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

## Answer to Reviewers' Comments

### November 22, 2023

We wish to thank the editors and reviewers for their valuable comments and advices 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 highlighted in blue all the changes in the revised manuscript. 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**

---

*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**

---

*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**

---

*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**

*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 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. Moreover, more motivation for a choice of such objective function is needed.*

**Answer R2.4**

*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, bit this becomes clear much later in the paper after introducing the full MINLP model). For the set $E_{ST}$, shouldn't we also add in the definition that these line segments should not cross any other barrier from B? When introducing constraints (N-C), please mention that $P_N$ are variables.*

**Answer R2.5**

## 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 $f_{PQ}^{ST}$ 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).*

**Answer R2.6**

Done.

*For all introduced constraints, please introduce quantifiers and variable domains for which these constraints are imposed. 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 the 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 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 R2.7**

*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 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 R2.8**

*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 R2.9**

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

**Answer R2.10**

## Computational results

*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 R2.11**

*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 R2.12**

*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.*

*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 R2.13**

*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 R2.14**