Department for Transport

NATIONAL TRANSPORT MODEL VERSION 2R PEER REVIEW

June 2020

Quality Control

Issue/revision	First issue	Revision 1	Revision 2	Revision 3
Remarks	Working document brought into WSP format	Additional material on QA added and minor post DfT changes.		
Date	22nd June 2020	29th June 2020		
Prepared by	John Bates Ian Williams	John Bates Ian Williams Andrew Stoneman		
Checked by	Andrew Stoneman	Chris Sanders		
Authorised by	Andrew Stoneman	Andrew Stoneman		
Project number	7006059	7006059		
Report number	1	2		

File reference	\\uk.wspgroup.com\central data\Projects\70006xxx\70006059 - T-TEAR NTM Futures\E Models and Drawings\05 Documentation

Contents

Tables

1. Background

1.1 The Department for Transport (DfT) has been developing a new spatially detailed National Transport Model, known as NTMv5, which has been reviewed separately. However, it was thought sensible in parallel to update the existing more aggregate model which has been in use for the last 19 years (referred to as NTMv2), and this update will be referred to as NTMv2R. As part of the NTMv2R project the DfT commissioned an independent Peer Review to be conducted by John Bates (John Bates Services), Ian Williams (Ian Williams Services) and Andrew Stoneman (WSP) which is the subject of this document.

1.2 It is the Peer Reviewers' understanding that the prime use of this revised model will be to produce the National Road Traffic Forecasts (and indeed it has already been used to produce the latest set for 2018 [RTF18]), though it is unclear whether this will in time be supplanted by the new model NTMv5.

1.3 In essence, the V2 model consists of a demand model ("PASS1") which takes trip ends from the DfT's National Trip End Model [NTEM] and allocates them to a set of aggregate area types by mode, together with an aggregate highway supply model ("FORGE") which, by means of an interface to convert travel demand on the highway to traffic ("mileage profiles"), makes use of Traffic Census data to model the speed response taking account of capacity. Other mode costs are provided, but do not change in response to different levels of demand.

1.4 For the purpose of the update, the Other Mode costs have been re-worked, the traffic database which supplies FORGE has been updated to include all 2015 hourly counts/ATC data & Annual Average Daily link Flow (AADF), minor changes have been made to FORGE relating mainly to the fuel cost functions, and the PASS1 model itself has been re-calibrated but without any change to its structure or segmentation. Note that the "Mileage Profiles" (which convert between PASS1 trips and FORGE traffic) have **not** been updated.

1.5 The available documentation has been produced in two reports on the DfT Website:

- NTM Future Model Development: NTMv2 recalibration – NTMv2R: Demand Model Implementation, April 2018;
- NTMv2R Demand Model Calibration and Validation, Feb 2018

In addition, a number of Technical Notes are available, and the DfT has carried out a programme of tests, related primarily to its forecasting ability under relatively extreme conditions ("stress tests") and its ability to reproduce past events by means of a "back-casting" exercise to 2003. This programme of tests is reported in the document:

- NTM Version 2R Analytical Review_DRAFT_v08, 12th June 2020

In an effort to make the documentation of the model and its update more accessible, we have produced an overview document “NTMv2R Overview of Structure and Update to 2015” to be published alongside this Peer Review.

This document constitutes a brief review of the PASS1 re-calibration process, followed by a more detailed assessment of the DfT’s stress tests, backcasting and quality assurance environment.

2. Review of the Re-calibration

2.1 The update of the PASS1 model requires:

- new demand data;
- updated cost/Level of service [LOS] data; and
- a procedure for re-calibrating the model parameters.

2.2 Once this is done, certain tests need to be carried out to provide some validation. In this section we review these elements.

2.3 As noted, the overall aim of the update was to maintain the structure and segmentation of the existing model. We consider this a reasonable approach.

Assessment of Demand data

2.4 The demand data has been provided from the National Travel Survey [NTS]. Given the need to retain adequate sample sizes for key categories, it was concluded that for distance band [DB] and destination choice for all purposes, it would be necessary to combine at least three NTS years (2012–2014) to get sufficient sample sizes. For the recalibration of mode choice, however, the aim was to get sample sizes of above 1,000 within each combination of purpose, household structure and distance band when all modes are aggregated, and this led to using nine years of data from 2006 to 2014. We are satisfied that these samples are adequate, and that the NTS data has been appropriately used.

2.5 At the implementation stage, the model makes use in the base year (2015) of the trip end data from NTEMv7.2. A comparison in Table 3.3 of the V2R Implementation Report with those in DfT forecasts for 2015 (based on NTEM 6.2) indicated a large minus 23% reduction of trips, even though the underlying mid-year population assumptions are little different. Checking the NTS through its table NTS0101, the observed overall trip rate has reduced by a maximum of -15% between the peak year 1996-98 (1,097 trips per person per annum), which would have been around the period from which the initial trip rates were calculated, to its minimum in the trough year 2015 with 934 trips. Subsequently, it has increased yearly, rising back up to 986 trips by 2018. Even taking the absolute maximum change, there is still a large gap between the observed -15% and the modelled -23% reduction. It would be helpful to have a more complete explanation of all of the source mechanisms generating the reduction of -23% in trips per person that are input to V2R.

Assessment of Cost data

2.6 Although the model structure is unchanged, considerable effort was required to update the inputs to 2015, and a number of data sources were consulted, including

NTS, mode-specific datasets (MOIRA for rail, National Express for coach), and the TAG databook for values of time and vehicle operating costs. Apart from the issue of the concessionary bus fares discussed below, we consider that this has been logical and comprehensive in its general approach.

2.7 Concessionary bus fares were applied in the model for children and for the 75+ age group. In reality the free bus pass starts at age 60, while in V2R it is only applied for those aged 75+. Accordingly, the majority of concessionary adult travellers are likely to be subsumed in the <75 group and so would be assumed in the model to pay full fares. The change in the age cut-off from 65 to 75 that has been implemented within V2R will have substantially increased the scale of this problem but unfortunately it is not clear from the results presented how large an error this simplification might introduce in estimated bus usage. An analysis using NTS data into the scale of the misclassification that this introduces would provide helpful guidance for interpretation of results from the demand model.

Assessment of Calibration and Realism tests

2.8 The calibration process was done by an iterative trial and error, with the general aims of achieving a good fit to the base NTS data as well as an acceptable elasticity response in line with recommended TAG realism tests. The step-wise process reported is not easy to assess, and we can only judge on the basis of the results.

2.9 It is noted in Section 2.5 of the V2R Calibration and Validation Report that, by agreement with DfT, the size terms (attractions), S, used for the distribution model have not been updated from V2, where they were defined as:

$$Size_{Iad}^{p} = -\frac{1}{\lambda_{A}^{p}} \ln\left[\frac{W_{Iad}^{p}}{Max_{I|adp}\left(W_{Iad}^{p}\right)}\right]$$

2.10 The notation has been changed to make it compatible with our overview document “National Transport Model Version 2 Revised”, and the derivation is given in Annex 1 of this document). While the re-calculation would certainly be tedious, the reason given (“the relative attraction of area types is unlikely to be significantly different from those in NTMv2.”) is not very convincing: quite significant relative changes in the zonal attraction variables are likely to have taken place between 1998 and 2015.

2.11 Aside from this, the five stages adopted within the calibration of the model provide an appropriate approach, though greater clarification would have been helpful for the following minor issues [references are to the V2R Calibration and Validation Report].

2.12 Page 11, bullet 3 at page end: "We also checked the predicted rail trips to London for NTMv2R and modified rail adjustment parameters to get a better match to NTS and census data." How has it been ascertained that it should be adjustments to the destination-end rail disutility that should utilised here, rather than explicit improvements to the representation of car performance and costs within Central and Inner London. This distinction could be important for future year forecasting and explorations of policy impacts.

2.13 Page 25 Table 19: the PT fare elasticities do indeed appear to be rather large for longer distance trips, particularly those for bus (coach). Page 26 states: "That is partly due to the NTMv2 model structure, which does not allow the implementation of a continuous cost damping approach." The contribution of this statement to this issue is not clear, because the parameter values differentiated by distance band [DB] included later in Appendix B indicate that a piecewise linear approximation to cost damping could be implemented directly, as was already noted in Section 3.5.1.

2.14 It would be helpful to understand to what extent the relatively low car journey time elasticity (see discussion below) results from the use of: (a) distance band alternative specific constants via the constraints; (b) cost damping effects; (c) other calibration issues.

2.15 In Appendix B, Table 31, the increase above that for shorter distance bands in the value of parameter $\boldsymbol{\lambda_A}$ for DB 11 for commuting (HBW) appears out of line with normal expectations for cost damping which would produce reductions in parameter values over increasing distance. It is unclear whether this is as a result of sampling errors (few commuting trips are >100 miles) or of shortcomings in the representation of modal supply characteristics. The same query arises for DB12 and DB13 for HBRec.

2.16 Chapter 4 presents some of the results at different stages of the calibration, in terms of differences in total trips and trip length, and modal distance distribution, by purpose and household type. It also presents the results of the “realism” tests, which are summarised below. Note that these are “first round” tests – no mention is made of possible supply effects via FORGE, so that (with the exception of Car time) they are not true realism tests as advised in TAG M2.1 §6.4.

Table 1 Modal Elasticities

Mode		**Fuel Price (Km) [Table 17]**	**PT Fares (trips) [Table 19]**	**Car time (trips) [Table 21]**
Walk		0.09	0.06	0.11
Cycle		0.35	0.24	0.49
Car Driver		–0.32	0.07	–0.20
Car Passenger		–0.19	0.17	–0.12
Bus		1.19	–0.98	0.60
Rail		1.26	–0.58	1.21
(combined PT)			–0.85	
TOTAL*			**0**	**0**

*Since the total number of trips is fixed, the trip-elasticities are constrained to be zero. There could still be an effect on trip-Km, but this has not been reported.

Table 2 Direct Elasticities for relevant mode, split by Purpose

Purpose	**Fuel Price (Km) [Table 18]**	**PT Fares (trips) [Table 20]**	**Car time (trips) [Table 22*]**
HB WORK	-0.14	-0.55	–0.27
HBEB	-0.12	-0.44	–0.27
HB EDU	-0.11	-0.43	–0.11
HB PB	-0.23	-1.00	–0.17
HB RVF	-0.38	-0.95	–0.15
HBHOLS	-1.08	-2.50	–0.30
NHB EB	-0.09	-0.41	–0.19
NHB OTHER	-0.48	-1.41	–0.22
ALL Purposes	**-0.32**	**-0.85**	**–0.20**

*For some reason the car time elasticities are not presented, but they have been approximately calculated based on other results in the Table.

2.17 In the Summary (§5), it is noted that “except for the bus fare elasticity test, which is slightly above the maximum limit recommended in TAG as a result of differences in the proportions of long-distance bus trips between the model and the TAG evidence.. [and given] the limitations of retaining the existing model structure, we believe that the calibrated model is fit for the purpose of modelling strategic policies on the roads network.”

2.18 In terms of the aims, it can be agreed that a model, based essentially on the early V2 PASS1 model, has been calibrated so as to reproduce the current base NTS data in terms of (sub-) mode split by distance, household type and purpose. While accepting that the elasticities are first round only, they appear to be generally in line with TAG recommendations.

3. Assessment in the light of the DfT Tests

3.1 The V2R model as reviewed above was used to provide the latest set of traffic forecasts (RTF 18). As noted, the DfT has carried out a programme of "stress tests" and a "back-casting" exercise to 2003, "designed to provide a wider understanding of the model's behaviour and methodologies". These are the subject of this section of the Review. We note that the current description is only in draft form, and we have already provided substantial comments on how the exposition could be improved.

The Stress Tests

3.2 The "stress tests" (§3 of the DfT document) are confined to car travel, so they do not report any response from freight vehicles (and other modes). In addition, the geographical scope is confined to England.

3.3 Included among the stress tests is a section (§3.3) entitled "Realism Tests Analysis". While the realism tests discussed above are carried out using the NTS data sample on which the re-calibration is based, the tests carried out by DfT make use of the implemented model which is driven by the NTEM v7.1 estimates of (personal) travel in the base year. Rather to our surprise, however, these turn out also to be "first round" tests: the impact of the changes introduced for the tests on demand is not further conveyed to FORGE – there is no feedback effect. We are informed that software constraints make it difficult to apply this in the base year.

3.4 Table 6 copies the outcome elasticities from the Calibration Report, before briefly describing how the realism tests were carried out on the implemented model. Although the resulting elasticities are not actually given in tabular form, they can be pulled off the various figures in the text, assuming the 10% increase in the underlying variable. They turn out to be remarkably similar to the values in Table 6: it may be assumed that any differences are essentially due to differences between NTEM and the NTS data used for calibration. While this similarity is encouraging, in that it suggests the implemented model is working as intended, it is unfortunate that the feedback effect is absent – typically this would be expected to reduce the outcome elasticities.

3.5 The stress tests have been carried out for two forecast years – 2030 and 2050. These tests relate partly to the impact of fuel price changes and partly to the impact of road capacity changes. The tests are carried out using the assumption, which was also a feature of the earlier V2 model, that Value of Time [VoT] will only rise half as fast as GDP/head: we discuss this below.

3.6 The fuel price tests involve changing the expected cost of fuel by ±10% in both forecast years and running the model to convergence. The elasticities of Car-Km are

–0.19 in 2030 and –0.15 in 2050 (these are based on the price increases; the results for price decreases are marginally smaller). The reason for the reduction over time is associated with the increased VoT, as we discuss below. But note that even the 2030 result is substantially lower than the "first round" elasticity for the base year realism tests of –0.32: this suggests that the capacity (supply) effects might bring down the true (base year) elasticity below the lower end of the TAG range (–0.25). This needs to be checked.

3.7 For both years, the variation in elasticity by journey purpose is plausible, with the highest values relating to the most discretionary purposes (Other), the lowest values for Employers' Business, and relatively low values for the mandatory purposes (Commute/Education).

3.8 VoTs for appraisal purposes are assumed to rise over time in line with GDP/head – thus an elasticity of 1 (TAG A1.3, paragraph 4.5.1). The guidance for **modelling** purposes is perhaps less clear, but it is standard procedure to make the same assumption. As has often been noted, this has the result that – given that it is usual to define generalised cost/utility in time units – cost elasticities will fall, and among other effects this implies a switch away from "active" modes.

3.9 As noted, NTMv2R makes the unconventional assumption that VoT will only rise half as fast as GDP/head. Thus some "damping" to the decline in the reduction in the fuel price elasticity over time is already present in the results reported above. Reverting to the more conventional assumption reduces the elasticities to –0.18 in 2030 and –0.13 in 2050 (again based on the price **increases**). From this it appears that the effects of the "damping" assumption appear relatively mild, indeed perhaps unexpectedly so.

3.10 A further experiment was carried out that scaled the increase in VoT relative to GDP/head upward by a factor of 2.5, rather than damping it. This was carried out for 2050 only, and had only a small effect on predicted traffic levels: the major response in this experiment is a large increase in bus distance travelled. The various charts and discussions then explain clearly the model mechanisms that have caused this to occur. In reality this response does not appear to accord with initial preconceptions or past trends. It would have been interesting to have had some further exploration of whether this response indicates shortcomings in the representation or the calibration of the bus mode within the demand model.

3.11 For the capacity tests, the capacity of rural and urban roads (both separately and for all roads together) was changed by $\pm$50% for the two forecast years, and the impact on total car traffic[1]and "congestion" (as measured by additional seconds per Km over and above free-flow conditions) was presented.

3.12 No description has been provided as to how the change in capacity has been implemented, but it may be noted that the total veh-Km at each FORGE "site" is distributed over different volume/capacity [V/C] bands in the base year (reflecting, among other things, variation by time of day), and the growth in traffic over time will change the distribution (though the way this is done is not very clear). Speeds are derived using relationships involving V/C ratios. Thus it can be assumed that the change in capacity has a direct effect on the distribution among V/C bands.

[1]While growth factors for LGVs and HGVs are part of the model system, demand responses as a result of changes in generalised cost are only modelled for cars.

3.13 Given the size of the modelled change, it can be expected that an increase in capacity will have a much smaller effect than a decrease, and this is borne out by the model results. In the test with the capacity increase applied to all roads, total traffic increases by 1.2% and 1.9% in 2030 and 2050 respectively, but for the capacity decrease, the traffic decreases are 16.2% and 51.5% respectively. Correspondingly, “congestion” falls by 37.4% and 40.1% for the capacity increase, but increases by 145.3% and 156.0% for the capacity decrease. Subsequent analysis indicated that out of the six capacity decrease scenarios (3 road groups x two years), only that for the rural road capacity reduction in 2030 produced a stable converged result, so that only this capacity reduction test was presented in detail. This test led to a -6% reduction in total car traffic, comprising: -8.4% on roads within rural areas, a smaller -2.5% reduction in conurbations and urban areas but surprisingly a -9.60% reduction on London's roads. As discussed later below this odd London result may indicate an unnoticed lack of convergence in traffic around London.

3.14 While the variation in impacts across the capacity changes by road type is intuitively reasonable, it is much harder to judge the level.

3.15 For the 50% **increase** in road capacity in 2030, Table 12 indicates that traffic on motorways always reduces below that estimated without the increase, though other road types experience traffic increases in almost all cases. A potential reason for this is given: "traffic on less congested roads is forced to travel at faster (and potentially free flow) speeds. This is particularly the case on Motorways where the increased fuel costs that result from the higher speeds may no longer be optimal for the lower value of time journey purposes. As an example, the value of the time saved from increasing speeds is less than the cost of the additional fuel consumed." p.53

3.16 We have understood this to imply that the removal of congestion on motorways has led to an increase in average motorway speeds and so to higher fuel costs per km on motorways. This response is contentious because in reality the slower speeds caused by congestion are likely to involve more acceleration and deceleration activity on motorways. Accordingly, congestion generated speed reductions will in reality increase rather than decrease fuel consumption and associated costs per kilometre. Fuel consumption as a function of speed should take account of whether the speed achieved is reasonably constant, as a function say of the speed limit on the road type, or alternatively is caused by congestion with associated inefficient fuel consumption due to fluctuating speeds. An enhancement for future models, explained in Annex 2 below, would adopt fuel consumption functions that are differentiated by the road type of a link and its associated speed limit for each vehicle type, while continuing also to be a function of vehicle speed. This improvement is particularly important for HGVs in providing more realistic estimates of fuel consumption and in improving their assignment onto motorways, rather than onto slower roads that do not encourage constant speeds or low fuel consumption.

3.17 Some consideration has been given to the convergence of the model in terms of the number of iterations carried out. Given that the standard assumption is to run 6 iterations, the change from increasing this to 10 has been examined, and at least at the aggregate level of total trips and total mileage, the effects seem small. In most models because non-convergence typically is focused into just a subset of local areas, it is these areas that are of most interest to analyse, without averaging them into the many other larger well-behaved areas. For this reason it is less surprising that the results for car trips to and from congested London have significant changes of around 5% between runs 6 and 10 that are clearly not locally converged. It would be instructive to examine convergence with greater segmentation focussing in on the most congested road categories and area type combinations to confirm that no significant oscillations are occurring between the final iterations in 2050. It is crucial for model users to have a detailed understanding of the locations and types of policy test for which convergence issues are likely to be most challenging. However, for the general aims of the model the level of convergence is probably adequate.

The Back-casting Exercise

3.18 At the outset it may be said that while the aims of back-casting are worthy, there is relatively little guidance on how it can be carried out in practice, nor are there many successful cases in the literature. Given that a reasonable elapsed time is necessary to allow the model to be reasonably tested, there are significant problems in establishing the "true" data relating to the back-cast year, especially when a considerable level of detail is involved.

3.19 It may be said, therefore, that the exercise is to be commended in a) establishing the key principles for carrying out a back-cast, and b) deriving the best assumptions both for the model inputs and the data against which the outcomes can be compared. The year 2003 was chosen because this was the base for the V2 model.

3.20 Essentially, given the predicted growth in total trips, segmented by the various purpose and socio-economic categories in NTEM, the essential test of the model is whether it can predict the distribution and modal split, in the light of changes in modal generalised cost and highway capacity, as well as the predicted mileage on different road types. However, it could be considered unreasonable to judge the model in the light of possible errors in the NTEM trip growth.

3.21 Although some effort is required in collating the data, allowing for general changes in generalised cost, in terms of vehicle operating cost and value of time, is relatively straightforward. Much less straightforward is an appropriate allowance for capacity changes: however, this is greatly facilitated by the fact that the traffic database was available for both years 2003 and 2015. This meant that a separate capacity growth factor, based on lane-Km, could be calculated for each combination of region, area type and road type, though particular problems were encountered in the London area. In this connection, some changes were made to the speed-flow curves for London. There is also a 1% reduction from 2003 to 2015 in motorway length indicated for the Eastern region in Table 22, the reasons for which are not immediately obvious.

3.22 After some analysis of different estimates of trip rate growth, it was decided to use NTEM 6.2 (released in 2011, using the original trip end methodology based on NTS data 1988-96), but to correct the NTEM 6.2 estimate of population for 2003 by a global factor to bring it in line with the actual population and to make some correction (via "spot trip rates") to allow for changes in trip rates over time. Unfortunately, at the time of writing no clear description of this latter process has been provided.

3.23 On this basis, the total car-Km forecast can be compared with that from the "Reference case" (though again, insufficient details are available as to how this was constructed), and the result is a slight underprediction (i.e. overprediction of the true increase between 2003 and 2015) of 3.7%. When further analysed by purpose, the results were generally close with the exception of HB Other (NB the definition of this purpose is not clear) where the actual growth seems rather less than that in the back-cast. This requires further investigation.

3.24 The close match of car-Km in the backcast for the urban and rural areas is impressive in Figures 64 and 65. The difference for conurbations (-7%) is less impressive, while that for London (-21%) indicates a significant problem there. It would have been informative to have additionally carried out a validation against actual observed car ownership rates in Inner and Outer London for both 2003 and 2015. This may perhaps indicate that some of the backcast difference found in London for car kms could relate to the differences from the observed levels in the car ownership rates that have been input to the demand model from NTEM for 2003 and 2015.

3.25 In terms of traffic by road type, there is a slight tendency to overestimate the growth on major roads. In the comparisons of the match for car kms across road types in Figures 68 and 69, it would have been informative to have further segmented between urban and rural road types. This would have identified whether the differences from the reference case indicated for motorways and A roads were primarily within the urban roads sub-category, especially London, in keeping with the poorer results already seen for London and the conurbations. This could have clarified how good the match is for the rural road network, where the model might be expected to perform better. Because observed traffic growth trends in recent decades have differed considerably between urban and rural roads of a given type, an urban/rural segmentation of road types should be retained throughout the analysis and should influence the discussion on the likely underlying sources of discrepancies.

3.26 The match of trips in total (-13%) and especially the match for modes walk (-41%), bus (+29%) and rail (+76%) is not impressive by comparison with the good match achieved in the backcast of car kilometres. Moreover, this poor match for each of walk, bus and rail is of a similar magnitude and direction across most trip purposes and distance bands, which suggests systematic sources of errors in their representation. Because these non-car modes in reality are most competitive with car within London and the conurbations, their poor performance in general may be playing a significant role in the less good performance observed for car kilometres within those areas. The reasons for this poor match should be explored more systematically so as to understand better their implications when forecasting car demand through into the future, as well as to identify where most effort should be placed in improving future model versions, if these should need to be developed.

4. Quality Assurance

Quality Management Report

4.1 The QA environment for the NTMv2R was described in "NTMV2R Quality Management Report" (QMR) which was produced in July 2017 and aligns itself with DfT's "Quality Assurance of Analytical Modelling[2]." The QMR sets out an extensive description of procedures and processes which cover a wide range of issues that arise during model building.

4.2 As it was produced to explain the QA environment for the recalibration work a number of the issues are discussed in the past tense – when the original model was built – and other issues are discussed in terms of actions that will be undertaken during the project. This is an understandable and acceptable position.

4.3 The QMR also sets out clearly which processes have been included in the scope of the NTMv2R commission and those which have been excluded. This is a welcome summary as it demonstrates that every aspect has been considered rather than exclusions having simply been overlooked.

4.4 One issue which is apparent is that the QMR identifies that the owner and author are the same person but does not provide any indication of whether it was agreed and approved by any other individual.

4.5 The QMR also provides commentary on which aspects of QA are to be undertaken by the consultants developing the core components of the model - described in the Quality Plan discussed below – and the DfT as it updates supplementary components of the model.

4.6 Of the DfT activities, evidence in the QMR indicates if these have been completed and signed-off, where appropriate, by the Senior Model Owner. Where components have not been completed the QMR indicates the reason for this confirming that sign-off was not required.

4.7 The QMR concludes with recommendations for an Analytical Assurance Statement and with a statement confirming that the QA processes have been followed as set out. This appears to be a reasonable conclusion to reach for the high level review.

NTMV2R Quality Management Report Annex v1.1

4.8 A subsequent annex was added to the QMR in May 2018 (QMR Annex) after an error had been identified in the computation of generalised costs due to work and non-work car non-fuel operating costs being transposed.

4.9 The QMR Annex also records QA processes which were applied to a series of other minor updates which the correction allowed to be implemented at the same time such that NTMv2R incorporated more up to date values for fuel costs and time as well as

[2] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/888350/qa-modelling-guidance.pdf

implementing the recommendation from TAG to remove non-fuel operating costs from the calculation of generalised cost.

4.10 The decision was taken to isolate a copy of the NTMv2R on a separate modelling server to implement changes such that the previously signed-off version was retained without risk of being corrupted by modifications and tests associated with correcting the error. This is a sensible precaution and mitigates a number of risks associated with version control and using incorrect (as in unsigned-off models).

4.11 The QMR Annex also shows the split of QA processes between the consultants and the DfT. For the DfT activities a schedule is provided to indicate which tasks were undertaken and when with a comment about the task status. Of the 12 tasks listed all but three are noted to be complete but there is no indication in the "sign-off column" that there has been an acceptance of any task completion.

4.12 Evidence of the checks which were undertaken and comments on those checks is provided in the "QA Summary Table" at the end of the document. The schedule appears to cover a good range of checks on data, operation and outputs from the NTMv2R. However, a vast majority of the checks appear to have been undertaken by one individual and there are several checks with question marks against the identified initials suggesting there is uncertainty about who undertook the check or whether it has been completed. To complete the QA processes it would be beneficial to complete this table with the confirmed checkers identified.

4.13 The QMR Annex includes recommendations for the Analytical Assurance Statement and a conclusion which states that the QA processes have been applied and the model is capable of providing "good quality results and can therefore be judged fit for the purposes of strategic policy analysis." This appears to be a reasonable conclusion to draw, subject to the components identified as being complete achieving an appropriate sign-off.

Output 02 – NTMv2 Recalibration Task 2.8 – Quality Assurance v3.0 (TN 2.8-01)

4.14 TN 2.8-01 captures the QA processes which the consultant team has undertaken during the model recalibration project. The note also refers to the DfT's "Quality Assurance of Analytical Modelling" as guidance for QA on modelling projects.

4.15 A QA process which was undertaken in addition to those recommended was re-running the NTMv2 upon receipt to make sure results which were obtained within the DfT could be replicated by the consultant team. This is a very important initial check in all modelling projects when a model is handed from one team to another.

4.16 Where inconsistencies were found between the results that were provided by the DfT and those which were created using the same files by the consultant team the source of the discrepancy was investigated and remedies were identified and implemented where possible or documented where, due to software or systems, implementation was not possible. This is good practice for handling models across organisations.

4.17 A further additional activity was the introduction of a working practice that added "README" sheets into spreadsheet workbooks. These cover sheets essentially replicate the QA sign-off process from reports capturing version numbering and changes as well as the contributing and approving members of the consultant team. This is another welcomed addition which improves the overall QA environment.

4.18 When new spreadsheets and tools were created specifically for the NTMv2R project these have been built README pages and internal checking processes and cells which allow for comparison to target data, for example NTEM trip end totals.

4.19 For the model implementation runs control spreadsheets were used to document the assumptions being implemented and a series of batch files which stopped runs if errors had occurred. Each completed run was subjected to a series of investigations to ensure it had completed using the correct inputs and process files. Reviews identified issues which were rectified and model re-run. This consistent production and review process concluded that the model runs were completed correctly.

4.20 Model convergence was reviewed to determine how well trip attractions by zone produced from model runs match input targets and the volume of trips in distances bands stabilises. The note explains how well convergence was achieved from four comparative runs of the NTMv2R and identifies plausible reasons for not achieving the perfect score of 1.0 across all purposes.

4.21 The note concludes that the QA environment agreed at inception has been implemented throughout the model recalibration work. Checks of inputs, processes and outputs have been documented and tests against specified acceptance criteria have been undertaken to reach the conclusion that the "the NTMv2R model provides an improved basis for implementing forecast assumptions to test a wide range of forecast scenarios."

Output 02 – NTMv2 Recalibration Task 2.8 – Quality Assurance (TN 2.8-02)

4.22 TN 2.8-02 was produced after an error was identified in the application of vehicle operating costs (VOC) which required some rework for the demand model recalibration. The opportunity was also taken to update some subsequently revised VOC values and other TAG values.

4.23 The VOC issue arose from an independent review of model results in an area of the model documentation which had been overlooked. The error was tracked back to a spreadsheet which employed a functional form based on VOC calculations which had been superseded. A second issue, of non-fuel costs being added to non-working time was also identified and addressed.

4.24 Two model runs were implemented to test the impact of reversing the transposed costs and then removing the non-fuel costs. In combination these tests resulted, as expected, in more car trips and car travel (vehicle kilometres).

4.25 The note then summarises the additional calibration work involved with the updates of TAG which was undertaken by Rand Europe and reviewed by Atkins. The conclusion of the review of these tasks was that the output calibration lamda values were "reasonable" and "generally within the ranges that might be expected."

4.26 TN2.8-02 concludes that the working procedures on NTMv2R did not need to change and the QA environment of the project was reasonable and proportionate. Whilst an error had been retrospectively identified, the QA process enabled its source to be isolated and corrective actions implemented efficiently.

V2R Implementation Report

4.27 This report, which was produced by Atkins, includes appropriate Quality Assurance sign off information identifying the originators of the content, the individuals charged with checking the source material is valid and calculations have been implanted correctly, document reviewers that have verified the material as appropriate and the individual authorising that the document can be issued to the client.

4.28 The Quality Assurance sign off sheet indicates how version control has been implemented and provides an indication of the reason to update the report and increment the version number. This evidence of quality processes being followed fulfils the QRM and QP requirements.

V2R Calibration and Validation Report

4.29 This report, which was produced by RAND Europe, does not include a standard quality assurance sign off sheet, but does identify the main author on the cover page and includes at Appendix C a record of the Quality Assurance processes which have been implemented.

4.30 The Appendix indicates that the “standard RAND Europe peer review” has been conducted by a named individual. In addition, there is a schedule of quality checks which have been undertaken within the project itself. This explicit schedule provides a useful insight into how the work has been reviewed and provides confidence that the outputs have been tested in an appropriate manner.

NTM Version 2R Analytical Review_DRAFT_v08

4.31 This report has been produced by the DfT itself. There is no Quality Assurance sign-off sheet, which is common with most DfT documents. As the report is presently in draft format there is an opportunity to include a statement which explains the quality assurance process and confirms it has been implemented, potentially indicating in a manner similar to the RAND Europe report, the tests which have been undertaken to verify the model outputs.

Summary of Quality Assurance

4.32 Overall, there is evidence that QA processes have been implemented throughout the model recalibration project. Roles and responsibilities are generally identified both within organisations and across the project team as a whole (DfT and its consultants).

4.33 The error which led to additional work having to be undertaken was outside the scope of the original QA processes but was rectified and its impact addressed. In itself, this demonstrates that even in a good QA environment errors can be missed but that a good QA environment makes tracking and isolating the causes an easier task.

5. Conclusions and Recommendations

5.1 Overall, the results are encouraging for the main use of the model – the prediction of traffic.

5.2 The general aim of the DfT to open up the national models to wider scrutiny and greater transparency is helpful. The publication of the results of the stress tests and the backcasting is a useful step towards this aim.

5.3 In relation to the stress tests for fuel price changes, the results are generally convincing, though they do make clear the need to carry out a proper base year realism test: if the current software is truly unable to do this, then appropriate changes should be made. In addition, while we are surprised that the model outcomes are insensitive to the assumption about the elasticity of VoT to GDP, in the light of these we feel it would be better to make the conventional assumption of an elasticity of 1, as is generally recommended by TAG.

5.4 The capacity tests raise more questions: it needs to be better explained what has actually been done to implement these. The tests involving a decrease in capacity appear to raise some issues of convergence, but perhaps the scale of the test is just too severe. In any case, however, it would be wise to investigate convergence at a more disaggregate level, even though the overall indicators are encouraging. We are surprised that the 50% **increase** in road capacity in 2030 leads to a decrease in traffic on motorways, and find it hard to accept the suggestion that the time saved is outweighed by higher fuel costs associated with higher speed.

5.5 The backcasting exercise provides a useful insight into the general performance of the V2R model, identifying: those aspects of good performance, such as the accurate backcasts of car kilometre changes outside London and the conurbations; as well as the poor performance for most of the passenger modes competing with car.

5.6 It would have been helpful to have provided increased spatial segmentation detail in the results that are presented differentiating between values for: major urban centres; other urban areas; and rural areas, so as to distinguish more clearly those types of areas where the model performs best and those where its results are less convincing.

5.7 It is important to put the results of this specific backcasting exercise in context. The backcasting analysis here is deliberately focused on just two components of the complete NTM system: the demand model and FORGE. It expressly excludes any backcasting of other parts of the complete NTM system, such as the trip end estimation (NTEM) or car ownership estimation (NATCOP), whose inputs to the demand model also play a major role within the complete procedure used to create road traffic forecasts. It is important for both 2003 and 2015 in all types of areas, particularly in London, that the car ownership levels by area type within NTEM match closely to those observed. Unless it is certain that all external data that are input to the demand model are fully consistent with the observed situation, apparent shortcomings within the demand model or FORGE may instead be a result simply of errors in inputs to these NTM component models.

5.8 It could be very informative to carry out a similar backcasting exercise for the NTEM and NATCOP components, validating them against observed data that is segmented by area type or some finer level of spatial detail. This expanded backcasting would improve understanding of the performance of the complete NTM system. It is of critical importance for any future model enhancement activities to ensure the correct identification of the real underlying sources of all discrepancies between the 2003 reference case and the results backcast from the model. Accordingly, this V2R backcasting exercise would benefit from a further stage that validates all of the input NTEM/NATCOP data against observed values.

5.9 The accuracy of the fuel consumption estimates and of the environmental costs that are output from the model, as well as the allocation of HGVs across road types, could be improved through enhancing the emissions formulae into a form that takes better account of the impact of congestion on emissions using the approach discussed in Annex 2 below.

Annex A: Size Terms

The original V2 Calibration is described in the ME&P (2002) Report "DFT: Integrated Transport and Economic Appraisal, EQUILIBRIUM MODELLING 2A, Model Implementation v2.0"

The following text is extracted from Section 7.3.1 of this report:

"*The first step in the calculation of the size terms is to derive the set of zone pair weights for each of the 13 regional Pass3*[3] *zoning systems as shown in equation (7.1).*

$$W_{odl}^{pk} = A_d^{pk} \partial_{odl} \qquad (7.1)$$ [4]

For attraction / destination zones d within the Pass3 region and all production / origin zones o in Great Britain.

A_d^{pk} trip attractions in Pass3 destination zone d from NTEM for purpose p and mode k

∂_{odl} = 1 if the travel distance between origin zone o and destination zone d $\in$ l

= 0 if distance between origin zone o and destination zone d $\notin$ l

The weights for the …Pass1 [model is] obtained from the basic weights as follows:

$$\text{Pass1 weights} = W_{IJl}^{p} = \sum_{k,i \in I} \left(\sum_{j \in J | \delta_{ijl} = 1, j} W_{ijl}^{pk} \right) \qquad (7.3)$$

where I = Pass1 trip production zone

J = Pass1 trip attraction zone

l = distance between I and J (obtained from Pass1 factor)

[3] Pass3 model was the route choice or road network component of the full implementations of NTM versions 1 and 3

[4] NB Normally the size term would be independent of the Production zone. However, the unusual structure of the model means that the term has to be appropriate for the distance band, which is dependent on the Production zone.

p[5] = trip purpose and mode (obtained from Pass1 factor)

7.3.2 Calculation of size terms

"Size terms for Pass1 model: $S^p_{Iad} = -\frac{1}{\lambda^p_A}\ln[W^p_{Iad}] - Min_{I|adp}\left(-\frac{1}{\lambda^p_A}\ln[W^p_{Iad}]\right)$ (7.4a)

Which is equivalent to: $S^p_{Iad} = -\frac{1}{\lambda^p_A}\ln\left[\frac{W^p_{Iad}}{Max_{I|adp}\left(W^p_{Iad}\right)}\right]$ (7.4b)

Where λpl is the parameter controlling the sensitivity of the trip distribution model to variations in zone pair disutilities and I, J and l are as defined in previous equations."

Note that it is not clear from the text what the parameter λpl refers to: "distribution" is used to refer to both the distance and attraction area type. However, in the V2R Calibration and Validation Report it is explicitly noted that the parameter is λpA.

Translating this into the current notation:

$$W^{pm}_{sd} = A^{pm}_s \partial_{sd} \qquad (7.1)$$

for attraction / destination zones *s* within the Pass3 region and all production / origin zones *r* in Great Britain.

A^{pm}_s trip attractions in Pass3 destination zone *s* from NTEM for purpose *p* and mode *m*

∂_{rsd} = 1 if the travel distance between origin zone *r* and destination zone $s \in d$

= 0 if distance between origin zone *r* and destination zone $s \notin d$

The weights for the …Pass1 [model is] obtained from the basic weights as follows:

Pass1 weights = $W^p_{Iad} = \sum_{m,r\in I}\left(\sum_{s\in a|\delta_{rsd}=1,r} W^{pm}_{rsd}\right)$ (7.3)

where *I* = Pass1 trip production zone

a = Pass1 trip attraction zone

d = distance between *I* and *J* (obtained from Pass1 factor)

pm = trip purpose and mode (obtained from Pass1 factor)

[5] Presumably this should be "pk"

Annex B: Recommendations for TAG on emission function formulæ

The accuracy of routing within assignments by type of road in FORGE, as well as the calculations of fuel consumption, could be improved by **revising the fuel consumption formula** specified in **Section 5 of TAG Unit A1.3 to be based also on link type** and not solely on vehicle speed - the current formulation. Analogous **enhancements to the emission functions** should also be applied for the various emission types covered in **Sections 3 and 4 of TAG Unit A3** on Environmental Impact Appraisal.

An important reason why long distance HGVs are concentrated onto the motorway network lies in their ability to achieve a constant speed there, which in turn generates relatively low levels of fuel consumption per kilometre. Uncongested motorway travel is more fuel efficient than driving on dual carriageways that generate regular deceleration/acceleration phases at roundabouts or traffic lights. Observed average HGV speeds on high quality roads that are significantly lower than the speed limit are generally a result of unavoidable variations in speeds caused by road conditions or road congestion. Consequently, these lower speeds will increase, rather than decrease fuel consumption rates. Travelling on uncongested high quality roads with a 40 mph speed limit would not create high fuel consumption rates, whereas travelling on a motorway also at 40 mph but now as a result of congestion would lead to very high rates of fuel consumption. Accordingly, the fuel consumption functions for a specific vehicle type on a link should be enhanced to take account of: its speed limit; its road type; and not solely the speed achieved by the vehicle.

The formulae currently used in TAG to estimate fuel consumption for individual road vehicle types were derived (Ricardo-AEA, 2014)[6] from speed emission factor curves that represent the change in emission factors as a function only of vehicle speed. These original curves, as well as those within the most recent update (Ricardo Energy & Environment, 2019[7]) have been derived from the widely used COPERT factors which were constructed using an **Average Speed Emissions** model. Average Speed type formulations were designed for use in the context of national or regional inventories - rather than for use with more localised network based traffic characteristics.

Over the last twenty years various research projects have developed alternative **Traffic Situation Emission** models that are more directly suited to the needs of TAG and of transport modelling in general. For example, the TRL coordinated the

[6] Ricardo-AEA (2014) Production of Updated Emission Curves for Use in the National Transport Model. Report to the Department for Transport. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/662795/updated-emission-curves-ntm.pdf accessed 06/05/20

[7] Ricardo Energy & Environment (2019) Production of Updated Emission Curves for Use in the NTM and WebTAG. Report for the Department for Transport: 1-650 P04102073 through Arup, ref no: 268492-13. https://www.gov.uk/government/publications/transport-modelling-and-appraisal accessed 02/06/2020

European-wide consortium ARTEMIS[8] that developed one such model that segments emission factors by traffic situations that were distinguished through a combination of: speed limit; type of road [12 - urban/rural by road category]; and traffic condition [4 - Free-flow; Heavy; Saturated; and Stop & Go]. This model covers a wide range of emissions including NOx, CO2 and PMs for a range of vehicle types. A traffic situation emissions model of this type, but which has been updated to take account of research developments over the last decade since ARTEMIS, could provide TAG with much improved emission factor curves that are differentiated by road type and speed limit, together with average speed.

The benefit arising from this enhancement should both improve the realism of route reassignments between road types, primarily for HGVs, within FORGE and other highway assignment models. It would also improve the accuracy of the environmental cost estimates created for each of the main vehicle emission types. In this way it would lessen the likelihood of drawing inappropriate conclusions from policy measures being tested within transport models such as the NTMv2R.

[8] Boulter P G and IS McCrae (2009) ARTEMIS: Assessment and Reliability of Transport Emission Models and Inventory Systems – final report. TRL Published Project Report PPR350. https://trl.co.uk/reports/PPR350 accessed 04/06/2020

Kings Orchard
1 Queen Street
Bristol
BS2 0HQ

wsp.com