

### Answer R2.12

All instances used in this work can be downloaded from Valverde and Puerto (2023).

*Are the authors using some kind of warm-start (sometimes called MIP-start) procedure to provide the solution found by the matheuristic to Gurobi? This should be better clarified.*

### Answer R2.13

**** Done??? **** Carlos comprueba que lo hemos explicado en el paper***

*What is the major idea behind reporting $Gap_{build}$ - why is it important to look at this metric? Is this a solution found by Gurobi without using matheuristic? Please clarify.*

### Answer R2.14

This metric shows the improvement found by the solver over the initial solution computed by our own matheutistic. This is important to assess the quality of the matheuristic and how it may help to close the gap in difficult instances of the problem.

*On page 13 the authors state "Note that the maximum gap between the best incumbent solution obtained by the matheuristic in a time limit of 100 seconds and the best by the exact formulation after an hour is 5.96%." However, this gap is much larger in the tables provided, so this should be clarified.*

> **Answer R2.15**
>
> Reorganisation of Section 5 has changed the results obtained.

*Finally, given that the problem is new and there are no "competitors", it would be interesting to see the difference in the performance between the SCF and MCF model mentioned above.*

> **Answer R2.16**
>
> Following the referee's advice we have restructured our computational results section including a full comparison between SCF and MCF models. In addition, we have included the analysis of the influence of the link distance in the difficulty of problem solving. Due to the large number of tables to compare all the factors in the problem, namely different formulations, weight $\omega_L$ for the link distance, number of neighbours and barriers, use of initial solution ..., we have organised the results in tables comparing statistical significance ($p$-values). We believe that this presentation is much more informative and compact.

## Minor concerns

- *Change on-the-self by off-the-shelf.*

- *Section 3.2., line 2: replace $\in E_X$ with "belongs to $E_X$".*

- *Please explain what in your notation $det(P_1|PQ)$ means.*

- *When introducing gamma variables, mention that they are also obtained as McCormick linearization.*

- *Please simplify the notation and improve the formatting of the MINLP model on page 9 (there is no need to call the set of vertices $V_{KMPHN}$, you can use much shorter names for a better readability).*

- *Please use wlog or w.l.o.g. consistently.*

- *Please use dots instead of comma for showing decimal numbers.*

> **Answer R2.17**
>
> Thanks for carefully reading the manuscript. The changes that we have made are:
>
> - We have changed *on-the-shelf* by *off-the-shelf.*
>
> - We have replaced $\in E_X$ by  *that belongs to $E_X$.*
>
> - We have introduced a separated subsection in the **Section 2** that summarises the notation used along the manuscript.
>
> - We have introduced a sentence for the $\varepsilon$ variables that are obtained as a standard McCormick linearisation.
>
> - We have made our efforts to improve the notation to make it more clear to the reader.
>
> - We have changed *w.l.o.g.* by *wlog.*
>
> - We have adopted the decimal point to show the decimal numbers.

# References

Arsanjani, J. J., Helbich, M., and de Noronha Vaz, E. (2013). Spatiotemporal simulation of urban growth patterns using agent-based modeling: The case of Tehran. Cities, 32:33–42.

Blanco, V., González, G., Hinojosa, Y., Ponce, D., Pozo, M. A., and Puerto, J. (2022). Network flow based approaches for the pipelines routing problem in naval design. Omega, 111:102659.

Butt, S. E. and Cavalier, T. M. (1996). An efficient algorithm for facility location in the presence of forbidden regions. European Journal of Operational Research, 90(1):56–70.

Hamacher, H. W. and Nickel, S. (1995). Restricted planar location problems and applications. Naval Research Logistics (NRL), 42(6):967–992.

Kariv, O. and Hakimi, S. L. (1979). An Algorithmic Approach to Network Location Problems. II: The p-Medians. SIAM Journal on Applied Mathematics, 37(3):539–560.

Klamroth, K. (2002). Single-Facility Location Problems with Barriers. Springer, New York, NY.

LaPaugh, A. S. (1980). Algorithms for Integrated Circuit Layout: An Analytic Approach. PhD thesis, Massachusetts Institute of Technology.

Lozano-Pérez, T. and Wesley, M. A. (1979). An algorithm for planning collision-free paths among polyhedral obstacles. Communications of the ACM, 22(10):560–570.

Maheshwari, A., Sack, J.-R., and Djidjev, H. N. (2000). Chapter 12 - Link Distance Problems. In Sack, J. R. and Urrutia, J., editors, Handbook of Computational Geometry, pages 519–558. North-Holland, Amsterdam.

Mangiameli, M., Muscato, G., Mussumeci, G., and Milazzo, C. (2013). A GIS application for UAV flight planning. IFAC Proceedings Volumes, 46(30):147–151.

Mitchell, J. S. B. and Papadimitriou, C. H. (1991). The weighted region problem: Finding shortest paths through a weighted planar subdivision. Journal of the ACM, 38(1):18–73.

Nickel, S. (1995). Discretization of Planar Location Problems. Shaker Verlag, Aachen.

QGIS Development Team (2009). QGIS Geographic Information System. Open Source Geospatial Foundation.

Valverde, C. and Puerto, J. (2023). Instances for the Hampered k-Median Problem with Neighbourhoods.

Wangdahl, G. E., Pollock, S. M., and Woodward, J. B. (1974). Minimum-Trajectory Pipe Routing. Journal of Ship Research, 18(01):46–49.