GRANT SMITH
Principal Consultant
TDG
grant.smith@tdg.co.nz

DAVE HUNTER
Senior Principal Transportation Engineer
TDG
dave.hunter@tdg.co.nz

How good are our models?

In 1993, the Cities of Wollongong and Shellharbour, in conjunction with the Roads and Traffic Authority built a three step transportation model covering the two local Government areas. The base year for that first model was 1991, but there were no future forecasts at that time.

The model was updated using the 1996 census data in 1998, with future years of 2006 and 2016. The 2001 census data was used in the 2004 update. The most recent update occurred in 2010 using the 2006 census data with futures of 2011/21/26 and 2036.

At each update improvements were made in terms of the modelling technology, the detail included in the road network, and the size of the zones, but the basic model form was similar in each case.

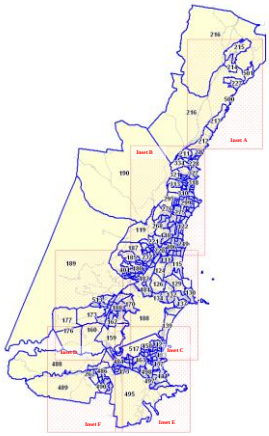
The paper looks at the performance of each of the three model updates, and the ability of the model to predict the traffic flows in each of the 'future' years that have already been reached. The effects of changing modelling technology (by using forward and back projections) and the accuracy of the land use forecasts are explored.

The paper concludes with a commentary on the forecasting ability of the models, with particular emphasis on the accuracy of the traffic flow forecasts, once corrections for land use forecasts have been made.

1. Introduction and Background

Wollongong City Council and Shellharbour City Council are two neighbouring local authorities some 80km to the south of Sydney on the eastern seaboard. Although there are two councils, the cities are one contiguous urban area.

The first transportation model in the area was commissioned by the two Councils in conjunction with the (then) Roads and Traffic Authority (RTA), now the Roads and Maritime Services (RMS) of Transport for New South Wales. That first model used data from the 1991 census, but was validated to 1994. There were no future year forecasts prepared using that model.



In 1993, when that first transportation model was built, it was logical for the model to encompass both Councils and included the full urban areas of both, as shown on the adjoining diagram. That model area has remained reasonably constant over most versions of the model since then, with the exception of 2001 which extended the network through to Central Sydney to

capture the effects of commuters to Sydney. The constant model area has the advantage that geographic differences have only a small effect when comparing model inputs and outputs.

The first update of the model took place in 1998, using the 1996 census data to provide the land use inputs. The first model had 196 zones, and the update expanded this to 376 zones.

After successful validation of that version of the model, a future year of 2016 was prepared by using land use projections, and an intermediate year of 2006 prepared by interpolation of the resulting demand matrices.

In 2003, the 1996 model was expanded to 470 zones and that model was updated using the 2001 Census as the base in 2004. A future year of 2026 was projected.

In 2010, the model was updated using the 2006 census and expanded to 550 zones, largely as a result of adding more detail into the Port. Future years of 2011, 2016, 2021, 2026 and 2036 were projected.

In summary, there is a 25 year history of the Wollongong/Shellharbour model, and several of the forecast years have come and gone. As a result, there is an opportunity to check how good the forecasts were.

2. Model form

All of the models were based on the traditional three step process of generation, distribution and assignment – by definition a vehicle driver model. In the 2001 update, a four step model was also introduced, as was an experimental dynamic assignment model form. Neither of these was progressed in the 2006 update. For the purposes of this paper, only the three step model has been considered.

The key component in terms of comparing the forecasting ability of the models is the generation step. The models all used a household category model for generation (which controls the total number of trips) and regression derived attraction models. In the 1991, 1996 and 2001 models, the generation model was a 3x3 category model based on employees per household and cars per household. The 1991 model used trips rates derived from New Zealand models, but all subsequent versions used trip rates calibrated from the Sydney household travel survey data.

In 2006, the category model was changed into a 5x4 category model of persons per household and cars per household.

Perhaps the most significant change in the model technology over the years has been in the assignment model. The early models (1991 and the first 1996 model) used an incremental[1] assignment with intersection delays calculated at each approach. In the second 1996 model, and all subsequent versions, the intersection delays have been calculated by movement. Signal cycle times and phase splits are calculated for each intersection according to the demand at each assignment increment.

[1] An incremental assignment was used because intersection delays are calculated using opposing flows. An equilibrium assignment cannot be used in this context as the mathematics requires a monotonically increasing function of link flow only – not the opposing volumes.

3. Base data

3.1 Households and Jobs

The two key land use indicators are households and jobs. Table 1 contains the household numbers used in each of the models and Table 2 the number of jobs.

Table 1: Households

	Number of Households						
Model	**1991**	**1996**	**2001**	**2006**	**2011**	**2016**	**2026**
1991	81,063						
1996 (376 zone)		83,092	*88,741*[1]	*101,875*[2]		120,658	
1996 (470 zone)		85,038	88,741				118,693
2001 (470 zone)			88,741				
2006 (550 zone)				92,609	98,661	105,613	121,116
2011 (Census)					96,354		

Notes: [1] Derived from the 2001 ABS census data. [2] Interpolated between 1996 and 2016

Table 2: Jobs

	Number of Jobs						
Model	**1991**	**1996**	**2001**	**2006**	**2011**	**2016**	**2026**
1991	62,266						
1996 (376 zone)		72,684	*76,935*[1]	*84,092*[2]		95,500	
1996 (470 zone)		68,040	77,168				101,918
2001 (470 zone)			77,168				
2006 (550 zone)				77,696	81,620	89,930	104,010
2011 (Census)					85,307		

Notes: [1] Derived from the 2001 ABS census data. [2] Interpolated between 1996 and 2016

There is an anomaly between the two 1996 models in that they both have been sourced from census data, probably with a different definition of ‘Household” but that difference was not able to be resolved when preparing this paper.

Clearly the land use forecasts made for the 1996 model were high for 2016, while those made for the 2006 model look to be about right, at least for households. It appears that jobs have been increasing more rapidly than expected but these are used for the attraction equations and do not affect the overall generation.

3.2 Trips

The land use, combined with the generation models give the number of trips. While the models cover the morning peak, inter peak and evening peak periods, only the morning peak is included in this paper. All of the models have a one hour assignment, although the

generation is normally two hours. Table 3 gives the total trips and trips per household for each of the models, where the trips are those made by all vehicles including Goods Vehicles

Table 3: Morning peak hour Trips and trip rates (trips per household)

	Trips and Trip Rates						
Model	**1991**	**1996**	**2001**	**2006**	**2011**	**2016**	**2026**
1991	51,702 (0.64)						
1996 (376 zone)		56,821 (0.68)	*59,883*[1] *(0.67)*	*70,092*[2] *(0.69)*		83,370 (0.69)	
1996 (470 zone)		52,482 (0.62)	54,681 (0.62)				76,631 (0.65)
2001 (470 zone)			54,681 (0.62)				
2006 (550 zone)				57,842 (0.62)	60,981 (0.62)	66,037 (0.63)	76,851 (0.63)
2011 (Census)					65,501 (0.68)		

Notes: [1] Derived from the 2001 ABS census data. [2] Interpolated between 1996 and 2016

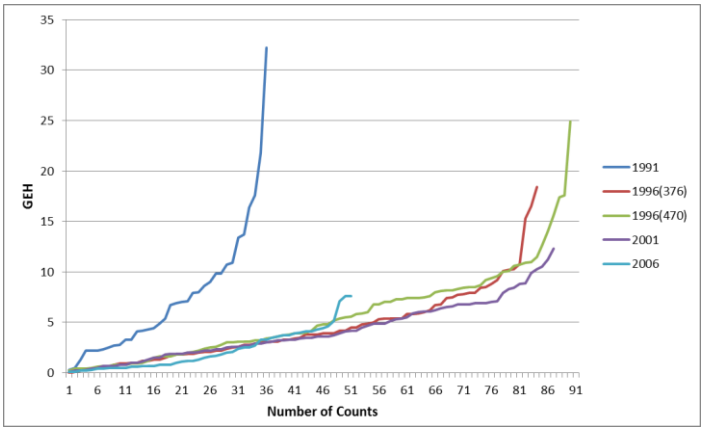
With the exception of the 1996 (376 zone) model, there is a remarkable degree of consistency in the trip rates, even though the generation models are very different.

4. Base Year Validation

Before discussing validation it is important to note that there is **no** matrix estimation used in any version of the model.

The first version of the model with a base year of 1991 was validated against 12 sets of directional counts (screenlines) comprising 36 comparisons when two way counts are included as well. In the interests of brevity, these have not been tabulated here. In summary, 17 of the 36 (47%) have a GEH less than 5.0, 22 out of 36 (61%) are less than 7.5, 28 (78%) are less than 10, and 30 of the 36 (83%) are less than 12.0.

The GEH values are plotted as the blue line on the graph.



These are not as good as would be expected today, but it should be appreciated that this was the first version and the traffic count programme had not been set up for model validation.

The 1996 (376 zone) version had

considerably more counts with 28 sets of counts comprising 84 comparisons. Of these, 55 (65%) have a GEH less than 5, 69(82%) less than 7.5 and 77 (92%) less than 10. Only 3 (4%) were greater than 12.0. this is shown as the red line on the graph

The 1996 (470) zone model was intended to fine up the zone system, and 30 sets of counts were available. Of the 90 comparisons, 74 (52%) have a GEH less than 5.0, 64 (71%) are less than 7.5, and 78 (87%) are less than 10.0. Although the validation should have improved with the finer zone system, it is fair to say that it is similar, but slightly worse. It is shown as the green line on the graph.

The 2001 validation is considerably better. There are 29 locations with 87 comparisons. Of these 58 (67%) have a GEH less than 5.0, 77 (89%) have a GEH less than 7.5 and 83 (95%) have a GEH less than 10. Only 1 of the 87 is greater than 12.0. It appears as the purple line on the graph

In 2006 the screenlines had become better defined, and the traffic counting programme had been reviewed so that the locations better matched the screenlines. There were 17 screenlines with 51 comparisons. Of these 47 (92%) were less than 5.0, and all (100%) were less than 7.6. as shown on the Cyan line on the graph.

Thus as the model matured, the trip generation module altered, and with better data available for calibration and validation checks, the performance of the base year models has improved markedly.

5. Comparisons available

5.1 Land use scenarios

As the models stand, there are only three land uses (years) that are comparable.

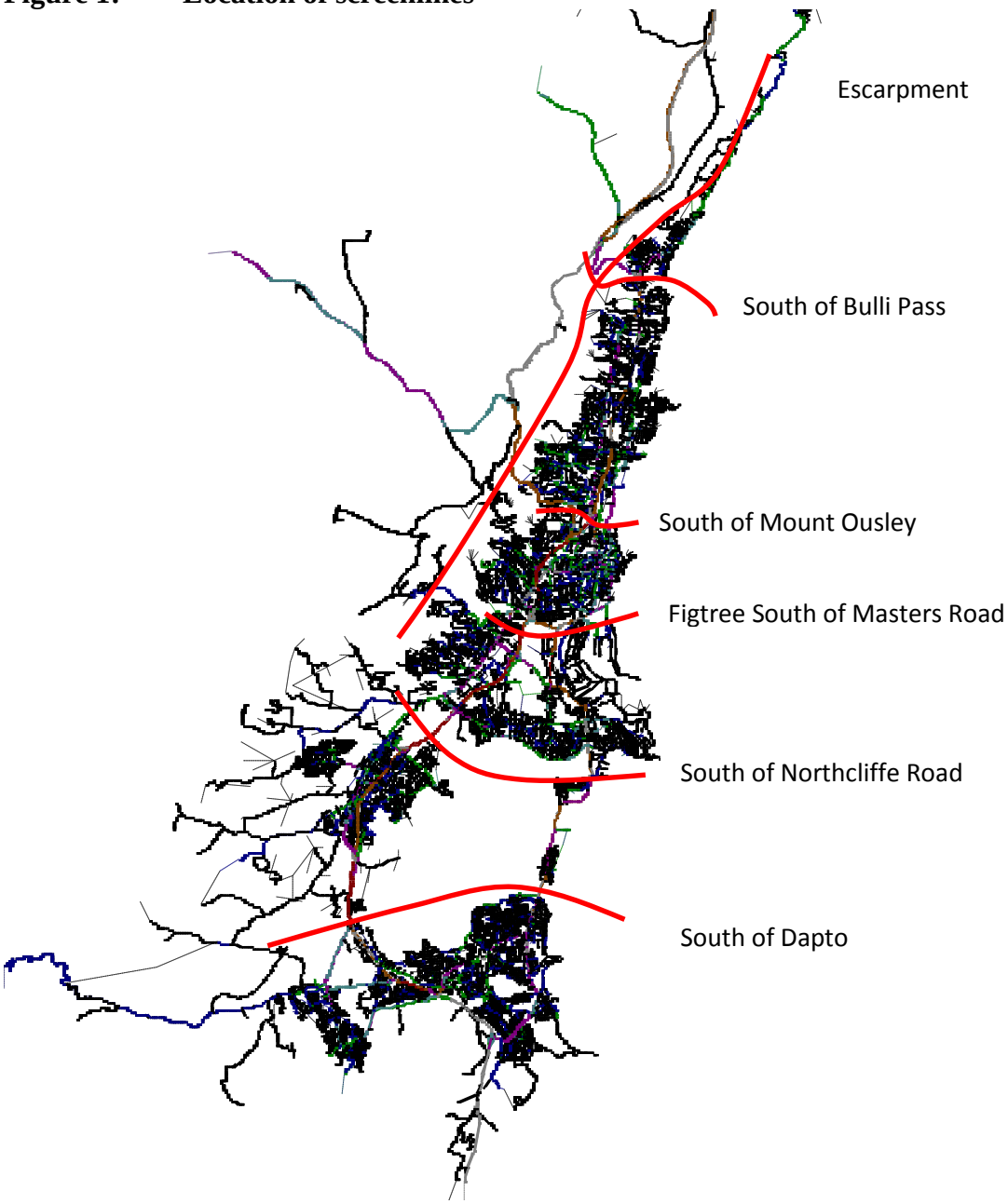
- The first is the 1996 forecast of 2006 against the 2006 base model
- The second is a comparison of the 1996 model with different zone systems
- The third is a comparison of the 2026 forecast from the 1996 (470 zone) base and the 2006 base. This comparison has not been done.

In addition to these, for the purposes of this paper, a 2011 land use file was prepared for each of the five models, and run through the models. The comparison of these effectively eliminates the land use forecasting issue, and enables testing of the central thesis that provided the land use forecasts are accurate, then the models will accurately forecast traffic flow.

In order for these comparisons to be made, a set of true screenlines were developed – ‘true’ in the sense that all roads on the screenline were included. The comparisons were ‘model against model’, rather than ‘model against count’ as the issue is the performance of the models with respect to each other. There were five screen lines running west – east, a screen that runs north –south on the escarpment, and a screen that included all of the roads crossing the railway line.

The screenlines are shown on Figure 1, with the exception of the railway screenline that – runs north/south through the centre of the city.

Figure 1: Location of screenlines



5.2 2006 – Forecast from the 1996 model against the 2006 validated base model

The 1996 base model forecast land for 2016, and the 2006 model was a linear interpolation between them, where half of the growth was added to the 1996 demands. Effectively, that means the 2006 land use sits halfway between 1996 and 2016 in the 1996 model. Table 4

summarises the difference between the 2006 models.

Table 4: 2006 Model year comparisons

2006 Comparisons			
	2006 Model	**1996 Model**	**Difference**
Households	92,609	101,875	+10%
Jobs	77,696	82,092	+6%
Trips	57,842	70,092	+21%
Screenlines			
South Bulli Pass	4,926	5,678	+15.3%
Escarpment	4,970	5,636	+13.4%
South of Mt Ousley	11,059	10,081	+9.7%
Figtree South of Masters	12,862	14,634	+13.8%
South of Northcliffe	9,981	12,987	+30.1%
South of Dapto	7,423	8,645	+16.4%
Railway	35,143	40,736	+15.9%

The 2006 forecast from the 1996 model 10-20% high with a screenline just south of Northcliffe Road 30% high. These are consistent with the higher number of trips – in part caused by the overestimate of land use activity, and in part by the change in the generation model.

5.3 1996 – The 376 zone system compared with the 470 zone system

As noted above there were two models validated at 1996, one using a 376 zone system and the other with the zones disaggregated to for 470 zones. There was a significant difference in the assignment with the 376 zone system using approach based delays at the intersections, and the 470 zone system using delays calculated for each movement. Table 5 summarises the difference between the 1996 model.

Table 5: 1996 Model year comparisons

1996 Comparisons			
	470 zone model	**376 zone model**	**Difference**
Households	85,038	83,092	+2.3%
Jobs	68,040	72,684	-11.2%
Trips	55,784	56,889	-2.0%
Trip length (minutes)	12.6mins	11.2mins	+12.5%
Screenlines			
South Bulli Pass	4,086	4,430	-8.4%

Escarpment	4,037	4,175	-3.4%
South of Mt Ousley	8,194	9,130	-11.4%
Figtree South of Masters	11,374	11,439	-0.6%
South of Northcliffe	9,789	9,332	+4.7%
South of Dapto	7,159	6,761	+5.6%
Railway	30,276	30,380	-0.3%

Thea are similar, except for south of Mt Ousley, and are largely explained by the longer trip time in the 470 zone system as the intersections have more detailed modelling.

5.4 2011 – Forecast from the 2006 model against Census Land use data

As noted above, the 2006 model was well validated, and is only five years away from the 2011 census. The first comparison that was made used the Census data against the 2011 forecast made in 2010. Table 6 summarises the difference between the 2011 model.

Table 6: 2011 Model comparisons at 2011

2011 Comparisons			
	Forecast land use	**Census land use**	**Difference**
Households	98,661	96,354	2.3%
Jobs	81,620	85,307	-4.5%
Cars	142,579	158,524	-11.2%
Trips	60,981	65,501	-7.4%
Trip length (minutes)	11.70 mins	11.40 mins	2.5%
Screenlines			
South Bulli Pass	3,937	3,913	0.6%
Escarpment	5,568	5,735	-3.0%
South of Mt Ousley	10,616	11,069	-4.3%
Figtree South of Masters	13,331	13,219	0.8%
South of Northcliffe	10,455	10,358	0.9%
South of Dapto	7,962	8,077	-1.4%
Railway	35,769	37,145	-3.8%

Although the forecast number of houses was slightly high, the underestimate of car ownership led to a net reduction in trips. This was at a time when the conventional planning wisdom was that increasing car ownership could not continue. Clearly, the census data showed the error of that view.

However, the longer trip length has a countervailing effect, which meant that the final flows from the two land use patterns were very similar.

6. Forecasts from the five models using the ABS 2011 land use

6.1 Forming the model land use files

This is the most comprehensive test that was possible. It involved taking the Australian Bureau of Statistics (ABS) Census Collector District data, coupled with the Bureau of Statistics and Analytics (BSA) land use data from the Sydney Strategic Transport Model (STM) and aggregating these into the four zone systems used by the five models.

The 2011 census data was obtained from ABS for the model area at SA1 level of detail for households and at SA2 level for employment. The methods used for determining household and employment model inputs were handled differently.

For household data, concordance tables from ABS relating the new SA1 zone system with the previously used CCD system were used to convert each of the model area SA1 household variable data into the old CCD system. Where SA1 zone boundaries crossed over multiple CCD areas, proportions of each SA1 were assigned to each CCD.

For employment, the SA2 employment data was first disaggregated into SA1 level zones using proportions based on previous modelled and BSA Travel Zone job distributions. The calculated SA1 jobs were then combined into equivalent CCD zones jobs based on the same concordance tables used for households.

Once all 2011 census data had been combined into equivalent CCD zones, the CCD to model zone concordance tables created for each of the 1991, 1996, 2001 and 2006 models were used to recreate the 2011 future land use for each of the models. For the 1991 and 1996 models, a number of CCDs needed to be combined to recreate the slightly older CCD boundaries used in the early models. It was also apparent that the 1991 model did not use the same definition of model area as the later models and a number of CCDs were not included in that model's land use. This resulted in a slightly lower number of households and jobs used for that model's 2011 future.

6.2 2011 using the five models

Table 7 shows the land use and screen line values for the five models.

Table 7: 2011 Land Use Model Comparisons

2011 land use applied to the various models					
	2006 model	**2001 model**	**1996 470 zone model**	**1996 376 zone model**	**1991 model**
Households	96,354	96,354	96,354	96,144	94,953
Jobs	85,307	85,311	85311	85,050	87,792
Cars	158,524	158,523	158,523	158,090	155,785
Trips	65,501	67,518	67,368	74,104	77,073
Trip length (minutes)	11.40 mins	13.95mins	13.81mins	11.64mins	11.66mins
Screenlines					

South Bulli Pass	5,609	5,616	5,468	5,844	4,242
Escarpment	5,735	5,655	5,675	5,649	3,210
South of Mt Ousley	11,069	10,696	10,722	11,493	11,048
Figtree South of Masters	13,219	12,507	12,167	13,009	12,443
South of Northcliffe	10,358	11,006	10,985	11,454	10,746
South of Dapto	8,077	8,540	8,523	8,911	8,108
Railway	37,145	37,744	37,303	38,416	40,607

With only minor exceptions as discussed above, all of the models were using essentially the same land use. However there are significant differences in trip making by the 1991 and the 1996 (376 zone) models when compared against the later models. Taking the 2006 model as the benchmark, the 1991 model has some 18% more trips, with car ownership forecast to be about 2% lower. The trip generation category model changed between the 1991 and 1996 models, and again significantly when the 2006 model was built.

In spite of this, at the screenline level, all of the models assign similar volumes of traffic. The 1991 model is a little lower on the two northern screenlines (by some 25-45%). The railway screenline for the 1996 (376 zone) model and the 1991 model is high, but that is at the lower edge of the model and prior to extension of the expressway past Oak Flats, meaning that there were more vehicles crossing the rail at Lake Entrance Road. Most other screen lines are within a few percentage points relative to the 2006 model.

It should also be noted that the 1991 model did not include any provision for the Northern Distributor north of Bellambi Lane which may well have influenced the amount of trip making between the CBD and the northern suburbs.

At the next level of detail, the 2011 assignment from the 2006 model was taken as being the base against which the other models could be compared, and the statistical measures commonly used in validation applied to each of the screen lines. However, before looking at these results, the following points need to be borne in mind.

- The 1991 model network consisted of 7,839 links and 3,002 nodes. The assignment calculated delays at the approach level.
- The 1996, (376 zone) model had 6,009 links, and 2,375 nodes. For some reason, much of the detail in Shellharbour that was included in the 1991 model was removed in the 1996 model, hence the fewer links and nodes. It also calculated intersection delays at the approach level.
- The 1996, (470 zone) model had considerably more detail with 12, 569 links and 4,989 nodes. This was the first model that had intersection delays calculated at the movement level during the assignment.
- The 2001 model was similar to the 1996 (470 zone) model with 12,508 links and 4,972 nodes, with intersection delays at the movement level.
- The 2006 model network was completely rebuilt from the (then) RTA GIS road centreline data. As such it contained much more network detail – every road was included, and the Port Kembla area zoning was refined. It is on a different co-ordinate system than the earlier models – although that makes no difference to the model operation. It has 19,506 links and 7,790 nodes.

Figures 2 and 3 show the networks used in the 2006 model and the 1991 model, with the difference in detail clearly obvious.

The statistical measures used to compare the models at each screenline are :

- Total flow
- The percentage difference
- Correlation coefficient
- Per cent RMS
- R^2
- GEH – overall, and less than 7.0 for each screenline

For the purposes of this paper, only the morning peak flows in both directions combined have been reported.

Figure 2: 2006 model network

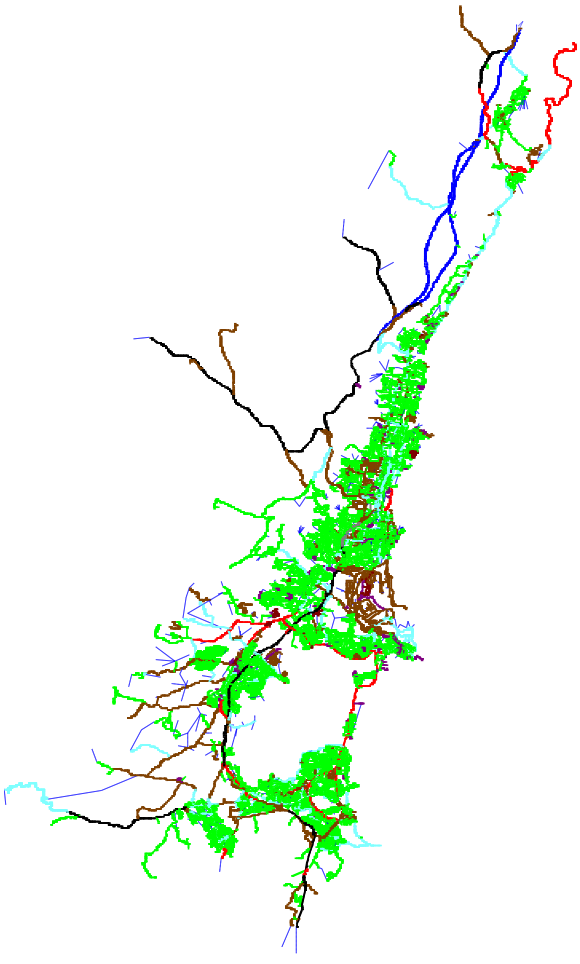


Figure 3: 1991 model network

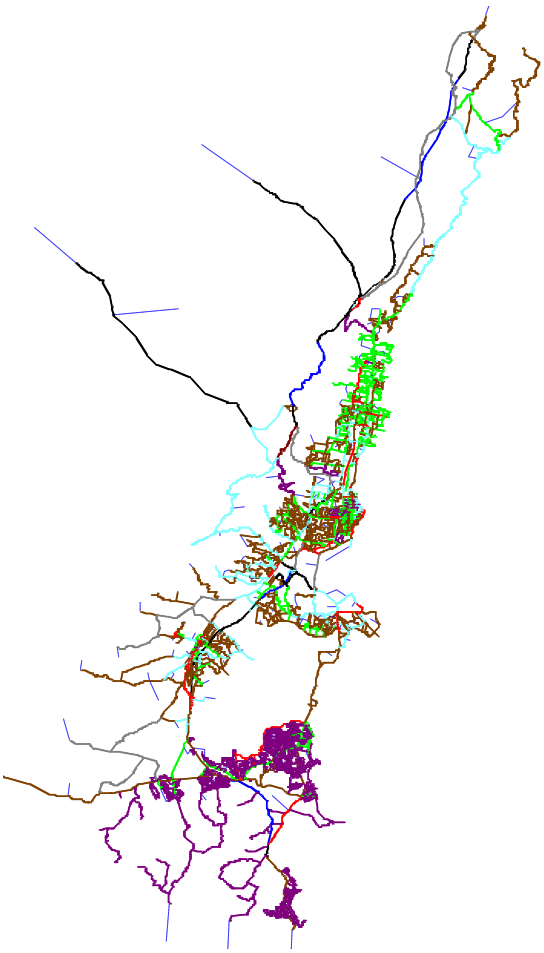


Table 8 shows the statistics for each screenline for each of the models. In general, with the exception of the 1991 model, the statistics fall within normally accepted Australian (RMS) and New Zealand (NZTA) modelling validation guidelines for strategic models. In many cases the correlation coefficients and r^2 statistic are not particularly relevant as there are too few points on the screenlines.

Table 8: Statistical measures by Screenline for each model at 2011

2011 forecast flows for each model relative to the 2006 assignment				
	2001 model	**1996 470 zone model**	**1996 376 zone model**	**1991 model**
South Bulli Pass				
Assigned flow	5,616	5,468	5,844	4,242
Percent difference	100.0%	97%	104%	76%
Correlation coefficient	1.00	1.00	1.00	1.00
% RMS	8.25	14.63	5.94	38.92
R^2	1.00	1.00	1.00	1.00
GEH overall	0.1	1.9	3.1	19.5
% GEH <7.0	100%	75%	100%	25%
Escarpment				
Assigned flow	5,655	5,675	5,649	3,210
Percent difference	99%	99%	99%	56%
Correlation coefficient	0.990	0.989	0.992	0.970
% RMS	20.84	22.35	18.0	87.96
R^2	0.980	0.978	0.984	0.940
GEH overall	1.1	0.8	1.1	37.8
% GEH <7.0	66.7%	83%	75%	67%
South of Mt Ousley				
Assigned flow	10,696	10,722	11,493	11,048
Percent difference	97%	97%	104%	100%
Correlation coefficient	0.974	0.995	0.939	0.838
% RMS	13.99	9.15	20.92	33.28
R^2	0.950	0.991	0.882	0.703
GEH overall	3.1	3.3	4.0	0.2
% GEH <7.0	80%	50%	60%	30%
Figtree South of Masters				
Assigned flow	12,507	12,167	13,009	12,443
Percent difference	106%	103%	110%	106%
Correlation coefficient	0.983	0.982	0.996	0.965
% RMS	23.09	20.46	26.08	30.63
R^2	0.966	0.965	0.992	0.913
GEH overall	6.6	3.6	11.1	6.1
% GEH <7.0	40%	56%	44%	11%

South of Northcliffe				
Assigned flow	11,006	10,985	11,454	10,746
Percent difference	106%	106%	111%	104%
Correlation coefficient	1.00	1.00	0.989	0.988
% RMS	16.63	18.23	31.02	36.56
R^2	0.999	1.00	0.979	0.975
GEH overall	6.3	6.1	10.5	3.8
% GEH <7.0	50%	50%	38%	63%
South of Dapto				
Assigned flow	8,540	8,523	8,911	8,108
Percent difference	106%	106%	110%	100%
Correlation coefficient	0.997	0.992	0.998	0.849
% RMS	14.67	17.59	19.38	43.6
R^2	0.994	0.983	0.996	0.720
GEH overall	5.1	4.9	9.0	.3
% GEH <7.0	50%	88%	88%	50%
Railway				
Assigned flow	37,744	37,303	38,416	40,607
Percent difference	102%	99%	103%	109%
Correlation coefficient	0.935	0.938	0.959	0.908
% RMS	43.97	42.22	34.98	51.61
R^2	0.874	0.879	0.919	0.825
GEH overall	3.1	1.9	6.5	17.6
% GEH <7.0	54%	57%	53%	49%

7. Discussion

A wise modeller once said “when you make a prediction it is always going to be wrong – the only question is when and by how much”. That comment is true here. All household forecasts are too high, even the ones made for 2011 for use by the 2006 model – and those were done in 2010. The 2011 forecast of jobs turned out to be low although the earlier forecasts were all high.

At the trip generation level, with the exception of the 1996 (376 zone) model, all of the models produced a morning peak trip rate of 0.62-0.63 trips, even though the generation models were significantly different. The 1996 model was about 10% higher.

As the models became more sophisticated, with the zone system made progressively finer, more detail in the network, and a very much improved assignment algorithm where all intersections were modelled in detail at each assignment increment, so did the base year validation. The 2006 validation was extremely good by any standard, to repeat, none of the models included any matrix estimation.

The Wollongong model has benefitted from its longevity – the first version was developed in 1994 (some 22 years ago) and the incremental changes made in each version after that have

led to the current model in which there is a high degree of confidence. It should be remembered that the model built in 1994 was, to some extent, constrained by the hardware limitations of the day. On that note, the 1991 model took over an hour to run when it was built. It now runs in just over a minute. The 2006 model (with twice as many zones nodes and links) takes 14 minutes to run.

When the model assignments are compared with a consistent set of land use data, there is good agreement among all of the five models at the overall screenline level.

However, the 1991 and to some extent the 1996 (376 zone) suffer from a coarser zone system and network such that the distribution of travel on the various roads in the screenline is not that good, as evidenced by the GEH values.

However all of the models from the 1996 (470 zone) and after have not only good agreement in total across each screen line, but also have good agreement on the roads that make up each screenline.

8. Conclusions

The conclusions that can be drawn from the analysis are:

- The finer the zone system and the more detail that can be included in the network, the better the validation. To some extent this needs to be balanced against available computer resources and run times.
- Forecasts of travel are only as good as the land use data that is used as input.
- If the land use data is accurate, then the forecasts of traffic flow will also be accurate. However the land use forecast is seldom accurate, and a prudent modeller will test a range of land use scenarios to test the robustness of the analysis.
- Provided that the base year validation is good, and that there is confidence in the land use forecast, then a high degree of confidence can be placed in the models as forecasting tools.