Simplified geodata models for integrated urban and public transport planning

AGILE: GIScience Series, 4, 32, 2023. https://doi.org/10.5194/agile-giss-4-32-2023
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.

Simplified geodata models for integrated urban and public transport planning

Flavia Lopes[1], Jorge Gil[1], and Ioanna Stavroulaki[1]

[1] Chalmers University of Technology, Department of Architecture and Civil Engineering, Gothenburg, Sweden

Correspondence: Flavia Lopes (flavia.lopes@chalmers.se)

Abstract. The current division between urban and transport planning is a significant obstacle to achieving sustainable urban development. To transform cities towards sustainability, both fields must adopt shared or at least compatible models of the urban systems, namely transport, street and public space networks for all users and urban activities. Although several models and tools have emerged in recent years to facilitate this integration, there are still usability gaps that hinder their wider adoption. One of the gaps is a lack of flexibility to operate at different stages of integrated planning. To address this gap, the study aims to develop a set of aligned and flexible multimodal urban network models and tools to support different stages of planning. This paper focuses on the public transport geodata models, which were built by aggregating a General Transit Feed Specification (GTFS) dataset at various spatial and temporal levels. The aggregation levels range from a baseline data model that is useful for detailed planning stages, up to a topological data model that is suitable for macro scale and strategic planning. By using this unified set of models, the dialogue between the two fields at different integrated planning phases can be facilitated, and decision-making can be enhanced.

Keywords. GTFS, data model, public transport, integrated planning, urban planning.

1 Introduction

The co-dependency between public transport and urban planning has been widely discussed in recent decades, due to its importance for the promotion of more sustainable cities (Bertolini et al., 2005; Curtis, 2011; Brömmelstroet and Bertolini, 2016). Well-managed cities, in the sense of creating better integration between the built environment and movement of people, can promote well-connected places capable of making the everyday journeys of individuals more effective, and can have a significant impact on the environment and the well-being of citizens (Cervero et al., 2017; Moriarty, 2022). However, the current division between urban and transport planning is still a challenge in both research and practice and poses a significant obstacle to achieving sustainable urban development.

To achieve sustainable cities, both fields must adopt shared or at least compatible models of the urban systems, particularly transport, street and public space networks for all users and urban activities. Without such integration, there is a risk of sub-optimization of the separate urban systems, not taking advantage of their potential for complementarity or overseeing possible conflicts. In recent decades, several tools and models related to urban planning, public transport planning or both have been developed (Miller and Goodchild, 2015; Orozco et al., 2021; Lovelace, 2021). However, there are still gaps in their usability (Curtis, 2011; Curtis and Scheurer, 2019), namely their lack of flexibility to operate at different stages of integrated planning. Some are very detailed but complex for strategic planning, others are simplified but are not based on open standards or geocomputational procedures. Therefore, the use of diverse incompatible models and tools that do not cover all phases of planning is required. This makes the process of integrating both fields difficult for researchers and practitioners worldwide.

To address this gap, a set of aligned and flexible multimodal urban network models and tools is being developed as part of a research project. This paper presents the public transport geodata models envisaged to address different phases and scales (from detailed to strategic) of urban and regional planning. The models were built by aggregating public transport timetable data in the General Transit Feed Specification (GTFS) format, a widely available and detailed data source, at distinct levels of detail. In the upcoming phases, the public transport geodata models will be connected to individual transport networks (i.e., car, bicycle, walk) and land use systems to implement the integrated multimodal approach.

The sections that follow present a brief state-of-the-art including a perspective on the GTFS standard, the methods used to build the geodata models, the models themselves, and a conclusion about this stage of the work.

2 Related Work

Tools and models related to urban and public transport planning have been created and optimized to support in a more efficient way analyses and decision-making. Such instruments offer several possibilities to understand complex problems regarding land-use development, public transport performance, accessibility, and its effects on the life of the individuals, such as differentiation in access to goods and services, social justice, socio-spatial segregation, and urban sustainability (Bertolini et al., 2005; Miller and Goodchild, 2015; Brömmelstroet and Bertolini, 2016; Lovelace et al., 2020).

To understand such complex problems, tools and models range from disaggregated spatio-temporal models for refined analysis and simulation in early diagnosis of the baseline conditions, and in later detailed planning stages, to more aggregated models useful for macro-analysis of a strategic nature.

Existing multimodal transport analysis packages, such as *OpenTripPlanner* (Morgan et al., 2019), R5R (Pereira et al., 2021), and R5py (Fink et al.,2022) are increasingly used in multimodal routing and detailed time-based accessibility analysis. To build the urban model, these packages require land-use data, individual transport network (walk, bike, cars) data extracted from OpenStreetMap (OSM) and the public transport network data from GTFS data. With a similar objective, other models that use data from diverse sources at a more aggregated level, have been created over the last three decades, within the scope of GIS-T (Mavoa et al., 2012; Tribby, Zandbergen, 2012; Salonen; Toivonen, 2013; Djurhuss et al., 2015; Gil, 2016; Tenkanen, 2017; Lopes, 2022).

While suited for calculating detailed analyses, these models do not adequately support macro-scale strategic planning, where working with simpler, large scale and aggregated data and discursive models can result in better analysis and decision-making (Curtis, 2011). In this sense, Curtis and Scheurer, (2019) developed the SNAMUTS model. This model is created from the combination of public transport networks data with land use to measure distinct types of centralities. Despite proving to be efficient, it does not consider the integration of the individual transport network with the public transport network, thus focusing on only a part of the city's movement.

On the other hand, several models for studying centralities in street networks have been developed over the last three decades such as space syntax (Hillier, Hanson, 1984) and multiple centrality analysis (Porta et al., 2006), including tools that connect street networks and land use, such as Place Syntax (Ståhle et al. 2005) and the Urban Network Analysis tool (Sevtsuk and Mekonnen, 2012). On a related direction, models have been developed integrating street network models with public transport network models (Gil 2012, 2016; Law et al. 2012; Lerman and Lebendiger 2017) to conduct multimodal centrality and accessibility analysis, to support sustainable mobility design in cities and regions or transit-oriented development solutions.

It is notable from this review that many tools and models exist. Despite their potential, these earlier works have not been extensively used. One of the reasons is that they have a specific domain of application and do not offer a flexible approach capable of covering several stages of planning in a single tool or workflow. Another issue is that some of the models do not comply with reproducibility and data standards or were not implemented into user-friendly analytic or design-support tools, which makes it difficult to apply the analyses in different urban and transport planning cases.

2.1 GTFS data

To improve the flexibility and reproducibility of models and tools, it is necessary to consider the type of input data used. In this sense, departing from open-source data with some level of worldwide standardization is advantageous (Lovelace, 2021). Following this strategy, the GTFS open standard was used for the construction of the models presented in the next sections of this work.

GTFS is a public transport data model that supports detailed information about stop locations, routes, trips, and times of operation. Essentially, a GTFS data set consists of a collection of text files for each class of the data model, describing a component of the public transport system. Currently, these data are used as standard by public transport operators in several countries, regions, and municipalities for publishing timetable information, and are used for multimodal routing and accessibility analysis and the construction of multimodal transport models.

In the context of this work, GTFS is useful for certain applications in early diagnosis and detailed stages of planning, but it is considered too spatially and temporally disaggregated for planning phases that require more flexible and discursive approaches, where alternative options and changes to the transport network need to be tested. For this reason, different models were constructed derived from the GTFS data, with different levels of spatial and temporal aggregation, aiming to be applied in distinct phases and scales of planning.

3 Methods

To create public transport models capable of supporting the distinct phases of urban planning practice, GTFS dataset were aggregated in different forms and levels. This paper presents four different public transport models, built from the GTFS dataset of the Västra Götaland Region in Sweden. The models were built in Python using Pandas (version 1.5.3), Geopandas (version 0.12.2), NumPy (version 1.24.2), Shapely (version 2.0.1), and OSMnx (1.3.0) packages. The notebooks showing how the models were built are available under the GPL-3.0 license in the following repository: https://github.com/FlaviaMLopes/simplified_GTFSmodels

3.1. Model 01: the 'baseline' data model

The first step was to create a 'baseline' data model, a geographically explicit representation of the GTFS as a network with nodes and links. The model was built through combination of geographical and temporal data from the GTFS and OSM datasets. It assigns a geographic location to every stop time and creates links between those stop times, thus retaining the entire timetable information contained in GTFS in a flat format.

The public transport nodes table was built by merging the GTFS stops with the geographic information about these stops' location from the OSM database via OSMnx. The links table was built joining the nodes with the routes, trips, and stop_times from the GTFS (Figure 1).

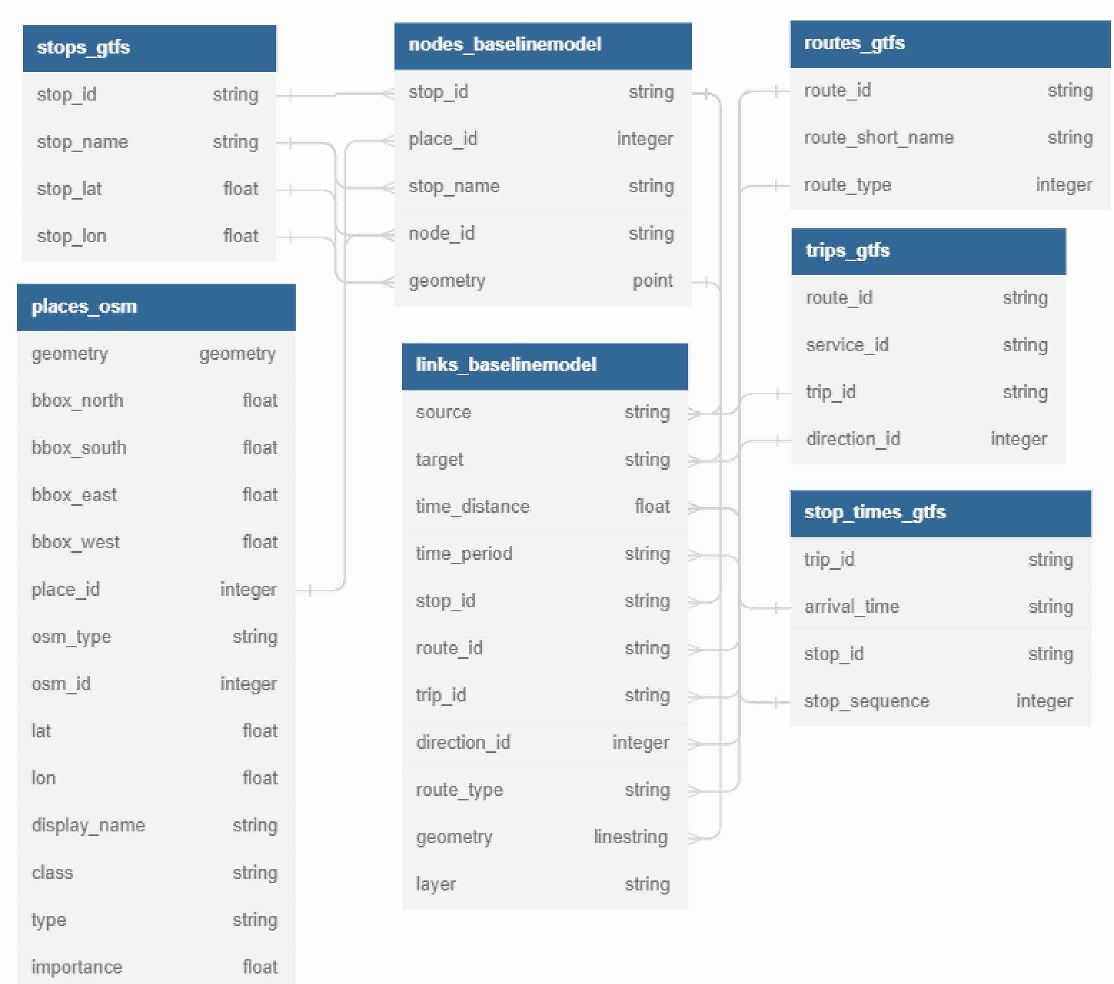

Figure 1. The 'baseline' data model.

3.2. Model 02: the 'frequency' data model

The second model is a first step in terms of GTFS data aggregation, departing from the 'baseline' model. The 'frequency' data model represents a temporal aggregation of the different public transport services of each route at different periods of the day, providing the frequency per hour of a route.

The 'frequency' model consists of classes with information on public transport stops ('nodes'), the links between these stops ('links'), and the possible transfers between public transport vehicles, whether these transfers are made at the same stop ('transf_samestop') or different ones ('transf_differentstop') in the same place (Figure 2).

The nodes table in this model results from selecting the unique stop ids from the nodes baseline model, thus dropping their temporal dimension. The links table was built by aggregating the baseline links by route keeping both directions and calculating the frequency with which vehicles circulate on each route direction. The frequencies were calculated using the 'calendar' and 'calendar_dates' datasets from GTFS.

For the transfers, two different tables were generated: one for transfers between the same stop, and the other for transfers between different stops. The first transfers table was built by joining the 'nodes' from the 'baseline' model with the links table from the 'frequency' model, to create links for the different routes passing through each public transport stop. The second transfers table has information on transfer possibilities between stops with the same name and similar locations. In this case, the join between baseline nodes and frequency links considered the possibility of walking between stops of the same name that are located close to each other.

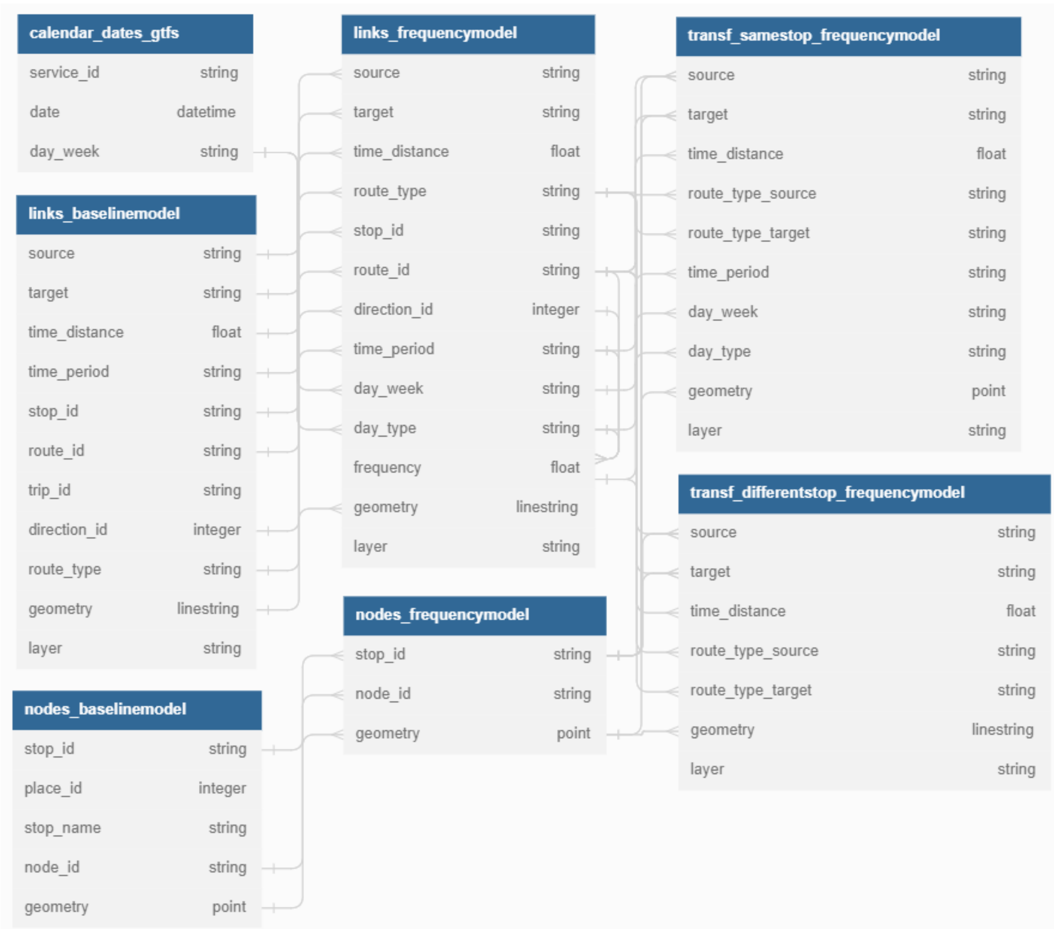

Figure 2. The 'frequency' data model.

3.3. Model 03: the 'modes' data model

In model 03, called the 'modes' data model, the simplification takes another step towards a model based on transport modes. In this model, the nodes are aggregated by their geographic location, while the links and transfers, and consequently the frequencies, are aggregated by the transport mode. This represents a geographical as well as temporal aggregation compared to the previous ones.

To build the 'modes' model, the tables (nodes, links, transfers same stops) from the frequency model were used. The nodes table was created by aggregating the nodes by name and location. The links table was created by aggregating their route id by the transport mode attribute. The transfers table was created by excluding from the frequency model transfers between routes of the same mode (Figure 3).

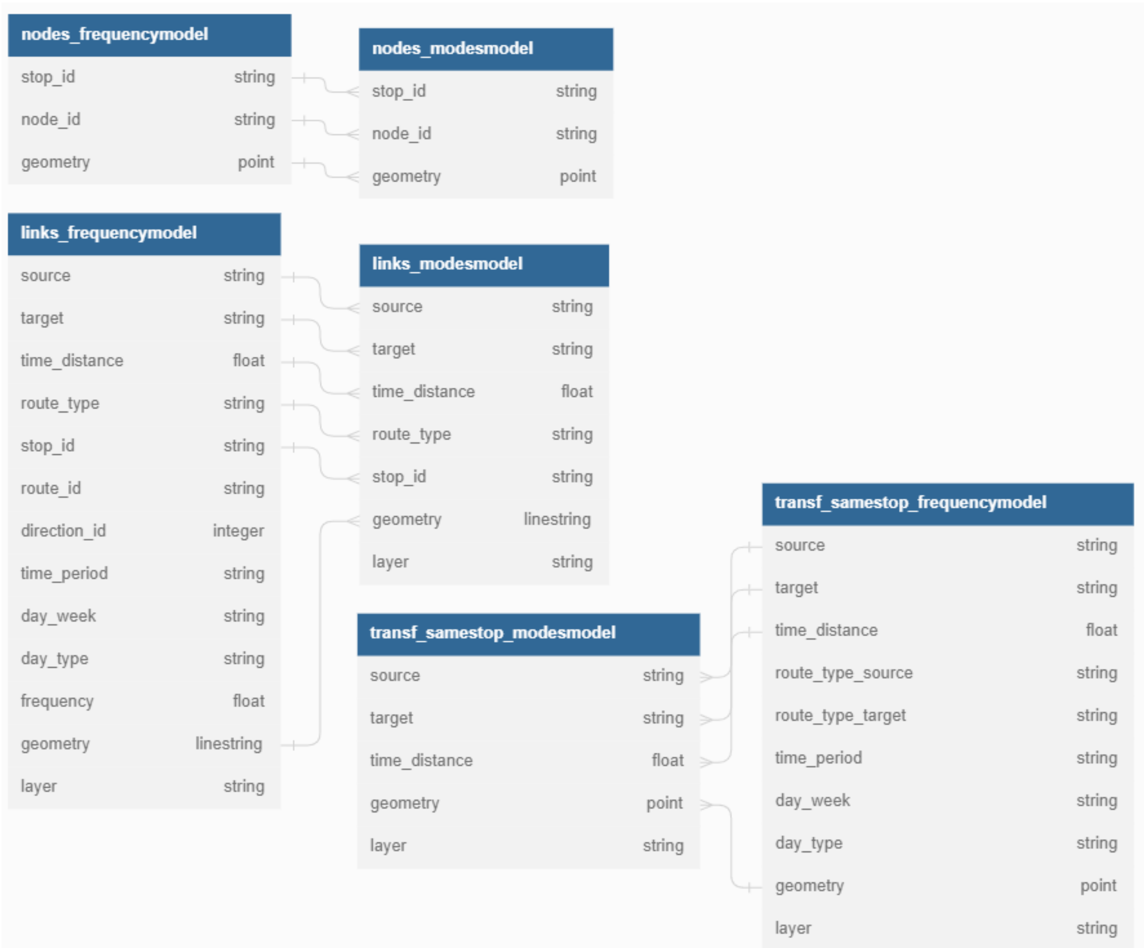

Figure 3. The 'modes' data model.

3.4. Model 04: the 'topological' data model

The last level of aggregation presented in this paper, is the 'topological' data model. This data model represents the structure of connectivity between locations (the topology) provided by the public transport network. At this level, the temporal dimension is excluded, the transfers disappear as stops are dissolved at their location, and the links between nodes are aggregated into a single system of links, without distinction of time, schedule, or mode.

To build this model, nodes and links tables are combined. The nodes table is the same used in the modes data model. The links table was built by aggregating the links from the modes model by their geography resulting in a network of single links between nodes (Figure 4).

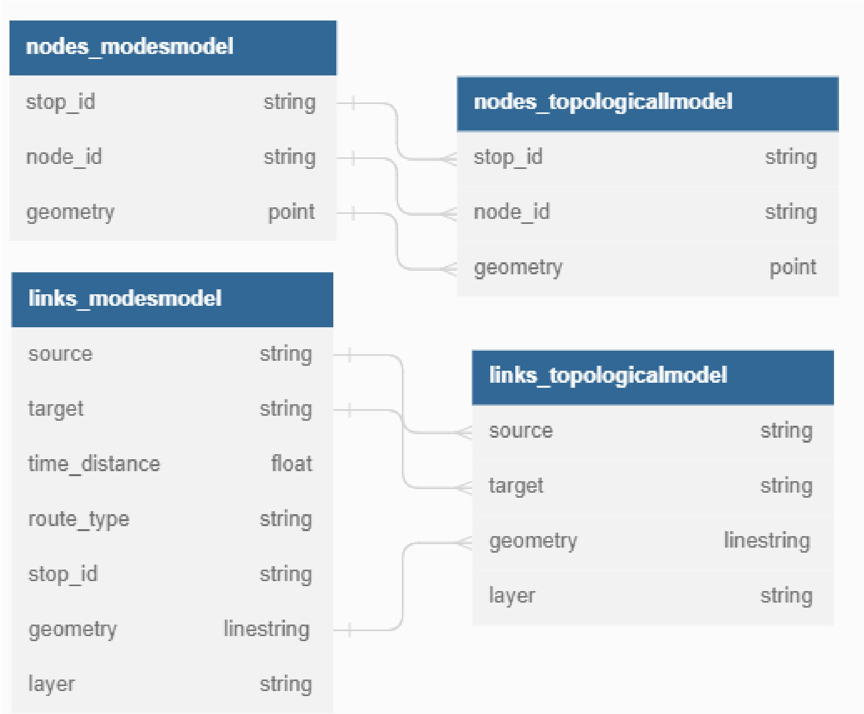

Figure 4. The 'topological' data model.

4 Results

The results are four different public transport models (Figure 5). Each model contains different information and analytical possibilities related to the different phases of integrated urban and public transport planning. Thus, it is possible to run different types of analyses based on different measures, such as time-based accessibility and measures of network centrality.

Among the models presented, model 01, the 'baseline' model, is the only one that does not relate to an analysis relevant for urban planning because of the level of the spatial and temporal disaggregation of data, identical to that of GTFS. However, it offers a baseline for building the other models.

Model 02, the 'frequency' model, provides the first level of analysis. It is possible to measure time-based accessibility due to the temporal attributes of routes and frequency of vehicles. And it does so with a much lighter model in the number of nodes and links, that does not require a specification of exact departure times as analysis done in similar tools based on GTFS data.

Model 03, the 'modes' model, is a hybrid model. From its analysis it is possible to obtain a time-based measure of accessibility at a more aggregated level, and measures of centrality of the multimodal transport network due to the level of links aggregation.

Model 04, the 'topological' model, is purely topological. With no temporal perspective, it only allows to analyse topological measures of centrality.

The measures are associated to the level of aggregation of the data for each model. In turn, these measures are related to different urban and transport planning phases. Centrality measures tend to be linked to strategic levels of planning, where the construction of different scenarios is more connected with the study of connections and integration of the structure than the planning of detailed journeys. On the other hand, simple time-based accessibility measures are related to detailed planning phases, where a more refined view of travel time is necessary, but detailed data on the public transport services and timetable is not yet available (Figure 6).

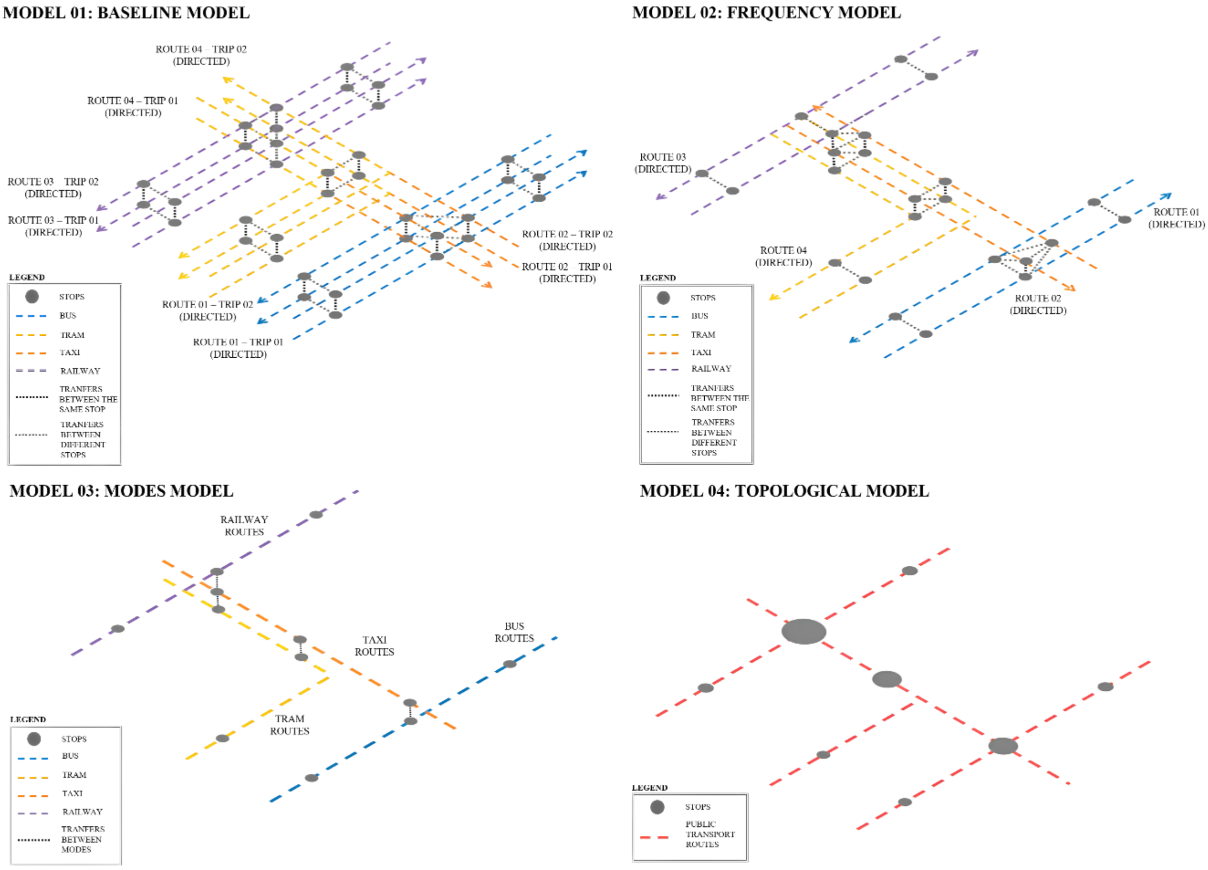

Figure 5. Schemes of the public transport models.

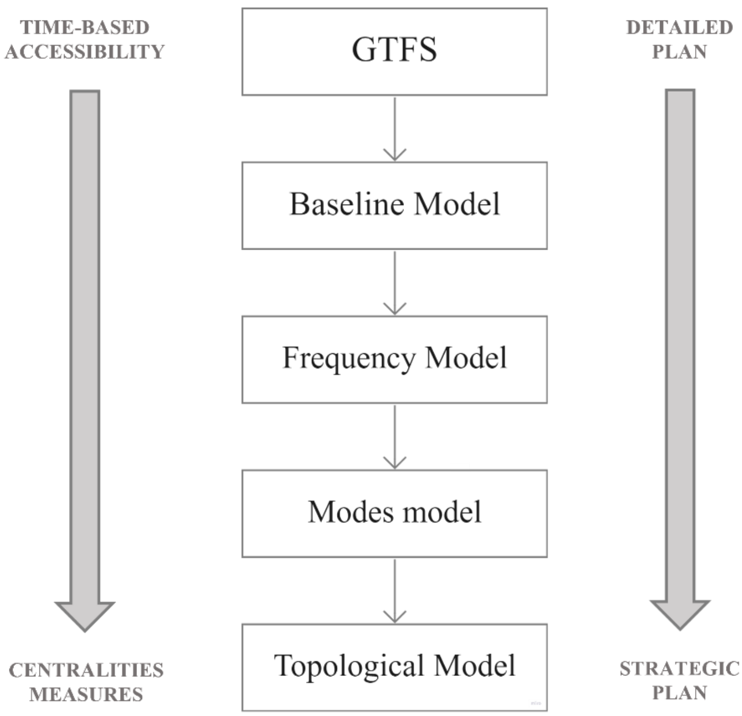

Figure 6. Schema showing the relation among the models, measures, and levels of planning.

5 Conclusion

This article presented four public transport data models built as part of ongoing work aiming to create a set of tools and models to support better integration between urban and public transport planning in practice. The data models were built in a flexible, systematic, automated, and unified workflow departing from the GTFS dataset from the Västra Götaland Region, Sweden.

Essentially, the GTFS was first aggregated in a 'baseline' model which provides geographic features to allow visualization of the GTFS as a combination of nodes and links between the stops. In the following steps, subsequent models - the 'frequency', the 'modes', and the 'topological' data models - were created derived from the 'baseline' model's attributes. Each of these subsequent models embraces the needs of the different phases (from detailed to strategic) of urban and public transport planning, subject to the possibilities of analysis that each offers. The models that are spatial and temporally disaggregated offer a possibility to measure time-based accessibility, used to calculate the travel opportunities of individuals in more detail. On the other hand, models that are more spatial and/or temporally aggregated are related to the measurement of centralities, associated with strategic levels of planning, where the understanding of the structure of the system is more important.

The models presented here have been influenced by previous works, such as Lopes (2022), Curtis and Scheurer (2019), and Gil (2012, 2016). The main contribution of this research is to build them in a systematic and reproducible workflow departing from an established open standard (GTFS) with widely available data. Moreover, this workflow offers flexibility, making it is possible to switch between models and run different analyses for the same case, where each level is related to a different level of integrated planning.

The next phase of this approach is to connect these public transport models to the individual transport network (i.e., walk, bike, and cars) and the land-use system, and test these integrated multimodal models using real needs to understand the possibilities they offer in more detail.

This flexible approach to modelling and articulation between models, can improve the collaboration and dialogue between researchers and practitioners from the urban and transport planning fields. By using the same approach to plan public transport and urban development, they can unify their efforts, improve the decision-making process, and consequently promote more sustainable cities.

Acknowledgments

This research has been partly funded by the Chalmers Transport Area of Advance through the project "Developing a design-support tool to integrate urban and public transport planning towards sustainable urban development", and the MISTRA Sport and Outdoors research and collaboration programme (https://www.mistrasportandoutdoors.se/en/).

References

Bertolini, L., le Clercq, F. and Kapoen, L.: Sustainable accessibility: a conceptual framework to integrate transport and land use plan-making. Two test-applications in the Netherlands and a reflection on the way forward, Transp. Policy 12, 207–220. https://doi.org/10.1016/j.tranpol.2005.01.006, 2005.

Brömmelstroet, M. te and Bertolini, L. (Eds.).: Transport Models in Urban Planning Practices: Tensions and Opportunities in a Changing Planning Context, Routledge, London. https://doi.org/10.4324/9781315540375, 2016.

Cervero, R. and Guerra, E., Al, S.: Beyond Mobility: Planning Cities for People and Places, Island Press, 2017.

Curtis, C.: Integrating Land Use with Public Transport: The Use of a Discursive Accessibility Tool to Inform Metropolitan Spatial Planning in Perth, Transp. Rev. 31, 179–197. https://doi.org/10.1080/01441647.2010.525330, 2017.

Curtis, C. and Scheurer, J.: The SNAMUTS Accessibility Tool in Action: Case Study Adelaide, Australia, Designing Accessibility Instruments, Routledge, London, 2019.

Djurhuus, S., Sten Hansen, H., Aadahl, M. and Glümer, C.: Building a multimodal network and determining individual accessibility by public transportation, Environ. Plan. B Plan. Des. 43, 210–227. https://doi.org/10.1177/0265813515602594, 2016.

Fink, C., Klumpenhouwer, W., Saraiva, M., Pereira, R. and Tenkanen, H.: r5py: Rapid Realistic Routing with R5 in Python, doi:10.5281/zenodo.7060437, 2022.

Gil, J.: Integrating Public Transport Networks in the Axial Model, Proceedings of the 8th International Space Syntax Symposium, Santiago, Chile, 2012.

Gil, J.: Urban Modality: modelling and evaluating the sustainable mobility of urban areas in the city-region, Ph.D. thesis, TU Delft, Delft, Netherlands,2016.

Hillier, B. and Hanson, J.: The social logic of space, Cambridge University Press, Cambridge, 1984.

Law, S., Chiaradia, A. and Schwander, C.: Towards A Multi-modal Space Syntax Analysis: A Case Study of the London Street and Underground Network, Proceedings of the 8th International Space Syntax Symposium, Santiago, Chile, 2012.

Lerman, Y.; Lebendiger, Y.: Visualizing Space Syntax Analyses for Decision Makers Lessons from Israel, Proceeding of the 11th Space Syntax Symposium, Lisbon, Portugal, 2017.

Lopes, F.: Minha Casa, Qual Vida Cotidiana? Investigando Perdas e Ganhos de Capital Espacial da Vida Cotidiana de Indivíduos Beneficiados pelo Programa Minha Casa, Minha Vida (PMCMV), Ph.D. thesis, Universidade Federal da Paraíba, João Pessoa, Brazil, 2022.

Lovelace, R.: Open source tools for geographic analysis in transport planning. J. Geogr. Syst, 23, 547–578. https://doi.org/10.1007/s10109-020-00342-2, 2021.

Lovelace, R., Parkin, J. and Cohen, T.: Open access transport models: A leverage point in sustainable transport planning, Transp. Policy, 97, 47–54. https://doi.org/10/ghfsg6, 2020.

Mavoa, S., Witten, K., McCreanor, T. and O'Sullivan, D.: GIS based destination accessibility via public transit and walking in Auckland, New Zealand, J. Transp. Geogr., Special Section On Child & Youth Mobility 20, 15–22. https://doi.org/10.1016/j.jtrangeo, 2011.

Miller, H.J. and Goodchild, M.F.: Data-driven geography, GeoJournal, 80, 449–461. https://doi.org/10.1007/s10708-014-9602-6, 2015.

Morgan, M., Young, M., Lovelace, R. and Hama, L.: OpenTripPlanner for R. J, Open Source Softw, 4, 1926. https://doi.org/10.21105/joss.01926, 2019.

Moriarty, P.: Making urban travel sustainable: Travel reductions are needed, Clean. Prod Lett, 3, 100010, https://doi.org/10.1016/j.clpl.2022.100010, 2022.

Orozco, L.G.N., Alessandretti, L., Saberi, M., Szell, M. and Battiston, F.: Multimodal urban mobility and multilayer transport networks, https://doi.org/10.48550/arXiv.2111.02152, 2021.

Pereira, R., Saraiva, M., Herszenhut, D., Braga, C.K., Braga, V., Conway, M. and Wigginton, M.: r5r: Rapid Realistic Routing on Multimodal Transport Networks with R5, R, Transp. Find, https://doi.org/10.32866/001c.21262, 2021.

Porta, S., Crucitti, P. and Latora, V.: The Network Analysis of Urban Streets: A Primal Approach, Environ. Plan. B, 33, 705–725. https://doi.org/10.1068/b32045, 2006.

Salonen, M. and Toivonen, T.: Modelling travel time in urban networks: comparable measures for private car and public transport, J. Transp. Geogr, 31, 143–153, https://doi.org/10.1016/j.jtrangeo.2013.06.011, 2013.

Sevtsuk, A., Mekonnen, M., Urban network analysis. A new toolbox for ArcGIS, Rev. Int. Géomat., 22, 287–305, https://doi.org/10.3166/rig.22.287-305, 2012.

Ståhle, A., Marcus, L. and Karlström, A.: Place Syntax: Geographic accessibility with axial lines in GIS. Proceeding of the 5th Space Syntax Symposium, Delft, Netherlands, 2005.

Tenkanen, H.T.O.:Capturing time in space : Dynamic analysis of accessibility and mobility to support spatial planning with open data and tools. Ph.D. thesis. University of Helsinki, Faculty of Science, Department of Geosciences and Geography, Helsinki, Finland, https://doi.org/10/work/39925712, 2017.

Trafiklab. Static GTFS Regional dataset [Data set]. In Tafiklab Static data. https://www.trafiklab.se/api/trafiklab-apis/gtfs-regional/static/, 2022.

Tribby, C.P., Zandbergen, P.A.: High-resolution spatio-temporal modeling of public transit accessibility, Appl. Geogr, 34, 345–355, https://doi.org/10.1016/j.apgeog.2011.12.008, 2012.