AGILE: GIScience Series, 4, 20, 2023. https://doi.org/10.5194/agile-giss-4-20-2023
Proceedings of the 26th AGILE Conference on Geographic Information Science, 2023.
Editors: P. van Oosterom, H. Ploeger, A. Mansourian, S. Scheider, R. Lemmens, and B. van Loenen.
This contribution underwent peer review based on a full paper submission.

Developing capacitated *p*-median location-allocation model in the `spopt` library to allow UCL student teacher placements using public transport

Nick Bearman [1], Rongbo Xu [2], Patrick J. Roddy [3], James D. Gaboardi [4], Qunshan Zhao [5], Huanfa Chen [6], and Levi Wolf [7]

[1]Geospatial Training Solutions and University College London, London, UK
[2]Urban Big Data Centre, University of Glasgow, Glasgow, UK
[3]Advanced Research Computing, University College London, London, UK
[4]Geospatial Science and Human Security Division, Oak Ridge National Laboratory, USA
[5]Urban Big Data Centre, University of Glasgow, Glasgow, UK
[6]CASA, University College London, London, UK
[7]University of Bristol, Bristol, UK

Correspondence: Nick Bearman (nick@nickbearman.com)

Abstract. Location-allocation is a key tool within the GIS and network analysis toolbox. In this paper we discuss the real world application of a location-allocation case study (approx 800 students, 500 schools) from UCL using public transport. The use of public transportation is key for this case study, as many location-allocation approaches only make use of drive-time or walking-time distances, but the location of UCL in Greater London, UK makes the inclusion of public transport vital for this case study. The location-allocation is implemented as a capacitated p-median location-allocation model, using the `spopt` library, part of the Python Spatial Analysis Library (PySAL). The capacitated variation of the p-median location-allocation problem is a new addition to the `spopt` library, which this work will present. The results from the initial version of the capacitated p-median location-allocation problem has shown a marked improvement on public transport travel time, with public transport travel time reduced by 891 minutes overall for an initial sample of 93 students (9.58 minutes per student). Results will be presented below and plans for further improvement shared.

Keywords. Location-allocation, `spopt` library, placement allocation, public transport

1 Introduction

Location-allocation is a key element of GIS and network analysis, and has a wide variety of applications across the public, private, and academic sectors. However, within location-allocation, there are many models, with different requirements for specific applications. This makes the broad application of location-allocation quite difficult, and many potential users of the technology are unaware of its potential application. Additionally, location-allocation is quite widely adopted with walking or drive time analysis, but much less so with public transport (Kotavaara et al., 2018). Part of the reason for this is the variety of data formats that public transportation frequency data is available in, and the trickiness of deciding on a 'typical' public transport journey.

This article details the development of a capacitated p-median location-allocation model for a specific case study involving student teachers travelling on public transport to their placement schools across Greater London, UK. We will provide a background of location-allocation problems, and our implementation of this, with details of the UCL case study. The app and library are currently in development, and we will report on their effectiveness thus far.

2 Location-Allocation

In numerous instances of public service infrastructure, such as hospitals, schools, fire stations, post offices, and libraries, the ideal scenario is for the facilities to be located near their users (Fredriksson, 2017). However, determining the optimal configuration of facilities based on demand can be a complex task. Unequal service and high costs can arise when the distribution of facilities is unreasonable, or when the number of facilities is either excessive or insufficient. To address these challenges in location problems, the location-allocation model is used to determine the optimal locations for facilities such as factories, warehouses, or service centres, and to allocate customers or demand to those facilities (Daskin, 1997).

This model takes into account factors such as transportation cost, facility costs, demand, and capacity constraints, and generates an optimised plan to realise a more efficient service system based on the existing facility location (Silver et al., 1998). The location-allocation model has been widely used in a variety of fields, including business (Venkatesan and Kumar, 2004), where it is used to identify the optimal location for new stores or distribution centres; and urban planning (Syam, 2008; Pan, 2010; Rahman and Smith, 2000). In healthcare, the location-allocation model can be used to identify the optimal locations for hospitals or clinics, while in disaster response, it can be used to identify the locations for emergency shelters or supply distribution centres (Horner and Downs, 2010).

Location problems are commonly classified into three categories: covering-based problems, median-based problems, and other problems (Daskin, 2008). The covering-based problem can be further divided into Location Set Covering Problem (LSCP), Maximal Covering Location Problem (MCLP), and p-centre location problem. Median-based problems include two main models: p-median location problem, and fixed charge facility location problem (Daskin, 2008). Among these models, LSCP, MCLP, p-centre, and p-median are the most widely used in public facilities (ReVelle, 1989; Coskun and Erol, 2010).

In the case of the UCL case study, the goal is to minimise the travel time for trainee teachers to reach their assigned schools. To meet the specific requirements of this case, we choose to design a new variant of the p-median model, which is one of the widely used models for public facilities. The p-median model is designed to minimise the total travel distance or time of customers to the nearest facility (Serra and Marianov, 1998). This model can provide an optimal solution for the placement of facilities based on the existing demand and location.

3 Case Study: UCL IOE Teacher Placement

Location-allocation has great potential across a wide variety of applications. However, the specificity of these applications can make explaining the application of location-allocation difficult and render applying location-allocation complex at best.

Nick Bearman was approached by the Institute of Education (IOE) within University College London (UCL) to see if there was a way of improving their student placement process. Each academic year, IOE's trainee secondary school teachers (teaching students aged 11–18) have to take two 60-day placements at a school teaching their subject to students. UCL IOE have approximately 800 students to place each year, with around 500 schools offering placements. Currently, this is done using a largely manual process, with the use of `Google Maps My Maps`[1] to locate students and schools, and the `Google Maps` routing tools to evaluate their potential public transport route (including trains, underground, buses, trams *etc.*) to try and allocate all 800 students fairly, with roughly equitable journeys.

This is a clear example of where location-allocation could be a useful tool, to calculate the minimum per-student journey time, whilst ensuring the minimum overall journey time. After some initial discussions with IOE and considering the literature, it was clear that this is possible, but the tools that exist do not meet the requirements (Murray, 2021; Chen et al., 2021).

A particular issue is a need to include public transport data, which is generally not in an easy-to-access form. Several location-allocation tools exist, but their main focus is on walking time or drive time analysis. `ArcGIS Pro` is probably the leading commercial software in this respect, and their tutorials[2] are a great resource for anyone new to this area. However, public transport data is not possible to be integrated into it. A number of other tools were considered, including the `FLP Spreadsheet Solver`[3] (Erdoğan et al., 2019), `PySpatialOpt` (Pulver, 2016), and `Allagash` (Pulver, 2019). However, the `FLP Spreadsheet` is not powerful enough for this application (it had a maximum number of facilities of 200), and `PySpatialOpt` and `Allagash` require a significant amount of work to be usable for this case study.

After some discussions and thought, it was decided to split the problem into two sections: public transport analysis and then location-allocation analysis. Some capacity for public transport did exist within existing tools, but it is not very well-developed.

We proposed that if the public transport option can be summarised to a weighted distance cost, then an OD (origin-destination) matrix could be produced that could then be processed by a location-allocation library. Within Greater London, Transport for London (TfL) run all the public transport (including London Underground, trams, buses and local trains). They have a journey finder tool on their website, which allows travellers to enter an origin, destination and time of departure/arrival and will produce several potential journey routes. The journey finder has an API and, as such, stage one is to take the origin-destination pairs and calculate journey times for all potential routes. UCL IOE approached UCL ARC (Advanced Research

[1] https://www.google.co.uk/maps/about/mymaps

[2] https://pro.arcgis.com/en/pro-app/latest/help/analysis/networks/location-allocation-tutorial.htm

[3] https://www.euro-online.org/websites/verolog/flp-spreadsheet-solver

Computing) to support this, and they provided 3 months of developer time provided by Patrick Roddy to develop the transport side of this work, and support integration of the app with the `spopt` library (Feng et al., 2021, 2022).

Once the public transport element was covered, the location-allocation library used to solve this was needed. After some initial discussions and research, Nick Bearman discovered the `spopt` library, which is actively developed and includes some facility location tools already. The library maintainers (including Qunshan Zhao, Levi Wolf, and James Gaboardi) are looking to expand these features, and together we applied for the funding initiatives from AGILE to develop the additional tools needed for this work. This funding was successful, and University of Glasgow appointed Rongbu Xu to work on this for 3 months. A condition of the funding is to present a workshop at AGILE 2023 on PySAL and spatial optimisation.

4 Current Status / Development of spopt

The Python Spatial Analysis Library (PySAL) (Rey and Anselin, 2007; Rey et al., 2021) is an open-source ecosystem for geographic data science comprised of 4 modules (`lib`, `explore`, `model`, and `viz`), each with related packages. The `spopt` (spatial optimisation) package resides within the `model` module and originated from the `pysal/region` package. With the introduction of facility location modelling, the original package was re-branded as `spopt` providing functionality in two domains: heuristic regionalisation models (`region`) and optimal facility location models (`locate`). Concerning `locate`, a handful of classic facility location models have been implemented in their mixed integer linear program (MILP) formulations, each with its own objective that can generally be thought of as optimally locating a set of facilities in either the most efficient or equitable configuration, or a combination of those two objectives. The initial set of models included the Location Set Covering Problem (LSCP) (Toregas et al., 1971), Maximal Coverage Location Problem (MCLP) (Church and ReVelle, 1974), p-median Problem (PMP) (Hakimi, 1964) and p-centre Problem (PCP) (Hakimi, 1965). The next phase of development introduced the Location Set Covering Problem with Backup coverage (LSCP-B) (Church and Murray, 2018), p-dispersion Problem (PDP) (Kuby, 1987), and Capacitated Location Set Covering Problem-System Optimal (CLSCP-SO) (Church and Murray, 2018).

All models, save the PDP, consider client locations (*e.g.* households), and the PMP and MCLP also explicitly consider client weights (*e.g.* members of the household). However, the inclusion of the CLSCP-SO was the first location model in `spopt` that constrains for each facilities' capacity. Moreover, the CLSCP-SO is much more of an equity-based approach (the objective is 100% client location coverage without a restriction on the number of facilities in the solution). Therefore, the capacitated variation of the classic PMP that is being developed in this work is the first efficiency-centric location model (due to the explicit constraint on the number of solution facilities) that explicitly considers not only client locations and weights, but also the capacity of facilities that will potentially serve the clients.

Future development on `spopt` will provide additional models and novel variations of existing models in both `region` and `locate`. Further, models focused on transportation and routing are also planned.

5 Development of the app and library

The organisation of the work was split in two, to match the design of the overall solution. UCL IOE needed a tool developed to allow them to automate as much as possible of their existing allocation process. This tool needs to make use of the `spopt` library to perform the location-allocation process. Fig. 1 shows the overall process and flows of data.

Patrick Roddy lead work on the development of the app for UCL IOE. *This may or may not ultimately be an app in the strictest definition, but we have used the term to distinguish it from the work taking place in the library.* The majority of this work was focused on pre-processing the data from UCL IOE into a suitable format, and using the TfL API to calculate and collate acceptable potential public transport routes for the students to travel to their placement schools. The TfL API is very flexible and the preferences and requirements of the students will be incorporated into this process in a later version. Currently, students can choose public transport, driving or cycling, but this will be developed to allow a wider range of options, including specific types of public transport (bus, underground, train, etc.) and maximum walking distances. The output journey data also needed to be reformed into a matrix suitable for use with the `spopt` library.

Rongbu Xu is leading work on the development of the existing `spopt` library, to include the new capacitated p-median functions within the library.

According to the requirements of UCL IOE, Rongbu Xu firstly constructed the capacitated p-median model. UCL IOE has approximately 800 students to place each year, with around 500 schools offering placements. The objective is to minimise the total travel time of all students, on the condition that every student is placed at a school, teaching the student's subject. UCL IOE has provided five types of constraints in this problem. We then categorised them into two groups: required and optional. The required constraint is the requirements that must be met, including subject constraint, capacity constraint, and school priority constraint. The subject constraint and capacity constraint refer to the subject the student teacher can teach, and how many placements are available (in that school for that subject). The school priority constraint refers to whether the

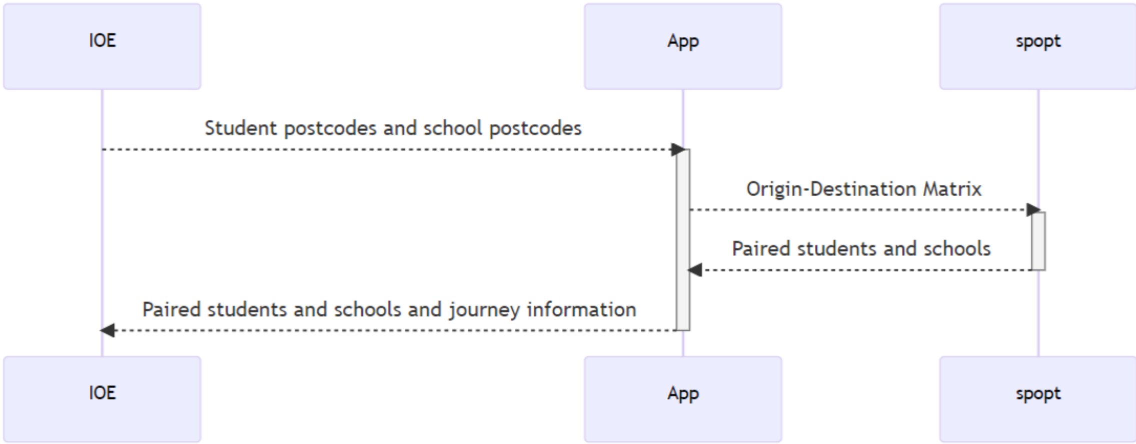

Figure 1. The flow of data between the source (IOE), the App and the `spopt` library.

placement at the school must be used or not. The optional constraints are either difficult to achieve or adding complexity to the model, including student priority constraint. Student priority refers to students being given a priority for shorter journeys for a number of different reasons, *e.g.* caring. After we checked the data and considered the difficulty of constraints, we decide not to include optional constraints in the model in this initial version.

The students and schools corresponding to each subject will form a group, and each group will use the model independently for optimisation. The capacitated p-median location-allocation in this case can be described as shown in the following equations.

In the model, demand points referring to students and are indexed by i. And I is the collection of all students. Facilities referring to schools and are indexed by j. And J is the collection of all schools. In the objective function, c_{ij} stands for the lowest travel cost from student i to school j. X_{ij} and Y_j are two decision variables. Constraint 2 ensures that each student is assigned to one slot, which means one subject slot in one school. Constraint 3 ensures that number of students assigned to each school for each subject does not exceed the available slots. Here b_j stands for the number of slots in school j. Constraint 4 ensures that the slots of priority 1 schools are fulfilled. Here p_j stands for the priority level of schools. There are in total 3 levels of priority among all schools, we set the value of their p_j as 1,2 and 3.

$$\text{Minimise} \quad \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}X_{ij} \tag{1}$$

Subject to:

$$\sum_{j=1}^{m} X_{ij} = 1, \quad \forall i \in I \tag{2}$$

$$\sum_{i=1}^{n} X_{ij} \leq b_j, \quad \forall j \in J \tag{3}$$

$$\sum_{i=1}^{n} X_{ij} = b_j, \quad \text{when } p_j = 1 \tag{4}$$

$$X_{ij} \leq Y_j, \quad \forall j \in J \tag{5}$$

Where:

$$X_{ij} = \begin{cases} 1, & \text{if student } i \text{ is assigned to school } j; \\ 0, & \text{otherwise.} \end{cases}$$

$$Y_j = \begin{cases} 1, & \text{if school } j \text{ is selected;} \\ 0, & \text{otherwise.} \end{cases}$$

Based on the above formulation, we determined that the package needed to handle capacity constraints for the facilities, as well as the option to include predefined facilities, such as the priority 1 schools in this case. To incorporate these changes, we modified the `from_cost_matrix` function in the `p_median.py` file. Specifically, we added three new parameters to the function: `predefined_facilities_arr`, `facility_capacities` and `fulfill_predefined_fac`. These parameters enable the user to input the capacity of all facilities, which can then be used to enforce capacity constraints when solving the capacitated p-median problem. The development of predefined facilities function can also help the user to deal with the situation where some facilities must be selected or assigned. The code containing these changes is currently available as a pull request, and it will be integrated in to the spopt library in due course.

To solve the optimisation problem, we use the default solver of `pulp`, which is COIN-OR Branch and Cut Solver (CBC). Take an example, when studying the business subject, there are in total 19 students and 38 schools waiting to be allocated, and 4 predefined schools that must be fulfilled. The calculation time is 0.2 seconds by using the new function in `spopt`, under the environment of Python 3.8.3 in macOS 10.15.5.

6 Data and Software Availability

Research data and code supporting this publication will be available via a `GitHub` repository[4] including a reproducible example. This will be linked to a DOI generated through `Zenodo`. The computational workflow supporting the location-allocation work in the `spopt` library will be available as a `Jupyter` Notebook. The original data we are using is confidential data from UCL IOE, as it contains personal details (location, priority information related to caring commitments, transport accessibility information related to disability) relating to students. A sample data set will be provided with anonymised and randomised data which can be used to demonstrate the usefulness of the app and library developments.

The work is currently being undertaken in Python, making use of the `pandas`, `geopandas`, `NumPy`, `gurobipy`, `spopt`, `filelock`, `pyarrow`, `pyrate-limiter`, and `requests` libraries.

7 Results and Conclusion

The results from the initial version of the capacitated p-median location-allocation problem has shown a marked improvement on public transport travel time, with public transport travel time reduced by 891 minutes overall for an initial sample of 93 students (9.58 minutes per student).

Initial feedback from UCL IOE has been positive, and they are pleased with the results generated by the tool. The results for our sample data of 93 students across Greater London are currently being compared in detail by IOE to see how the automated location-allocation varies from their manual method. There are improvements to make around different mode types, as well as further constraints to add.

Currently, the app does not handle students who choose to drive to their placement school. This accounts for approx-

[4] https://github.com/UCL/ioe-student-school-allocation

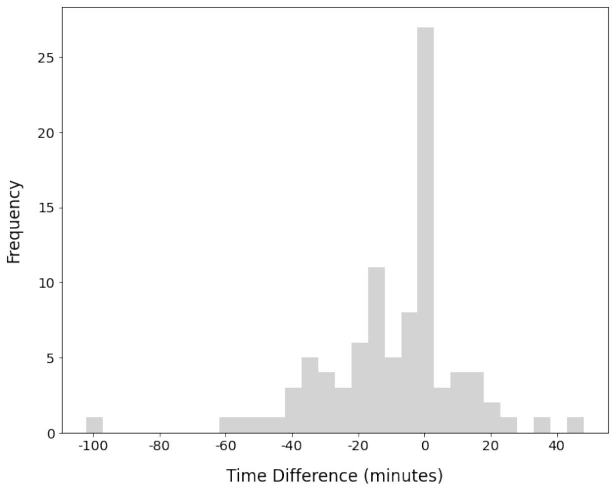

Figure 2. The histogram of the difference between optimised travel time and previous travel time. Negative values are the decrease in travel time for the students (compared with the IOE allocation) and positive values are an increase in travel time (compared with the IOE allocation).

imately half of our data, with the 93 students reported on all taking public transport. The app will be developed to include driving.

As Fig. 2 shows, the majority of the students have had their travel time decreased, but some students has had their travel time increased significantly. Among the total of 93 sample students, there are 53 students having shorter travel journey than before, 23 students without any change, meaning that the optimised facility choices are the same as IOE's manual decisions, and 17 students with longer travel time. We will investigate this further with IOE and make any necessary changes to the app.

The app will provide a graphic interface to show the school and student allocations, as shown in Fig. 3. Currently this shows the paired locations, but will be developed to allow the allocations to be adjusted manually, and the location-allocation process re-run on the remaining students. This will allow IOE to manually incorporate any additional requirements into the allocation process.

8 Acknowledgements

The authors wish to express their thanks to James Grindrod, Nick Mulliss in UCL IOE, Tim Machin in UCL ISD and Jonathan Cooper in UCL ARC for their involvement in the project and feedback on the initial versions of the app. This work was made possible by the generous support from AGILE funding initiatives. Dr Qunshan Zhao has received the ESRC's ongoing support for the Urban Big Data Centre [ES/L011921/1 and ES/S007105/1].

The following acknowledgement applies to James Gaboardi (affiliation 4) only:

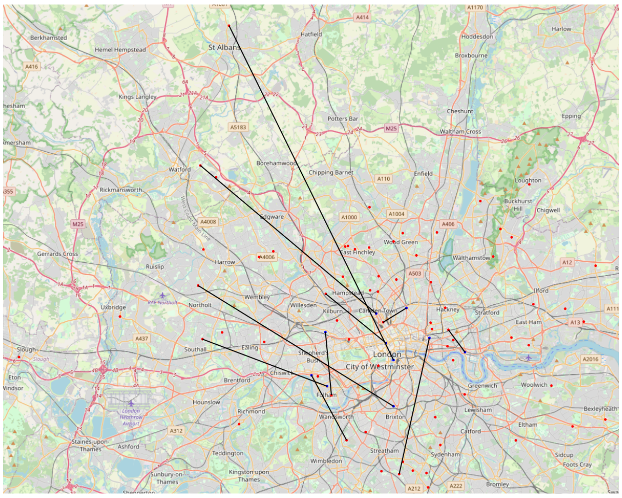

Figure 3. Visualisation of the matched school and student pairs, with lines showing connections. Students are shown in red, schools in black.

- This manuscript has been authored in part by UT-Battelle LLC under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a non-exclusive, paid-up, irrevocable worldwide licence to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

References

Chen, H., Murray, A. T., and Jiang, R.: Open-source approaches for location cover models: capabilities and efficiency, Journal of Geographical Systems, 23, 361–380, https://doi.org/10.1007/s10109-021-00350-w, 2021.

Church, R. and Murray, A.: Location covering models: history, applications and advancements, Springer, 2018.

Church, R. L. and ReVelle, C.: The Maximal Covering Location Problem, Papers in Regional Science Association, 32, 101–118, https://doi.org/10.1111/j.1435-5597.1974.tb00902.x, 1974.

Coskun, N. and Erol, R.: An Optimization Model for Locating and Sizing Emergency Medical Service Stations, Journal of Medical Systems, 34, 43–49, https://doi.org/10.1007/s10916-008-9214-0, 2010.

Daskin, M.: Network and Discrete Location: Models, Algorithms and Applications, Journal of the Operational Research Society, 48, 763–764, https://doi.org/10.1057/palgrave.jors.2600828, 1997.

Daskin, M. S.: What you should know about location modeling, Naval Research Logistics, 55, 283–294, https://doi.org/10.1002/nav.20284, 2008.

Erdoğan, G., Stylianou, N., and Vasilakis, C.: An open source decision support system for facility location analysis, Decision Support Systems, 125, 113 116, https://doi.org/10.1016/j.dss.2019.113116, 2019.

Feng, X., Gaboardi, J. D., Knaap, E., Rey, S. J., and Wei, R.: pysal/spopt, https://doi.org/10.5281/zenodo.4444156, 2021.

Feng, X., Barcelos, G., Gaboardi, J. D., Knaap, E., Wei, R., Wolf, L. J., Zhao, Q., and Rey, S. J.: spopt: a python package for solving spatial optimization problems in PySAL, Journal of Open Source Software, 7, 3330, https://doi.org/10.21105/joss.03330, 2022.

Fredriksson, A.: Location-allocation of public services — Citizen access, transparency and measurement. A method and evidence from Brazil and Sweden, Socio-Economic Planning Sciences, 59, 1–12, https://doi.org/10.1016/j.seps.2016.09.008, 2017.

Hakimi, S.: Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems, Operations Research, 13, 462–475, https://doi.org/10.1287/opre.13.3.462, 1965.

Hakimi, S. L.: Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph, Operations Research, 12, 450–459, https://doi.org/10.1287/opre.12.3.450, 1964.

Horner, M. W. and Downs, J. A.: Optimizing hurricane disaster relief goods distribution: model development and application with respect to planning strategies, 34, 821–844, https://doi.org/10.1111/j.1467-7717.2010.01171.x, 2010.

Kotavaara, O., Pohjosenperä, T., and Rusanen, J.: Integrated location-allocation of private car and public transport users — Primary health care facility allocation in the Oulu Region of Finland, p. 23, https://doi.org/10.5194/agile-giss-3-23-2022, 2018.

Kuby, M.: Programming Models for Facility Dispersion: The p-Dispersion and Maxisum Dispersion Problems, Geographical Analysis, https://doi.org/10.1111/j.1538-4632.1987.tb00133.x, 1987.

Murray, A. T.: Contemporary optimization application through geographic information systems, Omega, 99, 102 176, https://doi.org/10.1016/j.omega.2019.102176, 2021.

Pan, A.: The applications of maximal covering model in Typhoon Emergency shelter Location Problem, pp. 1727–1731, IEEE, https://doi.org/10.1109/IEEM.2010.5674577, 2010.

Pulver, A.: pyspatialopt: A library to create, solve, and analyze spatial optimization problems, https://github.com/apulverizer/pyspatialopt, 2016.

Pulver, A.: allagash: A spatial optmization library for covering problems, https://github.com/apulverizer/allagash, 2019.

Rahman, S. and Smith, D. K.: Use of location-allocation models in health service development planning in developing nations, European Journal of Operational Research, 123, 437–452, https://doi.org/10.1016/S0377-2217(99)00289-1, 2000.

ReVelle, C.: Review, extension and prediction in emergency service siting models, European Journal of Operational Research, 40, 58–69, https://doi.org/10.1016/0377-2217(89)90272-5, 1989.

Rey, S., Anselin, L., Amaral, P., Arribas-Bel, D., Cortes, R., Gaboardi, J., Kang, W., Knaap, E., Li, Z., Lumnitz, S., Oshan, T., Shao, H., and Wolf, L.: The PySAL Ecosystem: Philosophy and Implementation, Geographical Analysis, https://doi.org/10.1111/gean.12276, 2021.

Rey, S. J. and Anselin, L.: PySAL: A Python Library of Spatial Analytical Methods, The Review of Regional Studies, 37, 5–27, https://rrs.scholasticahq.com/article/8285.pdf, 2007.

Serra, D. and Marianov, V.: The p-median problem in a changing network: the case of Barcelona, Location Science, 6, 383–394, https://doi.org/10.1016/S0966-8349(98)00049-7, 1998.

Silver, E., Pyke, D., and Peterson, R.: Inventory Management and Production Planning and Scheduling, John Wiley and Sons, New York, third edn., http://books.google.com.tw/books?id=GI5jQgAACAAJ, 1998.

Syam, S. S.: A multiple server location-allocation model for service system design, Computers & Operations Research, 35, 2248–2265, https://doi.org/10.1016/j.cor.2006.10.019, 2008.

Toregas, C., Swain, R., ReVelle, C. S., and Bergman, L.: The Location of Emergency Service Facilities, Operations Research, 19, 1363–1373, https://doi.org/10.1287/opre.19.6.1363, 1971.

Venkatesan, R. and Kumar, V.: A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy, Journal of Marketing, 68, 106–125, https://doi.org/10.1509/jmkg.68.4.106.42728, 2004.