Proceedings of the Eastern Asia Society for Transportation Studies, Vol.9, 2013

Decision support tools in transport planning: from research to practice

Alex ERATH [a], Michael A.B. VAN EGGERMOND [b], Pieter J. FOURIE [c], and Artem CHAKIROV [d]

[a,b,c,d]*Future Cities Laboratory (FCL), Singapore ETH Centre, 138602, Singapore*
[a]*E-mail: erath@ivt.baug.ethz.ch*
[b]*E-mail: eggermond@ivt.baug.ethz.ch*
[c]*E-mail: fourie@ivt.baug.ethz.ch*
[d]*E-mail: chakirov@ivt.baug.ethz.ch*

Abstract: With the progress in transport demand modelling from aggregated, static and zone-based assignment models to fully dynamic agent- and activity-based transport simulations, the scope of data that needs to be collected, managed and maintained has changed dramatically. Similar applies to the type of data that is now available to transport planners and decision makers as for example the emergence of public transport smart card data shows. The type of analyses as well as potential analysts have also changed. Besides transport planners, we see tremendous potential for additional end-users to access and analyse such data: urban planners, policy-makers and the service industry.

In this paper, we present a framework of a decision support system designed to enable analysis of the wealth of information provided by agent-based transport demand models, travel diary surveys and automatically collected transportation data. We present a practical application of the framework using it for analyzing the MATSim model of Singapore in comparison with the Household Interview Travel Survey and public transport smart card transaction data.

Keywords: Decision support; Business Intelligence; Transport Planning Practice; Activity-based modeling; Agent-based modeling

1 INTRODUCTION

With the progress in transport demand modelling from aggregated, static and zone-based assignment models to fully dynamic agent- and activity-based transport simulations, the scope of data that needs to be collected, managed and maintained has changed dramatically. Conventional zone-based models require usually only a limited number of attributes for each zone describing trip generation and attraction. Depending on how many demand segments are to be distinguished, the attributes normally restrict for each zone to the number of residents in different demand categories to account for trip generation. To describe trip attraction, usually zone-specific information on the number of work and education places, shopping surface and indicators for potential leisure activities are used. Such data is usually collected from various sources and edited with conventional spreadsheet based software. Once fed in the transport demand model, the data is normally administered in the data structure stipulated by the demand modelling software. Generally, such transport demand models are updated every five to ten years despite the fact that key input data sets such as business census or inventory of shopping surfaces are updated continuously.

Agent-based transport demand models work on the level of individuals. This means that the transport modeler now works with a (synthetic) population as each agent represents an person living, working or passing through the modeled region. For each agent, socio-demographic details both on the level of individuals and households such as age, sex, income and mobility tool availability need to be maintained and accessible in various stages of the modelling and simulation process. Current implementations of agent-based transport demand simulations operate with 5 to 10 millions of agents of which each agent and household needs to be traceable. Furthermore, agent-based models usually work on a spatial resolution of parcels (Bekhor *et al.,* 2011), hectare raster (Balmer *et al.,* 2010) or individual buildings (Erath *et al.,* 2012). This demands that additional spatial information in the magnitude of 10^5 to 10^6 data points needs to be administered. At the same time, it is to note that each individual will perform several activities and trips over the simulated time period of one day which enhances the scope of data to managed depending on the desired level of detail by another magnitude of 10^7 data points.

In contrast to conventional transport demand models that either try to describe average daily or peak hour traffic, agent-based transport demand model are fully temporal dynamic. Taking full advantage of the activity based modelling approach, this means that information of preferred timing of activities on the side of agents and opening times on the side of facilities (i.e. locations where certain types of activities can be conducted) need to be managed. This can be extended to the description of the supply side, as certain elements of transport networks such as provision of dedicated bus lanes, road capacity and public transport schedule are of inherent dynamic nature, too.

At the same time, transport operation are more and more continuously monitored by systems such as loop detectors, road tolls, automatic public transport fare collection or mobile phone data. This data can be a valuable source for developing and calibrating such advanced transport demand models. Since the generated data exceeds by far levels that can be efficiently handled with conventional spreadsheet-based data management, the use of dedicated database solutions has become general practice. Similar applies to travel diary surveys, for which usually around a sample of 1% to 3% of the whole population is asked to report both activities and trips for a given day. In this case, databases that feature dedicated tables for person-, household- and trip-related attributes are widely used since decades.

Given the huge amount of data to be handled in agent-based transport demand models, current software packages such as MATSim (MATSim*,* 2013) or Transims (TRANSIMS Open Source*,* 2013) have defined dedicated formats to manage the respective data. Although certain interfaces for popular data formats have already been programmed or can be modularly added, so far no conventional standard as yet been established to allow swift data exchange from and to the model. Since such software frameworks have as open-source projects kept so far away from GUIs or GIS-based editing front-ends, model analysis has so far been restricted to the actual model developers or people possessing both the requires programming skills and knowledge of the code. Key audiences of such model as for example transport authorities, service providers (e.g. supermarket chains) or other entities are usually required to formulate the desired type of analysis in advance. Depending on the results, subsequent questions might arise which in turn require another iteration of data processing and analysis which prevents efficient, explorative data analysis and hence effective decision support.

In this paper, we present a data management solution designed to overcome the deficiencies outlines above and enable effective decision support tapping on the wealth of information provided by agent-based transport demand models. The remainder of the paper is structured as follows. In the subsequent section, we provide an overview on existing spatial database solutions and how business analytic software is used in other domains for explorative analysis of big data sets. This is followed by a presentation how a general framework allowing such analysis can be set up. Using the case study of MATSim Singapore, we present a practical application of such a framework. Exemplarily, we show how the decision support tool can be used for calibration and validation. We conclude by sketching other potential applications such as scenario analysis and proposing topics for research such as the development of a commonly used data structure and the potential of having a continuously updated database and transport demand model.

2 DECISION SUPPORT SYSTEMS: BACKGROUND & OTHER DOMAINS

Decision-support systems (DSS) evolved early in the area of distributed computing; prior to the mid 1960's it was very expensive to build up large-scale information systems. Powerful mainframe systems made it more practical and cost effective to develop management systems (MIS) in large companies. MIS focused on providing managers with structured, periodic reports (Power, 2003). Sol (1985) describes several subsequent stages of the evolution of Decision-Support systems. In the early 1970's a DSS was described as a 'computer based system to aid decision-making'. The starting point was found in interactive technology to managerial tasks in order to use computers for better decision-making. The cognitive focus of these systems was on the single decision-maker. From the mid to late 1970's the DSS movement emphasized interactive computer-based systems which help decision-makers utilize databases and models to solve ill-structured problems. The emphasis lies not on supporting the decision process, but on the support for personal computing with fast development tools and packages, for instance for financial planning. In the later 1970's to the early 1980's the DSS community provides technology to improve effectiveness of managerial and professional activities. User-friendly software is produced under the label DSS. Operations research and cognitive psychology are joining the DSS community. Sprague (1980) makes a distinction between three parts of a DSS: the database, model-base and the interaction with the end-user through dialog generation and management software. Models in the model-base mainly consist of econometric equations to support financial decision-making. In the early 1980's the role of a DSS is not only to generate reports based on a database, rather, the DSS contains expert knowledge and models. In the mid 1980's, a different type of decision-support sytems emerged: group decision support systems (GDSS). GDSS support searching for alternatives, communication, delibration, planning, problem solving, negotation, consensus building and vision sharing. GDSS are often applied in larger group situations, where group members are not necessarily cooperative, in situations where hidden agendas exist or where certain members seem overly dominant (Vogel and Nunamaker, 1990).

Executive Information Systems evolved from the aforementioned single-user, model drive DSS and complemented relational databases, also in the mid 1980's. These first systems used pre-defined screens or views on the data and were maintained for analysts for senior executives (Power, 2003). The term Business Intelligence stems from Proctor & Gamble, where sales information and retail scanner data was linked in a DSS (Nylund, 1999). BI describes a set of concepts and methods to improve business decision making by using fact-based support system. Business Intellingence systems are considered data-driven (DSS). In the early 1990's Decision Support systems built upon relational databases were actively promoted (e.g. Kimball and Ross, 2002), where data was stored in data warehouses. Simultaneously, a technology shift occured from mainframe-based DSS to client based DSS. These desktop tools contain On-Line Analytical Processing (OLAP) capabilities. With OLAP and data warehouses it was possible to interactively perform multidimensional analyses. Processes outlined by Kimball and Ross (2002) aim to build data warehouses. Data is periodically extracted from a range of databases, which may contain sales data, financial data and data from human resources. This data is cleaned and matched. On one hand, dimension tables are created in which 'fixed' elements are stored. These can be dates, cost centres and sales regions. On other, fact tables are created which contain individual sales or changes in human resources. Pre-defined OLAP queries make it possible to combine and query these fact and dimension tables. As every fact table contains a number of indexes to the dimension tables, queries are very quick as compared to queries on a relational database. With these OLAP queries it is possible to zoom in and zoom out, i.e. from year to month to day or from country to region to city. Also, these OLAP queries, written in languages such as MDX, were largely pre-defined and required a steep learning curve. In the mid 2000's a second technology shift occurred. Analysts, now familiar with OLAP tools and integration into spreadsheet programmes, required links to data sources available through the web as well as wanted to circumvent data warehouse specialists to speed up data analysis. Also, extra visualizations options were required, propagated by Tufte (1986) and Few (2012). Several software vendors, but also researchers, recognized these needs and have developed interactive visualization tools. Various examples from the Garnder BI survey (Schlegel *et al.*, 2013) include Tibco Spotfire, Tableau Software and SAS. Typically, these desktop tools allow connections to multiple data sources and databases simultaneously, interactive analysis of the data, contain state-of-the-art visualisation principles and the easy dissemination of findings through management dashboards.

3 REQUIREMENTS FOR A TRANSPORT PLANNING DECISION SUPPORT TOOL

3.1 Users

The decision support tool presented in this paper is geared to decision makers and researchers in the fields of transport planning and operations, spatial planning and spatial economics and geography. Generally speaking, it should serve professionals who are interested in mobility and spatial analysis and understand the principles of transport modelling but do not have the expertise to operate an agent-based transport simulation directly. Currently, we envision the following stakeholders and some example questions for our decision-support system. The list is by no means exhaustive and will be discussed with the different users in the near future.

- Transport planners

  How many trips occur where, when and what is the activity purpose?

  What are the socio-demographic characteristics of the these persons?
- Urban planners

  What are the temporal usage patterns of buildings and the surrounding neighbourhood?

  What is the flow from public transport stops to surrounding buildings?
- Policy-makers

  What are the costs and benefits of a new public transport service?

  Who are the winners and losers from constructing a new road?
- Public transport operators

  What is the breakdown of the ridership of certain bus lines?
- Service industry

  Which customers are in catchment areas, separated by mode?

Given the various training backgrounds of the stakeholders, the decision support tool should provide an intuitive data analysis front-end that allows explorative, and hence fast data analysis and visualisation. In addition offer interfaces to analysis tools that stakeholders are already familiar to work with such as GIS.

3.2 Functional requirements

As argued in the introduction, the objective of the decision support tool is to leverage from the wealth of information available from fully dynamic, agent-based transport demand models. At the same time, the tool should allow swift explorative analysis with low computational effort. To reach this goal, we argue that the data structure needs to be clearly and efficiently laid out to avoid duplications and not require vast amounts of storage since normally several different scenarios want to be compared, archived and shared with external parties.

Appraisal methods

In terms of functionalities, the decision support tool should allow to conduct classic transport appraisal methods such as cost/benefit analysis and evaluation of the spatial impact of transport infrastructure and policy measures. To this end, it should feature user- and trip-specific value of travel time savings and stage level attributes such as transfer waiting time as well as access and egress time. In addition, the tool should allow to fully tap in on the level of individual and spatial resolution of the agent-based demand model. This means that any analysis should be an aggregation of agent-specific information on performed journeys and activities. In addition, the tool should be able to perform any sort of spatial analysis on the finest level of granularity provided by the transport model; usually individual buildings or parcels as well as public transport stops and selected links such as count stations or tolled road segments.

Link level analysis

To allow evaluation on the level of individual links, the amount of data to be saved in the database would grow substantially, while at the same time GUI-based visualization packages such as Senozon VIA (senozon AG, 2013) readily allow such analysis. Therefore, we decided to abstain from it except for selected links which allow comparison with road counts and analysis of revenue of tolled links. However, given the specification above, detailed evaluation of any form of transport infrastructure and policy in form of winners and losers, differentiated both spatially and along socio-demographic strata can be specified with and derived from the decision support tool.

Temporal analysis

The tool should be geared to take advantage of the full temporal resolution offered by the agent-based transport demand model. Any analysis should be able to be performed for any pre-defined time interval. This is particularly relevant, as key performance indicators of any transport system such as travel times, speed and congestion are of a transient nature. Therefore, it is important that temporal dynamics of such indicators can be adequately represented by the decision support tool.

Spatial analysis

From the aforementioned requirements it becomes clear that spatial analysis capabilities are necessary. Here we envision possibilities such as selecting a current set of activities and showing the location of an activities.

Validation

Transport operations are more and more continuously monitored by systems such as loop detectors, road tolls or automatic public transport fare collection. These data can be a valuable source for developing and calibrating such advanced transport demand models. In order to compare simulation output with transport operation data data from these different sources should be stored according to similar definitions (i.e. in vehicle time, waiting time, journeys, stages & trips). Similar applies to travel diary surveys, for which usually around a sample of 1% to 3% of the whole population is asked to report both activities and trips for a given day.

3.3 Technical requirements

Concerning the abilities of the employed database, the requirements stated above call for a database which enables spatial queries. In line with the spirit of the main agent-based transport demand software packages, the decision support tool should also be based on open-source software. The tool's front end should allow interactive data exploration and offer intuitive interaction. Finally, the database should feature open interfaces to allow setups with various front-ends tools such GIS and business analytics software.

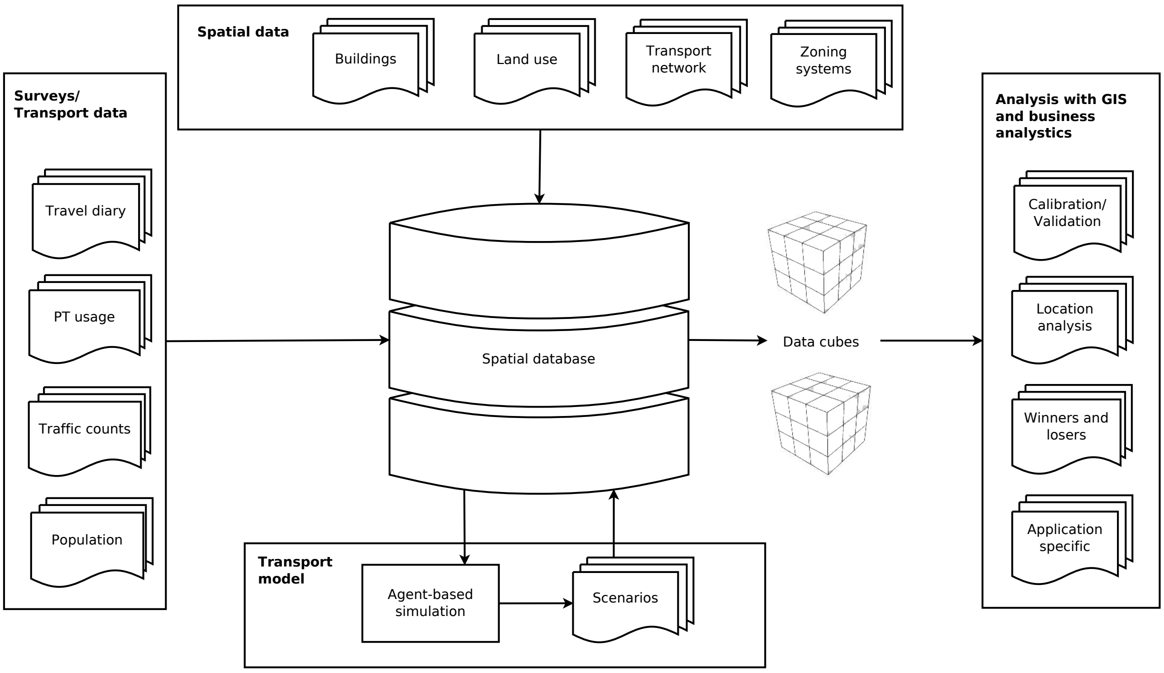
4 GENERAL FRAMEWORK OF TRANSPORT PLANNING DECISION SUPPORT TOOL

An overview on the general framework of a decision support tool to fulfill the requirements stated above is provided in Figure 1. At the core of the system is a spatial, object-relational database which is optimized to store and query data that is related to objects in space, including points, lines and polygons. In addition to conventional databases which operate with various numeric and character data types, spatial databases enable processing of spatial data types, typically called geometry or features. To facilitate efficient spatial query and interaction, a key requirement is that all spatially enable data sets are using the same spatial reference system.

Spatial data such as an inventory of buildings, land use zoning plan, transport network and zoning systems build the organisational reference system of any spatial data in the decision support tool. Any relevant information concerning buildings is stored in the data container 'Building'. This includes information on location, the type and number of activities that can be performed at the same time in a building and if applicable opening times. Depending on the type of building, additional information such as the number and type of individual units within the building building shape (either as 2d footprint or 3d shape), age, typology, ownership and many other attribute can be featured. The land use information can serve to further define building attributes, but mainly serves as functional aggregation layer to derive information such as distribution of accessibility of a certain land use types. The transport network describes transport facilities such as the road, pedestrian and public transport networks with nodes and links but also include features to reference further input data such as counting stations or road toll gates. As declared in the section above, we restrain from tracking and storing information from the transport simulation for all transport network elements, but include such network information to enable GIS-based network analysis and as raster for road count and tolling stations. Information on local zoning systems is mainly to be used to aggregate information to commonly know spatial entities such as neighborhoods or wards to facilitate communication in (political) decision processes.

Data collected in mobility surveys or automated systems such as fare collection or traffic

Figure 1: General framework



counts build the second main data source of the decision support tool and serve mainly two purposes: first, as key data source for the development of the transport model and second, to calibrate and validate the transport model. Travel diary surveys contain information on both the relationship between socio-demographic profiles of individual and households, performed activities and mobility decisions, normally recorded in individual, cross-related tables. To make optimal use of such information, it should be geocoded using the spatial data presented above as reference system. Depending on the availability of an existing, full population census, it additionally might serve for population synthesis. Similar to the travel diary survey data, the population information needs to be geocoded using to the employed spatial data structure. Data on public transport usage should be preferably also harmonised according to the structure of the other data sets. If, for example, detailed ridership information is available from smart card transaction data, it should adhere to the same definitions of stages, transfers and journeys as used in the transport simulation and travel diary survey which might requiring recoding. Cross-sectional information from loop detectors or road toll systems needs to be spatially referenced to the transport network.

The spatial database consisting of both spatial data, mobility survey and population information is the ideal group for the development of and agent-based transport demand model such as MATSim or Transims. Since such models use software specific data types and structures, data conversion algorithms are required to load from and feed back data sets to the spatial database.

Transport demand models are conventionally used to evaluate different scenarios which in turn are constituted of different combinations of infrastructure or policy measures. Therefore, several data sets of which each describes the transport system given a certain set of measures, will be feed back to the spatial database. To facilitate compatibility this data should again follow the same data structure as used when coding actual mobility observation data and use the same georeferences.

Depending on the type of analysis required, it might make sense to compile data cubes from individual tables stored in the spatial database to facilitate conceptually straightforward data analysis operations. For example, for comparison of actual and simulated ridership for individual public transport different tables might need to be unified in a way that allows efficient visual analysis.

As the actual front end of the decision support tool act software that allow intuitive and interactive data analysis. This can be Geographic Information System (GIS) packages and business analytics software. Given the wealth of information stored in the database, various types of analysis are possible. Similar in function to the comparison of different policy scenarios, the decision support tool can be used to calibrate and validate the transport demand model. By employing network based GIS analysis, the decision support tool can even serve to compute customised accessibility measures for location analysis without requiring direct access to the transport simulation model. Since all information on individuals and buildings is stored in disaggregated form in the spatial database, also detailed winner and loser analysis for testing the impact on social equity of transport measure can be performed out side the actual transport simulation. This ensures that results for transport simulations are readily accessible to a wider audience enhancing interdisciplinary planning processes.

5 CASE STUDY: DECISION SUPPORT TOOL FOR TRANSPORT PLANNING IN SINGAPORE

5.1 Data preparation

We use digital map data featuring building footprints with unique identifiers provided by the Singapore Land Authority for the year 2008 as the basis of any building related information. Information on the potential size of each building and its usage is derived from the parcel-based land use Masterplan 2008 provided by the Singapore Urban Redevelopment Authority. Due to special situation that in Singapore almost every building is attributed with an unique postal code, it was in addition possible to enrich this data set with information derived from various additional web resources such as the HDB InfoWeb and REALIS, URA's real estate information system. To allow spatially aggregated analysis various zoning systems including zones and subzones are used. In our case those are the zones and subzones used by URA's Developing Guide Plan (DGP) and Transport Analysis Zones (TAZ) provided by Singapore Land Transport Authority. The employed transport network is based on road navigation network provided by NAVTEQ where information on road capacity and speed as been updated according to the transport planning network provided by LTA. Whereas stations and links of the MRT and LRT network have been added as additional network layer, bus routes and stop locations were matches semi-automatically on the high-resolution navigation networks (Ordóñez Medina and Erath, 2011). To ensure interoperability of the various spatial data sets, we use the same spatial reference system for all spatial data, namely the Tranverse Mercator projected WGS 84 / UTM zone 48N (EPSG:32648).

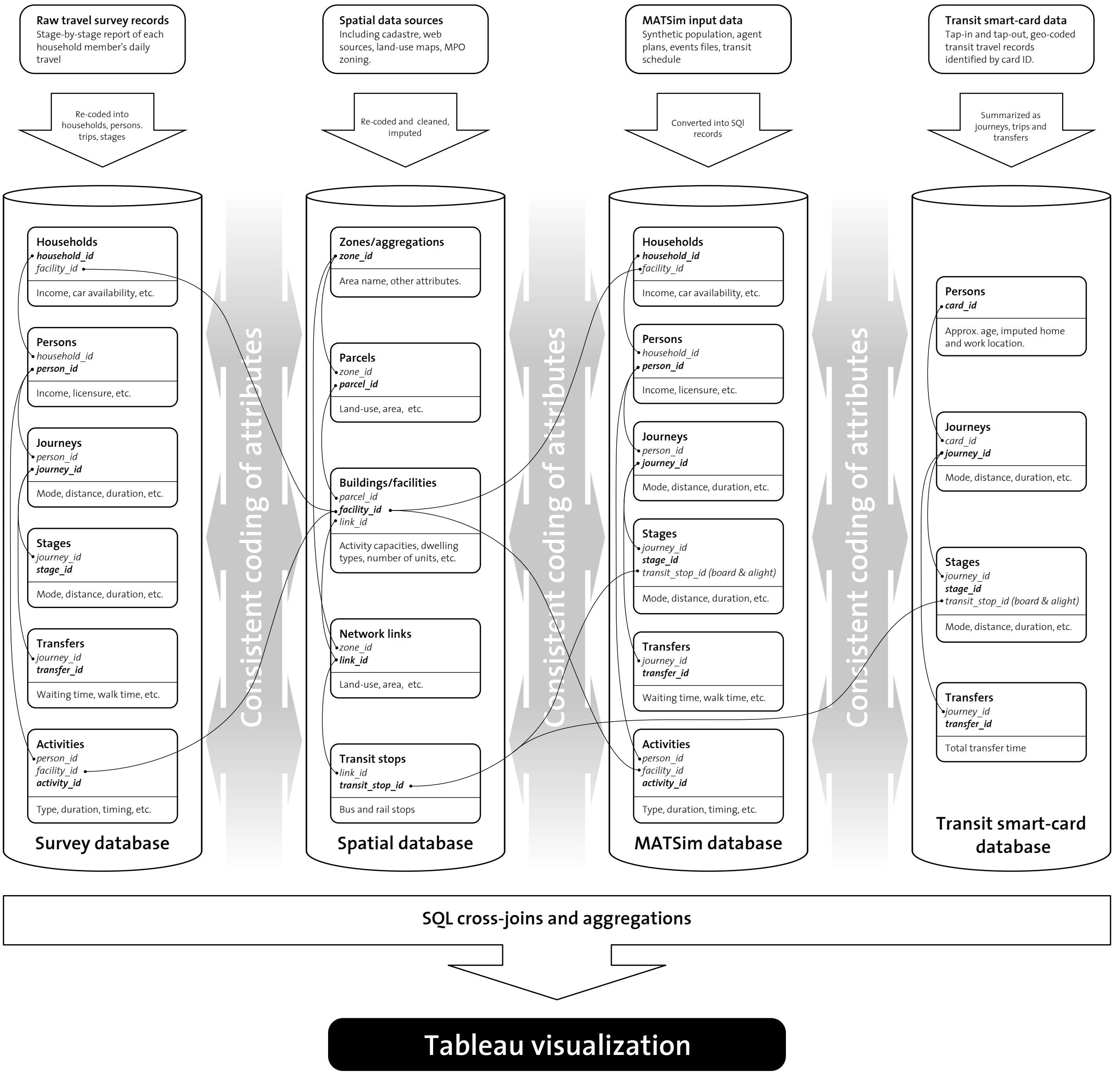
As key source of individual specific mobility and activity data, we use the Household Interview Travel Survey (HITS) (Choi and Toh, 2010) of the year 2008 but recode it according to the data structure requirements stipulated above featuring individual tables containing information on stages, journeys and transfers as well as on persons and households. All tables are connected through unique identifiers which allows combining any sort of available information. Since HITS is fully georeferenced, we were able to link it to the various spatial data sources, in particular the inventory of buildings and land-use parcels.

In equal manner, the CEPAS data has been edited to fit the commonly used spatial data sources. Additionally, we adjusted the definition of a journey deviant from its original definition. In CEPAS, any stage which starts prior than 45 minutes after the last alighting/system leaving record is allocated to the same journey to avoid wrongful double charging of initial journey fares. Reducing this threshold to 20 minutes as the maximal assumed average headway, we ensure comparability with information from HITS and the transport simulation. The CEPAS data features along with boarding and alighting location and respective time stamps for each stage also the public transport service number as well as three user segments, namely students, senior citizens and regular users. For bus journeys transfers can be tracked based on the data as registering both while boarding and alighting is mandatory. For MRT and LRT journeys however, transfers to segment observed journeys into stages are inferred based on the assumption of time-shortest path routing using the MATSim simulation environment.

As documented in Erath *et al.* (2012), the development of the agent-based transport model is tightly related to the information provided by the spatial database. The number of work places we expect in each building, for example, is derived from data mining CEPAS data and combining it with information from HITS, Masterplan 2008 and building information (**?**). In addition, the assignment of work locations draws heavily from HITS as it is used as source to estimate the gravitation model parameters or to generate contingency tables displaying the frequency of different occupations and land-use types. In addition, the MATSim model uses the same spatial objects (e.g. building and public transport stop identifiers) to ensure comparability.

MATSim model simulation output is automatically written into the spatial database. By using the same data structure featuring tables for person, households, activities as well as journeys, stages and transfers, we ensure consistent comparison of attributes. Households and activities are linked to individual buildings which in turn are related to network links. Public transport journeys are georeferenced based on the records of the boarding and alighting stops of its respective stages.

Figure 2: Structure of decision support tool for Singapore case study



5.2 Technological setup

5.2.1 Geospatial database

We use the object-relational database management system PostgreSQL (PostgreSQL Global Developement Group, 2013) as spatial database. It is available for many platforms and released based on an MIT license and thus free and open source software. Featuring a full data definition language (DDL), extensible data types, operators, index methods, functions, aggregates and procedural languages it is offers a very suitable environment for data management and manipulation. Equally importantly, we use the third party extension PostGIS (Refractions Research, 2013), which is also an open source software program, to expand the functionalities of the PostgreSQL database to support geographic objects and spatial queries.

5.2.2 Interactive analysis and visualisation tool

After evaluating various software solutions that allow both interactive analysis and visualisation tools, we decided to use the software Tableau Desktop (Tableau Software, 2013) despite being released under proprietary license. Besides a very intuitive GUI which allows to visualise data by simple drag and drop actions it feature the necessary connectors to link to the PostgreSQL database. Furthermore, Tableau's data engine is designed to query huge data sets hosted by dedicated servers using personal computing devices.

Alternatively, most GIS software, both open source and proprietary such as QuantumGIS (QGIS Development Team, 2013) and ArcGIS (ESRI, 2013), provide connectors to the PostgreSQL database allowing in particular advanced spatial analysis and visualisation options.

5.3 Applications

To showcase the versatility of the decision support system, we present two example applications, namely how it can be used for calibration and validation of the MATSim Singapore model and for explorative analysis of travel patterns observed in a travel diary survey. At the time of writing, both applications are not yet fully developed and only preliminary versions are presented in the following paragraphs.

5.3.1 Calibration and validation of agent-based transport demand model

The key data source to calibrate the transport model is HITS as it includes detailed information both on public and private transport journeys, activities and socio-demographic profile of individuals. Within the MATSim framework, the key behavioral parameters for transport decisions are the perceived disutility of the various elements of a journey. For private, motorised journeys, the employed utility function differentiates for traveling time, distance-based cost and road tolls. For public transport trips, access and egress distance, travel time, fare, number and duration of transfers and waiting time are scored individually. In addition, the model allows to penalise if activities are performed before or after a desired time. Hence, the employed analysis and visualisation dashboards should support immediate comparison between model results and HITS observation along those variables. For journey distance and travel time related variables this can be achieved by a dashboard as exemplarily presented in Figure 3. Since at the time of writing, data from actual MATSim scenario runs has not yet been feed into the spatial database, the underlying data refers to journey attribute as reported by respondents in HITS and re-routed using the MATSim Singapore model. The pyramid graphs on the top half of the dashboard compare the travel time distribution for travel car and public transport and various travel time elements. The display makes immediately noticeable that routed travel times both for car and public transport trips are shorter than as reported. At the same time, the two scatter plots below show clearly that indicated and routed travel distances are similarly distributed. Therefore, the verdict would be that travel times as routed in the MATSim network are systematically shorter than in reality or reported with systematic bias. The fact that the MATSim Singapore version with which we routed the HITS journeys does not account for delays at traffic lights nor includes background traffic demand from light and heavy good vehicle as well as taxis, suggest the first to be the more relevant reason. The distributions of walking time from and to public transport stops fit much better indicating that assumed parameter influence public transport route choice lies in the right range. The rather substantial difference in waiting time distributions can be either caused that in the MATSim simulation, public transport services ran more stable than in reality or that waiting time in HITS is reported with some bias as it is perceived longer than it actually is. Whereas it is very difficult to investigate the second assumption, the decision support tool would allow further analysis such as differentiation of waiting time for different time of day and mode of public transport right away. Like this, the actual problem can be efficiently confined and decided how the simulation potentially would need to be amended to obtain more realistic results.

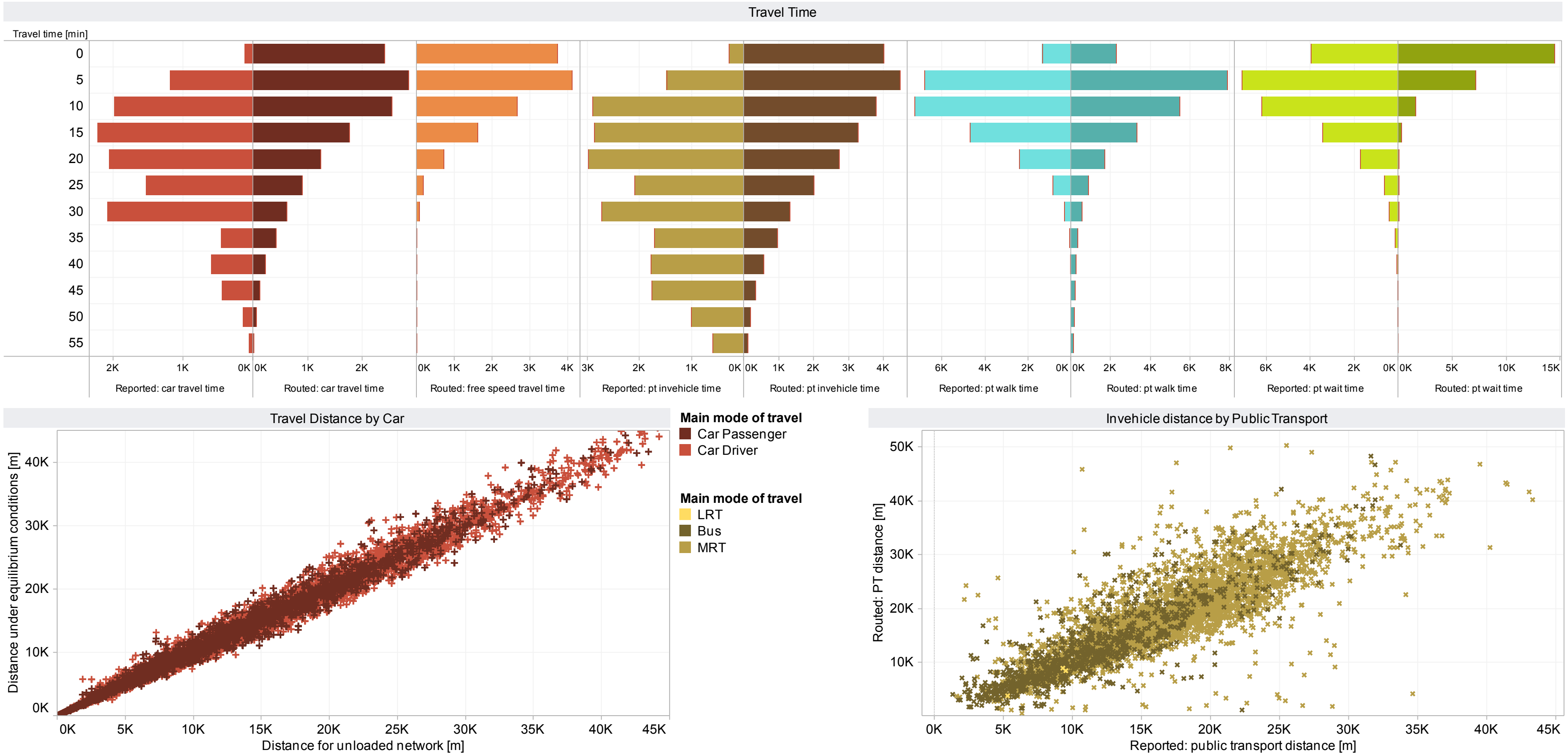
In addition to the graph-based analysis, key performance indicators to describe the overall goodness of fit between actual observations and simulation results for various aspects such as modal share for various user groups or distribution of activity durations can be defined and presented on further dashboards. However, when comparing actual observations and simulation results, it is important to clarify first potential issues in terms of comparability of the data as they might have a different scope. For example, the current version of MATSim Singapore does not include pick up and drop off activities, which are reported quite frequently in HITS. Therefore, the analyst must filter out such trips beforehand which, however, can be done very straightforwardly with the decision support tool.

5.3.2 Explorative analysis of travel patterns

The decision support tool also provides an ideal environment for explorative analysis of travel patterns either drawing from actual observed journeys (e.g. HITS and CEPAS) or MATSim model. Figure 4 shows a dashboard designed to interactively analyse journeys to work as simulated with the MATSim Singapore model. The four maps in the top pane serve as spatial selection interfaces providing filtering options either based on home or work location. Records can be filtered both based on predefined polygons or rectangular selection. All maps are interlinked meaning that given a selection of home locations only the related work locations are shown. Equally, all bar charts are automatically updated according to the spatial selection.

The bar charts shown in this dashboard are designed to discover patterns of the reported travel model to work with respect to the socio-demographic profile of the respondent. Each bar and bar segment can be interactively selected to update all other graphs an maps accordingly. In

Figure 3: Mock up of calibration dashboard to compare journey travel time and distance



addition, the filters arranged on the pane at the right side of the dashboard allows further filtering options.

Again, the presented dashboard should only exemplarily show the functioning and abilities of the decision support system. The main advantage is that such dashboards can be developed rapidly according the type of analysis needed once the data is available in the spatial database in the proposed structure.

6 OUTLOOK

With the progress in transport demand modelling from aggregated, static and zone-based assignment models to fully dynamic agent- and activity-based transport simulations, the scope of data that needs to be collected, managed and maintained has changed dramatically. Furthermore, the type of analyses as well as potential analysts have changed. Whereas traditionally transport modelers would analyze data and create static maps, pre-defined graphs and tables we envision other end-users: transport authorities, architects, policy-makers and the service industry. These analysts can utilize a decision-support tool, based on the output of an agent-based model as well as analyze their own datasets and fully analyze spatial and temporal agent behavior without requiring continuous support from a transport modeler.

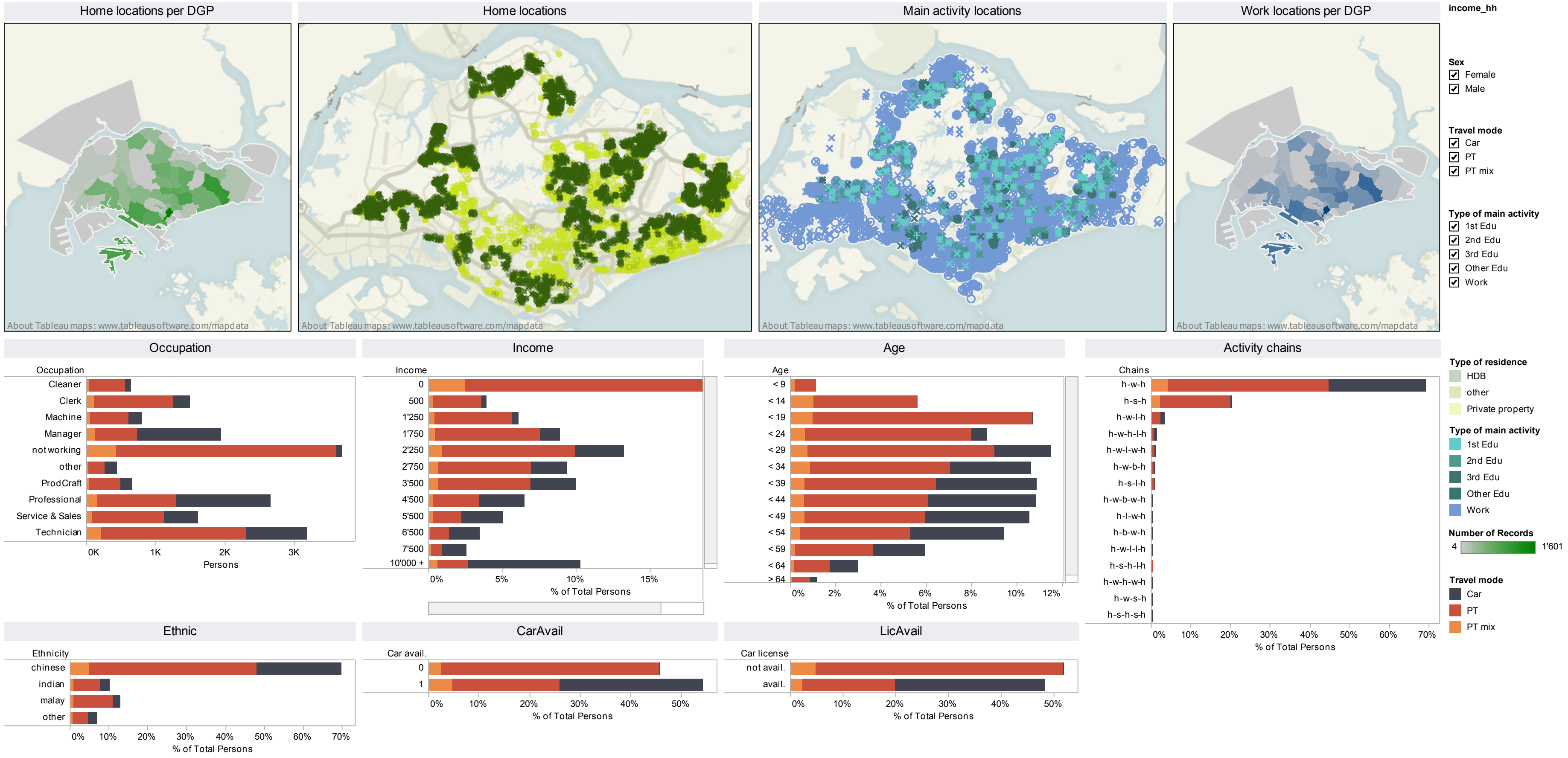
In this paper, several potential questions a user might ask our decision-support tool as well as a number of requirements were sketched for several user groups. We will further refine these questions and requirements in a workshop with relevant stakeholders. Based on these requirements, data aggregations consisting of dimensions and measures as well as a set of data dashboards will be designed. On one hand, these data aggregations and dashboards will serve as a validation tool for our MATSim model of Singapore. Transport and land-use authorities can compare the results of the model against surveys, smart-card data, traffic counts and of course common sense. Once trust in the MATSim model has been gained, these aggregations and dashboards can be used to analyze different pre-defined scenarios, not only in a single stakeholder setting but with multiple stakeholders to facilitate group decision-making.

Whereas in first instance we will focus on the evaluation of pre-defined scenarios, in the next stage we intend to expand the scope of our MATSim model. Currently, this scope is defined by a 24 hour time horizon with a fixed initial demand, consisting of a population and a wide range of activity opportunities. This scope will be expanded to create a 3D city model that continuously ’lives’. Data feeds enter the MATSim model, such as new residential locations, new work locations and a changing population. This living model of a city will support decisions across disciplines, thus facilitating truly interactive and inter-disciplinary policy-making.

REFERENCES

Balmer, M., K. Meister, R. A. Waraich, A. Horni, F. Ciari and K. W. Axhausen (2010) Agenten-

Figure 4: Dashboard for explorative analysis of journeys to work as modelled MATSim (1% sample)



basierte Simulation für location based services, *Final Report*, F&E Förderung: Science to Market, **KTI 8443.1 ESPP-ES**, Datapuls AG, IVT, ETH Zurich, Zurich, February 2010.

Bekhor, S., C. Dobler and K. W. Axhausen (2011) Integration of activity-based with agent-based models: an example from the Tel Aviv model and MATSim, paper presented at the *90th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2011.

Choi, C. C. and R. Toh (2010) Household interview surveys from 1997 to 2008: A decade of changing travel behaviours, *Journeys*, **5** (May 2010) 52–61.

Erath, A., P. J. Fourie, M. A. B. van Eggermond, S. A. Ordóñez Medina, A. Chakirov and K. W. Axhausen (2012) Large-scale agent-based transport travel demand model for Singapore, paper presented at the *13th International Conference on Travel Behaviour Research (IATBR)*, Toronto, July 2012.

ESRI (2013) Arcgis, software, `http://www.esri.com/software/arcgis`.

Few, S. (2012) *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, 2nd edn., Analytics Press, USA.

Kimball, R. and M. Ross (2002) *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, Wiley.

MATSim (2013) Multi Agent Transportation Simulation, webpage, `http://www.matsim.org`.

Nylund, A. L. (1999) Tracing the BI family tree, *Knowledge Management*, July 1999.

Ordóñez Medina, S. A. and A. Erath (2011) Semi-automatic tool for map-matching bus routes on high-resolution navigation networks, paper presented at the *16th international conference of Hong Kong Society for Transportation Studies*, Hong Kong, December 2011.

PostgreSQL Global Developement Group (2013) PostgreSQL: Release 9.2, software, April 2013, `http://www.postgresql.org/`.

Power, D. J. (2003) A brief history of decision support systems, *DSSResources.Com*.

QGIS Development Team (2013) Quantum gis, software, `http://qgis.org/`.

Refractions Research (2013) PostGIS, software, April 2013, `http://postgis.net/`.

Schlegel, K., R. L. Sallam, D. Yuen and J. Tapadinhas (2013) Magic quadrant for business intelligence and analytics platforms, *Gartner*.

senozon AG (2013) via – visualization and analysis tool, webpage, `http://www.senozon.com/matsim/analysistool`.

Sol, H. G. (1985) Expert systems and artificial intelligence in decision support systems, in H. G. Sol, C. Takkenberg and P. de Vries Robbé (eds.) *Proceedings of the Second Mini Euroconference*, Kluwer Academic, Lunteren, the Netherlands, November 1985.

Sprague, H. J., R (1980) A framework for the development of decision support systems, *Management Information Systems Quarterly*, **4** (4) 1 – 26.

Tableau Software (2013) Tableau desktop, software, `http://www.tableausoftware.com/products/desktop`.

TRANSIMS Open Source (2013) Getting Started with TRANSIMS, webpage, `http://code.google.com/p/transims/wiki/GettingStarted`.

Tufte, E. R. (1986) *The visual display of quantitative information*, Graphics Press, Cheshire, CT, USA.

Vogel, D. and J. Nunamaker (1990) Group decision support system impact: Multi-methodological exploration, *Information & Management*, **18** (1) 15 – 28.