Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 39 (2012) 400 – 408

Procedia
Social and Behavioral Sciences

The Seventh International Conference on City Logistics

Process validation of urban freight and logistics models

Rick Donnelly[a*], Russell G. Thompson[b], Marcus Wigan[c]

[a]*Department of Civil and Environmental Engineering, University of Melbourne, Australia*
[b]*Institute of Transport Studies, Department of Civil Engineering, Monash University, 3800 Australia*
[c]*Centre of Excellence in the Governance and Management of Transport, University of Melbourne, Australia*

Abstract

A number of innovative modelling approaches for the analysis of urban freight demand and its impact upon the built environment and transport infrastructure have been proposed over the past several years. These range from new and more robust synthetic models to tour-based formulations based on truck survey data to agent-based microsimulation models. As impressive as these contributions are, most have only included nominal validation efforts, typically limited to comparing the flow estimates to observed traffic counts. In many cases in both research and practice the quality and quantity of these counts are disappointing, and definitive conclusions about model validity and accuracy are difficult to draw from them. Fortunately, increasing the number of counts is far from the only option open to modellers. A far more expansive practice known as process validation can not only overcome the limitations of count data, but admit a far wider spectrum of information, data, and knowledge to the task. This paper illustrates how the process was applied to a tour-based microsimulation model of urban freight, and offers suggestions how it can be more widely applied to freight and logistics models.

Keywords: Validation; freight modelling; agent-based modelling

1. Introduction

The paucity of behavioural data from which to build robust and comprehensive urban and freight and logistics models has long been lamented (Wigan 1971, Meyberg & Mbwana 2002, Wigan & Southworth 2006, to cite but a few). Most researchers and practitioners have devised novel methods to overcome the

* Corresponding author. Tel.: +1-505-881-5357; fax: +1-505-881-7602.
E-mail address: rick.donnelly@gmail.com

1877-0428 © 2012 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of 7th International Conference on City Logistics
doi:10.1016/j.sbspro.2012.03.117

lack of data, to include the use of synthetic data, derivation of demand matrices from count data, use of theoretical or expert knowledge, and synthesis of composite trip rates and other behavioural data by fusing multiple surveys. The Quick Response Freight Manual (Cambridge Systematics 2007) is an example of a widely used source of such data in the USA. Even when travel survey data are available they are often small in size, expensive to collect, and reveal large variances among otherwise similar shippers and carriers. Faced with scant data about a population exhibiting a high degree of variability most modellers use all available data for model estimation or calibration, leaving little or none for model validation. As a consequence most models that report validation results do so by comparing modelled flows to observed ones. In too many cases even the count data are sparse compared to the magnitude and extent of freight flows within an urban area or region (Turnquist 2008, Donnelly 2008), making comparisons to them suspect.

Even when adequate count data are available the entropy maximization basis of most deterministic urban freight models admits the strong possibility that several different combination of inputs could have given rise to modelled flows that resemble reality (de la Barra 1989, Wilson 2006). Even models not based upon entropy maximization methods, such as agent-based or activity-based travel models, may reveal several plausible outcomes whose derived network flow patterns resemble reality. Donnelly (2007, 2099), using an agent-based microsimulation of urban freight in Portland, Oregon (USA) found that repeated simulations gave rise to substantially similar network flow patterns. Thus, it would appear that almost any model of urban freight is capable of mimicking the flow patterns revealed by comparing a model to observed counts alone.

A more expansive approach to model validation can help overcome both problems. Admitting information other than just counts to the validation process can overcome the problem of two few counts, or an adequate number of them in unhelpful places. A wider spectrum of data can also ensure that the validation process is applied to all parts of the model, not only at the end of the modelling chain. However, the literature in transport modelling in general, and freight modelling in particular, appears largely devoid of innovative or expansive validation techniques. Fortunately, contributions in other realms can be adapted to the practice of freight modelling with good results. One such technique is *process validation*, pioneered in the fields of systems dynamics and software engineering. However, the technique does not appear to have been applied in transport modelling to date. This paper describes its adaptation to urban freight modelling and offers suggestions about it might be more widely applied.

2. Fundamentals of process validation

Process validation is an elegant relativist framework for validating complex models, and is particularly useful in cases where data are scant, assumptions are numerous, and behaviour is modelled piecemeal. It is based upon the premise that not only is the reproduction of behaviour necessary in order to consider a model valid, but that how a model gives rise to such behaviour is equally important. Barlas (1996) argues that, "what is crucial is the validity of the internal structure of the model." Process validation stands as a viable alternative to subjective or neglected assessment of model performance in cases where a statistical approach is not appropriate. Whilst others have written about process validation (Cook & Wolf 1999, Janssen 1995) they are not as well-structured and comprehensive as the approach laid out by Barlas.

Process validation is a qualitative process, although it can easily admit quantitative measures. That does not distract from the rigour of the approach, which is as systematic and comprehensive as formal statistical testing. The framework includes two groups of structure tests and two groups of behaviour tests, as shown in Figure 1. The first two groups examine the internal validity of the model, using a combination of direct and indirect tests. Once confidence is gained in the structural test outcomes the second group of tests focus on emergent behaviour.

The overall process is illustrated in Fig. 1, and examined within the context of freight modelling in the following sections.

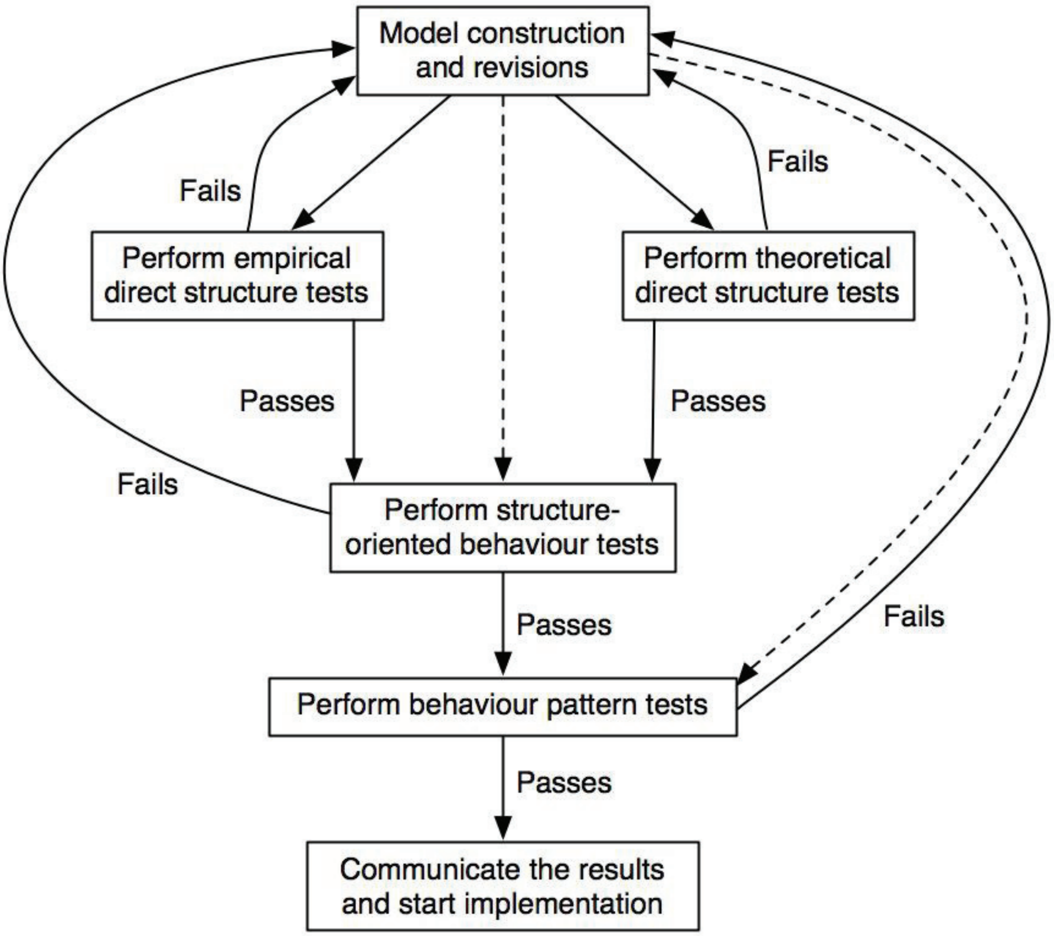

Fig. 1. Steps in process validation (Source: Barlas, 1996)

As illustrated in the Figure, the process is iterative, running throughout the life of a modelling project. An important aspect of the process is that internal validity (structure tests) must be proven before behavioural testing, a cycle repeated many times in the construction of most models. Process validation ideally begins early during model specification, long before code is written or analyses carried out, and influences the conceptual form of the model. Conducting the tests along the way will likely uncover errors that would be more difficult to detect and correct later in the development process. Moreover, waiting until the end of the process and focussing solely on behaviour pattern tests would likely result in incorrect diagnoses, which in turn would have invited "fixes" to the model that would not have addressed the root problem.

Process validation is an ideal framework for agent-based models, which almost by definition are not amenable to the standard statistical tests used to validate transport models. Process validation, as illustrated here, has the ability to impose far more rigour into freight modelling than is currently expected. The results can include more transparent and trustworthy models, fewer defects, and greater confidence in the outcomes they suggest.

2.1. Comparison to purely statistical approaches

Process validation does not seek to supplant the statistical testing of model performance. Statistical methods have been widely used in modelling and simulation across all realms (Kleijnen 1998, van Dijkum & van Kuijk 1999), and remain the dominant approach used in transport modelling. Relying solely upon statistical measures in freight modelling is untenable for several reasons. Many of the data used to characterize urban freight patterns and choices violate two of the primary assumptions required to use classical statistical techniques (Kennedy 1993, Sheskin 2003):

- The data are free from cross-correlation or collinearity.
- The data are normally distributed.

Many of the relationships in freight models, from trip length frequency distributions, mode choice propensities by distance, distributions of shipment and vehicle weights, intermodal cost functions, and other artefacts are not normally distributed. Most follow a generalised beta or similar distribution. Means can be drawn from such distributions for use in deterministic models with arguably little resulting bias, but not when sampling from such distributions in agent-based or microsimulation models. In such cases many of the data are by definition auto-correlated and cross-correlated, from which there is no recourse without resorting to an entirely different modelling approach.

The consequence of violating the assumption regarding normal distribution of the data is that inferential statistical tests cannot be carried out (Casella & Berger 2001, Sheskin 2003). Such tests allow the researcher to assess the likelihood that that the observed and modelled data belong to the same population. However, the loss of inferential properties does not necessarily render statistical summaries of the data worthless. They most certainly have a place in the overall validation framework, to include in process validation. The important point to be made here is that they cannot solely form the basis for validating models of emergent behaviour, which is why process validation admits them later in the process.

2.2. Empirical direct structure tests

This first group of tests compares the structure of the model to knowledge gained through direct observation or measurement of the system under study. This information can come from a variety of sources, such as travel surveys, expert panels, syntheses such as the aforementioned Quick Response Freight Manual, and models imported from elsewhere. Irrespective of their source their analytical derivations must be judged to be correct in mathematical and algorithmic terms. Such evaluations should be carried out for each model component (e.g., trip generation, mode choice) as well as for the entire model. The tests performed included:

- *Parameter confirmation* involves testing each of the estimated parameters of each model component. As Barlas points out, this is done, “both conceptually and numerically.” The former requires that the form of the model must symbolically represent processes and relationships in the real world, while the latter ensures that the parameter estimates are robust. Because most models are implemented in computer code this part of the testing can also involve verification that the software implements the processes as intended.

- *Extreme condition* testing is used to assess the effect of changing the parameter values over the range of plausible values. For example, varying parameter values of the gravity model across a wide range can test their impact on trip distribution. The results should emerge as expected (e.g., average trip distance increases without discontinuities as impedance is increased). This test is of reduced value when testing stochastic model elements that sample from observed distributions, as there were no mathematical representations to evaluate.

2.3. Theoretical direct structure tests

Taken together, the validated empirical structures mimic a system in the real world. The theoretical structure tests compare the overall model structure with knowledge of how such systems actually work. This knowledge is contained in the literature, as well as in the minds of domain experts. However, there is no general theory of urban freight transport, although the works of Nagurney et al (2002) and Xu et al (2003) come closest to putting it into a holistic context. Both posit the relationship between financial, supply chain, and transport networks, with information flowing between them. These linkages extend well beyond a single urban area, with the supply chain being global in some instances. Drewes Nielsen et al (2003) outline a similar approach, albeit not from a modelling standpoint. They also propose a multi-level frame- work in which changes in production and consumption influence logistic patterns, which in turn are revealed in transport measures and indicators and environmental impacts.

Many freight models represent these major factors and institutions affecting the demand for freight and their collective impact upon the urban environment. They do so by postulating that all freight activity is generated by the trade between firms and households at the level indicated by urban or regional economic accounts. Monetary flows are translated into commodity flows using transform functions, and carried to their destination using carrier optimisation techniques widely reported in the literature. While the internal structure of supply chains are not often considered in such models, their interface with the transport system as distribution centres and points of production and consumption are entirely consistent with general models of supply chains.

Absent a generally accepted theoretical model of urban freight one can turn to domain experts for theoretical structure validation. This commonly takes the form of peer review or expert panel consultation, where the members are acknowledged industry leaders and respected academics. The external examination of academic research is another means of obtaining such feedback. The results of this test are qualitative, in that reviewers can affirm or reject the proposed framework in varying degrees.

2.4. Structure-oriented behaviour tests

The second category of tests proposed by Barlas deals with the behaviour generated by the model, which entails examining model outputs rather than individual components. In agent-based modelling the focus of such tests are upon the emergent outcomes, whereas in deterministic models average outcomes are expected. In this phase of validation the emphasis remains upon uncovering structural flaws in the model. *Extreme condition testing* is advanced as the most important test in this category. It involves assigning extreme values to model parameters and assessing the resulting model response. Balci (1994) calls this stress testing of the model, designed to isolate discontinuities in the model response surface outside the range of the values used to estimate the model. Setting the bounds (extremes) of any given parameter requires assertion of the lowest and highest plausible values the parameter would be expected to take on in the real world. This test is equivalent to the sensitivity testing described earlier, and its findings are directly relevant here. When used within process validation it is useful to include a quantifiable measure of success, such as the percentage of tests that resulted in the expected outcome.

Other tests are relevant in this category of testing. A *phase relationship test* can be used to measure the relationship between pairs of variables in the model. It is generally thought that such testing is possible in a simulation with only a few key variables, but perhaps impractical in large-scale models (Hamby 1994, Fraedrich & Goldberg 2000). It is probably even more difficult to measure in simulations and agent-based models, where closed form solutions are not possible and where the outcomes (both individual and collective) vary between runs. Unfortunately, except when using stated preference experiments such real world phase relationships are difficult to isolate from traditional travel surveys.

Rykiel (1996) suggests Turing tests for structure-oriented behaviour testing. Experts are asked if they can discriminate between simulated and real system outputs. A demonstrated ability to do so may reveal structural deficiencies in the model.

2.5. Behaviour pattern tests

Behaviour pattern tests are the next to last step in process validation, completed only after confidence is gained in the results of previous tests. In contrast to previous steps these tests are quantitative, although they need not be exhaustive. The goal is to compare the major behaviour patterns generated by the model to those in the real world. Barlas notes that, "it is critical to note that the emphasis is on pattern prediction… rather than point estimates." Moreover, those patterns must be revealed by data other than those used to develop the model. This step is equivalent to the traditional practice of validation in transport modelling, which typically ignores the previous steps.

A model that passes all four groups of testing is considered valid for the purpose for which it was constructed. Because it validates against so many different types of tests that exercise the model at different levels it is a far more powerful assessment than traditional statistical testing alone. Indeed, in the absence of adequate data for statistical analyses a model that satisfies the other requirements noted above can be found valid and quite useful for its intended purposes.

3. Application to an agent-based urban model

Donnelly (2007, 2009) included process validation in a hybrid microsimulation model of urban freight transport demand, which serves to illustrate its broad potential for this type of modelling. Lacking a holistic theory or data about the formation of truck tours in urban areas, an agent-based simulation was developed by fusing many disparate sources of data that each provided only a glimpse of overall freight system in Portland, Oregon (USA). The chosen approach started with the translation of urban economic activity into freight transport flows, and their allocation to individual firms within an urban area. Particular emphasis was placed on the development of behaviourally rich representations of truck tours and routing decisions, as well as their use of distribution centres and warehousing.

This model could not be constructed from a single holistic source of data, as such does not exist. Instead, heterogeneous and otherwise incompatible data from several sources were fused to create emergent behaviour of urban freight demand. Each piece of the data revealed a different part of the overall puzzle. The resulting model was a hybrid formulation that included statistical sampling, rule-based decision-making, and spatial allocations.

The resulting model was tested using data from Portland, Oregon (USA). Process validation was employed throughout model development, testing, and application. Each of the tests described above were employed in a sequential manner. When a model component failed a certain test the defects were analysed and corrected, and the process was restarted from the beginning. It is interesting to note that the majority of errors in specification and coding were found in the first stage (empirical direct structure tests) rather than in the quantitative behaviour pattern tests later on. The overall outcomes are shown in Table 1.

Several aspects of the behaviour pattern tests deserve mention. The Portland model used a destination choice model only loosely condition by impedance. Instead, economic input-output relationships were used to link producers and consumers of each commodity, where relative sizes mattered more than distance. That is, firms producing large quantities of certain goods were more likely to be matched up with firms consuming large quantities of them rather than one-to-many relationships. Trip length frequencies by commodity emerged from the simulation but did not constrain it. This model form was suggested by expert reviewers, and likely would not have been considered without their input during the theoretical direct structure tests. Observed trip length distributions from local truck surveys were compared to the model. Not surprisingly the model did not do as well replicating such patterns for light goods vehicles, for which information were largely lacking in model development and about which the local experts were least familiar. However, the model did surprisingly well for larger truck types despite their relatively smaller incidence for local shipments.

The last test used in model validation – comparison of observed to modelled network flows – is often the only one used in most transport models. In this case it was the final test of model performance, assessed only after a high degree of confidence was gained in earlier tests. The model performed acceptably in this regard, gaining a higher degree of correlation between the two than possible using the traditional trip-based truck model this model replaced.

Table 1. Process validation tests and outcomes

Category	Test	Outcome
Empirical direct structure tests	Parameter confirmation	*Passed:* Embedded parameters tested for formal mathematical functions. Statistical distributions checked for correspondence to observed data or literature
	Extreme condition	*Passed:* Each parameter individually checked over wide range of values to ensure that resulting values fall within expected range.
	Model alignment	*Passed:* Each model component prototyped in R statistical language or spreadsheet model before implementation in Python, with results of each compared to one another.
Theoretical direct structure tests	Structure confirmation test	*Passed:* Comparisons to other holistic frameworks proposed in the literature confirms multi-level structure of current model.
	External examination	*Passed:* Examination of model by acknowledged experts in the domain to determine feasibility; thesis examination is typically most rigourous of external examinations.
Structure-oriented behaviour testing	Extreme condition (stress) testing	*Passed:* Systemic effect of variations in key parameter values checked as part of sensitivity testing of overall model.
	Turing test	*Indeterminate:* Attempts to have experts distinguish between real and modelled outcomes could not be completed because the latter was obvious to them.
Behaviour pattern tests	Pattern prediction tests	*Passed:* High degree of correlation between observed and modelled testing trip length frequency distributions and individual link flows.
	Overall summary statistics	*Passed:* High degree of correlation between observed and modelled system-wide network volumes (vehicle kilometres of travel by roadway category, etc.).

4. Conclusion

Process validation, as illustrated here, has the ability to impose far more rigour into freight modelling than traditional statistical approaches to validation, especially those limited to comparison of observed to modelled link flows. The more expansive process validation approach results in more transparent and trustworthy models, fewer defects, and greater confidence in the outcomes they suggest. Process validation is particularly attractive in that it does not prescribe the exact tests to be performed or their interpretation, allowing the framework to be adapted to a wide range of freight models. Like the data used in urban freight models, the process admits information from a wide variety of sources, each providing only a glimpse of the total picture. The sum is truly greater than the sum of its parts.

References

[1] Balci O. Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research* 1994; 53: 121–173.
[2] Barlas Y. Formal aspects of model validity and validation in systems dynamics. *System Dynamics Review* 1996; 12: 183–210.
[3] Cambridge Systematics. *Quick response freight manual II*. Report FHWA-HOP-08-010, Federal Highway Administration, Washington, D.C.; 2007.
[4] Casella G, Berger RL. *Statistical inference*, 2nd edition. Duxbury Press, Duxbury, MA; 2001.
[5] Cook JE, Wolf AL. Software process validation: quantitatively measuring the correspondence of a process to a model. *ACM Transactions on Software Engineering and Methodology* 1999; 8: 147–176.
[6] De la Barra T. *Integrated land use and transport modelling: Decision chains and hierarchies*. Cambridge University Press, Cambridge; 1989.
[7] Donnelly R. Evaluation of practice today. In: Hancock K, editor. *Freight demand modelling: Tools for public decision-making*, Transportation Research Board, Washington, D.C; 2008, p. 27-35.
[8] Donnelly R. A hybrid microsimulation model of freight flows. In: Taniguchi E, Thompson RG, editors. *City logistics V* , Institute of City Logistics, Kyoto; 2007, p. 235-246.
[9] Donnelly R. *A hybrid microsimulation model of urban freight transport demand*. Unpublished PhD thesis, University of Melbourne; 2009.
[10] Drewes Nielsen L, Jespersen PH, Petersen T, Hansen LG. Freight transport growth—a theoretical and methodological framework. *European Journal of Operational Research* 2003; 144: 295–305.
[11] Fraedrich D, Goldberg A. A methodological framework for the validation of predictive simulations. *European Journal of Operational Research* 2000; 124: 55–62.
[12] Hamby DM. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment* 1994; 32: 135–154.
[13] Janssen PHM, Heuberger PSC. Calibration of process-oriented models. *Ecological Modelling* 1995; 83: 55–66.
[14] Kennedy P. *A guide to econometrics*, 3rd edition. MIT Press, Cambridge, MA; 1993.
[15] Kleijnen JPC. Experimental design for sensitivity analysis, optimization, and validation of simulation models. In: Banks J, editor. *Handbook of simulation: Principles, methodology, advances, application, and practice*, John Wiley & Sons, New York; 1998, p. 173-223.
[16] Meyburg A, Mbwana J. Data needs in the changing world of logistics and freight transportation: conference synthesis. New York State Department of Transportation, Albany, NY; 2002.
[17] Nagurney A, Ke K, Cruz J, Hancock K, Southworth F. Dynamics of supply chains: A multilevel (logistical – informational – financial) network perspective. *Environment and Planning B* 2002; 29: 795–818.
[18] Rykiel EJ. Testing ecological models: the meaning of validation. *Ecological Modelling* 1996; 90: 229–244.
[19] Sheskin DJ. *Handbook of parametric and nonparametric statistical procedures*. Third edition, Chapman & Hall/CRC Press, Boca Raton, FL; 2003.
[20] Turnquist M. Characteristics of effective freight models. In: Hancock K, editor. Freight demand modelling: Tools for public decision-making , Transportation Research Board, Washington, D.C; 2008, p. 11-16.
[21] Van Dijkum C, Van Kuijk E. Validation of simulation models in the social sciences. In: Dijkum C Van, Tombe DD, Van Kuijk E, editors. *Validation of simulation models*, SISWO Publication 403, Netherlands Universities Institute for Coordination of Research in Social Sciences, Amersterdam; 1999, p. 7-29.

[22] Wigan MR. *Benefit assessment for network traffic models and application to road pricing*. Report LR 417, Road Research Laboratory, Crowthorne, Berkshire (UK); 1971.
[23] Wigan MR, Southworth F. What's wrong with freight models, and what should we do about it? *Transportation Research Board Annual Meeting* 2006 paper 06-1757.
[24] Wilson AG. Ecological and urban systems models: Some explorations of similarities in the context of complexity theory. *Environment and Planning A* 2006; 38: 633–646.
[25] Xu J, Hancock K, Southworth F. Simulation of regional freight movement with trade and transportation multinetworks. *Transportation Research Record* 2003; 1854: 152–161.