ON THE STATE-OF-THE-ART DEMAND FORECASTING MODEL DEVELOPED BY NETHERLANDS RAILWAYS

Bert de Vries
Senior Researcher at Netherlands Railways (NS)

Jasper Willigers
Senior Researcher at Significance

Paper presented at the European Transport Conference
Glagow, Scotland, UK, 10-12 October 2011

1. INTRODUCTION AND ABSTRACT

NS (Netherlands Railways) and Significance have developed a new model system to provide train travel demand forecasts for NS's strategic decision making. The model is among others used for the evaluation of timetable alternatives, forecasting of volumes of new stations, evaluation of tariff policies, all under different exogenous scenario's for demographic, economic and car-related developments. This paper focuses on the development and design of this model system from two, seemingly conflicting, perspectives.

The first perspective is about the user requirements which were formulated and relate to how the model should work in practice, such as:
1) the model should have a clear structure and be transparent in how its results have been calculated,
2) input data and model coefficients should be up-to-date,
3) the model should be able to interact and be consistent with other, complementary models currently in use by NS, and
4) the model should be user friendly and easy to maintain.

In contrast, from the second perspective the model system has to have sufficient technical detail, complexity and flexibility to provide forecasts for different scenarios, time horizons and study areas and to be adequate in its forecasting precision. It is this trade-off between easy-to-understand forecasts and the necessary complexity of the model to be adequate in its forecasting results which has been the main challenge of this project. As a solution a modular system has been developed to fulfil all these requirements, while finding a trade-off point where both perspectives might oppose each other.

Significance contributed by developing a new module to the model system for forecasting demand volumes per station pair for different exogenous scenarios. The module consists of four sub-modules, each with a clear function and output that can be viewed and confirmed.

The model system's design has been greatly influenced by the precondition of transparency: the contribution of separate modules in the final forecasting result can be easily visualised. This has contributed to the understanding and acceptance of the model results by the decision makers.

2. NS WAS IN NEED OF A NEW DEMAND FORECASTING MODEL

For many years NS have been using models in order to make appropriate forecasts of future demand volumes. Forecasts of either the number of trips or the passenger kilometers play an important role in the strategic decision making of NS. Forecasts are being used at different levels of detail and for different time horizons. For example, short-term forecasts at Origin-Destination level (OD-level) are being used for next-year timetable design. Mid-term forecasts at the more aggregate route level are being used for the determination of the necessary rolling-stock volumes and types. Long-term forecasts are necessary for rail-infrastructure investment programs of the Dutch government.

More examples of forecasts being used by NS are:

- Volumes at new stations (i.e. total passenger access and egress) to be built in 5 to 10 years;
- Number of passenger kilometers for a new part of rail-infrastructure to be constructed in about 20 years and for different timetable alternatives;
- Number of passenger kilometers at the aggregate national level for the next three years;
- Growth of corridor passengers kilometers after a timetable change (see figure 1 as an example of this).

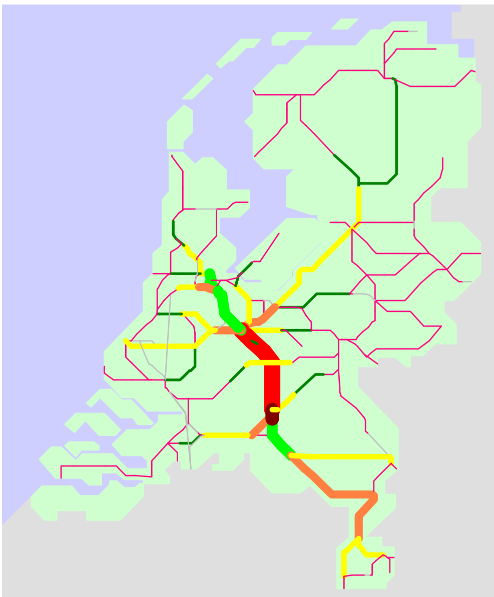

Figure 1. Growth of corridor passenger kilometers

These examples show that a wide variety of forecasts are being used for different study areas (stations, corridors, national) as well as different time horizons (from one to 20 years).

NS used to have different models and databases for the different studies. Examples are:

- Origin-Destination-(OD)-matrix: every year an OD-matrix of the number of trips is constructed from many NS-databases (e.g. sales-figures);
- TRANS: allocates, for all OD-pairs, and based on a specific timetable, all trips to specific lines and corridors. Results can be aggregated;
- PINO: specific model for forecasting volumes at new stations;
- Tariff model: for the evaluation of a range of alternative tariff-structures;
- ELMO: a macro elasticity model for forecasting the yearly trends in volume based on endogenous variables as well as many exogenous variables (economic, demographic and car-related parameters);
- Promise: a long-term forecasting model at OD-level, with many endogenous and exogenous input variables, but with some disadvantages in the way results are presented. The model gives a forecast for just one future year for each model run and because of the complexity of the calculation algorithms results are frequently not easily understood.

Main disadvantages of the former state of affairs were:

- Different models may result in inconsistent forecasts, e.g. both PINO and Promise give volumes for new stations, but by using a different approach. There is no guarantee that results will correspond to each other;
- In general it is not easy to understand forecasts from complex models like Promise, because the structure of those models are in such a way that they serve more or less as black box calculations: the input is a base OD-matrix, subsequently many model calculations are done based on endogenous and exogenous input values and finally the result is a forecast OD-matrix. No insight can be supplied by such models to understand the forecasts, for example by an indication of which variables had the main impact on the forecasting results. For model forecasts to be really used in policy making it is very important that there is great confidence in the model calculations and results.

Forecasts play a strategic role in NS policy and this will only increase in the future. The past way of working was not optimal for this role. It was therefore decided that NS should develop a new state-of-the-art forecasting model.

3. DEVELOPMENT PROCESS OF THE MODEL SYSTEM

First step in the development process of the new model system was a thorough definition study, consisting of the following aspects:

- Compilation of an inventory of the way of use, the forecasting quality and performance of the models already in use as perceived by the users of the forecasts: this gave insight into the actual situation as it was;

- Broad investigation (by means of interviews) within NS on the up-to-date requirements concerning forecasts in a wide range of disciplines: this gave insight into the optimal future situation;
- Benchmarking study to learn from other international railway companies the way which, and how, forecasting models are being used;

The most important user requirements, besides the functional specifications were mainly about the use of the forecasts and the way they should be interpreted. The bottom line from these interviews was that forecasts, even based on complex models, should be in a way easy to understand and made plausible. This transparency is the most important condition for forecasts to be really used by the decision makers.

Secondly it was learned from the internal investigation as well as from the benchmarking study that NS has already many advanced algorithms and databases available, although mainly as separate models. Examples are:

- NS has already a well-defined algorithm to construct a yearly OD-matrix, which can be used as a base-matrix for the model system;
- TRANS contains adequate allocation algorithms for translating OD-volumes into specific lines and corridors;
- The structure of the elasticity model ELMO, which calculates growth-factors for different input-variables, can be used, although the elasticity model would be used at the OD-level instead of for forecasting a total national passenger volume.

To fulfil the main requirement of transparency of the model outcomes and also make optimal use of the appropriate algorithms and models already available within NS, NS decided to build a modular model system. The modules were either already available (i.e. the available separate models after some minor modifications) or were specially constructed for this purpose.
Characteristics of the modular model system are:

- The system partly consists of modules already existing within NS: TRANS, Tariff-model, ELMO-structure;
- Other modules have been built specially, e.g. the Access-Egress module and the Exogenous module;
- The most recently constructed OD-matrix is used as the most up-to-date base-matrix;
- Forecasts for all modules are made at OD-level. Each module calculates its own year-to-year growth factors for each OD;
- Modules can be run separately, but the different results are integrated into one forecast (for each OD) within the system;
- Because different forecasts are available from the different modules it is possible to give insight in the specific contribution of the modules in the final forecast, which thus gives very transparent forecasting results;
- Separate NS models are integrated in one system and are therefore automatically made consistent with each other.

4. THE MODEL SYSTEM: THE WAY IT WORKS

The final model system consists of the following modules:

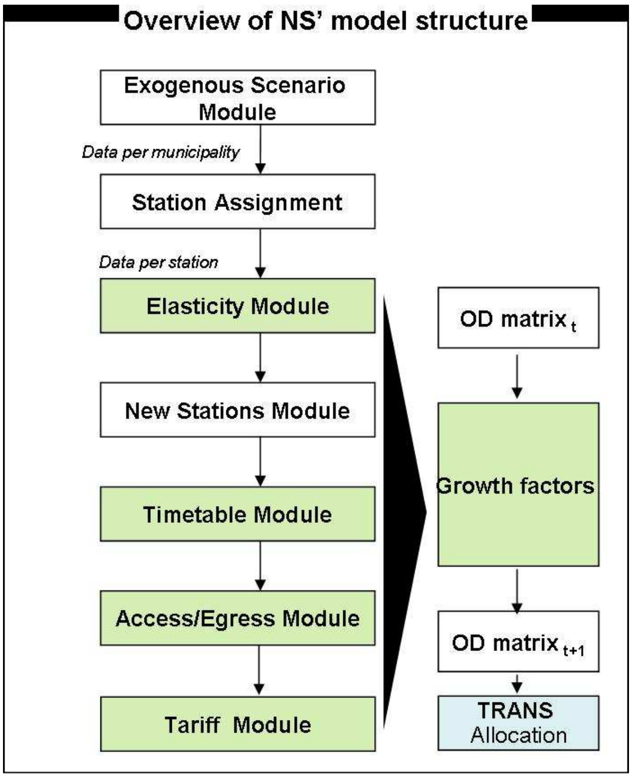

Figure 2. Modules in the model system

The modules in the system:

- The first four modules mentioned (Exogenous Scenario Module, Station Assignment, Elasticity Module and New Stations Module) are in fact four modules of the Exogenous submodel. This submodel allocates data of exogenous scenarios (demographic, economic and car-related variables) from several geographic levels (region, municipality, postal-code) to NS-stations. Further, based on the yearly growth of those variables at OD-level and in combination with estimated elasticities, growth-factors at OD-level are determined. More details on the Exogenous submodel, which is the backbone of the system, are given in section 5;
- The Timetable Module calculates the effect of timetable changes at OD-level. For each OD the Generalized Journey Time (GJT) is calculated, for the old timetable as well as for the new one. Growth-factors are based on de relative change in GJT and the elasticity;
- With the Access/Egress Module it is possible to calculate at station level the effect of improvements in all kind of aspects of public transport (e.g. comfort, travel-time), availability of parking for cars or bikes, etc. on the increase (or decrease) of rail demand;
- The Tariff Module calculates the effect of tariff-policy at OD level;
- It is possible to add growth-factors from other models or calculations on an ad hoc basis. For example, growth-factors as a result of recent

government policy, based on variables not yet in the model, can be separately estimated and integrated with the model results.

All modules calculate growth-factors at OD-level and for each future year. These factors can be multiplied over the modules to get a forecast one year ahead and can be multiplied over the years to get future-year forecasts.
The system's algorithms can thus calculate year-to-year forecasts. However the merit of those year-to-year forecasts depends heavily on the quality of the year-to-year input: appropriate input must be available on a year-to-year basis.

Traditional models usually also make use of several modules for the different functions in the model. Mostly, they have built-in iterative loops between those modules and the results cannot be separated anymore. Because we have our main emphasis on the transparency of the model we had to make a trade-off between being theoretically completely correct, but with complex algorithms on the one hand and being theoretically less correct but with the possibility to give some insight in the forecasts on the other hand. To make a reasonable trade-off decision the modules were defined in such a way that variables which were statistically independent from each other, i.e. with hardly any interactions, were put in different modules, but variables with large correlations are in the same module. In this way the modular system closely approaches the theoretically completely correct models, but with the great advantage of transparency.

The different modules all give their own growth-factors. We can see them as separate results that are independently calculated and integrated afterwards. Those separately calculated effects lead to individual contributions to the final forecast, initially at OD-level, but aggregations are also possible. Those individual contributions are then easily presented in a waterfall-chart between two years, often the base year and a future year.

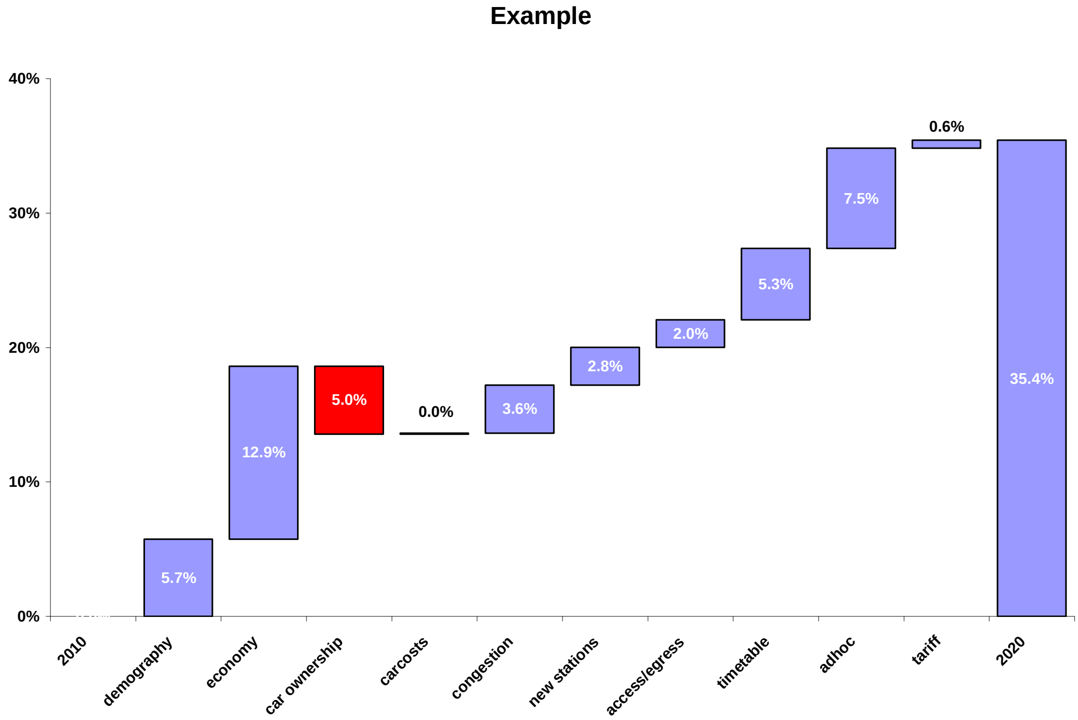

Figure 3. Example of waterfall-chart as output of the model system

5. THE EXOGENOUS SUBMODEL IN DETAIL

As mentioned in section 3, the model system was built around a number of models that were already operational. Between the functionalities of the existing models some gaps still existed, which needed to be filled by new modules. One of these gaps was a submodel that could forecast station-to-station trip matrices based on exogenous scenario input for demographic, economic and car-related variables as well as on policy variants for timetables and fare setting.

5.1 Preconditions for Design

The new submodel needed to satisfy a number of preconditions that are consistent with the requirements for the model system as a whole. The most important of these preconditions are summarized as follows:

- Demographic and economic variables must be treated in a detailed way;
- Transparency: no complex iteration mechanisms should be used;
- The new software must be complementary to existing modules (e.g. TRANS, PINO);
- The submodel should be build upon the experiences with existing forecasting tools (particularly the ELMO tool);
- The submodel should provide forecasts for 1 to about 15 years ahead, making a distinction between short- and medium-term effects;
- Model coefficients are to be based (estimated or calibrated) on empirical data;

- The software should have a limited runtime (several minutes for each future year).

The model system's design has been greatly influenced by the precondition of transparency. For travel forecasting on a national scale the NS had positive experience with the ELMO tool. ELMO had gained confidence among the users, as it was easy to understand and gave logical results. Yet its capabilities and level-of-detail were too limited to be used within the new model framework. Still the positive experience with this tool has led to the decision to adopt an approach using elasticities for the new submodel. The new submodel applies elasticities at a more disaggregate level (per Origin-Destination station pair), which required new solutions for dealing with competition between stations and for providing forecasts for newly opened stations.

5.2 Structure of the Submodel

The submodel consists of four modules, each with a clear function within the framework:

1. The Exogenous Scenario Module uses external input for demographic, economic and car-related variables to create a year-by-year dataset on the municipal level.
2. A Station Assignment Module uses the municipality data set to create a station level dataset of these same variables.
3. An Elasticity Module uses growth factors for scenario and timetable developments to forecast demand volumes at OD-level, separately for each of six travel purposes.
4. A final module forecasts demand volumes for newly opened stations: the New Stations Module.

The figure below shows how the four modules interact. For each model year the four modules are run in the above sequence (from 1 to 4). The processes modelled within each of these modules are explained in the next sections.

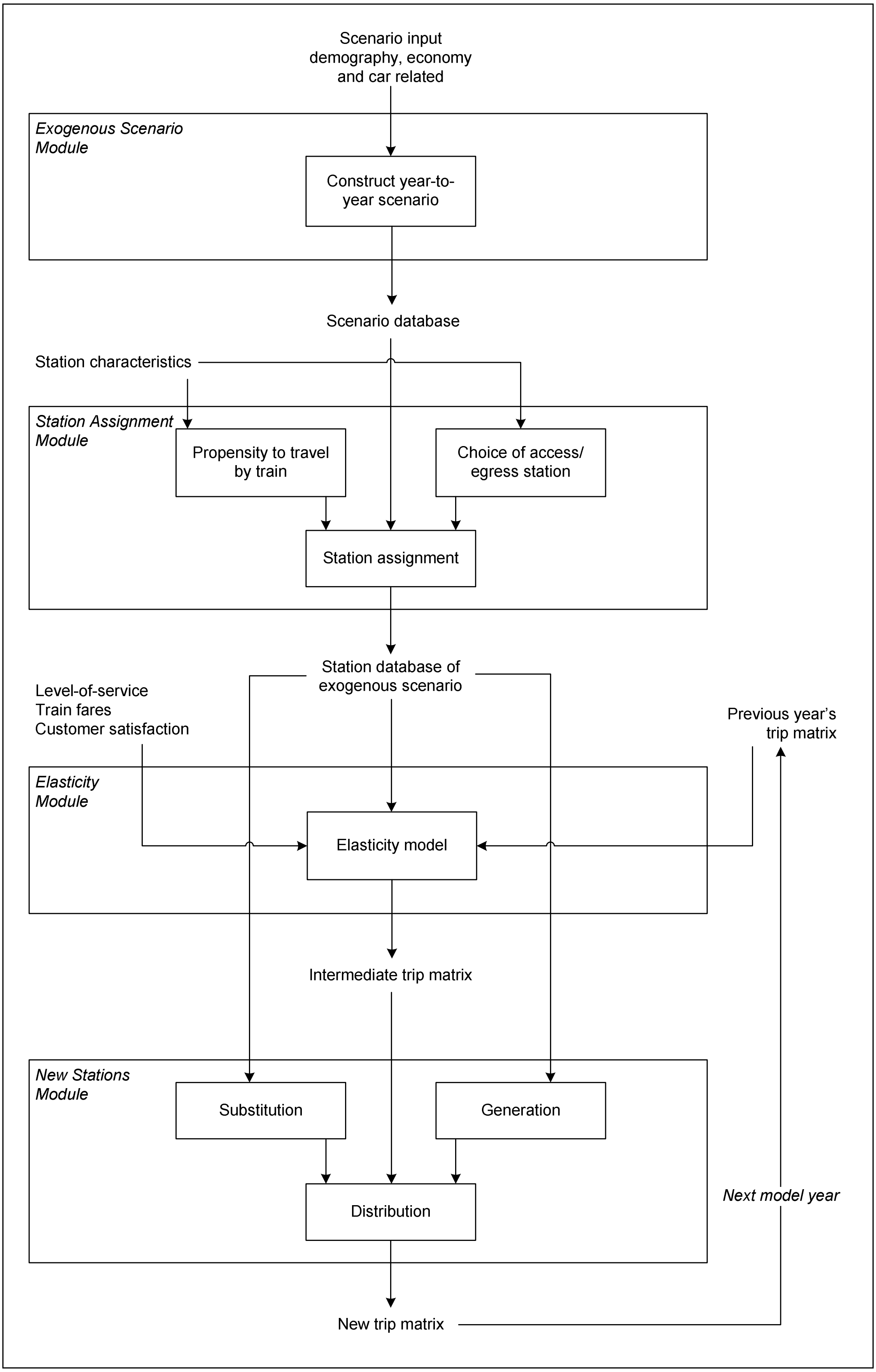

Figure 4. Structure of the Exogenous submodel

The contribution of the four separate modules in the final forecasting result can be easily assessed. Each module in itself produces results that are a logical consequence of the input. Results can be verified since no complex mechanisms, such as iterative procedures, are used within the modules.

The submodel is still capable of modelling the complex processes that are relevant for forecasting rail passenger volumes. The next sections discuss how the submodel deals with lagged response to timetable changes, competition between stations and the opening of new stations.

5.3 Elasticity Module with Lagged Response

The submodel aims to provide forecasts for the short and medium term, from one to a maximum of about fifteen years ahead. In practice the magnitude of the effect of for example a timetable change is seen to increase gradually over several years before passenger volumes have fully adapted. There are several causes for this lagged response, such as the use of yearly subscriptions, habitual behaviour of travellers, the time required to change jobs, buy a car, etc. (e.g. Dargay and Vythoulkas, 1999; Kitamura, 2008) It is therefore required to distinguish between short and medium term effects by taking account of lagged responses.

The elasticities for the Elasticity Module have been estimated from a time series of trip matrices. This allows lagged coefficients for some of the most important explanatory variables to be estimated: travel time, fare and customer satisfaction. Coefficients have been estimated for lags of up to three years.

Figure 5 shows an example of how this works in a practical application. In this example a change of the timetable led to a considerable improvement in the level-of-service at a station, both in service frequency and in travel time to some important destinations. The model forecasts most of the effect as occurring in the first year of implementation: a 21 percent increase of the number of daily boardings. For the second year a further 5 percent increase is forecast, which is the lagged effect of the timetable change.

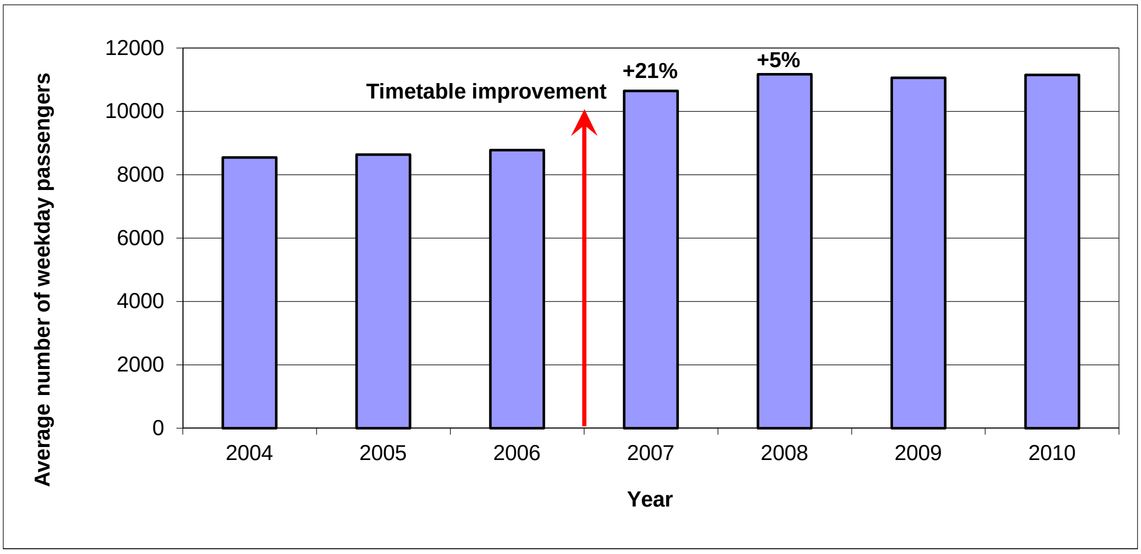

Figure 5. Example of the modelled station-level effect of a timetable change

Lagged effects can be substantial, as seen in this example. The year-by-year forecasts thereby provide insights into the timing and development of dynamic processes.

5.4 Station Competition

The Station Assignment Module models competition between stations by a model of station choice. On a spatially detailed level (4-digit postal codes, almost 4000 zones within the Netherlands) the probability of train travellers using a station for access or egress to the railway network is calculated by a multinomial logit model (e.g. McFadden, 1981). Explanatory variables are the level-of-service at a station and the distance to a station; coefficients are estimated from NS' yearly survey.

For assigning the municipal data to stations, the station choice probabilities per postal code area are aggregated to the municipal level. Hereby the postal code areas are weighted by the propensity to travel by rail. Travellers use the train more often when they have easier access to the railway network. The propensity to travel in the model is dependant on the distance to railway stations and the level-of-service provided by these stations.

Figure 6 shows an example of the station assignment. The coloured postal code zones indicate the station with the highest choice probability and the red circles indicate the propensity to travel by rail. The blue circles are the railway stations; their service frequency is indicated by the size of the circle. It can be seen that the smaller stations only serve a local market, while the main stations attract passenger from the whole region as they are well connected with the regional bus network. The main stations therefore have a much larger area that influence their passenger forecast.

When the service frequency at a station is improved, both the local propensity to travel and the served area are increased.

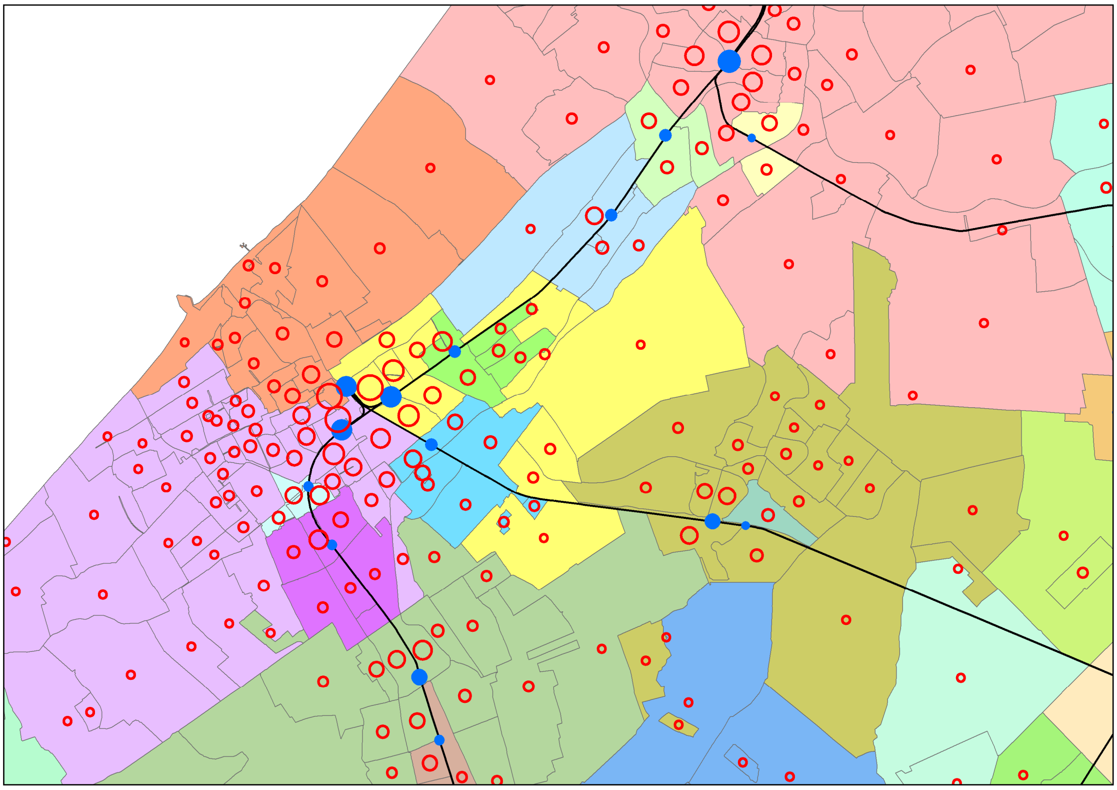

Figure 6. Station assignment: station choice (zone colour) and propensity to travel by train (red circles) depend on station service frequency (blue circles)

5.5 New Stations

The New Stations Module gives forecasts for the volume of travel to and from newly opened stations. Thereby a distinction is made between the substitution and generation effects of new stations.

The substitution effect is the number of travellers to/from the new station that would otherwise use another, previously existing station. Substitution is in fact already accounted for by the Station Assignment and Elasticity Modules. Opening a new station reduces the assignment to the existing stations, which will lead to a decrease in the forecast number of trips at these stations when the elasticities are applied. The New Stations Module uses the difference between station assignments with and without a new station to calculate the substitution effect.

The generation effect is the number of trips that would not have been made without the new station. The New Stations Module gives a forecast of the generation effect based on among others the population from the station assignment, the accessibility of destination stations from the new station and the distance to the nearest competing station. Coefficients are estimated from a dataset with passenger counts at stations newly opened in 1996 or later.

The substitution and generation effects both have considerable time lags. From the dataset of new stations it is calibrated that newly opened stations show a more than average increase in demand for three years after service has begun.

When the number of trips due to substitution and generation is known, the New Stations Module calculates the distribution of these trips across destination and origin stations. A multinomial logit model is used for modelling origin/destination choice, with as explanatory variables the travel time, travel fare and the station assignment (population size, employment, etc., depending on the travel purpose).

The figures below show what happens when a new station opened, based on a backcast using the Exogenous submodel. Figure 7 shows the new station assignment for the first year after opening. This figure can be compared to Figure 6, which shows the situation for one year earlier. The arrow points at the newly opened station and the surrounding green area gives an impression of the area is that served by this new station. In this first year after opening (2006) the service frequency is still quite low: four departing trains per hour (two per direction). In its second year the service frequency is doubled to eight trains per hour. This leads to a further increase of the served area and the propensity to travel.

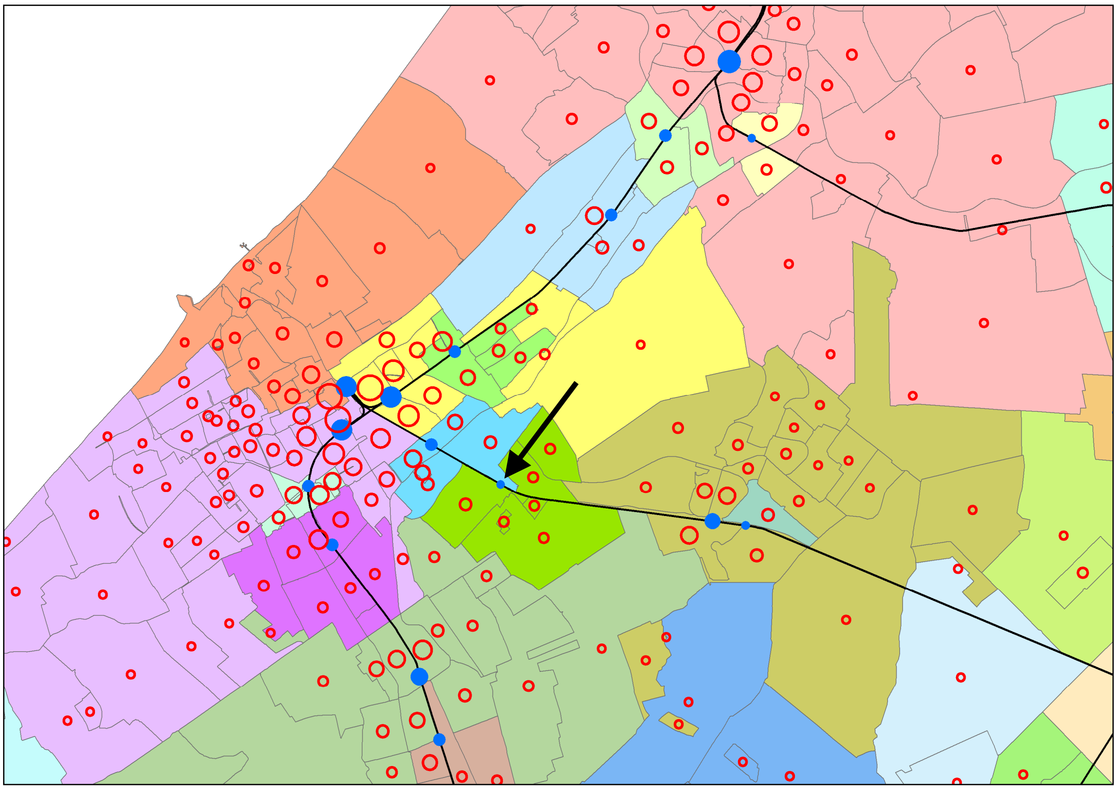

Figure 7. Station assignment with the new station

From Figure 8 it can be seen how the forecast passenger volume evolves for the first years after opening. The passenger forecast continues to grow until the fourth year after opening. In this particular case, the service frequency increase in 2007 leads to an additional increase in passengers, especially in 2007 but also in 2008 and 2009.

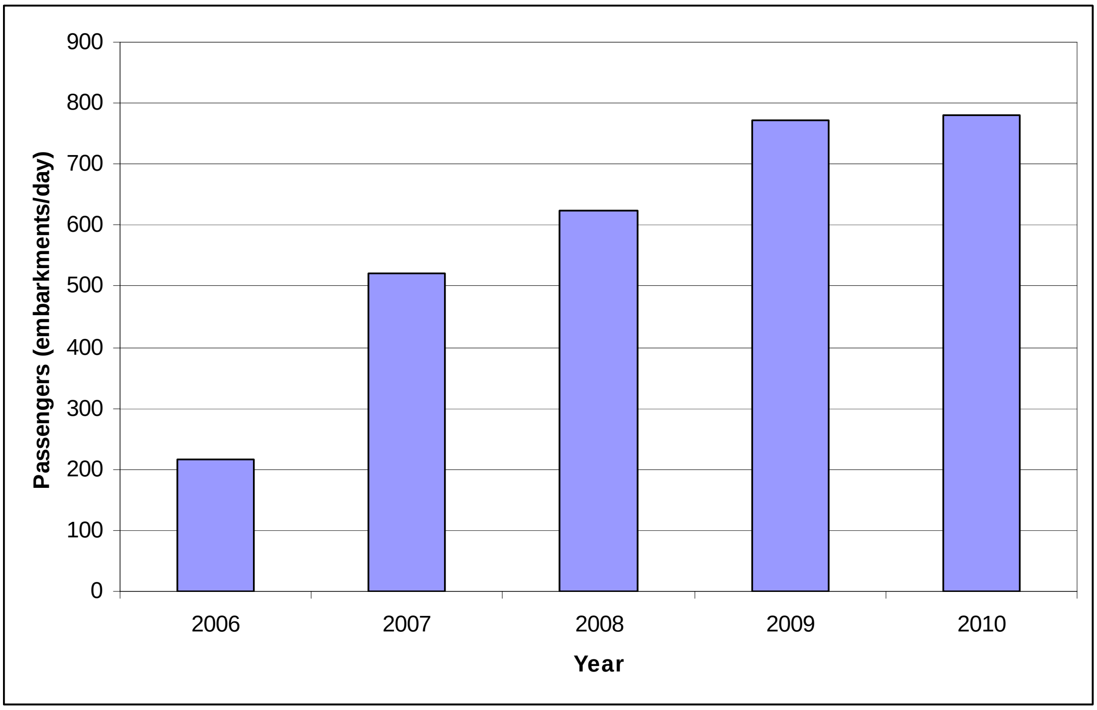

Figure 8. Growth of a new station's passenger volume

In this section we have described in more detail the Exogenous submodel as one of the important modules of the new NS' model system. The submodel is capable of modelling complex processes, such as competition between stations and lagged response to timetable changes. The year-by-year forecasts thereby provide insights into the timing and development of dynamic processes. Because the calculations are based on a multiplicative elasticity structure it is easy to get growth-factors for the different demographic, economic and car-related input-variables. Their individual contributions to the overall forecast can therefore be easily separated.

6 THE ROLE OF THE NEW MODEL SYSTEM IN NS DECISION MAKING

After completion of the first version of the model system in 2009, it has been used for a wide range of forecasting projects, varying from infrastructural studies for Dutch government to minor timetable studies for NS. Examples:

- A large study for the Dutch government, where demand was forecast for the year 2020. The main goal was to evaluate the effect of many different high-frequency timetable concepts on the travel demand for a few specific corridors as well as nationwide. Sensitivity analyses have been performed for different demographic/economic scenarios to gain insight into the range of uncertainty in the forecast. In discussions of the results waterfall-charts has been proven essential;
- NS Business-Planning forecasts for five years ahead. Demand forecasts of the future-year OD-matrix are produced. The TRANS model allocates, for all OD-pairs, those forecast number of future trips to specific lines, based on the expected timetable. The results are input for model-calculations of required future rolling-stock;

- Minor timetable changes for increasing the accessibility of the Northern part of the Netherlands. To construct a business case for minor timetable changes, besides insight in additional costs, also forecasts of expected demand growth for trips towards the North are used. With the new model system's Timetable Module and the use of GJT's the effect of the timetable change on all OD's can be evaluated.

In the communication of the forecasting results of those projects the built-in transparency has been proven to play a significant role.

In order for the model system to be up-to-date we are currently working on a re-estimation of parts of the model system, especially the Exogenous submodel. Whereas the first version was based on data series up to 2006, the second version will also include the later years 2007-2010.

The model system has also been implemented with a user-friendly interface. Four user-groups have been defined: the *administrator*, *power members* for complex forecasting where many options are available, *members* for minor projects with some predefined options and *visitors* for viewing results only.

The new NS' demand forecasting model system meets all the predefined requirements: it is transparent, up-to-date, consistent and user-friendly.

BIBLIOGRAPHY

Bruyn, M.O. and Spijkerman, P. (2011) Business cycles and future demand: how use of a leading indicator improves forecasts or rail passenger transport*, ETC Glasgow*

Dargay J.M. and Vythoulkas, P.C. (1999) Estimation of a Dynamic Car Ownership Model, Journal of Transport Economics and Policy, 33 (3) 287-302

Hensher, D.A., Rose, M.J. and Greene, W.H. (2005) *Applied choice analysis: a primer*, Cambridge University Press, Cambridge.

Kitamura, R. (2008) Longitudinal methods, in Hensher D.A. and Button, K.J., eds., *Handbook of Transport Modelling*, 2nd edition, Elsevier, Amsterdam/London, pp. 133-150

Mandel, B., Gaudry, M. and Rothengatter, W. (1997) A disaggregate Box-Cox logit mode choice model of intercity passenger travel in Germany and its implications for high-speed rail demand forecasts, The Annals of Regional Science, 31 99-120.

McFadden, D.L. (1981) Econometric models of probabilistic choice, in Manski C.F. and McFadden D.L., eds., *Structural analysis of discrete data with econometric applications*, MIT Press, Cambridge, pp. 198-272.

Rietveld, P. Bruinsma, F.R., Van Vuuren, D (2001) Coping with unreliability in public transport chains: a case study for The Netherlands, Transportation Research, 35 539-559.

Wardman, M. (1998) The value of travel time: a review of British evidence, Journal of Transport Economics and Policy, 32 285-316.