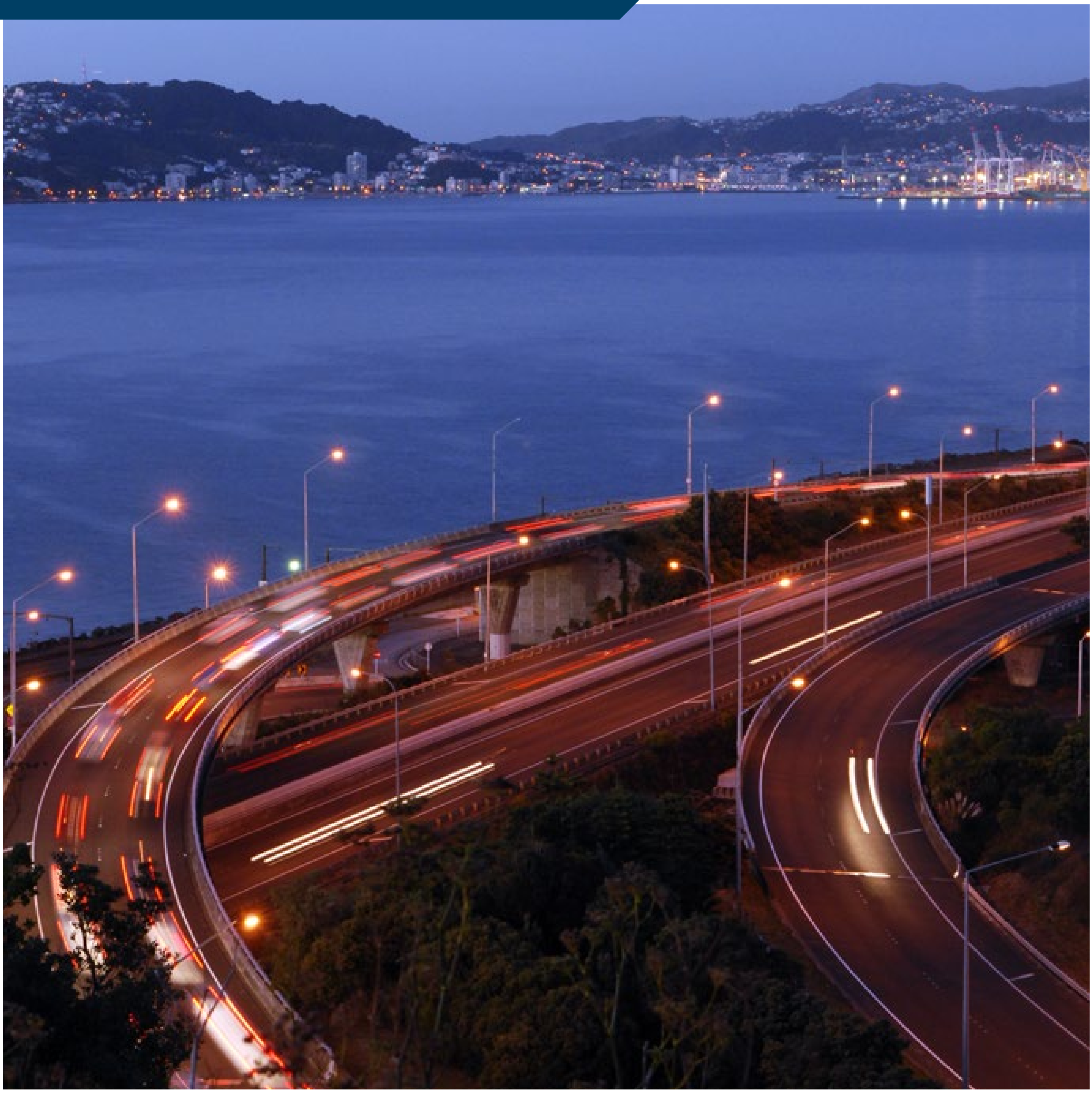
TRANSPORT MODEL DEVELOPMENT GUIDELINES

The key purpose of this document is to provide guidance for the comparisons carried out between observed and modelled outputs, commonly during the base model development phase of a project.



WAKA KOTAHI
NZ TRANSPORT AGENCY

New Zealand Government

Waka Kotahi NZ Transport Agency
This edition published September 2019

ISBN 978-0-478-41943-6 (online)
ISBN 978-0-478-41944-3 (print)

If you have further queries, call our contact centre on 0800 699 000 or write to us:

Waka Kotahi NZ Transport Agency
Private Bag 6995
Wellington 6141.

This publication is also available on the Transport Agency's website at **www.nzta.govt.nz**

CONTENTS

DOCUMENT MANAGEMENT PLAN

1. Purpose

This management plan outlines the updating procedures and contact points for the document.

2. Document information

DOCUMENT NAME	Transport model development guidelines
DOCUMENT NUMBER	O1
DOCUMENT AVAILABILITY	This document is located in electronic form on the Transport Agency's website at www.nzta.govt.nz
DOCUMENT OWNER	Tony Brennand
DOCUMENT SPONSOR	Pete Simpson

3. Amendments and review strategy

All corrective action/improvement requests (CAIRs) suggesting changes will be acknowledged by the document owner.

ACTIVITY	COMMENTS	FREQUENCY
Amendments (minor revisions)	Updates incorporated immediately they occur.	As required
Review (major revisions)	Amendments fundamentally changing the content or structure of the document will be incorporated as soon as practicable. They may require coordinating with the review team timetable.	As least every 2 years
Notification	All users that have registered their interest by email to xxx @NZ Transport Agency.govt.nz will be advised by email of amendments and updates.	Immediately

4. Other information (at document owner's discretion)

There will be occasions, depending on the subject matter, when amendments will need to be worked through by the review team before the amendment is actioned. This may cause some variations to the above noted timeframes.

5. Distribution of this management plan

Copies of this manual management plan are to be included in the NZ Transport Agency intranet at the next opportunity and sent to:
xxx

RECORD OF AMENDMENT

AMENDMENT NUMBER	DESCRIPTION OF CHANGE	EFFECTIVE DATE	UPDATED BY
1	Typographical changes to improve readability	23 September 2019	Tony Brennand

SECTION 1: INTRODUCTION

1.1 GUIDELINE PURPOSE AND OBJECTIVE

Transport and traffic models are developed for a range of purposes and are applied to a variety of assessments. For example regional land-use and infrastructure planning, forecasting transport demands, evaluating the impact of new infrastructure, through to design and analysis of intersections. Models commonly play a key role in the design and/or evaluation of large state highway projects, corridor plans or strategies.

As described in section 1.2 below, a key step in the development of a model is comparison with observed data – these are calibration and validation steps.

The key purpose of this document is to provide guidance ranges for the comparisons carried out between observed and modelled data, commonly during the base model development phase of a project. These comparisons are generally undertaken as part of the wider requirement to ensure that the model is suitable for its intended application. A core aspect of this guide is the target criteria levels defined by transport assessment purpose category contained in the various tables within sections 5, 6, 7, and 8.

The text and additional sections provide important context to these comparisons and the processes involved in a transport project featuring the application of a model. Also included is discussion and guidance on important aspects that need to be considered through a project:

- Key outcomes of model scoping.
- Involvement of key personnel.
- Considerations relating to the specification of models.
- Considerations which are key to specific forms of projects.

Application of this guideline should ensure modelling work associated with transport projects in NZ deliver products which represent quality and value for money.

This guide is not intended as box ticking exercise. Meeting calibration and validation criteria does not automatically mean that a model is robust and/or fit-for-purpose. Meeting criteria within a competent whole model development and application approach improves confidence in the model outcomes and reduces the risk to decision makers.

This document is intended to sit alongside other documents which use model outputs for specialised purposes.

1.2 MODEL DATA COMPARISONS

The lifecycle of a transport model will generally require the development of a base, or base-year, model representing the existing on-street transport environment. Commonly completed as the first stage of the modelling task, or as an update of a legacy model, this model forms a foundation upon which analysis will be carried out.

Development and verification of the base model requires comparisons of modelled information with observed travel data (eg records of trip making activities, counts, journey times). These processes are used to calibrate and validate the model – key steps in gaining confidence that the base model is an appropriate foundation for its intended application.

Comparisons with observed data are important for most transport models, with the possible exception of isolated intersection modelling at lower volumes, and are used to establish and check:

- that the description of the model transport environment (network) is appropriate
- that the parameters used in the development of the model are appropriate for the study area
- that the base demand information (eg trip generation, distribution, matrices, profiles) offers a robust representation of transport movements through the study area
- and that the combination of the above offers a robust representation of the operation of the study area.

1.3 PROJECT STAGES AND DOCUMENT STRUCTURE

A core aspect of this guide is the target criteria levels defined by purpose category contained in table 1 through to table 9 within sections 5, 6, 7, and 8. Figure 1 provides a high level overview of the typical steps involved in a transport modelling project and how the sections of this guideline fit within this process.

The steps shown in the lighter boxes are not covered in detail within this guideline, but background guidance is provided in several key areas. These include important scoping decisions made in liaison with the client and peer reviewer which affect the application of this guideline through the project, and 'hands-on'/'how to' modelling tasks.

Along with the sections outlined in the figure below, the introduction outlines existing guidance documents, the intended application of the guide and the source of the target levels contained within this document, section 2 outlines the roles of key personnel, and appendix A contains important definitions.

FIGURE 1: APPLICATION OF THIS GUIDELINE TO THE STAGES OF A TRANSPORT PROJECT

1.4 RELEVANCE AND EEM REQUIREMENTS

Historically the *Economic evaluation manual* (EEM) established requirements for transportation models. The target levels and guidance within this document cover the range of common model applications, including areas currently most relevant to the existing EEM. The historical EEM requirements have been used as a reference source for this document and model applications which are currently less applicable to the EEM requirements will be subject to more exacting requirements. This means that the meeting of the requirements of this document will in itself ensure that the requirements of the EEM are also met..

1.5 OVERSEAS MODEL REQUIREMENTS

The UK's *Design manual for roads and bridges* (DMRB)[2], now moving to webTAG[3], is referenced occasionally in NZ and includes relatively comprehensive information on transport modelling (production of trip matrices, etc), including traffic count and journey time calibration criteria which have stood up to scrutiny in the UK for project traffic models. Similar to the EEM, the DMRB calibration criteria are somewhat specific – relating to trunk road (NZ state highways) scheme assessments. Of note, some of the US's Federal Highway Administration (FHWA) and Austroads modelling guidance uses the DMRB calibration criteria as a source.

1.6 SOURCE OF GUIDANCE

This document has been developed in collaboration between the NZ Transport Agency and the NZ (transport) Modelling User Group (NZMUGS), a sub-group of the Engineering NZ Transportation Group. NZMUGS is a modelling industry interest group which has a specific objective of the development of guidance for the NZ industry. This guideline incorporates work completed by the NZMUGS group, reviewed by the Engineering NZ Transportation Group.

This guideline sets out comprehensive criteria, in the form of target levels, for each set of observed and modelled comparison carried out during the development of base models in NZ. The range of target levels in the criteria are based on model purpose categories defined in section 3 below.

The form of the traffic count and journey time comparisons has been initially generated from the existing EEM and DMRB guidelines. The combination of these two sources has been deemed by the authors as offering a comprehensive set of comparisons. The target levels for these comparisons have been set according to the experience of the NZMUGS members in developing and applying a wide variety of transport and traffic models.

The public transport comparisons and target levels have been based on the current Industry understanding associated with the development of regional models in three main centres, Auckland, Wellington and Christchurch.

The document also provides guidance on presenting before and after demand matrix adjustment comparisons. This guidance has been generated from industry practice and the experience of the NZMUGS members and key Transport Agency staff in developing and applying models.

1.7 APPLICATION OF GUIDANCE

These guidelines have been developed and set to be applied with some flexibility. It is anticipated that the comparisons undertaken and the associated target levels would be identified during the model scoping phase in discussion with the client and/or peer reviewer (refer to section 2.2). The model purpose categories provide some context and guidance to the target ranges presented in the comparison criteria tables and are designed to be interpreted with a degree of flexibility.

It is anticipated that those involved in determining which purpose or purposes are most appropriate for the model, which criteria to check the base

1 NZ Transport Agency *Economic evaluation manual*, Worksheet 8 Transport modelling checks
2 *Design manual for roads and bridges*, Volume 12 Section 2 Chapter 4 Para 4.4.42 Validation Acceptability Guidelines
3 www.dft.gov.uk/webtag/

model against, and the resulting appropriateness of the calibration/validation levels would apply their professional judgement in making these assessments. The criteria presented in this document are not intended or anticipated to be a simple pass/fail checklist. Models which meet all criteria may not be fit-for-purpose, and conversely models which do not meet all criteria may be considered fit-for-purpose (see section 5.8 for further information).

A reasonably comprehensive set of comparisons is provided in this guideline and again some flexibility is anticipated in selecting appropriate calibration and validation comparisons. For example, for a comprehensive urban strategic network, a full set of observed and modelled screenline total and individual directional link flow comparisons would commonly be anticipated (including GEH, percentage, and absolute difference comparisons). However comparisons with turning movements may not be necessary. In contrast, a corridor model is likely to focus on turning movement comparisons and may not include screenline comparisons.

In many cases (work in regional centres, smaller geographic models, straightforward project work etc.) it would not be expected to produce all of the calibration and validation comparisons described in Tables 1 through 9 below. The key comparisons are typically count GEH checks, XY scatter plot checks, and travel time comparisons. The additional comparisons (e.g. absolute and percentage count differences) are typically provided for higher profile projects, in the development of legacy models (base model within intended lifespan of greater than two or three years), and to provide further information when the key comparisons do not provide a suitable level of information on the level of calibration/validation (eg results are poor, volumes are atypical).

The model purpose and focus should be considered when selecting the data used for, and the resulting comparisons presented as, calibration and validation.

For example, a model with a focus on forecasting traffic flows is likely to include several traffic count validation comparisons whereas a model with a focus on representing existing turning movements for design and operational investigations may only include count calibration comparisons. For further information, see section 3.3 and the definitions in appendix A.

The comparisons outlined in this document focus on the stage of model development where travel demand is assigned to the transport network. They do not include calibration/validation comparisons conducted in the earlier stages of regional modelling (eg between trip making and distribution relationships and household interview data). The definitions of calibration and validation given in appendix A may differ from historical definitions relating to the comparisons completed in these earlier stages of regional model development.

The purpose categories and associated ranges of the target criteria levels are appropriate to all widely used transport modelling techniques and model forms (ie macroscopic, static, mesoscopic, microsimulation etc).

This document does not provide 'how to' or extensive good practice guidance on scoping, building, and applying transport models. Some background information and important context is provided in appendices.

The Transport Agency intends that this will remain a live document, and therefore periodic updates will be issued to address issues and queries as they arise. A system for the submission of issues and queries and updates of the document will be developed in collaboration between NZMUGs and the Transport Agency. It is acknowledged that as with any new guidance, this document is a work in progress and changes may be expected as the criteria continue to be tested in practice.

1.8 NZ TRANSPORT AGENCY PROJECT CLASS DEFINITION

This guideline makes specific reference and has particular relevance to 'large' Transport Agency projects. For the purposes of this document this will mean a project or activity that has an expected cost of $5 million (capital) or more developed by the Transport Agency and its justification will be dependent on the model in question. Further refinement of 'large' projects into three classes is given in figure 2 below.

FIGURE 2: NZ TRANSPORT AGENCY PROJECT CLASS DEFINITION

CLASS OF PROJECT OR ACTIVITY	1	2	3
Value in $ million	5-50	50-200	>200

SECTION 2: PERSONNEL & PROJECT INVOLVEMENT

4 This section refers to persons and project work associated the Transport Agency. Similar comments could relate to persons within territorial local or regional authorities.

2.1 KEY PERSONNEL

A number of people are likely to have involvement through a transportation project which includes the development and/or application of a model. The key personnel referred to within this guideline include:

- Client: The client/authority who receives the outputs of a project.
- Lead analyst: Person(s) undertaking the project work for the client. This refers to persons involved in the development and application of models, and to project managers responsible for coordinating this.
- Peer reviewer: Person(s) reviewing the work of the analyst on behalf of the client.

2.2 PERSONNEL COMPETENCY

The development and application of transport models are not simply exercises in following processes and procedures. It is dependent on the skills and experience of the analyst and peer reviewers. It is also important that the client takes ownership of the model's fitness and the outcomes of its application.

Transport Agency specialist

Generally speaking, project managers are not suitably skilled or experienced in transportation modelling. Transport Services will assist in appointing a suitably qualified person to provide support for state highway projects. Other authorities may wish to have the assistance of the Transport Agency's modelling capability in this way, referred to as the Transport Agency specialist.

Transport Agency project class competency levels

The section below outlines competency level suggestions in relation to application of models (testing and assessment) to NZ Transport Agency projects. The development or re-development of base year models for more generalised purposes may consider differing levels of competency, particularly when the form of model development is unique and/or not carried out often. This is at the discretion of the Client.

Peer reviewers need to be involved from the start of the project as does an appropriately qualified and experienced Transport Agency person.

For all transportation models the analyst, peer reviewer and the Transport Agency modelling specialist must have relevant modelling experience. As an illustration, appropriate experience could be judged as a minimum level of 10-15 comparable project applications.

All three parties must agree that the model is generally 'fit for purpose' in terms of the guidance outlined in this document.

'Relevant modelling experience', however judged, is a minimum requirement and the qualifications of the analyst and peer reviewer for higher classes of projects and activities will be more stringent.

For the various classes of large project or activities the expected level of experience of the analyst and peer reviewer is outlined in the table below:

FIGURE 3: NZ TRANSPORT AGENCY PROJECT CLASS AND COMPETENCY LEVELS

CLASS OF PROJECT OR ACTIVITY	1	2	3
Requirement for the analyst and peer reviewer*	Both the analyst and peer reviewer should each have a minimum relevant modelling experience, for example 10-15 or more comparable project applications.	Either the analyst and/ or peer reviewer robust foundation of relevant experience across a breadth of model forms and the other should have a minimum of relevant modelling experience, eg 10-15 or more comparable projects applications.	Both the analyst and peer reviewer should be a robust foundation of relevant experience across the full breadth of model forms (regional-to-operational) and experience in at least two comparably sized projects.

* These requirements, particularly for large projects, are unlikely to be met by individuals (particularly breadth of experience across a range of model forms for comparable projects). Consideration may be given to achieving competency levels through analyst/peer reviewer 'teams', particularly where the team has well-established and proven record.

2.3 PERSONNEL RELATIONSHIPS

It is intended that the analyst, peer reviewer and the client should work as a collaborative partnership. This requires the establishment of an ethos that drives behaviours that seek to deliver what is 'best for the project'. It is essential that this partnership be established at the beginning of the modelling project.

A key outcome of this partnership is not only the development and application of a robust model but also the identification of what aspects of the modelling process and model design represent a risk to the client's business. This requires this partnership to focus on the risks to the client's business, their relative importance and the management of those risks rather than debates on esoteric modelling philosophy.

Another important feature of this relationship is that the client will be active in this partnership rather than 'hands off'. It is important that client personnel take responsibility for its business both in ensuring we get a best for the project outcome and managing the risk to its business.

2.4 PROJECT INCEPTION AND MODEL PURPOSE

At the inception of a project, it is anticipated that the Client, Peer Reviewer, and Analyst would identify and agree the key intended model purpose/application. This is discussed further in section 3.1 below.

2.5 MODEL OWNERSHIP AND INTELLECTUAL PROPERTY

Models developed and paid for, either in totality or part, by the Transport Agency are the intellectual property (IP) of the Transport Agency and are to be available for use by all agents working on behalf of the Transport Agency. The Transport Agency will take all reasonable steps to protect matters of methodology, techniques and skill invested by its suppliers in building or reviewing a model but ultimately the model must be available for the Transport Agency use as the Transport Agency sees fit.

SECTION 3: PURPOSE CATEGORIES

This section broadly defines seven categories which are based on the intended purpose for which the model would be applied and geographic coverage – two elements which are generally interrelated. The categories are not specific to any particular modelling software or technique. These categories and definitions should not be considered absolute and some crossover may exist for certain study areas/projects. These categories have been defined to create a suitable range for the target levels of comparison criteria.

The reason for separating calibration and validation targets into categories is that it is not feasible to develop a single model which covers a wide range of transport assessment purposes (model applications), particularly as geographic coverage and traffic volume levels increase. Therefore models typically need to be designed and developed to achieve a calibration level which is suitable to their intended purpose.

3.1 IDENTIFYING MODEL PURPOSE

When initiating the development of a new model, the adaptation or upgrade of an existing model, or the application of an existing model to a new project, it is important that an explicit statement of the purpose is made. This must identify clearly what the model is designed to do and what it is to be used for. In particular, this requires a careful definition of the problem statement to be addressed and a clear understanding of what questions the model is designed to answer. This will also require the level of confidence that the model is required to deliver to be defined.

It is expected higher levels of confidence will be demanded where the applications of the model are envisaged to support activities involving larger levels of expenditure. In practice larger levels of expenditure are likely to relate to larger schemes and as such larger and more complex models. This would appear to produce a contradiction; larger more complex models are likely to have correspondingly lower levels of calibration and validation (as per the scaled calibration/validation targets based on the categories defined below) unless significant investment is placed in the model development. A level of pragmatism is anticipated in relation to project expenditure and model confidence, eg a balance of investiture in the model development to achieve a level of confidence appropriate to the risk and expenditure of the project.

The reason for separating calibration and validation targets into categories, and for making a clear statement around the purpose of the model application as outlined above, is that it is generally not feasible to develop a single model which covers a wide range of transport assessment purposes (model applications), particularly as geographic coverage and traffic volume levels increase. Due to this, models typically need to be designed and developed to achieve a validation/calibration level which is suitable to their intended purpose.

Applications of models, particularly subsequent to an initial base model development or update, also require consideration of purpose and calibration/validation level appropriateness. This ranges from being as straightforward as simply considering/confirming that the application fits within stated purposes and achieved levels of the base model, through to considering the need for updating the calibration/validation levels or model for the intended application.

3.2 LIST OF CATEGORIES

PURPOSE TYPE A: Regional transportation assessments.

Regional transportation assessment purposes require representation of land-use activities, demographics etc. Purposes include the assessment of strategic impacts of land-use changes, larger scale transport and PT projects, and the effects of policy changes on wider regions. These models are typically 3, 4 or more stage or activity based model.

PURPOSE TYPE B: Strategic network assessments.

A strategic network assessment is likely to be focused on strategic links such as motorway corridors, the state highway, and/or the arterial route network across a wider geographic area. Assessments include major transport infrastructure changes, eg large-scale motorway schemes, bridges. These models are commonly 'traffic assignment' models.

PURPOSE TYPE C: Urban area assessments.

Urban area assessments focus on the operation of urban conurbations, city centres, and other urban style environments. This is a potentially wide range of applications which may include local authority planning, development strategy, urban traffic management and road schemes, infrastructure and policy change assessments, ITS etc. These models are typically of varying form.

MODEL TYPE D: Transport Agency scheme assessment / project evaluation (within area of influence/focus).

This category, and associated guidance, could be applied to any road controlling authority scheme/project at their discretion. Could be a model or models of any form and scale. Where larger, eg regional, models are applied to a scheme within sub-region of the model, criteria/target levels in this guide relate to the area of influence/area of focus of the assessment.

PURPOSE TYPE E: Small area with limited route choice/corridor assessment.

Assessment focussed on an urban area with limited route choice, commuter corridors, smaller towns, and rural areas. Applications and assessment may be similar to larger urban area models but are likely to be focused more on traffic management testing than transport planning.

PURPOSE TYPE F: Single intersection/short corridor assessment.

Intersection or short corridor (around 3 intersections) assessments commonly focus on the performance of movements and approaches at intersections under different design layouts and/or traffic conditions (growth, development scenarios etc).

PURPOSE TYPE G: Special case high flow/high speed/ multi-lane corridors assessment.

Assessment and analysis of high flow, high speed, and/or multilane corridors such as motorways may require special treatment, eg detailed data collection and higher levels of model calibration and validation. Testing may include detailed motorway design, ITS, incident management, lane management, the effects of 'soft' policies etc.

3.3 APPLICATION TO PROJECTS

The categories above have been defined based on the purposes for which the models tend to be developed and geographic coverage. Figure 4 below gives an indication of how the purposes, and the models associated with them, overlap and apply to transport projects.

This should not be considered as a guide to selecting a modelling approach for a project, it is provided to offer further information on the purpose categories. The application of the model classifications to different projects has been broadly graded as:

- **S** – STRONG CORRELATION
- **P** – PARTIAL CORRELATION
- **N** – GENERALLY NO CORRELATION

FIGURE 4: LIKELY PROJECT APPLICATION OF PURPOSE CATEGORIES

PROJECT APPLICATION	PURPOSE CATEGORY						
	A: REGIONAL	B: STRATEGIC NETWORK	C: URBAN AREA	D: NZ TRANSPORT AGENCY LARGE PROJECT	E: SMALL AREA /CORRIDOR	F: INTERSECTION / SHORT CORRIDOR	G: HIGH FLOW, SPEED, MULTI LANE
AREA DEMAND RESPONSES, LAND-USE/TRANSPORT PLANNING, POLICY INVESTIGATION	S	N	N	N	N	N	N
LARGER TRANSPORT SCHEME FEASIBILITY, SCOPING STUDY	S	S	P	P	N	N	S[5]
LOCAL AUTHORITY TRANSPORT INTERVENTION & LAND-USE PLANNING	S	S	S	P	P	P	N
OPTION TESTING, DESIGN REFINEMENT, ECONOMICS	P	P	P	S	S	S	S
DEVELOPMENT FORECASTINGS AND/OR IMPACTS	P	P	P	S	S	S	S
DETAILED DESIGN, TRAFFIC MANAGEMENT	N	N	P	P	S	S	S
ITS, INCIDENT MANAGEMENT, ACTIVE MODE DESIGN AND IMPACT	N	N	P	P	S	S	S

Purpose category D (Transport Agency scheme/ project assessment), and by association the application of this guideline to Transport Agency projects, is principally applicable to 'large' NZ Transport Agency projects[6]. Section 1.7 above defines 'large' projects as costing over $5m for the purposes of this guideline and refines the definition into three classes based around investment. It is anticipated that application of target comparison criteria as defined in the chapters below is likely to include greater rigour for higher project classes, i.e. with greater investment and associated risk to Transport Agency. This can be achieved through consideration of the criteria adjacent to purpose category D.

5 Assumes model type-G has been built to cover an appropriate area of influence of the scheme.
6 This distinction could also apply to other road controlling authority projects.

SECTION 4: OBSERVED TRAFFIC DATA

4.1 TYPES OF DATA

There are many types of data used in transport planning and analysis, for example land-use data drives travel demand, and transport infrastructure data (PT services, road layouts, signal timings etc.) relates to supply. For the purpose of this document, two broad types of travel data have been categorised below based on current practises in collecting observed data for comparisons with transport/traffic models at the stage where travel demand is assigned to the transport network.

- Demand data: Data which relates to the movement of transport components through and within the study area. Includes origin-destination information, passenger counts, vehicle turning and link counts, vehicle classification data etc.
- Performance data: Data which relates to the performance of the transport network. Includes travel times, delays, queue lengths, speeds etc.

Demand and performance data may be used for calibration and/or validation. For guidance see the definitions in appendix A and discussion below.

4.2 INDEPENDENT DATA AND DEVELOPMENT OF DATASETS

LOCAL/PROJECT PURPOSES, CATEGORY C–G

As noted in the definitions (appendix A), independent data is data which is not used in any iterative model adjustment process (ie calibration). The majority of models will require the development of the specification of demand, which includes comparisons between observed and modelled demand data. For regional models, category A, comparisons between observed and modelled count data are generally undertaken following the development of the trip making models. Count data is usually considered as independent, and these comparisons are commonly considered as validation.

For a local or project-based model (category C through G), specification of the base year demand which is undertaken through iterative comparisons between observed and modelled count data is considered part of the model calibration process. For strict validation of these forms of models to traffic count data some form of independent count data would be required. To achieve this either:

- good quality observed data could be withheld from the calibration/demand development process
- poorer quality observed data[7] not used in the calibration could be compared
- or a complete independent count dataset could be collected.

Withholding good quality data from the demand development process is likely to result in poorer demand development, comparing poor quality data is likely to give a misleadingly low level of validation, and collecting a complete independent count dataset is likely to be too expensive. Additionally if a process (such as matrix estimation) is used to refine input demand matrices, the less observed data used in this process the easier it is for this process to manipulate and distort the demands to match the observed targets and to 'hide' errors in the specification of the model network (eg route choice). For further discussion, see section 8.1.

It may be accepted that many local/project-based models, ie non-regional, are calibrated and not validated to count data as this is a practical approach – producing the most robust base year trip matrices, a better calibration, and as a result offering the best model and value-for-money. Alternatively, consideration could be given to developing and maintaining separate traffic count datasets for thorough and comprehensive modelling, and/or where greater emphasis is placed on the forecasting/predictive capabilities of the model as opposed to requirements to reproduce the existing traffic conditions. A separate count set could be maintained for the demand development/calibration process, and an independent set for validation. Some potential examples of where this may be appropriate include regional models, area models (sub-regional) with a direct tie to the regional model, area models not linked to land-use but developed to include (limited) traveller responses (eg peak spreading, redistribution of trips, links between mode), or special case models (eg where multiple day-types are represented).

For regional modelling, back projections (ie comparing an historical land-use scenario with an associated count dataset) may also provide a mechanism for comparison with independent data.

7 Poorer quality data could include historical counts, data collected through less robust methods, non-directional non-classified link counts etc.

4.3 CALIBRATION AND VALIDATION

Different target level criteria for calibration and validation are not provided in this guideline. It is anticipated that if validation was carried out to a separate dataset, as outlined in the paragraph above, this would be reported in addition to the calibration results. Some relaxing of the target levels may be appropriate for the validation comparisons. The rationale for the reduction should be justified based on knowledge of the validation dataset.

Irrespective of the maintenance of a number of datasets, this guideline anticipates that any reporting of observed and modelled data identifies whether the comparison is calibration or validation. For guidance, see definitions in appendix A.

Re-representing data used in a calibration in a different format is not considered validation. Examples of this include use of individual movements across screenlines for calibration and summing this data to present screenline totals as validation, or use of summed turning movement counts as link flows for calibration and then presenting the individual turn counts as validation (or vice versa). The data has been used, and is therefore no longer independent. More importantly the same data used in a different comparison will include a similar level of equivalence between observed and modelled values, and it is therefore likely to misrepresent true validation levels.

This does not prevent calibration data being presented in a number of ways (as calibration), this is recommended as a number of comparisons help build a broader picture of the level of equivalence between observed and modelled data. It may be possible to re-represent comparisons using some calibration data (i.e. mixing data used in calibration and other independent data). The extent of the use of each dataset would need to be identified any reporting.

4.4 PROJECT MODEL SURVEY DATA

The development of project models (categories B through F) will commonly require the collection of observed data specifically for the purpose of the project. It is important that the analyst, peer reviewer and NZ Transport Agency modelling specialist are satisfied that such surveys are consistent with the surveys undertaken to build any parent (regional) model. This may require the adjustment of survey data to account for the day of the week or the time of year or both.

SECTION 5: TRAFFIC DEMAND COMPARISONS

5.1 OBSERVED AND MODELLED COMPARISONS

The statistics outlined in the sections below are generated by comparing observed and modelled counts. Where possible the observed data used in comparisons should be the raw surveyed count, not factored data (some seasonal adjustment may be acceptable when using a large dataset collected across a long timeframe). The reporting of calibration/validation statistics to factored data (eg data that has had some balancing applied to improve consistency with adjacent counts) may present a misleadingly high level of calibration/validation. The locations, source, any factoring of observed data, and all comparisons should be presented as transparently as possible within modelling reports.

The focus of many studies and hence the level of calibration/validation is commonly the peak periods (eg AM, PM and inter-peak periods). Statistics should generally be provided for intervals of no longer than one modelled hour across each modelled time period directly with the equivalent observed data. Where the modelled time period is longer (eg 3+ hours), there is likely to be a focus around the central, busiest hours, with target levels being relaxed for shoulder hours. Regional models category A (3 or 4 stage and similar) are often developed to cover 2-hour modelled periods and comparisons commonly undertaken with 2-hour data.

The GEH statistic, a commonly used threshold measurement (described in section 5.3), may be calculated for any observed and modelled values, ie hourly turn counts, screenline totals, two hour or full period link counts etc. When comparing larger values, greater than roughly 2000 units, it should be noted that GEH error tolerance reduces[8].

Generally observed and modelled count comparisons are carried out by comparing total volumes (ie all vehicle types). Reporting comparisons by vehicle classification or other disaggregation may be important in some circumstances, eg in locations where heavy vehicles have an atypical proportion.

Various forms of count comparisons are included within this guideline. Not all comparisons are anticipated or appropriate for all forms of models/projects, some expectations are broadly outlined in Section 1.7 *Application of this guideline*. Figure 5 expands on this and provides an indicative guide as to what comparisons are generally anticipated for the various purpose categories, graded as:

E	**P**	**N**
GENERALLY EXPECTED	POTENTIALLY USEFUL	GENERALLY NOT EXPECTED

FIGURE 5: LIKELY PROJECT APPLICATION OF PURPOSE CATEGORIES

PROJECT APPLICATION	PURPOSE CATEGORY						
	A: REGIONAL	B: STRATEGIC NETWORK	C: URBAN AREA	D: NZ TRANSPORT AGENCY PROJECT	E: SMALL AREA /CORRIDOR	F: INTERSECTION / SHORT CORRIDOR	G: HIGH FLOW, SPEED, MULTI LANE
SCREENLINE SUMMARY & MVT GEHS	E	E	E	E	P[8]	N	P[8]
INDIVIDUAL TURNING/LINK GEHS	N	P	E	E	E	E	E
SCREENLINE SUMMARY & MVT COUNT BANDS	E	E	P	P	P[8]	N	P[8]
INDIVIDUAL TURNING/LINK COUNT BANDS	N	P	P	P	P	P	E
XY SCATTER PLOTS	E	E	E	E	E	E	E
RMSE STATISTIC	E	E	E	E	P	N	N

8 To achieve the commonly used threshold of GEH = 5, at an observation of 2000 units the modelled value needs to be within 11.5%.
9 Where this comparison is considered useful, eg where parallel route choice and overall travel patterns are important cosiderations, and no targets are defined in the following tables. The targets for an adjacent model type may be used as a guide.

5.2 POTENTIAL COUNT COMPARISONS ANTICIPATED FOR MODEL SCREENLINES

Screenlines should be used in the majority of network modelling tasks to present comparisons of directional observed and modelled link counts. Presenting count comparisons on screenlines should highlight key travel patterns through the study area.

Screenlines should be defined based on the key travel directions through the study area, commonly north–south, east–west, inbound–outbound etc, on any geographical divides such as rivers and railway lines, and on the availability of observed count data. They should be carefully set up so that comparisons are not distorted, eg by double counting movements across single screenlines. Observed turn counts may be summed and presented as link counts although each count should only be used once in this process.

For regional models screenlines should be as complete as practically possible, ie capture all movements on as many of the links crossed by the screenline as feasible, to verify overall demand patterns regardless of route choice. For project models this is less critical and generally as many screenlines should be defined as can be supported by the observed data. Generally for project models every modelled link across a screenline should have an associated observed traffic count and this will tend towards a focus on the routes (links) carrying the heaviest traffic volumes and movements most affected by schemes or projects.

5.3 OBSERVED VS. MODELLED HOURLY LINK AND TURN COUNT GEH COMPARISONS

The GEH statistic is a form of Chi-squared statistic that can be used to compare observed and modelled counts. It is a useful for these comparisons because it is tolerant of relative and absolute errors, eg larger percentage differences on lower counts and larger absolute differences on higher counts. The GEH formula is given below.

$$GEH = \sqrt{\frac{(m-o)^2}{\frac{(m+o)}{2}}}$$

Where **m** is the modelled count and **o** is the observed amount

GEH comparisons should be summarised in bands, eg < 5, < 7.5 etc, and the percentage of modelled counts achieving each band presented. Table 1 presents the hourly GEH comparison criteria for observed vs. modelled total screenline counts, individual link counts on screenlines, and individual turning movements and provides guidance on target levels for each purpose category.

TABLE 1: HOURLY GEH COUNT COMPARISON CRITERIA[10]

COUNT COMPARISON	PURPOSE CATEGORY						
	A: REGIONAL	B: STRATEGIC NETWORK	C: URBAN AREA	D: NZ TRANSPORT AGENCY PROJECT	E: SMALL AREA /CORRIDOR	F: INTERSECTION / SHORT CORRIDOR	G: HIGH FLOW, SPEED, MULTI LANE
TOTAL DIRECTIONAL COUNT ACROSS SCREENLINE:							
GEH<5.0 (% OF SCREENLINES)	>60%	>75%	>85%	>90%	NA	NA	NA
GEH<7.5 (% OF SCREENLINES)	>75%	>85%	>90%	>95%	NA	NA	NA
GEH<10.0 (% OF SCREENLINES)	>90%	>95%	>95%	100%	NA	NA	NA
INDIVIDUAL DIRECTIONAL LINK COUNT ON SCREENLINES:							
GEH<5.0 (% OF COUNTS)	>65%	>80%	>85%	>87.5%	NA	NA	>90%
GEH<7.5 (% OF COUNTS)	>75%	>85%	>90%	>92.5%	NA	NA	>95%
GEH<10.0 (% OF COUNTS)	>85%	>90%	>95%	>97.5%	NA	NA	100%
GEH<12.0 (% OF COUNTS)	>95%	>95%	100%	100%	NA	NA	100%
INDIVIDUAL TURNING MOVEMENTS AND / OR DIRECTIONAL LINK COUNTS:							
GEH<5.0 (% OF TURNS)	NA	>75%	>80%	>82.5%	>85%	>95%	>85%
GEH<7.5 (% OF TURNS)	NA	>80%	>85%	>87.5%	>90%	100%	>90%
GEH<10.0 (% OF TURNS)	NA	>85%	>90%	>92.5%	>95%	100%	>95%

The raw data underlying the summary table should commonly be presented in an appendix within modelling reports.

10 Sources include 2010 EEM criteria for individual link flows (worksheet 8, Transport Model Checks, section 3 pg 5-36) and DMRB assignment validation criteria for individual flows and screenlines (volume 12, section 2, part 1, chapter 4 Traffic Model Development, table 4.2).

5.4 OBSERVED VS. MODELLED HOURLY LINK AND TURN COUNT BAND COMPARISONS

To support GEH comparisons observed vs. modelled hourly counts should also be summarised in count bands. This provides an additional check of observed and modelled flows; a comparison which meets the GEH comparison and doesn't meet the targets below may require explanation or further investigation and vice versa.

Table 2 shows a recommended approach for presenting hourly count band comparison criteria for observed vs. modelled total screenline counts, individual link counts on screenlines, and individual turning movements and provides guidance on acceptability levels for each model type.

TABLE 2: HOURLY COUNT BAND COMPARISON CRITERIA[11]

COUNT COMPARISON	PURPOSE CATEGORY						
	A: REGIONAL	**B:** STRATEGIC NETWORK	**C:** URBAN AREA	**D:** NZ TRANSPORT AGENCY PROJECT	**E:** SMALL AREA /CORRIDOR	**F:** INTERSECTION / SHORT CORRIDOR	**G:** HIGH FLOW, SPEED, MULTI LANE
TOTAL DIRECTIONAL SCREENLINE COUNTS:							
WITHIN 10% (% OF SCREENLINES)	>70%	>80%	>85%	>90%	NA	NA	NA
WITHIN 15% (% OF SCREENLINES)	>80%	>90%	>92.5%	>95%	NA	NA	NA
INDIVIDUAL DIRECTIONAL LINK COUNT ON SCREENLINES:							
<700VPH WITHIN 100VPH (% OF COUNTS)	>70%	>80%	>85%	>90%	NA	NA	>90%
700-2,700VPH WITHIN 15% (% OF COUNTS)	>70%	>80%	>85%	>90%	NA	NA	>95%
>2,700VPH WITHIN 400VPH (% OF COUNTS)	>70%	>80%	>85%	>90%	NA	NA	100%
INDIVIDUAL TURNING MOVEMENTS AND/OR DIRECTIONAL LINK COUNTS:							
<400VPH WITHIN 50VPH (% OF MOVEMENTS)	NA	>70%	>77.5%	>85%	>90%	>95%	>80%
400-2,000VPH WITHIN 12.5% (% OF MOVEMENTS)	NA	>70%	>77.5%	>85%	>90%	>95%	>80%
>2,000VPH WITHIN 250VPH (% OF MOVEMENTS)	NA	>70%	>77.5%	>85%	>90%	>95%	>80%

The raw data underlying the summary table should commonly be presented in an appendix within modelling reports.

11 Source DMRB assignment validation criteria for individual flows and screenlines (volume 12, section 2, part 1, chapter 4 Traffic model development, table 4.2).

5.5 OBSERVED VS. MODELLED COUNT XY SCATTER PLOTS

An XY scatter plot of all screenline link counts, or counts on key corridors, or turn and link count information may indicate how well the modelled counts fit the observed counts. Each individual observed count should be compared with the same modelled count. Each observed count should only be compared once, eg observed turn counts should not be summarised together and presented as link counts within the same dataset or to generate two comparisons.

XY scatter plots may be presented for hourly data or across each modelled time period depending on the focus of the project.

The equation of the line of best fit (trendline) should be displayed on the graph and this indicates how well the modelled counts represent the observed counts. The trendline should be forced to pass through the origin.

The equation of the trendline generally identifies if the model is trending low or high compared to the observed data.

The R squared value should also be presented as this gives a measure of how well the estimated trendline compares to the set of input data. A poor R squared value can indicate a spread of observed data, e.g. use of historical counts, or a poorly defined model network /demands. The R squared value should always be considered along with the equation of the line of best fit – a high R squared value can be obtained for a dataset which includes a consistently poor trend, eg an R squared value of 0.9 for a line y = 1.5x.

XY scatter plots are most effective where there is good count coverage, e.g. turn counts at a series of adjacent intersections and limited leak between intersections. Where there is less observed data this form of comparison may be distorted. Table 3 provides guidance on acceptability levels for R squared values and the line of best fit for XY scatter plots based on the model types.

Where a network includes areas carrying significantly different traffic volumes consideration should be given to splitting the dataset and presenting several XY scatter comparisons. For example, a network including both urban intersections with turning counts approx. 1000vph and less, and motorway links carrying higher volumes. Distinct geographic divisions may also be presented separately.

TABLE 3: OBSERVED VS. MODELLED COUNT COMPARISON XY SCATTER CRITERIA[12]

STATISTIC	PURPOSE CATEGORY						
	A: REGIONAL	**B:** STRATEGIC NETWORK	**C:** URBAN AREA	**D:** NZ TRANSPORT AGENCY PROJECT	**E:** SMALL AREA /CORRIDOR	**F:** INTERSECTION / SHORT CORRIDOR	**G:** HIGH FLOW, SPEED, MULTI LANE
R SQUARED VALUE	>0.85	>0.9	>0.95	>0.95	>0.95	>0.95	>0.95
LINE OF BEST FIT	y=0.9x - 1.1x	y=0.9x - 1.1x	y=0.9x - 1.1x	y=0.925x - 1.075x	y=0.95x - 1.05x	y=0.97x - 1.03x	y=0.97x - 1.03x

12 Source 2010 EEM criteria for observed and modelled flow comparisons (worksheet 8, Transport model checks, section 2 pg 5-35).

5.6 OBSERVED VS. MODELLED ROOT MEAN SQUARE ERROR (RMSE)

The root mean square error (RMSE) is a measure of the predictive success of the model and is a commonly referenced as providing an indication of the error of a model. It is usually calculated across all key observed and modelled count data points and may be calculated for each modelled hour, the key peak hour, or across the full modelled period depending on the focus of the project.

The percentage RMSE is defined as;

$$\% \text{ RMSE} = \sqrt{\frac{\Sigma(m-o)^2}{(N-1)}} \Big/ \Sigma \frac{o}{N}$$

where **m** is the modelled count and **o** the observed count and **N** is the number of modelled/observed data points. This should be calculated from the raw observed and modelled values. Table 4 provides guidance on acceptability levels for RMSE values.

TABLE 4: OBSERVED VS. MODELLED COUNT COMPARISON RMSE CRITERIA[13]

STATISTIC	PURPOSE CATEGORY						
	A: REGIONAL	B: STRATEGIC NETWORK	C: URBAN AREA	D: NZ TRANSPORT AGENCY PROJECT	E: SMALL AREA /CORRIDOR	F: INTERSECTION / SHORT CORRIDOR	G: HIGH FLOW, SPEED, MULTI LANE
RMSE							
ACCEPTABLE	<30%	<25%	<20%	<17.5%	<15%	NA	NA
REQUIRES CLARIFICATION	30 -40%	25-35%	20-30%	17.5-27.5%	15-25%	NA	NA
UNLIKELY TO BE APPROPRIATE	>40%	>35%	>30%	>27.5%	>25%	NA	NA

5.7 PRESENTING COUNT SPREAD

The spread, or range, of observed and modelled count data may be presented where the traffic model form is capable of presenting this information and/or a robust sample of observed data is available in key locations of the study area. This may be beneficial in scenarios where there is an interest in the variable conditions of key areas of the network, and/or where the model application may include more detailed (eg statistical) investigations into network reliability/variability issues.

5.8 COUNT COMPARISON CRITERIA APPLICATION

A model that does not meet the target levels in table 1, table 2, table 3, and/or table 4 may still be suitable for application (ie fit-for-purpose) if the discrepancies are acceptable due to known, noted, and accepted issues (eg observed data limitations) and any larger discrepancies are concentrated away from the areas of most importance to the appraisal. Conversely a model which passes the suggested acceptability levels but has significant discrepancies in key areas may be unacceptable. In these circumstances the analyst would be required to consider the implications for the intended application of the model and discussed these issues with the peer reviewer.

To expand on the above point and reiterate early comments; for the Transport Agency project class category in the above tables, the target levels in some circumstances are relevant to the area of interest/ scheme influence rather than the whole model. This category and approach may be used for schemes/ projects by other road controlling authorities if desired.

13 Source 2010 EEM criteria for percent RMSE (worksheet 8, Transport model checks, section 4 pg 5-36).

SECTION 6: PERFORMANCE COMPARISONS

Performance data is commonly important in providing an understanding of the on-street network performance and verifying the operation of the transport model.

Queues are volatile and subjective in nature and point speed data from loop counters only provides information at specific points in a network omitting knowledge of complete journeys. Directional journey time surveys along key routes through the network with intermediate timing points recorded for sections generally provide the most robust data relating to the performance of the transport network. Queues and point speed data may provide useful comparisons in situations where robust route journey time data is unavailable or cannot be easily measured. For isolated intersection modelling, approach and/or movement delay surveys may be more appropriate.

6.1 TOTAL ROUTE JOURNEY TIME COMPARISONS

The total observed and modelled journey times should be compared along each route including travel time through intersections, in each direction for the key peak time period, commonly peak hour but may be longer depending on the study area operation and length of journey time routes. Table 5 provides acceptability levels for comparing the total observed and modelled direction route journey times.

As with other comparisons, it would be expected that those involved in reviewing and assessing the appropriateness of the outcomes of these comparisons would apply their professional judgement in making these assessments.

Again, models that meet all the criteria may not be considered fit-for-purpose and conversely models that do not meet all the criteria may be. Particular checks relating to travel times include comparisons on key corridors, whether the model is consistently faster or slower, and whether the model is significantly different to the observed on any particular routes (eg more than 30-40% different).

TABLE 5: TOTAL JOURNEY TIME ROUTE COMPARISON CRITERIA[14]

TOTAL ROUTE DIRECTIONAL PEAK JOURNEY TIME	PURPOSE CATEGORY						
	A: REGIONAL	B: STRATEGIC NETWORK	C: URBAN AREA	D: NZ TRANSPORT AGENCY PROJECT	E: SMALL AREA /CORRIDOR	F: INTERSECTION / SHORT CORRIDOR	G: HIGH FLOW, SPEED, MULTI LANE
WITHIN 15% OR 1 MINUTE (IF HIGHER) (% OF ROUTES)	>80%	>85%	>85%	>87.5%	>90%	>90%	>90%
WITHIN 25% OR 1.5 MINUTES (IF HIGHER) (% OF ROUTES)	>85%	>90%	>90%	>92.5%	>95%	100%	100%

14 Source DMRB observed and modelled journey time comparison criteria (volume 12, section 2, part 1, chapter 4 Traffic model development, table 4.2).

6.2 JOURNEY TIME VS. DISTANCE GRAPHS

Cumulative observed and modelled journey times and distances should be plotted on an XY scatter graph, showing each timing section for each direction along each route. The slope of the lines indicates the level of congestion and demonstrates how the model is performing along key sections of the model network.

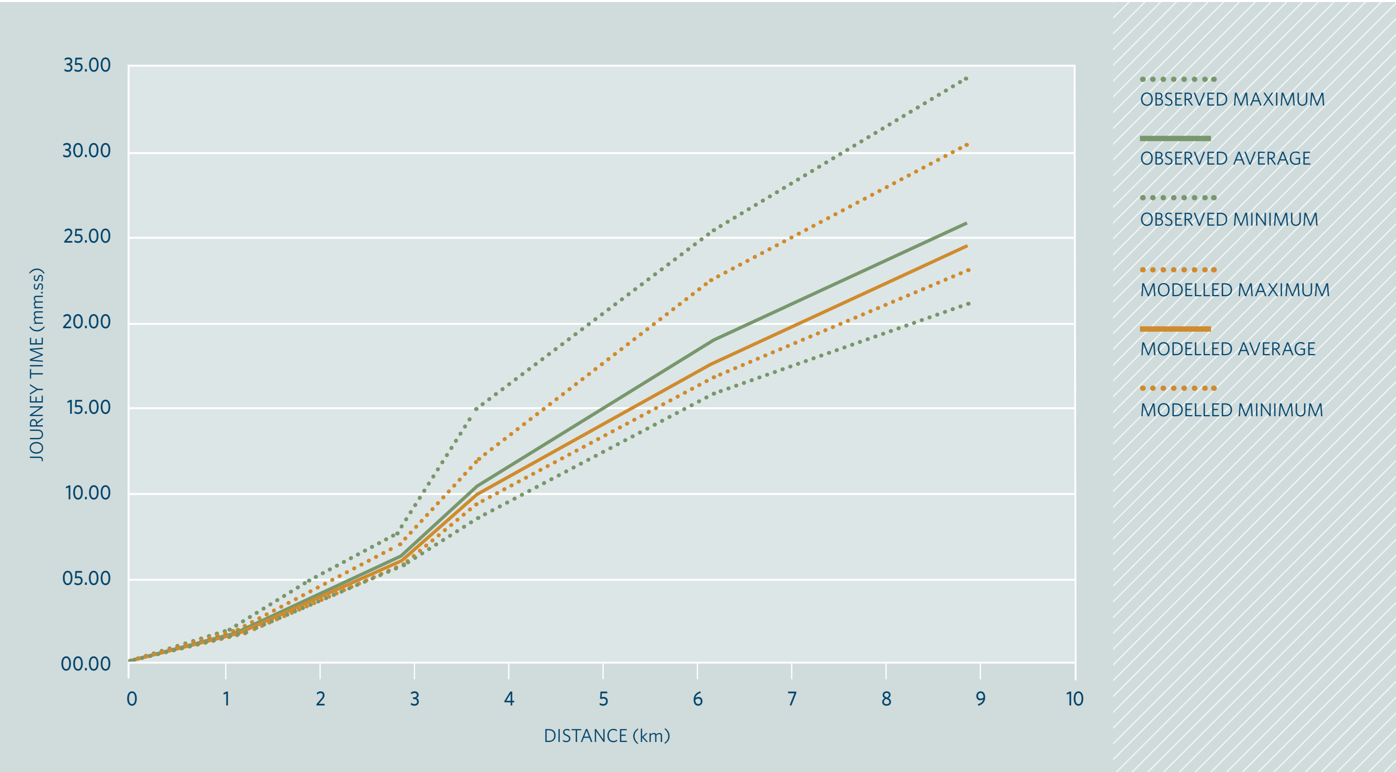
These graphs should be generated for the same period as the total route comparisons, commonly the peak hour but this may be longer depending on the study area operation and the length of journey time routes.

The graph should demonstrate the spread of observed journey times. Where the model form is capable of generating the same information the spread of modelled journey times should also be presented. For example the minimum, maximum (or other statistics indicating the spread of data) and average path journey times may be presented. An example is shown in figure 6.

The pattern of congestion along the sections of route (demonstrated by the slope of the graph) should be represented within the model. The spread of observed and modelled data should broadly correspond, ie ideally the modelled average should fit within the range of observed data and (where it is possible to present) the modelled range should show similar patterns to the observed data. If the modelled journey times are outside of the observed range then the potential impacts of this should be presented and discussed.

When presenting minimum and maximum journey times, consideration may need to be given to comparing the slowest/fastest individual sections or the slowest/fastest route. For example, the fastest/slowest journey times recorded in each section combined and presented as the overall route minimum/maximum may result in unrealistic values being presented (i.e. a fastest overall route journey time that could not be achieved in reality). The converse may also be true, if limited journey time sampling has been conducted the fastest/slowest overall route journey time may misrepresent the actual spread/range of journey times along the route.

FIGURE 6: JOURNEY TIME VS. DISTANCE XY SCATTER GRAPH EXAMPLE[15]



15 Source DMRB Local model validation report (volume 12, section 2, part 1, appendix B, figure B2).

6.3 QUEUE COMPARISONS

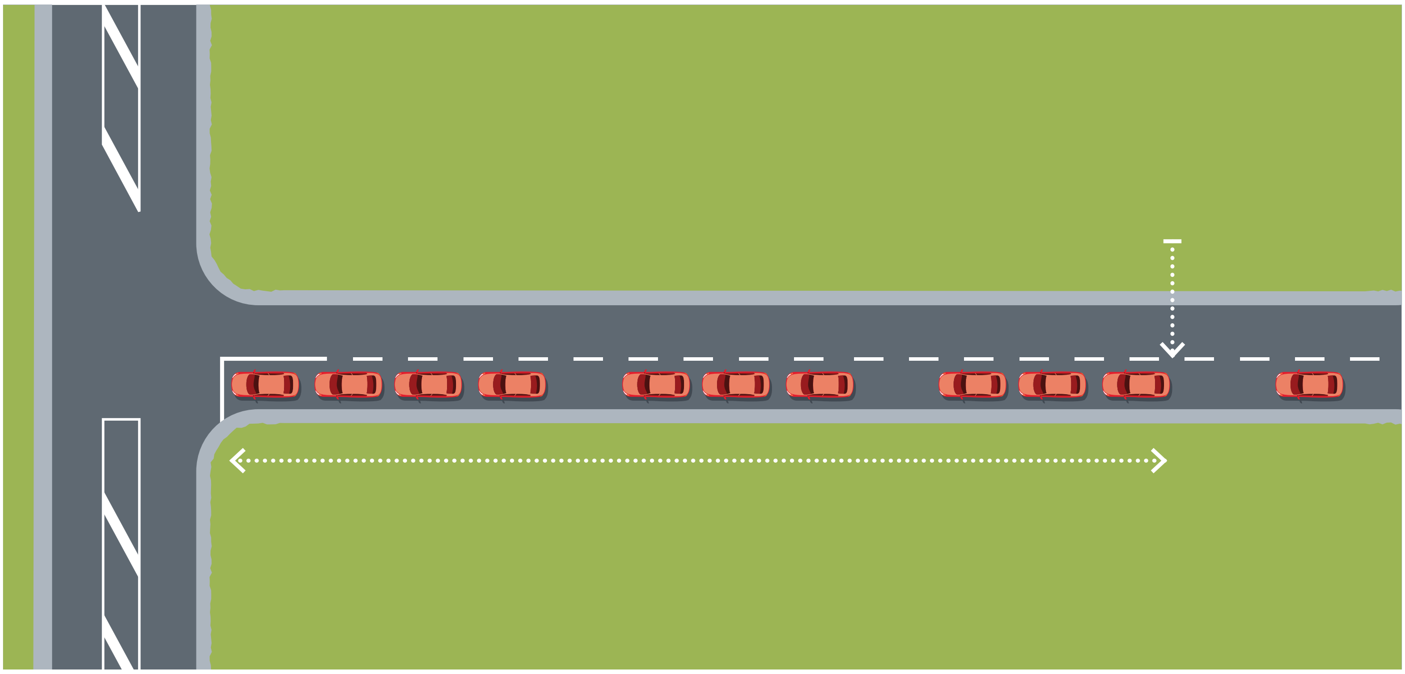
Queues are volatile and subjective by nature therefore defining numerical model queue length comparison criteria are generally not considered appropriate. Consequently, the main focus on model performance is typically journey times.

In some circumstances though, eg modelling a single intersection or a small set of adjacent intersections, or in some locations in the study area it may be that journey routes are restrictively short to provide reasonable data (a journey time section of less than 30 seconds is difficult to obtain robust data for) or prohibitively expensive to collect (eg along every route modelled). Queue comparisons may be used to support the model calibration in these situations/areas of the network.

If queue comparisons are undertaken the main interest for stakeholders is the back of the queue. As discussed above queues are subjective, this document assumes that the length of a queue is defined by the last vehicle travelling below a certain speed and within a certain distance of the vehicle in front. This length is measured back from an intersection stopline and the 'last vehicle' ignores any queue shockwaves, ie intermediate pockets of queues along the intersection approach. This definition is shown diagrammatically in figure 6 below. A queue of 10 trucks is longer than 10 cars. Therefore queue lengths should generally be recorded and presented as the maximum length in metres over each observation/measurement interval.

Observed and modelled queue comparisons may be qualitative, eg the back of queue was observed to extend approximately to location X, or concentrate on the build-up and dissipation of queues through time where the model form provides this functionality.

FIGURE 7: QUEUE LENGTH EXAMPLE



6.4 POINT SPEED DATA

Point speed data, eg from permanent motorway loop detector sites or dual tube counters, may be used as an additional check of model performance where this data is available. For example, by comparing observed and modelled speeds through the modelled time period or through speed/flow relationship comparisons.

Comparisons with point speed data should be used with caution. As the information is only recorded at a point, it omits any knowledge of the performance over the section of road carriageway. This can lead to difficulties e.g. a situation where the model is performing well over the section of road but the precise location of point-to-point operation differs from the on-street data resulting in an apparently poor point data comparison.

6.5 OVERALL MODEL OPERATION AND CONVERGENCE

Global model statistics such as average trip time and length will vary from study area to study area and it is therefore not possible to define criteria to judge whether these are appropriate.

Measures of model stability do differ depending on the form of the model. Some model forms iterate and converge, commonly this is based on global statistics (referred to below as 'equilibrium-style'). Microsimulation models generally do not converge and stability can relate to the change in operation from one run to the next.

Generally for equilibrium-style modelling some degree of convergence needs to be demonstrated. This is typically required to reach a consistent, reproducible assignment.

For microsimulation models instability would occur when there is a large difference in operation between runs (ie differing random seed runs) of the same model using the same inputs. The main microsimulation transport model software packages are generally considered to be inherently stable. There is generally no set requirement to demonstrate the model stability during the calibration phase of microsimulation models. Where instability does occur, eg gridlock occurring in the occasional run, it is commonly due to model error. Every effort should be made to correct instability issues if they arise in both base and future year scenarios. Errors may be related to demand inputs (ie unrealistically high demands) and/or poor model intersection and network representation (ie grossly inadequate network capacity, inappropriate route choice responses etc).

SECTION 7: PUBLIC TRANSPORT COMPARISONS (REGIONAL MODELLING)

For some specific forms of modelling (4 stage or public transport assignment models) comparisons between observed and modelled public transport (PT) data are generally required. This could include:

- screenline counts of PT trips (person based)
- PT demand (origin-destination) trip information
- trip breakdown data – number of trips, number of boardings, number of transfers
- PT journey times (service based and ideally person based to reflect interchange times).

Comparisons with trip making will generally be required for regional models. Service journey time data and comparisons between observed and modelled times may be required for both regional models and models used in more detailed PT scheme assessment work, eg in support of significant corridor bus.

7.1 PT PATRONAGE SCREENLINE COUNT COMPARISONS

Table 6 below presents the GEH comparison criteria for observed vs. modelled PT patronage screenline counts and provides guidance on acceptability levels for each model type. These comparisons would commonly be carried out across each modelled peak period.

TABLE 6: OBSERVED VS. MODELLED PT PASSENGER SCREENLINE COUNT COMPARISON CRITERIA

COUNT COMPARISONS	PURPOSE A: REGIONAL
TOTAL PT PAX ACROSS SCREENLINE:	
GEH<5.0 (% OF SCREENLINES)	>60%
GEH<7.5 (% OF SCREENLINES)	>70%
GEH<10.0 (% OF SCREENLINES)	>80%
GEH<12.0 (% OF SCREENLINES)	>90%
LINE OF BEST FIT	y=0.9x - 1.1x
R SQUARED	>0.85
INDIVIDUAL DIRECTIONAL PT COUNTS:	
GEH<5.0 (% OF LINKS)	>50%
GEH<7.5 (% OF LINKS)	>60%
GEH<10.0 (% OF LINKS)	>70%
GEH<12.0 (% OF LINKS)	>85%
LINE OF BEST FIT	y=085x - 1.15x
R SQUARED	>0.80

7.2 PT BOARDING COMPARISONS

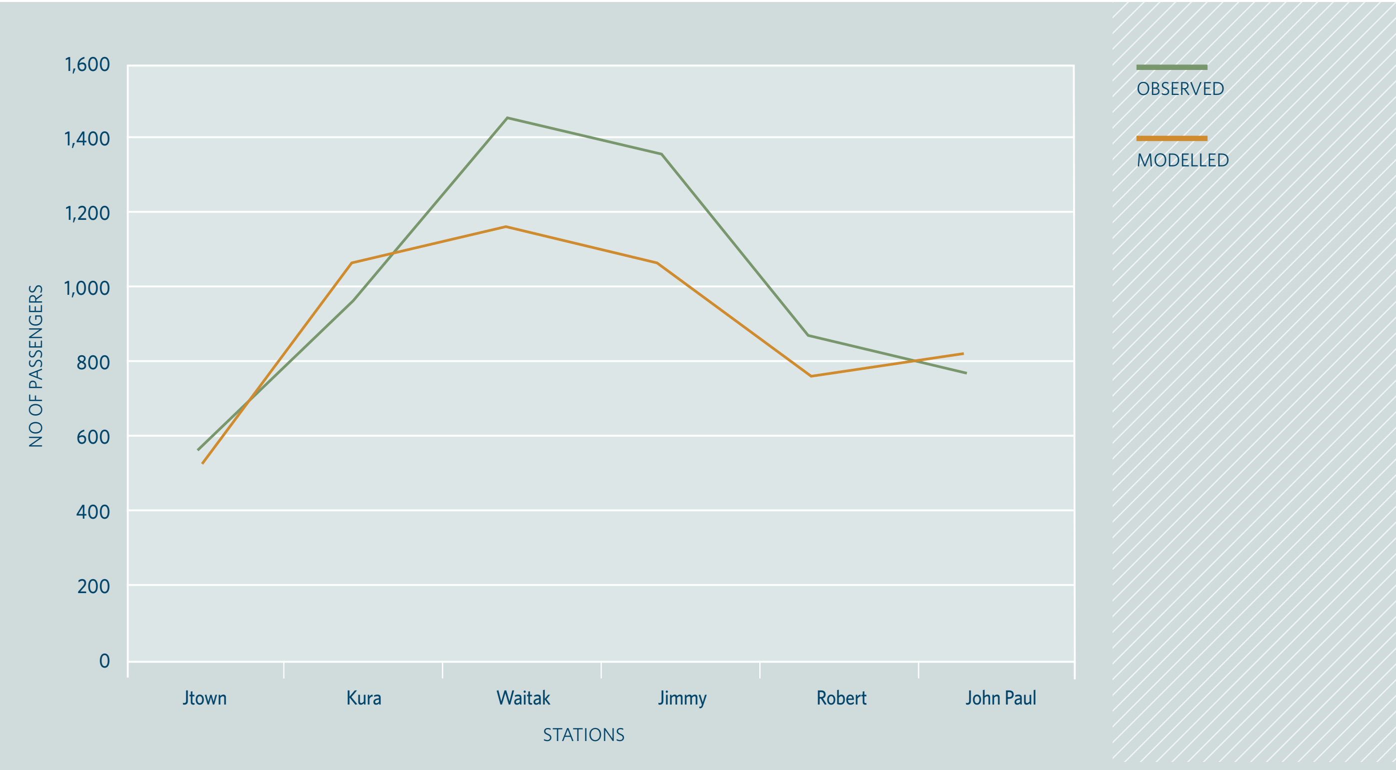
Comparisons between PT passenger boarding numbers may be made in several different ways. Common examples include the number of boardings along a corridor/route and the profile of boardings along key routes/corridors.

Where more comprehensive PT modelling is required, eg in a region featuring more extensive on-street PT usage such as a high use rail corridors, more detailed PT boarding comparisons may be made. For example the profile of passenger boarding numbers along key routes. Figure 8 shows an example this form of comparison.

TABLE 7: OBSERVED VS. MODELLED PT CORRIDOR/ROUTE BOARDING COMPARISON CRITERIA

BOARDING COMPARISONS	PURPOSE A: REGIONAL
CORRIDOR / ROUTE COMPARISONS	
GEH<5.0 (% OF CORRIDORS / ROUTES)	>50%
GEH<7.5 (% OF CORRIDORS / ROUTES)	>60%
GEH<10.0 (% OF CORRIDORS / ROUTES)	>70%
GEH<12.0 (% OF CORRIDORS / ROUTES)	>80%
LINE OF BEST FIT	y=0.85x - 1.15x
R SQUARED	>0.80

FIGURE 8: EXAMPLE PT BOARDING PROFILE COMPARISON



7.3 PT JOURNEY TIMES

It is anticipated that PT journey time comparisons would be broadly similar to vehicle comparisons, see table 5 and figure 5 in section 6, with some potential relaxing of guideline target levels depending on the focus and purpose of the model. This could be service based or person based depending on data availability. Person based comparisons enable transfer/interchange times to be confirmed.

7.4 PT DEMAND COMPARISONS

For PT demand modelling, it is anticipated that PT OD demand comparisons, and matrix adjustments, would broadly align with the guidance in section 8 below. Section 8 should not be considered explicit to vehicle demand matrices.

SECTION 8: MATRIX DEVELOPMENT PROCESSES & ADJUSTMENTS

8.1 PREDICTIVE MODELLING

The previous sections set out guidance for comparisons between observed and modelled counts (traffic demand comparisons section 4, PT comparisons section 6). There are many techniques that can be used to adjust the model and inputs to improve these forms of comparisons. Models are in essence 'predictive tools'. It is important that the model is not developed using techniques which constrain the model's predictive capabilities, eg in developing the network - prescribing lane choice based on vehicles network destinations or prescribing route choice through the extensive use of route choice cost/link type factors.

Adjusting the demand description (synthetic trip-making relationships or matrix directly) to improve the observed vs. modelled comparisons is a reasonably common step in the model development process. Adjusting the matrix directly is generally considered an important step where there is a strong requirement to represent the current local demand movements to a high level of equivalence, eg testing intersection design improvements. It is also one step where predictive and forecasting abilities of a model can be significantly eroded – by distorting the trip patterns in the demand matrix to 'fit' against the model's inaccurate representation of route choice combined with observed data.

A significantly distorted base year matrix results in several notable issues:

- Linkage/association with land-use is broken down resulting in difficulties in producing reasonable forecast future flows (generated from land-use change).
- If a more significant change is introduced in the network, eg a new link, the response of the model may be inaccurate (eg inaccurate forecast of the use of the new link and/or changes to volumes on links throughout the network).

When considering the use of and/or applying demand matrix adjustment techniques such as matrix estimation (ME), the description of the model network (notably route choice predictions), and achieving observed vs. modelled comparison criteria, all need to be carefully considered against the forecasting and predictive abilities of the model. It may be appropriate to relax comparison criteria where it is necessary to maintain links to land- use for forecasting, fixed trip generation in certain locations, and a predictive description of the road network. Figure 9 gives an indication of how this consideration is likely to apply to the various purpose categories, graded as:

R	B	N
KEY REQUIREMENT	BALANCED CONSIDERATION	NOT AS CRITICAL

Mathematical techniques (eg ME, furnessing) that are used to adjust the matrix are inherently 'dumb' techniques. This places a lot of emphasis on the analyst undertaking these processes using suitable techniques. This guidance does not extend to 'how to' advice on undertaking these procedures. Outlined in the following sections is guidance on reporting on the outcome of matrix adjustments.

FIGURE 9: LIKELY PROJECT APPLICATION OF PURPOSE CATEGORIES

PROJECT APPLICATION	PURPOSE CATEGORY						
	A: REGIONAL	**B:** STRATEGIC NETWORK	**C:** URBAN AREA	**D:** NZ TRANSPORT AGENCY PROJECT	**E:** SMALL AREA /CORRIDOR	**F:** INTERSECTION / SHORT CORRIDOR	**G:** HIGH FLOW, SPEED, MULTI LANE
STRONG PREDICTIVE FORECASTING CAPABILITY	R	B	B	B	B	N	N
LINK WITH REGIONAL FORECAST, ABILITY TO ACCOUNT FOR CHANGING LAND-USE	R	R	R	R	R	B	B
STRONG REPRESENTATION OF (CURRENT/ SHORT-TERM) ON-STREET TRAFFIC FLOWS	N	B	B	B	B	R	R

8.2 PROCESS AND DEFINITIONS

The representation of trips and travel is a key part of developing transport models in all but the simplest example of modelling single intersections.

For a regional model this includes comparisons between an observed trip matrix (eg from household survey, road side interview, journey-to-work data from Census) and the matrix generated by the synthetic trip making relationships. This would include trip generation, trip distribution, and modal split (for 4 stage models) comparisons in addition to the assignment checks of modelled vs. observed counts. As far as feasible, a demand refinement process carried out on the assigned OD traffic matrix to provide a better match with observed traffic counts (ME, furnessing and similar) should be avoided in the development of a regional model.

For an assignment model this usually involves the refinement of some form of source or prior matrix via a more direct process (eg matrix estimation (ME), furnessing, infilling, manual manipulation). This can involve 3 steps, firstly extraction of the 'raw' source matrix, secondly improving/correcting/updating of this matrix (eg infilling missing trips, furnessing, trip-end factoring) to provide a more appropriate representation of the study area and a more robust input to the potential third step – running through a final refinement process (manual adjustment, furness, matrix estimation etc).

A lot of adjustment may be completed at step 2, with the final step being a small-scale 'tweak'. This leads to three potential matrices as defined below:

I. Source matrix: Matrix obtained for external source, eg higher tier model, observed data, seeding or gravity generated.

II. Prior matrix: The updated source matrix, used as an input to a further demand refinement process.

III. Final matrix: The finalised matrix as used in the reported base model.

The checks that are outlined in the sections below involve comparisons between these three matrices. A regional transport model process will usually involve checking matrices I and III, ideally this will not include any matrix estimation or other refinement process on the assigned OD traffic matrix. Checks of an assignment traffic model may involve I and III, or II and III, or all three matrices.

For Transport Agency projects, if matrix estimation (or a similar matrix refinement process) is employed then it requires specific agreement from the analyst, the peer reviewer and the Transport Agency specialist and a clear rationale recorded in model reporting.

8.3 LEVEL OF REPORTING

The comparisons outlined below will provide a client or reviewer with a degree of information related to the changes made to a matrix. It is perhaps easier, notably for a reviewer, if the processes that have been used are described to a reasonable level in any model reporting. Where appropriate, this should include descriptions of the following key information:

- Level of network development prior to demand refinement (eg any isolated intersection calibration processes, development of profiles, application of speed/flow curves).
- Source of matrices.
- Adjustments to source matrices (description of method(s), notes on quality of source matrices).
- Details of adjustment to prior matrices to produce final matrices:
  - Description of method used, eg matrix estimation.
  - Inputs to method (survey data, routeing information, matrix used).
  - Key settings, eg number of internal ME iterations.
  - Use of constraints, ie bounds set on factors or changes to cells or trip ends.
  - Segments of the prior matrix protected from adjustment, by freezing or masking.
  - Weighting areas of the prior matrix.
  - Checks carried out on results.
  - Processes undertaken to improve outcomes.
- Decision process on when process was deemed complete.

The above list relates mainly to assignment models. The reporting relating to the demand development (trip making proxies) process in a regional model is anticipated to be significantly more exhaustive.

8.4 MATRIX ADJUSTMENT COMPARISONS

The sections below outline outputs that can be reported comparing the final matrices with source matrices. In many cases the focus may be justification on why notable significant changes are deemed acceptable.

8.4.1 MATRIX TOTALS

The matrix total – sum of the all the trips in the matrix – is a useful indicator of the overall change made to the matrix(es). Tolerance levels will depend on the source, quality, and level of sampling in the source/prior matrix, table 8 provides guidance levels.

8.4.2 SECTOR-SECTOR ANALYSIS

This is often the most revealing test and is recommended for all models. Sectors are groups of zones, commonly defined by areas of key land-use activity and/or geographic divisions, eg CBDs, suburbs, major geographic regions. Sector-to-sector analysis identifies travel patterns and volumes between these key areas.

The sector-to-sector analysis should be presented to show both the overall change in trips as well as the overall change in trip patterns. It is not possible to set guidance on target levels of acceptable change for individual sectors, as these will depend on the context of the individual models. The comparison should include discussion and explanation relating to why the largest absolute and percentage changes have occurred.

Where count data is available checks should be undertaken to determine whether there are systemic issues at a sub-regional sector level that is leading to a under or over representation of trip making. This can often be achieved via, and is an important function of, screenline checks. This may require some factoring of trip numbers associated with particular sectors. Where such factoring does occur there should be some investigation leading to a justification of why such factoring should take place.

8.4.3 MATRIX CELL ANALYSIS

Many network models will have in excess of 100 zones, making for vast numbers of individual matrix cells (100x100). Analysis of the distributions of cell changes (both as absolute changes and ratios) can be useful, but it may be difficult to identify useful information from the extensive amount of data. It is recommended that the largest cell changes are identified, and/or a frequency analysis conducted – for example, the no. of cells changing by >50%, 40 – 50%, 30 – 40% etc. The comparison should include discussion and explanation relating to why the largest cell changes have occurred and the extent of the changes.

8.4.4 TRIP RATE DIFFERENCES (SPECIFIC APPROACH)

Where project models use a trip generation and distribution approach to develop prior matrices the trip rates from or to a zone should be checked to see that they are broadly consistent with the land use that exists within each zone. This check is likely to be worthwhile for all models covering larger areas. This check should be focused on significant generators, particularly those using or directly influenced by any project/scheme that the model is developed to assess and/or in key areas of the network. Reputable sources of trip generation data such as that provided by the Trips Data Bureau or other should be used to check trip generation. Trip rates data can be provided as daily or specific time periods and this may require some manipulation for suitable comparison.

The comparison may be assisted through a XY plot of modelled trips by zone versus the selected source. This should be undertaken separately for departure and arrival volumes (ie row and column 'trip-end' totals). Such a plot would highlight anomalous zones and the slope of a line through the data points can show if there are any systemic bias in terms of trip generation.

TABLE 8: MATRIX TOTAL COMPARISONS

SOURCE OF COMPARISON MATRIX	PURPOSE CATEGORY						
	A: REGIONAL	B: STRATEGIC NETWORK	C: URBAN AREA	D: NZ TRANSPORT AGENCY PROJECT	E: SMALL AREA /CORRIDOR	F: INTERSECTION / SHORT CORRIDOR	G: HIGH FLOW, SPEED, MULTI LANE
HIGHER QUALITY, HIGH SAMPLE, GOOD CORRESPONDENCE BETWEEN SOURCE AND MODEL	<3%	<4%	<5%	<5%	<5%	<7%	<10%
LOWER QUALITY, LOW SAMPLE, POOR CORRESPONDENCE BETWEEN SOURCE AND MODEL	<6%	<8%	<10%	<10%	<10%	<15%	<20%

8.4.5 TRIP END CHANGES

The zone totals (inbound and outbound) are readily processed and analysed for any size of model and provides very useful data on changes to individual zones. The trip end totals should be presented as the percentage within GEH bands, eg <2.5, <5, <7.5, <10.0 and >10.0. Discussion and explanation should be included in relation to the extent of changes and in particular any higher GEH values (>10.0).

8.4.6 TRIP LENGTH DISTRIBUTION

The change in trip length distribution is an important indicator in assessing the appropriateness of any matrix adjustment process, and particularly to identify if the process has targeted certain movements (such as very short trips). The average trip length and the trip length distribution (percentage of trips by trip-length bins, plotted as a histogram) should be provided comparing the source and/or prior matrices to the final matrices. An example of a trip length frequency histogram is shown in figure 10.

A coincidence ratio (CR) is a measure of how well the trip length distribution of the trip matrix matches the observed trip length distribution. CR is defined:

$$CR = \sum[M\,in(fm, fo)] / \sum[M\,ax(fm\,, fo)]$$

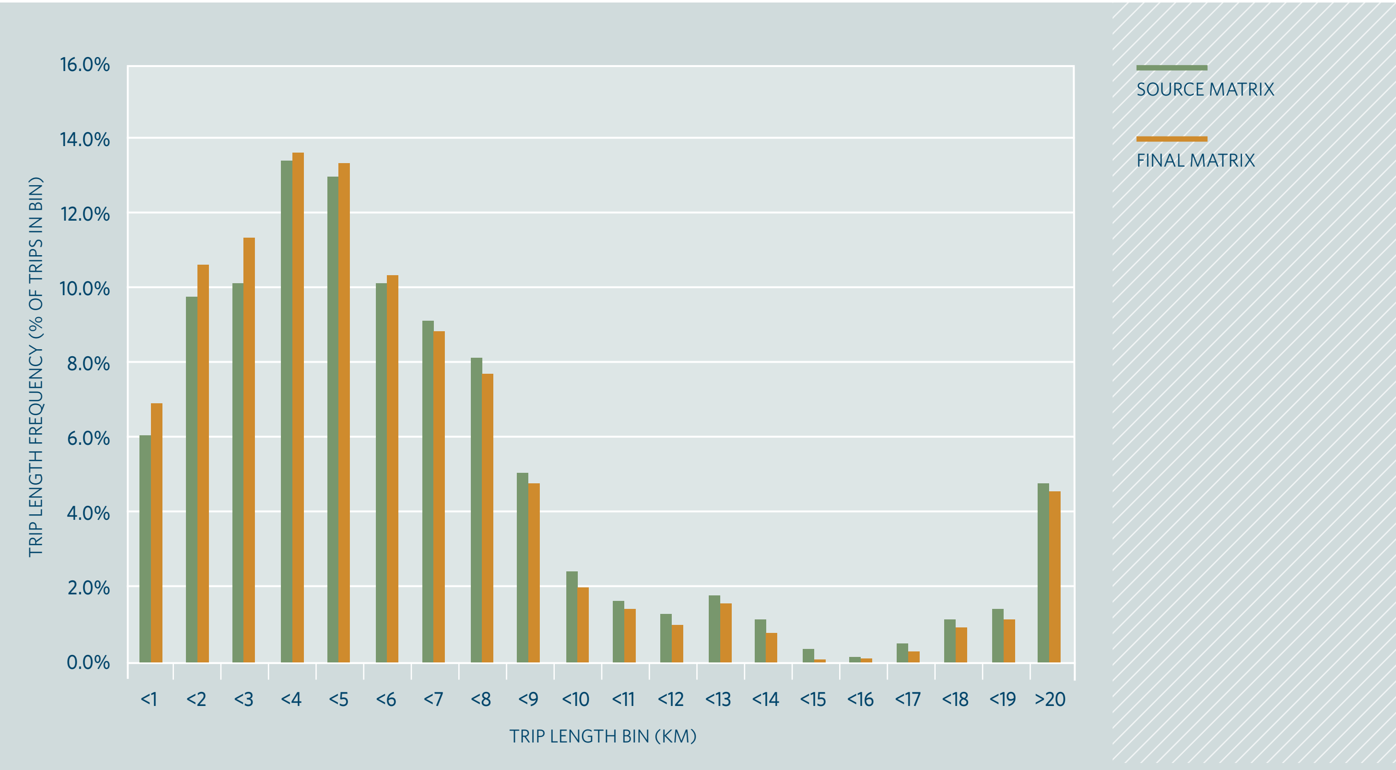
where the summations are over the number of bars in the histogram or distance ranges (n), fm is the modelled trip length frequency for a given distance range and fo is the observed trip length frequency for a given distance range.

The individual difference between the modelled and observed trip length frequency can be measured by the normalised trip length frequency deviation ND which is defined by:

$$ND = |fm - fo| / \sum[M\,ax(fm, fo)]$$

Guidance on the desirable range for the coincidence ratio and any normalised trip length frequency deviation are provided in table 9.

FIGURE 10: EXAMPLE OF TRIP LENGTH HISTOGRAM



Discussion and explanation should be provided along with these comparisons focussing on changes to the trip- length distribution, for example increases in shorter trips or any distortion to the overall distribution.

Trip length distributions vary by trip purpose and user class and so if the data is available these should be disaggregated and separately examined.

8.4.7 TRIP END BALANCING

Matrix adjustments can be sensitive to specific routing patterns and to the quality of the count data, potentially leading to inappropriate changes to the matrix. Often it is difficult to gauge the appropriateness of the individual changes, analysis of the balance of inbound and outbound trips can be revealing, especially during inter-peak periods. During such periods it could be expected that the inbound and outbound trips are balanced, yet inappropriate adjustments can distort one direction more than another due to non-symmetrical routing or count data. A comparison of the inbound/outbound trip ends before and after adjustments is recommended for inter- peak models, especially if flows from such models are expanded to estimate daily flows (where any imbalance would be exaggerated).

TABLE 9: TRIP LENGTH DISTRIBUTION MEASURES

MEASURE	PURPOSE CATEGORY						
	A: REGIONAL (commonly by trip purpose)	**B:** STRATEGIC NETWORK	**C:** URBAN AREA	**D:** NZ TRANSPORT AGENCY PROJECT	**E:** SMALL AREA /CORRIDOR	**F:** INTERSECTION / SHORT CORRIDOR	**G:** HIGH FLOW, SPEED, MULTI LANE
COINCIDENCE RATIO (CR)	>0.60	>0.65	>0.70	>0.75	>0.80	NA	NA
NORMALISED DEVIATION (ND)	<0.8/n	<0.7/n	<0.6/n	<0.5/n	<0.33/n	NA	NA

SECTION 9: FURTHER AND ADDITIONAL MODEL CHECKS

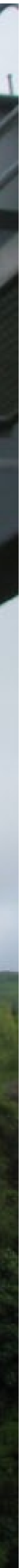
During the development or update of a base model, it may be appropriate to test the models responses to transport-related changes. This involves carrying out checks which are representative of how the model is likely to be applied to transport assessments and investigations.

These types are checks may be more appropriate for regional and legacy models, ie models with anticipated lifespans of several years or longer often covering key areas of a network or larger geographic areas.

Some forms of these tests are:

- Testing model operation and responses to increased traffic demand volumes through either applying estimated growth forecasts, or assuming a blanket growth rate on the base year matrix (e.g. 5%, 10%).
- Testing model operation and predictions (volumes, travel times etc.) by removing and/or adding a significant piece of transport infrastructure such as a key link in the network.
- Retrospectively testing a recent significant transport project, i.e. by removing it from the base year network and comparing with historical 'without' observed data (e.g. counts and travel times).
- Similar to above, transport models may be developed from a dataset which is recorded at a certain point in time (e.g. during a Census) and completed a period of time after this. If major transport projects have occurred since the observed dataset was recorded, then it may be possible to add these schemes to the model and measure the model's predictions against observed data (e.g. counts and travel times).
- For regional modelling, it may be appropriate to test elasticity responses in the model, e.g. changing Public Transport headways, fares etc. and review the model response. This may be benchmarked against International elasticities.
- For regional modelling, back-casting traffic 'forecasts' using previous census data to create historical demand scenarios and comparing against historical count data may be appropriate. Note, this may be a resource intensive and costly exercise.
- Similar to the above principle, for an assignment model it may be possible to create a historical 'forecast', particularly where forecasts are created from / linked to a regional model and historical scenarios are available, and compare to the assignment model outputs to historical observed data (e.g. counts and travel times).

APPENDICES



APPENDIX A DEFINITIONS

Study area: The area of influence defined for the study. Depending on the study, it may be limited to the area influenced by the transport scheme being investigated, or a wider region encompassing most of the land uses that generate demand for travel in the area.

(Transportation related) model: A single network including either the assigned demand or the mathematical equations representing trip-making relationships.

Network: The transport infrastructure (links and services) and features that provide for, control, and influence travel across a (study) area.

Demand: The representation of movements (vehicles, persons etc.) across the area. Commonly in the form of origin – destination trip matrices and the profile of demand through the time period modelled.

Project/assignment traffic model: A traffic model which assigns vehicle demand to a network and has no direct incorporation of land-use, demographics etc. This form of model requires the demands to be specified as an input. They may be developed with a specific task or transport project in mind. These models range from single intersections to entire inter-urban areas and are built using a wide range of model forms. These models will generate travel times between zones, vehicle emissions, queues etc.

Regional transport model: A transport model comprises a traffic assignment model with a demand model. These models are concerned with the movement of people and goods and provide information on mode share, induced travel, volume of freight carried etc. Transport models are built on relationships between land-use activity, demographics, etc. and commonly cover the movement of transport demand across an entire region. The forms of these models can be 3 or 4-stage, or activity based. May also be referred to as strategic, macro(scopic), or demand models.

Land-use model: A model which looks at the formation of households, migration of people, business start-up etc. (Often responding to accessibility provided by the network, showing the resulting demand for travel, and complementing with a transport model to form a land-use interaction (LUTI) model).

Calibration: The process of altering the model to reflect the observed data as well as necessary to satisfy the objectives of the study. Commonly an iterative process.

Independent data: Observed data not used in any model calibration process, ie data compared once and not used in any iterative adjustment of the model.

Validation: Observed and modelled comparisons using independent data that has not been used for calibration.

Matrix adjustments: The processes involved in developing the representation of demand movements through the network, notably model or direct adjustments made to improve some form of observed or starter matrix.

APPENDIX B MODEL SPECIFICATION

OVERVIEW AND BACKGROUND

The development of a new model, or significant update/adaption to an existing model, requires the specification of a wide range of aspects and consideration of a number of factors. This appendix provides a short note of considerations of several important aspects of model specification. Most of these aspects are particularly relevant to the development of project models, but broadly applicable to most common modelling projects.

KEY AREAS OF MODEL SPECIFICATION

The purpose of this discussion is to highlight areas of the model development which need to be considered and are likely to require discussion between the analyst, peer reviewer and/or client. The Analyst will develop a model specification which essentially defines the attributes of the model so that the model's purpose is achieved. The model specification will need to be agreed by the peer reviewer and client (in the case of the Transport Agency, the Transport Agency modelling specialist). In agreeing to this specification all parties are determining the attributes that the model will need to possess in order that model can do what the model is designed to do within the specified levels of confidence. This will include agreeing the data comparison criteria that the model will need to satisfy, ie by identifying the purpose or purposes and the subsequent category(s) that apply to the model based on the definitions in section 3.

The model specification will include a variety of important aspects, several key elements and important considerations are highlighted below.

Study area: the definition of the modelled study area is commonly based on the area of influence for the study. Considerations can include the area of influence of the transport scheme being investigated, key routing decision points, coverage of key trip origin/destination locations, coverage of congestion (extent of queuing), affects on/from adjacent network features (eg intersections).

Time periods: the determination of the model time periods commonly relates to the model purpose, the operation of the transport network, and in particular the issues that are to be addressed. Decisions relate to the time and day-type, ie weekday morning peak, inter-peak, afternoon peak, off-peak, weekend etc. and the length of these periods.

Base and forecast years: the base and forecast years are commonly selected based around key milestones in the study/project programme. For an economic assessment, as forecast traffic volumes and economic benefits stream are typically non-linear and evaluation periods can be lengthy it is common to have a base year and at least two forecast years.

Zone system design: there are a range of considerations relating to the detail and design of the zone system. These include aspects relating to the form of the network, detail and coverage of the study area, the demographic and economic make-up of the study area, any requirements to represent homogenous areas of land-use/physically segregated areas, car-parks, larger trip generators, access from zones onto key links etc. Consideration should also be given to how zones load traffic onto the network.

Network detail: some model forms can support high levels of network detail, ie the representation of individual time-dependent transport elements, road network links etc. Modelling excessive network detail can lead to issues in the application of the model, eg lack of clear outcomes, stability issues, inefficiencies. Consideration may need to be given to the level of network detail required to satisfy the objectives of the study and the observed data that is required/available to support network detail.

Assignment methodology and key settings: various modelling approaches include a range of assignment (route choice) methods. Consideration may need to be given on how re-assignment is to be achieved and what settings are likely in advance of the model development process. Decisions on this element are usually linked with the definition of the study area and necessary network detail.

Method for developing base matrices and forecast matrices: how the base year demands are to be developed is a key consideration in the development of the majority of models. This needs to be carefully considered, particularly in relation to the available data, potential input from other higher tier models, and the application of any base year demand refinement processes, eg matrix estimation. How the base year demands are developed is likely to inform and influence the method for producing forecast matrices, therefore decisions relating to the development of demands should consider both the base year and forecast method together.

Regional model cordoning: there will be occasions where a model is built by cordoning down from a larger regional model built for another purpose. This will require the analyst, peer reviewer and Transport Agency modelling specialist to determine whether the specification of the larger regional model will allow a lower level model to be built that fulfils the model's purpose. Consideration should cover the impacts of applying the regional model 'as is' and/or the benefit/value-for-money in developing a project model (particularly using a model form which offers additional features relative to the regional model) comparative to a regional model cordon. If the new model's purpose can be achieved, its specification realised, and the benefits of this approach are reasonable then this cordoning down from the larger regional model can proceed.

Approach to calibration and validation: as described in several sections in this guide, separate observed datasets may be maintained for strict model calibration and validation. Consideration may need to be given to the requirements to achieve calibration levels balanced against the predictive objectives of the model and producing validation comparisons. See sections 3.3 and 7.1 and for further discussion.

Observed data requirements: a vital aspect to consider is the on-street, observed and other forms of traffic data required to support the model development, notably the key elements described above. This is likely to be an iterative or interconnected process. In some instances the availability of data may inform the specification, eg the availability of a regional model is likely to strongly influence zoning and demand development methods for a project model. In other circumstances, the project requirements may dictate a specification which requires specific data to be collected, eg a project investigating traffic signal strategies is likely to require collection of detailed on-street signal data.

Sensibility tests: consideration should be given to the need to carry out sensibility/sensitivity tests on aspects of the base model specification and on the forecast demands.

The model peer reviewer and the client will ideally need to agree and sign off the model specification, including key aspects such as those described above, before any substantive work on the model development commences.

APPENDIX C DEVELOPMENT OF BASE YEAR TRIP MATRICES (PROJECT MODELS)

BACKGROUND

The development of base year demand information is likely to be the most important and resource intensive stage of the base model development, except in the simplistic application of modelling a single intersection. The importance of the demand data, and particularly the observed data that is used as input to this process, may sometimes be overlooked and too much focus or emphasis placed on the model network, settings and parameters, and/or software. Developing robust demands from suitable data is likely to be crucial in obtaining an appropriate level of calibration and applying the model to scenario testing.

For large projects it is generally appropriate to have separate trip matrices for light vehicles and heavy vehicles. This assists in separately identifying benefits for freight which is an important strategic driver and allows more rigor in the model development and validation.

REGIONAL MODEL INPUTS

Where there is a regional (eg validated multi stage strategic) model available a project model will commonly be required to integrate with this model. The regional model will provide key inputs to the project model such as travel pattern data, travel demand and forecasting data, and reference to regional planning (land-use development etc.) and future transport scheme objectives.

A key step in the integration between the two models is the transformation of trip numbers from matrix cells based on the likely coarser zones of the regional model to matrix cells based on the likely finer zones of the project model. A methodology will need to be developed to carry out this process and the Analyst will need to demonstrate that this trip transformation process is robust.

TRIP GENERATION AND DISTRIBUTION

Origin-destination inputs will need to be developed through a purpose-built process where there is no overarching regional model available. This could involve a wide range of techniques such as the collection and collation of observed trip making data (number plate matching, road side interviews, person/household surveys, Bluetooth data, GPS data etc), simple gravity modelling techniques in combination with land-use data / study area knowledge, through to simpler trip generation and distribution analysis. Such processes will require the analyst to demonstrate modelled trip volumes and trip length distribution satisfactorily match observed data (ref matrix comparisons section 8.4).

The practice of borrowing trip rates from other geographical areas has many pitfalls and if used should be carefully considered. If this approach is used then it is important that the analyst, peer reviewer and client (Transport Agency modelling specialist) demonstrate that the trip generation rates used are representative of the study area being considered. Large-scale projects or activities are such that the risks to Transport Agency's business are large enough that investment in determining trip rates corresponding to the geographical location is warranted.

If the project or activity that might arise out of any investigation with this model is significant this may lead to undertaking a variable trip matrix analysis. This is discussed further in the next section. However, the requirement to undertake a variable trip matrix analysis may affect the way base year matrices are constructed and should be considered in order to ensure the development of forecast trip matrices as efficient as possible.

APPLICATION OF MATRIX REFINEMENT TECHNIQUES

The extensive use of a manual and/or iterative matrix refinement process (eg matrix estimation) as a means to 'correct' matrices is not considered good practice. All efforts should be applied to ensure that the input data (ie data used to develop prior or 'starter' matrices) is robust before applying techniques such as matrix estimation.

In actuality, a process such as matrix estimation is best employed as an error checking method; to check, identify, and correct errors in the OD inputs, the observed count data, and the model network/assignment. For example, this could include identification of errors and optimisation of trip generation and distribution inputs when employing a purpose-built OD input process.

An imperfect trip matrix, ie one that potentially reaches lower levels of observed count data calibration and or validation, can be useful in a forecasting model if the reasons for the limitations in the matrix are understood. The balance of achieving higher levels of calibration against forecasting requirements is discussed in section 8.1.

If matrix estimation (or similar) is used then care must be taken to ensure that the resulting trip matrices maintain a strong correspondence with observed travel patterns. This provides a more robust basis for carrying out forecasting, applying the model to test scenarios, and the subsequent project outcomes. Methods to control and limit the manipulation of the OD inputs by ME include:

- limiting the number of internal matrix estimation iterations that are run, eg less than 10
- use of constraints to control individual cell, row, and/or column totals
- controlling the difference in the total number of trips when comparing the prior matrix and the final matrix
- ensuring that the change in the trip length distributions for the prior and final matrix are small – examination of the CR and ND statistics can be useful
- either ensuring comprehensive network coverage by traffic count data used as targets (to ensure ME cell manipulation is limited to OD trip volumes which correspond with observed traffic volumes), and/or maintenance of a robust count set for validation (ie traffic count data not used as an input to ME)
- ensuring that the matrix estimation process does not rely on fixing counts only but also includes some zone total productions and attractions as well
- ensuring that the changes in the prior matrix and final matrix cells are plausible in terms of the known land use distribution.

APPENDIX D OD FORECASTING (PROJECT MODELS)

APPROACH

Future year traffic forecasting is not an exact science and there are complex underlying mechanisms that inherently produce uncertainty in the forecasts. This being the case it is unreasonable to take a single view of the future as it implies a degree of certainty that is not real and does not provide information on uncertainty nor the drivers of that uncertainty.

It is recommended that a small number of plausible future scenarios are developed that represent the likely range of key variables such as population, employment and other. This will contribute to the understanding of the likely range of modelling outputs and the key variables that influence the range of modelling outputs.

A key attribute of the forecast scenarios developed is that they must be plausible. There appears to be a prevalence of an underlying 'optimism bias' with forecasts throughout the country with various jurisdictions assuming that their jurisdiction will attract high levels of growth in population, employment, economic activity and other. This inflates the 'issues' and the requirements for investment in these jurisdictions and when summed across the country implies unrealistic national growth and required infrastructure investment. In this respect the analyst, peer reviewer and client (Transport Agency specialist) must be satisfied that the growth implied by the future scenarios must be realistic in terms of the overall regional and national context.

This will also require careful consideration of underlying demand for activities. Just because land is available to provide for various activities does not mean there is a demand for these activities nor that the use of the land in the way proposed is commercially viable. These projections will require evidence of an underlying demand and that the land costs or rents involved can make the land development economically viable.

In particular, the basis for population and demographic forecasts should not be beyond the range provided by the Statistics New Zealand projections determined by half way between the low and medium growth and half way between the medium growth and high growth projections. A similar range should be applied to employment and regional GDP forecasts. More detailed guidance is available from the Transport Agency's report on traffic growth forecasting.

VARIABLE TRIP MATRICES

Large projects have the potential to change demand particularly if there is a significant change to generalised cost. Changes to demand can result from trip distribution changes, mode choice changes, trip retiming, relief of trip suppression and other mechanisms. Accordingly it may be important that trip matrices and forecast data corresponding to other transport modes are available from a larger regional multi modal model and used.

There is potential to change demand where improvements are proposed:

- Where there are high levels of congestion.
- Where the proposal achieves a significant reduction in travel time and/or journey length.
- Where there is an adjacent passenger transport service carrying a significant mode share of trip makers.

Careful consideration of these matters indicates that these conditions may be more applicable or less applicable depending on whether the proposed project or activities is located in an urban or rural area.

Variable trip matrices allow the number of additional trips induced by a scheme to be estimated and to modify the change in generalised cost from the 'do minimum' scheme and the 'do something' scheme to account for these additional trips. The recommended thresholds for applying variable trip matrix analyses relate to the reduction in travel time savings per vehicle to the additional trips in a preliminary analysis. A 7 to 10% or greater reduction in travel time savings per vehicles warrants consideration of the application in variable trip matrices.

Variable trip matrices may be developed using a three or four-stage regional demand model or using elasticity methods.

CALCULATION OF FORECAST YEAR TRIP MATRICES

Forecast future year trip matrices can be calculated using an approved regional model or by using manual growth methods. The outputs of manual growth methods should be tested to see that they comply with the plausibility, underlying demand and economically viability requirements.

If a matrix refinement process (eg matrix estimation) has been used to develop base year matrices it is important that the process to calculate future year matrices has a rationale that allows the shape of the forecast matrices to be affected by network changes and anticipated land use development recognising that some range of plausible future land uses should be considered. A range of techniques are possible and several are discussed below.

A simple method is to apply the absolute difference between the relevant base year and forecast year regional model matrix (commonly a cordon) to the refined base year project model matrix. This guarantees that the land-use and network changes which produce changes in the forecast travel patterns (relative to base year travel patterns) are brought through into the forecast project model matrices. This method is likely to require some simple checks and adjustments, eg to ensure negative trips aren't produced in the forecast project model matrices.

Relative growth and furnessing (eg by furnessing the base year matrix to the trip-end growth predictions from the regional model) are two alternative methods. These approaches should be used with some care, notably it is undesirable to distort the predicted pattern of traffic growth (eg high growth in the northern region of the study area and low/no growth in the east) based on the distribution of trips in the base year project model matrices.

FORECAST MATRICES CHECKS AND SENSITIVITY TESTS

As noted above, it is generally important that project model forecast matrices incorporate the predicted changes in future travel patterns. It is recommended that the pattern of sector-to-sector growth predicted from the regional model is checked against the sector-to-sector growth pattern in the project model. For simplicity, the raw regional model cordon matrices could be used, large/distinct sectors defined, and the percentage growth sector-to-sector compared.

It is useful to undertake sensibility testing with forecast future year trip matrices. A sample of model runs should be undertaken to ensure the outputs are reasonable. Where counter intuitive outcomes are observed these should be investigated to ensure that the model is behaving appropriately or there is a rational explanation for the outcome. Route choices, locations of high congestion, long queue lengths and others are useful to investigate for this purpose. Of particular interest is to see whether there is a significant deterioration of performance in areas of the network.

APPENDIX E MODEL CODING AND SETTINGS (PROJECT MODELS)

MODEL LINK PARAMETERS

In urban areas intersections almost always determine the network performance as they are usually the determinant of capacity. Intersection modelling, and the techniques/parameters which influence intersection performance, is carried out in a variety of ways depending on the form, size, and requirements of the model. It is therefore difficult to provide generalised advice/guidance on approaches to modelling intersections. However, there may be occasions where intersection capacity exceeds link capacity on some links in urban areas.

For macroscopic model forms speed flow curves are a key descriptor of the network and give the vehicle response to increasing traffic volumes on links.

Microscopic (microsimulation) models do not use these descriptors, speed flow curves are an output (prediction) of these models. Vehicle response to increasing traffic volumes on links is based on the physical network characteristics (link length, no. of lanes etc), the road rules (speed limit, keep left etc) and three areas of driver behaviour (distance to the vehicle ahead or gap acceptance, lane choice, and target speed).

In rural areas link capacity is often the key determinant of network performance and the correct accounting for capacity and speed response to traffic volume can be important.

Further, the performance of heavy vehicles on the network should be accounted for correctly. In particular gradient and curve radii will impact on speed profile.

Ideally free flow speeds and capacity (ie observed speed/flow data) should be measured to determine whether the defined speed flow curves and behaviour distributions are appropriate. Care needs to be taken when using observed point data (eg loops) to establish speed/flow relationships as the shape and form of speed flow curves varies through areas of the network. In high-flow, high-speed corridors speed- flow relationships can vary reasonably significant over relatively short distances depending on how bottlenecks function. However, for macroscopic models the *Highway capacity manual* provides procedures for calculating free flow speeds and effective capacity to account for a range of variables including road cross-section, curve radii, numbers of side accesses, terrain, gradient and other variables. For microscopic models, the driver behavioural distributions should have been verified through historical application of the software across a range of studies; these distributions can be checked where data is available.

A good estimation of travel speed is important as many of the Transport Agency's levels of service aspirations for state highways are framed in terms of speed.

TRAVEL TIME AND DISTANCE WEIGHTING IN THE GENERALISED COST FUNCTION

Typical values for the time to distance weighting are 1 : 0.3 and 1 : 0.6 for light and heavy vehicles respectively. However, these values are guides and may be subject to variations over time or whether the model area is urban or rural. It is possible to optimise the value of the distance weighting relative to the time weighting to improve the overall goodness of fit of the model. However, such a process should not be treated as purely an optimisation exercise. Departures from these suggested values need to have a rational explanation as to why such a departure is sensible in terms of the particular environment surrounding the modelling exercise.

Regional models may provide a good source of these weightings for project models developed within the same area.

Outliers from the assignment process should be investigated to ensure the route choices in the model are sensible.

A sensitivity test which varies the distance weighting relative to the time weighting should be undertaken. This test should confirm that the chosen weightings are reasonable and the magnitude of the response to varying the weightings is reasonable.

MODEL ADJUSTMENTS AND DEVELOPMENT

To develop the model, correct errors, and improve the fit of modelled predictions against observed data, (generally referred to as 'modelled calibration') the analyst can employ a range of processes. It is important to recognise that calibration is not purely a mathematical exercise aimed at optimal fit. It is important that the processes used and any adjustments made to parameters are sensible in terms of the model context and it's intended application and carried out to a degree of consistency throughout the study area (so that parameters/coding can be applied in a similar fashion to new infrastructure/scenarios in test models).

The analyst, peer reviewer and the Transport Agency modelling specialist should agree that the calibration adjustments are reasonable and sensible.

As a matter of philosophy the emphasis of the calibration should be to ensure a good fit for those movements that are critical to the model's purpose in particular. For example if it is intended to investigate the impacts of an expressway running north-south then the north-south movements and other flows that contribute to that movement need to be well represented in the model. More generally, the model should be specified and calibrated to represent the mechanisms which it is expected to reflect, eg mode choice where modal share is an important objective. Inadequacies in these components may not be apparent from the model's fit to the base year observed traffic data.

It should be recognised that count and travel time data inherently have a statistical distribution and a degree of randomness. This means that it is not reasonable to expect a perfect match of observed and modelled data. However, outliers should be investigated to ensure there are not modelling issues and if they can be identified as outliers it is reasonable to expect that they can be removed from data comparisons. This would require documentation and supporting information in model reports.

NZMUGS
New Zealand Modelling User Group

The NZ Transport Agency acknowledges and thanks the New Zealand Modelling User Group (NZMUGS) for their contribution and support of this document.

If you have further queries, call our contact centre on 0800 699 000 or write to us:

NZ Transport Agency
Private Bag 6995
Wellington 6141

This publication is also available on NZ Transport Agency's website at **www.nzta.gov.nz**

NZ TRANSPORT AGENCY
WAKA KOTAHI

New Zealand Government

14-120 April 2019