VALIDATION OF COMPLEX AGENT-BASED MODELS OF SOCIAL ACTIVITIES AND TRAVEL BEHAVIOUR

Nicole Ronald, Theo Arentze, Harry Timmermans

Eindhoven University of Technology

PO Box 513, 5600MB Eindhoven, The Netherlands

n.a.ronald@tue.nl

ABSTRACT

Recent travel forecasting models are based upon the fact that travel is derived from the activities in which people participate. The underlying architecture of these models are becoming more behavior- and individual-driven, and as a result, researchers are now looking towards agent-based modelling techniques for model development. This move provides new challenges with respect to validation. We describe a validation process, developed using techniques from transport modelling, agent-based modelling, and social simulation, for a complex model of joint social activity generation and scheduling. The tests focus on the activities predicted by the model, in particular when and where, and measure the extent of their similarity with theory, expert opinion and data collected in Eindhoven. We provide a recommendation of how agent-based travel behaviour models can be validated and the resources required.

Keywords: activity-based modelling, validation, social networks, agent-based modelling

INTRODUCTION

Models of travel demand are becoming more disaggregate and behaviour-driven. As a result, agent-based modelling is gaining popularity and has been used for, amongst other applications, modelling pedestrian behaviour (Hakley et al., 2001) and route choice behaviour (TRANSIMS, 2010; MATSIM, 2010).

This shift from single-facet aggregate models to multi-faceted disaggregate models involves a substantial increase in complexity, which in turn has major implications for the estimation and validation of these models. Consider the example of activity-based models: simultaneously estimating all parameters has been shown to cause various problems.

Consequently, models such as CEMDEP (Pinjari et al., 2007) consist of suites of separately estimated models which are linked in the context of a micro-simulation. Similarly, models such as Albatross (Arentze and Timmermans, 2001) assume a process model, linking the various choice facet models, estimated separately according to this sequential process model. The validation of these complex models is also considerably more complicated and raises questions such as should every model be validated or only the overall model? For example, Han et al. (2009) argued that it may not be possible for complex dynamic models of activity-travel behaviour to estimate and validate every parameter of the system, but that some partial estimation and validation and some calibration may be necessary. The approaches, standards and protocols of estimation and validation that the field has developed for aggregate and disaggregate choice models may not be fully applicable. What then can we do?

The problem of estimation and validation of multi-agent models tends to be more complicated. A model such as Feathers (Bellemans et al., 2010), for example, assumes that individuals have context-dependent aspiration levels and learn. How should this and similar models be estimated if data on all included concepts are not available?

The aim of validation is to describe what the model is capable of doing, and then the user can determine whether it is suitable for its purpose (Amblard et al., 2007). Empirical validation has traditionally been dominant in transport modelling, where model outputs are compared to data collected from real systems. However, these statistical techniques are not always applicable to agent-based models due to lack of data and possible chaotic/non-linear behaviour in the system (Klügl, 2008). Despite this, several methodologies have been proposed, which include a combination of face and empirical validation tests. The processes within the model are also inspected as well as the model outputs.

This paper describes a validation process, developed using techniques from transport modelling, agent-based modelling, and social simulation, for an agent-based model of joint social activity generation and scheduling. The paper is meant to trigger thought on validation issues for the increasingly more complex models that the transportation community is developing. We therefore begin with a discussion of the challenges of and approaches for validating social simulations and agent-based simulations, followed by a review of validation techniques for activity-travel models. Our case study is described, along with an overview of the available data, and a suggested validation process is presented. Although our process is tailored to our model, the steps may be appropriate for other individual-based activity-travel models.

WHAT IS VALIDATION?

> Modellers often plunge into the difficulty of setting up a set of rules and building a model. Yet the process of validation requires a clear view of what the model is attempting to explain and for what purpose. What are the key facts that the model needs to explain and how well must it do it? (Ormerod and Rosewell, 2006)

Validation is an important part of the modelling process and cannot be ignored until after the model has been developed. It differs from model verification in that verification is about checking that the model has been built correctly following a specification, whereas validation deals with whether the right model has been built.

Gilbert (2004) explains that validation is often considered to be only about comparing outputs to observed data at the expense of being able to use the model to increase understanding of a system or, as is generally seen in the transport field, to experiment with policy changes. This latter criterion is just as important. Validation is not concerned with showing *whether* the model is useful, but how (Louie and Carley, 2008).

Carley (1996) states that "the level of validation chosen depends on the model's purpose", therefore the purpose of the model will often lead to the amount of validation possible. From the other point of view, the data available could also determine the possible validity.

VALIDATION ISSUES AND CONCEPTS IN AGENT-BASED MODELLING

Agent-based models "consist of a system of agents and the relationships between them" (Bonabeau, 2002). The agents perceive their environment and other agents, make decisions following some rules, and act, possibly changing the environment in the process. Agents can also evolve over time, learning about their past experiences. Modelling at such a low level is sometimes more "natural" than attempting to, for example, create flow equations.

These components mean that the model can become very complex very quickly and as a result, validation of agent-based models is not a simple task. Making small changes at the micro-level, for example to the decision rules or interactions between agents, could produce very different results at the aggregate level. As a result, it becomes necessary to validate at both the macro/aggregate and micro/disaggregate level and both are difficult, if not impossible (Gilbert, 2004). Batty et al. (2004) note that as we move to richer and more detailed model structures, that they cannot be wholly validated.

Klügl (2008) lists several problems with agent-based models that may hinder validation:

- it is not as easy to collect descriptive values at the individual level for comparison to data as it is at the aggregate level;
- nor is it easy to collect the real-world data for comparison at the individual level;
- methods for validating the dynamics of an agent-based model are underdeveloped;
- models may exhibit chaotic behaviour as a result of feedback and non-linearity in the system, which is difficult to validate;
- the time and other resources required to obtain the necessary model outputs should not be underestimated; and
- too many parameters used with an automated optimising calibration should be able to fit the data in some way, so it is not possible to reject the model.

Types of validation

Troitzsch (2004) defines three types of validity following Zeigler:

- replicative: matches data already collected;
- predictive: matches data before they are collected;
- structural: matches data and the processes of the real system.

This was further elaborated by Klügl (2008), who proposed the following categories in two dimensions:

- the approach: face (human-observable tests) vs. empirical (statistical tests);
- the observed element: behavioural (studying the input and output of the model; encompasses replicative and predictive validity) vs. structural (the relations and reasoning in the model)

Returning to the idea of "how" a model is useful, she notes that both informal and formal validation techniques are required, and that different tests are needed for different purposes (see Table 1).

Types of validity	Behavioral v.	Structural v.	Both
Face v.	illustration	Understanding	training
Statistical v.	regression-like forecast	Prediction	what-if forecast
Both	more reliable forecast	more reliable prediction	strategic advice

Table 1 "Relation between type of validity and simulation objective" (Klügl, 2008)

Validation processes

Klügl (2008) describes a process for validating agent-based models, beginning with face validation, in particular animations, assessing output, and tracking single agents. Once this is satisfactory, sensitivity analysis should be undertaken. Parameters should then be calibrated, and after this stage, statistical validation can be completed. In the final step, all data should be used to thoroughly test the model.

Barlas (1996) presents process validation, which is focussed on structural validity. There are several types of tests described:

- direct structure tests: the structure of the model is compared with knowledge about the real system structure. Tests include comparing with information gathered from the system being modelled (qualitative and quantitative) and from theory, comparing equations with knowledge, and extreme value testing on individual equations with a comparison with the real world.
- structure-oriented behaviour tests: the structure is indirectly testing by looking at behaviour. Tests include extreme condition testing (is the real world also sensitive to

the same parameters?), relationships between variables (phases), and modified testing (if a real system can be modified in some way and data collected, does the model change in the same way?).

- behaviour pattern tests: moving on from the structure, are the outputs sensible? In particular, patterns (periods, frequencies, trends etc.) are more important than replicating data points. This can be done statistically (means, variances etc.) or visually.

Both processes comply with Gilbert's recommendation to validate at several levels.

VALIDATION OF ACTIVITY-BASED MODELS

Activity-based models differ from previous transport modelling approaches by "modelling relationships between individuals, households, and cities" (Buliung and Kanaroglou, 2007), which is not dissimilar to our definition of an agent-based model. Instead of only modelling the trips or trip-chains (tours), the activities of individuals are modelled and the travel derived from the activity-chain. This implies that a disaggregate approach is required in order to model individual behaviour and the associated spatial and temporal constraints.

Several models have been developed and are at various levels of maturity. The structure of models varies: some are based on utility maximisation and optimisation of daily schedules (e.g. PCATS), while some are rule-based (e.g., ALBATROSS).

These models are built for predictive purposes. In comparison to earlier transport modelling approaches, activity-based models are considered to be more behavioural and are therefore more transferable to different regions (Pendyala et al., 2004), however the focus has been on predictive validity and empirical comparison to data. Kurth et al. (2008) recognise that the separate components of an activity-based model (e.g., tour-level mode choice, tour-level mode choice, tour-level time-of-day choice, trip-level time-of-day choice) require testing. Sensitivity is also considered as important, however is more of a reasonableness test. Two overall sensitivity tests have been proposed in Kurth et al. (2008): comparing temporal outputs to a calibrated four-step model, and policy-oriented tests. Kurth et al. (2008) propose changes to the spatial environment only (e.g., different development densities, urban spread), whereas Pendyala et al. (2008) also mention that socio-demographic changes and pricing policies should be explored.

Buliung and Kanaroglou (2007) state that validation of activity-travel models is usually aggregate and consists of residual analyses. They point out that PCATS takes an average result over several runs and compares predicted and observed means for several activity-travel indicators, which is an accepted practice for validation of agent-based models.

Pendyala et al. (2004) used data from the 1999 Southeast Florida Household Travel Survey to validate the PCATS model. They presented mean comparisons of model outputs and observed data for the following variables:

- average daily trip rates (worker, student, other);
- average daily fixed and flexible activities (worker, student, other);
- first home departure time and final home arrival time (worker, student, other);
- modal split (single occupancy vehicle (SOV), HOV driver, HOV passenger, transit, other) to fixed and flexible activities.

Pinjari et al. (2007) describe the validation techniques used for CEMDAP, an activity-based model based on discrete choice. Activity diaries containing information about the type of activity, location, start time, end time, and transport mode were used. The measurements used were percentage shares (for discrete choices) and distributions (for continuous choices). Pattern-, tour- and trip-level characteristics are investigated. Comparison with the four-step model is also undertaken.

A type of sensitivity analysis was also carried out, in that several scenarios were tested: an increase in in-vehicle travel time, increase in costs, and increase in population. The outputs discussed (at pattern level) were the number of worker tours and stops, work start and end times, trip chaining, and average daily duration of activities.

The first round of validation for ALBATROSS used data from an activity-travel survey carried out in the Rotterdam region in 1997. The data included, for each activity, the purpose, the start and end time, the location, the mode used and travel time, and if others were involved in the activity. The outputs of the model that were compared were the mode for work, activity with whom, activity duration, start time, trip-chaining, and activity locations.

The re-estimation of ALBATROSS looks at the model components separately. At the tree-level, goodness-of-fit, chi-square (discrete) and F-stat (continuous) tests are undertaken. Predicted activity-travel patterns are compared with observed using string alignment techniques at the disaggregate level. At the aggregate level, several model outputs, such as number of work/secondary fixed/flexible activities, tours, activities in tour mode, activity type, mode of first link, activity duration, travel party, trip chaining, distance, and work duration are compared to observed data using chi-square/frequency differences and t-test/mean differences.

This brief summary of validation attempts of activity-based model shows that although the model are more complex than trip and tour-based models, the principles of validation have not fundamentally changed. The reason is that the activity-based models described in this section restrict themselves to modelling variables based on observed data. They do not include more abstract behavioural concepts or principles and they are not dynamic and except for Albatross, they are founded on full behavioural outcomes as opposed to an assumed underlying process that leads to emerging aggregate patterns. In that sense, validation of the majority of activity-based model implies a test of goodness-of-fit, and not a validation of the underlying behavioural process. However, the latest generation of *dynamic* activity-based models focus on social networks, aspiration, dynamic choice sets and similar concepts. For the reasons discussed, existing validation protocols may not be sufficient and not be applicable to these kinds of models.

CASE STUDY: JOINT SOCIAL ACTIVITIES

Social networks are of interest to transport modellers as people often travel to visit friends and activities are often undertaken together with non-household members. Joint activities with household members have been modelled (e.g. Anggraini et al., 2009). The ways in which social networks are being incorporated range from long-term social network dynamics and the influence on travel modes and housing location to daily activity scheduling. We focus on the latter, in particular by building a multi-day model of activities. As the model is multi-day, and not a single day optimisation model generally used in activity-based modelling, the social network needs to be explicitly modelled so that there is some consistency in terms of the activity participants and the locations people are visiting.

Recent travel forecasting models have focussed strongly upon the fact that travel is derived from the activities in which people participate, such as work, school, shopping, sport, leisure, and social events. Non-discretionary activities such as work and school can be partly explained by the traveller's sociodemographic characteristics and generalised travel costs. Participation in and scheduling of other activities is not as easily predicted.

Participation in social activities is determined by the groups that one is a member of, i.e., their household, their workplace/school, sporting groups, voluntary organisations and clubs. These groups form part of an individual's social network, which is a representation of the people one interacts with, and sometimes also contains an indication of how people are connected and how strongly.

The main processes associated with social networks are selection and influence. Networks evolve over time, as people leave jobs, move, join new clubs, and meet new people. The strength of the connections between people also change over time. Interaction with others is required to maintain relationships over time. These activities may not necessarily involve travel and meeting with people in person, especially in the current environment where email, phone, instant messaging, and social network sites are widely accessible. However, this project is mainly interested in day-to-day/short-term social activities that do have a travel component.

Individuals influence each other by sharing information, both intentionally and unintentionally. Relevant examples include telling a colleague about a new location, which may then become an option in the colleague's choice set for future activity locations with other friends.

Understanding the social network that lies on top of the spatial network could lead to better prediction of social activity schedules and therefore better forecasts of travel patterns, in particular for social and leisure activities. Taking into account the purpose of validation, we are interested in modelling who people are meeting, where they are meeting and how often.

The model generates social activities for individuals, who are represented as utility agents. Activities consist of participants, location, day and time, as well as the mode used for each participant. In order to make decisions about activities, the agents firstly use a utility function to choose who to interact with, then follow an interaction process to decide when and where to meet. During this process, information about locations is shared, and if all agents agree on the day, time and location, an activity is scheduled for a later date.

Social networks

Social networks are a representation of individuals and relationships between them. The relationship between two individuals can be defined in a number of ways, for example how similar they are, how they are related to each other, whether they interact or how often they interact, or how information flows between them (Borgatti et al., 2009).

The network can be represented in two ways: complete or personal. A complete network contains all of the relationships for all the individuals in the network, for example, all the friendship links between students in a class. Personal networks contain the relationships for a particular individual (known as the ego), however the attributes of the people they name (known as alters) are provided by the ego rather than the alter themselves. It is not guaranteed that the personal networks of egos in the sample will intersect.

As Newman (2003) recognised, research has been slow in understanding the actual workings of networked systems and the focus has been on structural form and analysis. As a result, there are many measurements for generating and comparing static, complete networks. As research moves towards dynamic networks, then the limitations listed by Klügl appear, in particular the requirement for individual level data. Collecting social network data can be expensive, depending on the domain and collection process.

Data

As mentioned earlier, the data available can determine the amount of empirical validation that can be undertaken. Several different surveys of social activities and/or social networks have been undertaken, however the techniques are still being refined and therefore the regions involved and data collected differs. Axhausen (2007) notes that existing methods for activity data collection can be used, however some extra details for each activity is required, namely social purpose, the beneficiary of activity, party composition (travelling), composition of party at event, locations of travellers/participants prior to activity, and distribution of activity costs. Some general activity data collections already contain with whom the activity was undertaken.

The majority of the data collections are concerned with the social networks in the population and either do not collect detailed activities or investigate networks and travel habits over a longer period of time.

In the former category is the Connected Lives data collection undertaken in Toronto, Canada in 2004-5, during which meeting frequencies were collected. Participants were also asked to describe several recent/typical activities, in particular the activity purpose, time/day and duration, location, who else was involved, mode, and how it was planned (Carrasco et al., 2008).

In the latter category, the focus is on where people went to school/university or where they have worked, where they have lived and why did they choose to live there, typical travel modes throughout their life, people they have stayed in touch with or lost contact with, and how they keep in touch with people (Ohnmacht and Axhausen, 2005). Following on from this survey is a snowball survey currently underway in Switzerland (Kowald et al., 2009), in which respondents are asked similar questions about their social networks, including placing their friends in casual groups, followed by the completion of an 8-day activity diary. The friends mentioned in the survey are then contacted and asked if they can complete the same survey.

Van den Berg (2008) collected data in the Eindhoven region which will be the source for our model. For each person, we have details about their socio-demographic data (age, gender, household composition, education level, income, occupation), their access to transport modes, their house location and therefore neighbourhood details, and access to and use of means of ICT (computer, mobile phone, etc.)

We also have access to their interactions over a two day period. For each person they interacted with, we have details about the:

- Characteristics of the alter (age, gender)
- Nature of relationship with the alter (social category: friend, relative, colleague, etc)
- Distance between homes
- Strength of the tie between ego and alter
- Frequency of social interaction by different communication forms
- How long they have known each other

For each interaction, we have the communication medium (e.g., face-to-face, phone, email, IM), the purpose of the interaction, group size (broken down by gender), the category of the location (e.g., home, shops, cafe/restaurant etc.), how far the person travelled to the location and how, and whether the interaction was preplanned, routine or coincidental.

Depending on the purpose of the model, the different data sets can be used for *partial* validation. For instance, the interaction data set and associated social network data cannot be used to test the long-term dynamics of the social network. In order to validate the individual decision-making process, individual stated-preference type surveys may be needed. This has already been recognised by the FEATHERS group, who are planning several different data collections to collect the required data for a dynamic activity-based microsimulation (Bellemans et al., 2010).

Related verification/validation work

Similar to Hackney and Marchal (2009), sensitivity analysis can be undertaken. Four types of inputs were varied: the starting social network (none, a random graph, and a random graph with addition and deletion of links), social interactions (none or sharing one location with a friend per time step), utility function, and replanning (varying proportions of changing route, changing activity time, changing locations based on agent knowledge or the whole environment). The outputs described in detail are:

- the number of people travelling at a particular time of day;
- the degree distribution of the social network; and
- the distance between home locations of connected pairs.

They also collect overall values for average trip distances, average trip duration, and the number of clusters and components in the social network. This shows an overlap between the outputs generated for activity-travel and transport models and social network models. By generating aggregate characteristics of social networks about which data are available as part of a sensitivity analysis, the changes induced by changing parameters can be checked to see if they are consistent with behavioural assumptions, which can be derived from theory, statistical analyses or even qualitative research.

PROCESS

In case of complex agent-based model, process validation is arguably just as critical as outcome validation and sensitivity analysis. The following set of approaches may be used for that purpose.

Direct structure tests

Structural validity is required if we want to say anything about the explanatory nature of the model, as output-only validity is only useful for prediction. We should be looking for “agent reasoning” and “causal relationships between variables” (Klügl, 2008).

If the structure of the model more closely reflects the real system, then if the real system changes then the model should be able to adapt. McNally (2000) noted a limitation of the four-step model in that it had “inadequate specification of interrelationships between travel and activity participation and scheduling” and therefore could not handle changes in, e.g., work hours, peak hours.

The equations and processes can be checked against the literature (theoretical structure). The equations can also be checked individually to ensure they perform as intended. For example, does the utility function capture time-of-day preferences? Processes, such as interaction protocols developed to make agreements, can also be tested individually. Again, in simple models, the behavioural of the model can be proven analytically. However, in complex model, emerging patterns and system response cannot be analytically derived. This is the very reason that in the agent-based research community often numerical simulation is used not only to illustrate the workings of the model, but also to examine alternative trajectories.

As an example, our model contains a utility function to determine who to undertake an activity with. One of the variables is the time since an individual last saw a friend, which has an associated parameter. Each individual will evaluate the utility of meeting each of their alters, and the alter with the highest utility will be chosen for an activity. The model was run many times with different parameters for one shot at choosing a friend. Figure 1 shows the relationship between the parameter and the time since last seen for the chosen alter. This shows a relationship between the parameter and the variable and is useful for checking that the model is performing as expected.

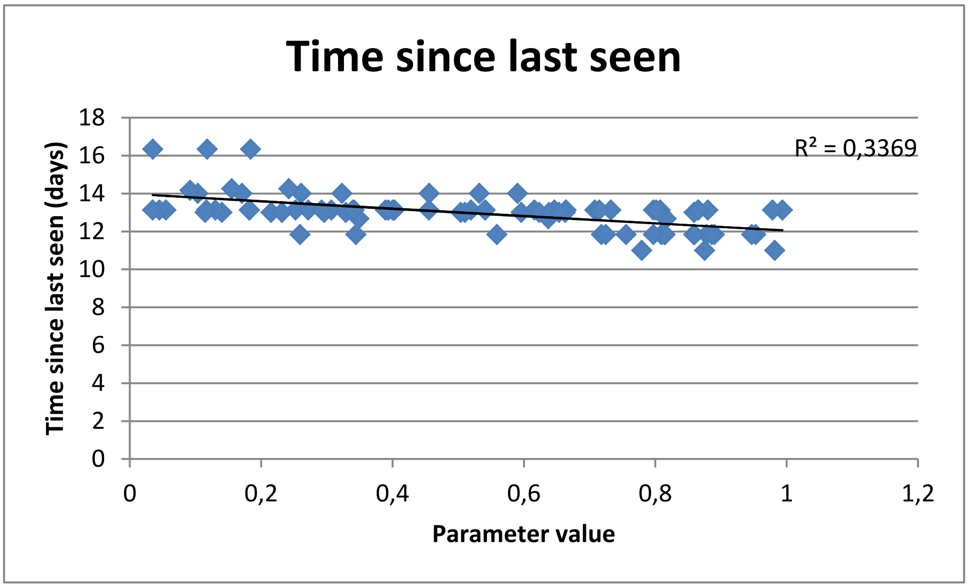

Figure 1 Testing the sensitivity of a parameter on an utility function

Face validation

The three techniques advised for face validation by Klügl (2008) are inspecting animations of the model, looking at the output of the model, and looking through the eyes of one agent in the model (immersion). Gilbert (2004) notes that face validity may be the only option available to micro-level validation in the absence of data on how individuals make decisions.

We can observe the dynamics of the social network over time, looking at the links created and deleted. However, social network dynamics are difficult to validate as the data collected is only a snapshot and also the time range of our model is too short-term to require explicit modelling of events that could change a social network such as marriage, children, and commencing a new job or at a new university.

The best approach for the social network is to check that it looks like a social network based on the theory regarding social network properties. As an example, Hamill and Gilbert (2009) mention properties such as low whole network density, personal networks with limited size and different sizes, fat-tailed distribution of connectivity, assortativity on degree (i.e., many links connect to others with many links), high clustering/homophily, communities, and short path lengths.

Another possibility is to use the methods described in the exponential random graph (e.g., Robins et al., 2007) and dynamic network literature (e.g., Snijders et al., 2010). This uses statistical methods to look for patterns in networks, such as the number of triads in the network, in order to compare to real-world networks. However, both of these may need to be adapted to work with personal or non-complete networks.

We are interested in the change and correlation in location knowledge and whether this is affected in some way, however we have no data regarding this. Theory on influence and rumour-spreading can be used to check if the information exchange is reasonable.

Looking at a trace of a particular agent will be useful to look at their decisions. This could be done either as text, a static map, or animation. For social interactions, it is also useful to look at a pair of individuals or a dyad. Figure 2 shows a static map which represents where two individuals have travelled in an environment. Locations they have both visited can easily be seen, as well as locations they visit often.

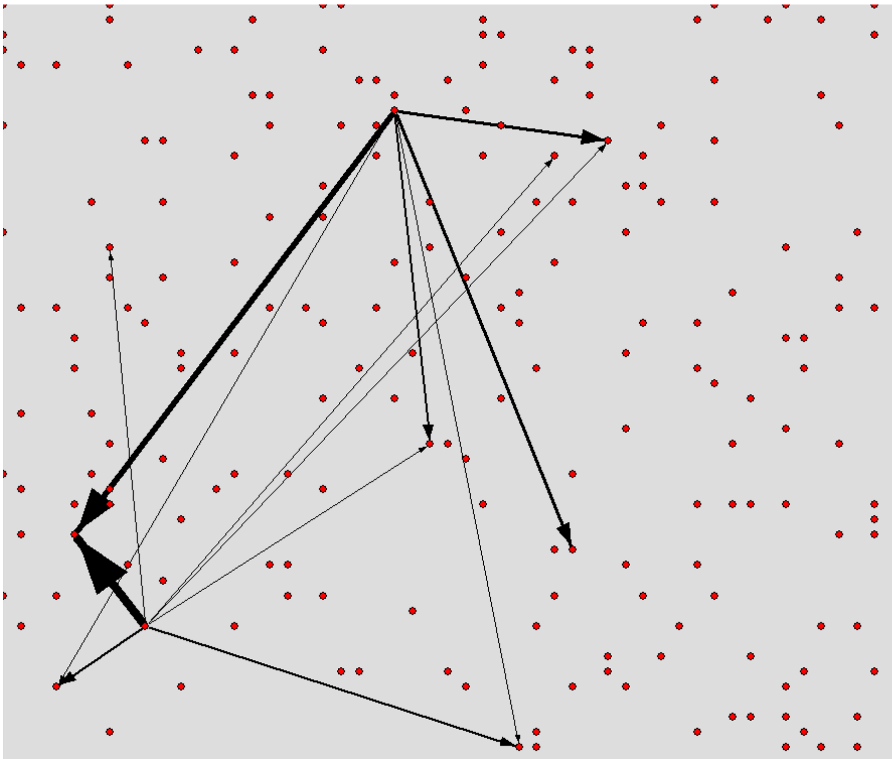

Figure 2 The locations visited by two individuals from their homes. The thickness of the lines indicates frequency.

Structure-oriented behaviour

Sensitivity analysis

In sensitivity analysis, we want to determine the effects of different parameter values and inputs. It can be undertaken before calibration, in order to determine if any parameters are insignificant enough to be removed, or afterwards, in order to explore the effects of policy changes.

A first step is to test the seed variability by varying the seeds. A number of runs will be undertaken (~50) with the same parameters. This will work best with the outputs that provide a single value for a run, such as total/average activities and total/average travel.

As an example, the number of activities for a run can be counted as a single number. By running the model many times with different seeds, the mean activities across all runs can be measured. This shows the central tendency in the runs. Another measure is the ratio of the standard deviation to the mean. Figure 3 shows sample figures for the number of activities for our prototype model. The y-axis on the left graph is limited to one standard deviation from the final mean, so the variation of the number of activities is minimal.

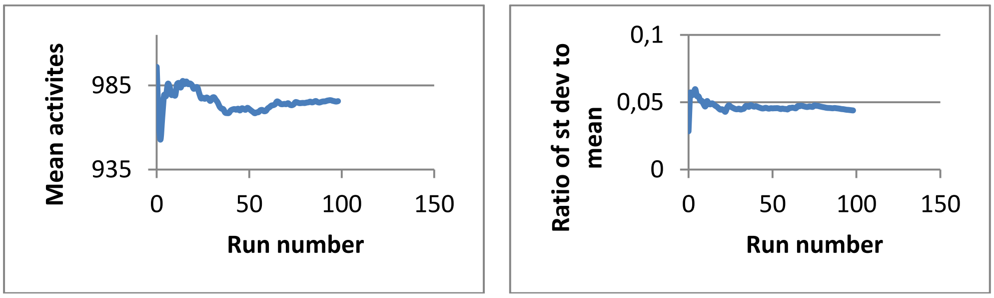

Figure 3 A sample seed variability test: the mean number of activities (left) and the ratio of the standard number of activities to the mean number of activities (right).

Sensitivity analysis can take the form of either altering parameters given a “base scenario" (i.e., changing single inputs) or providing different scenarios (e.g., increase in car ownership). For our model, different parameters, such as thresholds for utility functions and input matrices, can be altered to see the effects. Extreme values or bounds are also of interest.

Kleijnen (1995) notes that Design of Experiments is commonly used for sensitivity testing, but often in an inferior manner: only one parameter is changed at a time. Ideally interactions between parameters are also required, however this is time-consuming. We are looking for the influences of parameters on outputs, for example in the form of elasticity coefficients (e.g., a change of x% is observed for a particular output when an input is changed by y%) (Chattoe et al., 2000).

The latter option of changing scenarios can also be used. As with other transport models, the environment can be altered. For models that include a social network component, the initial network can also be altered, along with the strategies for individuals in terms of decision making.

Behavioural pattern tests

For this stage, we are looking to match patterns found in the observed data. This is a form of statistical analysis as defined by Klügl (2008).

Calibration

Calibration involves finding the parameter values that produce appropriate outputs. One method of doing this is to create a model population based on the surveyed population and taking a two-day sample from the model of their behaviour. The modelled aggregate behaviour can then be measured against the actual aggregate behaviour. We will need to use statistical measures that are parameter-sensitive. Traditional forms of calibration used in transport modelling (e.g., log-likelihood) are not as appropriate here because of the non-linearity in the system.

This can also be undertaken with various samples (train-and-test). A possibility is to use the technique of k-fold cross validation, where k is generally 10 (Kohavi, 1995). The dataset is split into ten parts, and then trained on nine parts and tested on the remaining part. This is repeated ten times.

Patterns

For the two-day sample, we can also match against frequencies of activities on particular days. A possible list of outputs could include:

- number of activities (per person, per location, per day of week)
- frequency/interval of activities (per person, per location, per type)
- distribution of activity group sizes/types of groups
- amount of travel
- social network properties: network density, size of personal networks, clustering, path lengths

An indication of frequencies for particular activities could be useful in place of a detailed activity diary. In this manner, the activity generation for locations types could be validated, however the distances travelled could be modelled incorrectly.

External validation

In order to test the model more thoroughly on previously unseen data, a general activity data set can be used, especially if it contains group size and the type of group (e.g., household, non-household).

CONCLUSION

Transport modellers are building models that are more concerned with modelling individual and dynamic behaviours than previously. This necessitates a shift in the modelling approach to individual- and agent-based models, which has implications for estimation and validation, due to both the complexity of the model and the data available.

This paper has investigated validation methods for agent-based models in the activity-travel context, in particular focussing on models incorporating joint social activities. The principles of validation are similar for traditional, agent-based, and transport-specific simulations, however the vocabulary differs slightly.

We present a proposed approach for the validation of a model of social activities. The process has been developed based on the process validation methodology described by Barlas and in connection with the available data. It covers several different validation techniques which will provide more insight into the usefulness and appropriateness of the model. Readers will not that many elements of the approaches are not very different from validation of simple non-agent-based models. However, as only particular empirical validation is possible the approach also includes some non-conventional elements.

As also noted by Gilbert, it is important to validate agent-based models at different levels, both individual level and aggregate level. However, this may be difficult, especially at the disaggregate level either because no data are available or because the model uses abstract behavioural concepts. In our case, we have data from the real world at an individual level, but we do not have information about how people made a particular decision to meet with a certain group of people or undertake a particular activity, nor do we have information about individuals' long-term plans. Different data collections are necessary to provide adequate data for validation of both the activities and social processes. On the other hand, the fact that agent-based models tend to generate emerging aggregate patterns from individual behavioural principles as opposed to using aggregate input offers validation potential.

In general, the type and amount of validation required depends on the aims of the model, the model setup, the data available, and the level of confidence/validation/accuracy desired, which should be determined before model development begins. Although this paper has not provided concrete solutions to the validation issue, we hope that it will generate further discussion. In the activity-demand context, it may be that a shift in expectations is required, from both end-users and modellers, regarding what can be validated and how. Future work involves testing and refining the process on our model, and providing a set of recommendations/lessons learned for similar models.

BIBLIOGRAPHY

Amblard, F., Bommel, P., Rouchier, J. (2007). Assessment and validation of multi-agent models. In: Phan, D., Amblard, F. (Eds.), Agent-based Modelling and Simulation in the Social and Human Sciences. The Bardwell Press, Ch. 4.

Angrainni, R., T.A. HArentze, H.J.P. Timmermans, T. Feng. (2009). Modeling household activity participation in a rule-based system of travel demand: decision of two household heads. In: Proceedings EATS09, Surabaya, Indonesia.

Arentze, T., Timmermans, H. (Eds.), 2000. ALBATROSS : a learning based transportation oriented simulation system. EIRASS, Eindhoven, Netherlands.

Axhausen, K. (2006). Social networks, mobility biographies and travel: The survey challenges. Tech. Rep. 343, Institut für Verkehrsplanung und Transportsysteme.

Barlas, Y., 1996. Formal aspects of model validity and validation in system dynamics. System Dynamics Review 12 (3), 182 – 210.

Batty, M., P. Steadman, Y. Xie (2006). Visualization in spatial modelling. In J. Portugali (Ed.), Complex artificial environments (pp. 49–70). Berlin: Springer.

Bellemans, T, D. Janssens, G. Wets,T. Arentze, H. Timmermans (2010). Feathers Activity-Based Simulation Platform. Proceedings of the 89th TRB Annual Meeting.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the United States of America 99, pp 7280 – 7287.

Buliung, R. N., Kanaroglou, P. S., March 2007. Activity-travel behaviour research: Conceptual issues, state of the art, and emerging perspectives on behavioural analysis and simulation modelling. Transport Reviews 27 (2), 151 – 187.

Carley, K.M. (1996). Validating Computational Models. Working Paper, Carnegie Mellon University, September 1996.

Carrasco, J.A., B. Hogan, B. Wellman, E. J. Miller (2008). Collecting social network data to study social activity-travel behaviour: An egocentric approach. Environment and Planning B, 35.

Chattoe, E., Saam, N. J., Moehring, M., 2000. Sensitivity analysis in the social sciences: Problems and prospects. In: Suleiman, R., Troitzsch, K. G., Gilbert, N. (Eds.), Tools and Techniques for Social Science Simulation. Physica-Verlag, pp. 244 – 273.

Crooks, A., Castle, C., Batty, M., 2008. Key challenges in agent-based modelling for geo-spatial simulation. Computers, Environment and Urban Systems 32(6), pp. 417-430.

Gilbert, Nigel (2004). Open problems in using agent-based models in industrial and labor dynamics. *Advances in complex systems*, 7(2), 285-288.

Hackney, J., F. Marchal (2009). A Model for Coupling Multi-Agent Social Interactions and Traffic Simulation. Proceedings of the 88th TRB Annual Meeting.

Haklay, M., D. O'Sullivan, M. Thurstain-Goodwin, T. Schelhorn , 2001, "So go downtown": simulating pedestrian movement in town centres. Environment and Planning B: Planning and Design 28(3) 343 – 359.

Hamill, L., Gilbert, N., 2009. Social circles: A simple structure for agent-based social network models. Journal of Artificial Societies and Social Simulation 12 (2), 3. URL http://jasss.soc.surrey.ac.uk/12/2/3.html

Han, Q., T,A, Arentze, H.J.P. Timmermans, D. Janssens and G. Wets (2009), Developing dynamic models of activity-travel behavior: principles, mechanisms, challenges in data collection and methodological issues. Paper presented at the TRB Annual Meeting, Washington, D.C. (Cd-rom)

Kleijnen, J. P. C., 1995. Verification and validation of simulation models. European Journal of Operational Research 82 (1), 145 – 162.

Klügl, F., 2008. A validation methodology for agent-based simulations. In: Proceedings of the 2008 ACM Symposium on Applied Computing. pp. 39 – 43.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. pp. 1137–1143.

Kowald, M., A. Frei, J. Hackney, J. Illenberger, K.W. Axhausen (2009). Using an ascending sampling strategy to survey connected egocentric networks: A field work report on phase one of the survey. Tech. Rep. 582, Institute for Transport Planning and Systems, Swiss Federal Institute of Technology, Zürich.

Kurth, David L., S. Childress, Erik E. Sabina, Thomas F. Rossi (2008). Proposed Validation and Sensitivity Testing of Denver Region Activity-Based Models. Innovations in Travel Demand Modeling: Summary of a Conference. Volume 2: Papers.

Louie, M. A., Carley, K. M., 2008. Balancing the criticisms: Validating multi-agent models of social systems. Simulation Modelling Practice and Theory 16 (2), 242 – 256.

MATSIM (2009). http://www.matsim.org/

McNally, M.G. (2002). The four-step model. Handbook of Transport Modelling (Hensher and Button, eds.), Chapter 3. Pergamon.

Newman, M. E. J., 2003, The structure and function of complex networks, *SIAM Review* 45, 167 – 256.

Ormerod, P., B. Rosewell (2009). Validation and Verification of Agent-Based Models in the Social Sciences. Lecture Notes in Computer Science 5466, p . Springer,Berlin.

Ohnmacht T., K.W. Axhausen (2005). Entwicklung des Forschungsdesign und der Erhebungsinstrumente fu« r das Projekt Mobilitätsbiographien, Mobilitätswerkzeuge und soziale Netze. Tech. Rep. 289, Institute for Transport Planning and Systems, Swiss Federal Institute of Technology, Zürich.

Pendyala, R. M., Kitamura, R., Kikuchi, A., 2004. Famos: The Florida activity mobility simulator. In: Proceedings of Progress in Activity-Based Analysis, Eindhoven University of Technology. Eindhoven.

Pinjari, A., Eluru, N., Srinivasan, S., Guo, J. Y., Copperman, R., Sener, I. N., Bhat, C. R., 2007. Cemdap: Modeling and microsimulation frameworks, software development, and verification. University of Texas, Austin.

Robins, G., T. Snijders, P. Wang, M. Handcock, P. Pattison (2007). Recent developments in exponential random graph (p*) models for social networks. Social Networks 29 (2), 192–215.

Snijders, T.A.B, G.G. van de Bunt, C.E.G. Steglich (2010). Introduction to stochastic actor-based models for network dynamics. Social Networks 32 (1), 44-60.

TRANSIMS (2009). http://tmip.fhwa.dot.gov/community/user_groups/transims.

Troitzsch, K. G., 2004. Validating simulation models. In: Horton, G. (Ed.), Proceedings of the 18th European Simulation Multiconference.

van den Berg, P., Arentze, T., Timmermans, H., 2009. ICT and social interaction: Modeling communication mode choice and its effect on travel behavior. In: Proceedings of the 12th International Conference on Travel Behaviour Research.

Windrum, P., Fagiolo, G., Moneta, A., 2007. Empirical validation of agent-based models: Alternatives and prospects. Journal of Artificial Societies and Social Simulation 10 (2), 8. URL http://jasss.soc.surrey.ac.uk/10/2/8.html