Transport Policy 67 (2018) 77–84

Contents lists available at ScienceDirect

Transport Policy

journal homepage: www.elsevier.com/locate/tranpol

Trip distribution model for regional railway services considering spatial effects between stations

Rubén Cordera *, Roberto Sañudo, Luigi dell'Olio, Ángel Ibeas

Department of Transportation, Universidad de Cantabria, Av. de Los Castros s/n, 39005 Santander, Cantabria, Spain

A R T I C L E I N F O

Keywords:
Railway
Trip distribution models
Poisson regression
Spatial filtering

A B S T R A C T

The railways are a priority transport mode for the European Union given their safety record and environmental sustainability. Therefore it is important to have quantitative models available which allow passenger demand for rail travel to be simulated for planning purposes and to evaluate different policies. The aim of this article is to specify and estimate trip distribution models between railway stations by considering the most influential demand variables. Two types of models were estimated: Poisson regression and gravity. The input data were the ticket sales and the prices between stations on a regional line in Cantabria (Spain) which were provided by the Spanish railway infrastructure administrator (ADIF – RAM). The models have also considered the possible existence of spatial effects between train stations. The results show that the models have a good fit to the available data, especially the gravity models constrained by origins and destinations. Furthermore, the gravity models which considered the existence of spatial effects between stations had a significantly better fit and provided a more realistic journey pattern in a future scenario than the Poisson models and the gravity models that did not consider these effects. The proposed models have therefore been shown to be good support tools for decision making in the field of railway planning.

1. Introduction

The European Commission transport roadmap (European Commission, 2011) gives priority to the railways because of their proven safety and environmental sustainability compared to road transport. One of the Commission's stated future goals is the creation of a unique European railway space, the introduction of new technological solutions and the construction of new infrastructure financed and priced intelligently.

In order to reach these goals, the European Commission has highlighted the need to evaluate transport projects to guarantee their social profitability and the added value they give to the EU. This evaluation needs to be supported by the available evidence and transport demand models which allow user behaviour to be accurately simulated.

Among the group of transport demand models are trip distribution models which allow the interaction between origin and destination points to be simulated. The most well-known and widely used distribution model has traditionally been the gravity model which, based on the analogy with Newtonian physics, has later been theorized from a probabilistic perspective as a maximum entropy model (Wilson and Bennett, 1985). The state of the art provides many calibration techniques for the parameters of both origin and destination as well as for travel cost (Ortúzar and Willumsen, 2011). Other researchers have insisted on the need to use Poisson type regression models given the discrete and non-negative nature of the journeys (Flowerdew and Aitkin, 1982).

This article proposes the estimation of trip distribution models based on the boarding and alighting data of passengers on a regional railway line. The data used has been obtained from ticket sales on the line provided by the Spanish railway infrastructure administrator (ADIF – RAM). The models were estimated based on two methods: a Poisson type nonlinear regression without any kind of constraint and a Wilson type gravity model doubly constrained to origins and destinations. Both types of models are compared by considering their goodness of fit with the data, in order to determine if the greater number of parameters estimated in the gravity models really does provide greater significance. The models have also been estimated with additional variables to consider the existence of spatial effects between stations to determine if these effects are significant and increase the explanatory capability of the models. Finally, the models have been applied to a rising demand future scenario to check their performance. The results show that gravity models restricted to origins and destinations with additional variables which consider spatial effects like contiguity between stations have a significantly better goodness of fit to the data and provide a more realistic

* Corresponding author.
E-mail address: corderar@unican.es (R. Cordera).

https://doi.org/10.1016/j.tranpol.2018.01.016
Received 22 July 2016; Received in revised form 21 June 2017; Accepted 18 January 2018
Available online 1 February 2018

journey pattern in the future scenario.

A brief review of the state of the art in the field of trip distribution models and distribution models applied to the railways is presented in the following section. The methodology followed is summarised in Section 3 concentrating on Poisson type regression models and doubly constrained gravity models. Section 4 provides a description of the study area and presents and discusses the results obtained by the models. Finally, the conclusions drawn are summarised in Section 5.

2. State of the art about trip distribution models

Spatial interaction models were applied very early on in multiple fields of study for simulating the effects of spatial interaction such as the movement of people between urban areas (Ravenstein, 1885) or commercial flows (Huff, 1959). The first models proposed were based on an analogy with Newtonian gravity theory with the sizes of origins and destinations and the distances between them as explanatory variables. This type of model has a reasonably good fit to the data although they lacked theoretical justification. The theoretical base was provided by Wilson (1970) who showed the possibility of deriving a great number of models from the principle of maximum entropy by which the most probable distribution matrix is the one which maximises the microstates of a given macrostate (Fotheringham et al., 2000). Cochrane (1975) later proposed a derivation of the gravity model from the principle of utility maximization. Exponential gravity models, as derived by Wilson, have proven to be particularly useful for precisely representing the macro level behaviour of a wide variety of micro level interactions (Sen and Smith, 2012). Other authors have later insisted on the convenience of using Poisson type non-linear regression models given their greater adaptability to the trip generation and distribution phenomena (Flowerdew and Aitkin, 1982; Winkelmann and Zimmermann, 1995).

The currently available trip distribution models can be classified into two large groups depending on the data used: models based on aggregate data which uses, for example, ticket sales information and models based on surveys which use disaggregate data on an individual level. Cascetta et al. (2007) also differentiated mixed distribution models which incorporated characteristics of both aggregate and disaggregate models.

The specification of the travel cost function plays a key role in the distribution models so that the predictions fit as closely as possible with the distribution of the observed journeys (Tiefelsdorf, 2003). The most commonly used functional forms in practise are the potential, the exponential and the combined (also known as the Tanner deterrence function) (Cascetta, 2009). While the combined travel cost function is the most appropriate for urban environments where an increase in journeys may occur for small travel costs, in more extensive environments the potential and exponential functions should give a better fit. Travel cost is usually represented through a generalised cost which may include variables like journey time and fare. In the case where the generalised cost is expressed in terms of money, the journey time parameter can be interpreted as the value of time for users (Ortúzar and Willumsen, 2011).

Distribution models have been widely applied in the field of transport planning. Wang et al. (2016) applied a distribution model based on linear regression to the journeys obtained from the entrance and exit validations of contactless tickets at stations on the Beijing metro (China). The model allowed the authors to provide estimations about how journey distance and land use distribution affected journey patterns without having to estimate a complete four stage transport model. However, the authors did not consider the possible existence of spatial effects in the data, which is something that could affect the estimated parameters. In contrast, de Grange et al. (2011) estimated a gravity type distribution model considering spatial correlation. The model was estimated with data from the bus service of the city of Santiago (Chile). The authors concluded that explicitly considering spatial effects in a gravity model could significantly increase its explanatory and predictive capabilities.

In the field of trip distribution models relating specifically to railways, these models allow different planning alternatives to be evaluated. Among the aggregate models based on ticket sales, Wardman (2006) proposed an unrestricted generation-distribution model using time series data for the United Kingdom in the 1990s. The estimated model presented variables corresponding to the characteristics of the origin such as the population, GDP and the rate of motorisation, as well as to the journey such as the overall cost. The author found that GDP was the most important factor in explaining the growth of journeys, even though in a complete four stage model these types of variables are usually introduced into trip generation models. In a similar work applied to railway journeys to and from airports, Lythgoe and Wardman (2002) estimated a demand model based on linear regression which calculated elasticities for different variables like GDP, the fare or journey time.

Where disaggregate data is available, models based on user surveys allow researchers to simulate individual choices considering personal characteristics (age, gender, income, etc.) and transport service characteristics as well as origins and destinations (Ben-Akiva and Lerman, 1985). However, this type of disaggregate model based on random utility theory require greater effort during the data collection phase because they are generally estimated using fewer data than models based on ticket sales.

3. Methodology

Different authors have highlighted the specification problems involved in using a multiple linear regression model (MLR) to estimate the generation and distribution of journeys in a study area (Flowerdew and Lovett, 1988; Thill and Kim, 2005). The dependent variable in distribution models is of a discrete nature, whereas the MLR model assumes a continuous distribution. Therefore, it is desirable to use a model specified with a qualitative dependent variable such as the Poisson regression model (Gujarati and Porter, 2009). This model takes the form:

$$P(Y_i) = \frac{\mu^Y e^{-\mu}}{Y!} \tag{1}$$

The Poisson regression assumes that each dependent variable Y_i is extracted from a Poisson type discrete distribution with the distribution parameter μ_i of (1), logarithmically linked to a linear combination of explanatory variables:

$$\ln(u_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_k X_{ki} \tag{2}$$

Where:

β_k are parameters to be estimated
X_{ki} are the independent variables

The Poisson model cannot be made linear, meaning that the parameters cannot be estimated using Ordinary Least Squares (OLS). Alternative estimation methods such as maximum likelihood implemented through algorithms as reweighted least squares have been proposed, producing good results (Green, 1984).

A particular case of the Poisson model appears when all the independent variables are specified as dummy variables. In this case the Poisson model is equivalent to a log-linear model as both the dependent variable and the independent are qualitative. Log-linear models are more frequently used for modelling contingency tables (Agresti and Kateri, 2011). This type of model can be specified as totally saturated, in other words, with a perfect fit to the data as a parameter is specified for each observation. Willekens (1983) has shown how log-linear models are equivalent to the gravity models if they are conveniently scaled, usually by equalling the equilibrium factors of the first origin and destination to 1.

The fit of a Poisson model can be evaluated through different indicators as the Akaike information criteria (AIC), the log–likelihood or through the difference in the log-likelihood of the model estimated with respect to the totally saturated model, in other words, using a likelihood ratio test (LR) of the following kind:

$$LR = -2\left[L\left(\widehat{\theta}_0\right) - L\left(\widehat{\theta}_s\right)\right] \tag{3}$$

Where:

$L(\widehat{\theta}_0)$ is the log – likelihood of the estimated model
$L(\widehat{\theta}_s)$ is the log – likelihood of the saturated model

This type of test asymptotically distributes χ^2 with r degrees of freedom. In this case r is the difference between the number of estimated parameters in the not saturated model and the number of parameters estimated for the saturated model. The LR test can only be used to compare the fit between general models and their constrained versions with fewer parameters which is the case of the Poisson models estimated without and with constraints on origins and destinations.

The variables to be introduced into the model will vary according to the problem being addressed. A trip distribution model estimated using a Poisson regression is usually specified with three variables: a variable of the trips produced by the origin, a variable of trips attracted by the destination and a travel cost variable between both zones, where the variables of the produced and attracted trips are usually extracted from a trip generation model (Hall, 2012). Therefore, this type of model would not present any kind of constraint although it could have problems of spatial autocorrelation in the origins or destinations which would be convenient to address to guarantee the reliability of the estimated parameters (Griffith, 2007). One of the techniques which is available for addressing this spatial autocorrelation in nonlinear models is Spatial Filtering (Tiefelsdorf and Griffith, 2007) where the spatial effects are separated from the rest of the non-spatial effects, thereby eliminating the possible correlation present in a neighbourhood matrix.

The Poisson regression can also be specified with constraints on the origins or destinations by estimating a different parameter for each zone. The case of a doubly constrained model with a travel cost variable leads to the well-known gravity distribution model derived from the principle of maximum entropy (Wilson, 1970):

$$T_{ij} = A_i O_i B_j D_j \exp\left(-\beta C_{ij}\right) \tag{4}$$

Where:

T_{ij} are the trips between zones i and j
O_i are the trips produced by zone i
D_j are the trips attracted by zone j
C_{ij} are the costs between zone i and zone j
β is an parameter to be estimated

The travel cost parameter β can be estimated using different procedures like the method proposed by Hyman (1969) or using a log linear model (Dennett, 2012). There is also the possibility of estimating the model with a combined travel cost function with two parameters which would offer a better fit in urban areas. The balancing factors A_i and B_j are codependent, meaning they need to be estimated iteratively using, for example, the method proposed by Furness (1965):

$$A_i = \frac{1}{\sum_j B_j D_j \exp\left(-\beta C_{ij}\right)} \quad B_j = \frac{1}{\sum_i A_i O_i \exp\left(-\beta C_{ij}\right)}$$

Given the constraints on the origins and destinations of the model, the resulting fits are usually high. However, it is possible to introduce new variables into the model in order to consider other spatial effects. Flowerdew (2010) has proposed inserting dummy variables into the model to consider zonal contiguity, as depending on the type of trip being modelled, the contiguous zones may be a more or less likely destination than the rest of the areas. This type of spatial effect may help in improving the fit of the models by adapting them to the peculiarities of each study area.

4. Study area and results

4.1. Available data and the study area

The trip distribution models have been estimated using data provided by the Spanish railway organisation ADIF – RAM about ticket sales on a narrow gauge regional line in Cantabria (Spain). The ticket sales provide information on both the origins and destinations of the passengers meaning the trip matrix gives an exact representation of travel on the line. ADIF – RAM also provided the prices of tickets needed to travel between each pair of stations.

The studied line has a total of 23 stations being the two terminals located at Santander and Cabezón de la Sal (see Fig. 1). The data obtained corresponds to the week from 19th to 25th January 2015 and counted 26,371 passengers. The stations with the highest production and attraction trips were the two largest towns in the region, Santander and Torrelavega, which accumulated more than 50% of the passengers given their higher demographic weight.

The variables contained in the database can be seen in Table 1. Between all the O-D pairs there is an average of 52.5 trips with a maximum of 2,900 trips corresponding to the Santander – Torrelavega pair. The travel cost between the pairs has been specified through a generalised cost (C_{ij}) measured in Euros which combines the journey time between the stations (in minutes) with the fare variable between the stations. The value of time was provided by a previous study based on surveys asked to regional train users with a final weight of 0.25 € per minute of journey time (Grupo de Investigación de Sistemas de Transporte, 2008). This case assumes a fixed value of time because the line is of limited length (about 45 km). In cases where the lines were much longer it would be reasonable to assume a variable value of time which increased as a function of distance (Wardman, 1998).

Two dummy variables were also included in the database to consider the possible presence of spatial effects. A variable of contiguity between stations taking a value of 1 if the stations are adjacent, and a variable which takes a value of 1 in the Santander – Torrelavega and Torrelavega – Santander pairs. This latter variable could be important because, as can be seen in Fig. 2, the number of trips in the cost interval of 10–15 euros increases with respect to the interval 5–10 Euros due largely to the journeys produced between the two towns.

The possibility of estimating the model with a combined travel cost function was tested, but in this case the results were no better in terms of the fit to the observed journeys. This fact is certainly due to the interurban nature of the modelled journeys, without an increase in their number for low travel costs.

4.2. Results and discussion of the models

The parameters estimated for the seven models are summarised in Table 2. The first four (P-1 a P-4) correspond to Poisson type regression models, while the three latter are Wilson type gravity models.

The P-1 model was specified with the totals produced and attracted by the origin and destination stations, using the generalised cost between them as independent variables. The production and attraction parameters were identical and had a positive sign, whereas the travel cost parameter was, as expected, negative. Furthermore, all the parameters were clearly significant. The parameters show, using the transformation $100^*(e^\beta - 1)$, that one unit change in production and attraction generates, ceteris paribus, 0.05% more trips. However, an increase of one euro in the generalised cost implies about a 9% reduction in the number of trips being made. According to the AIC index the model had a fit of 20,279 and an R^2 of 0.85 for the estimated journeys compared with the observed journeys. The P-2 model adds to the variables of the P-1 model, the dummy variable of contiguity between stations, which showed a

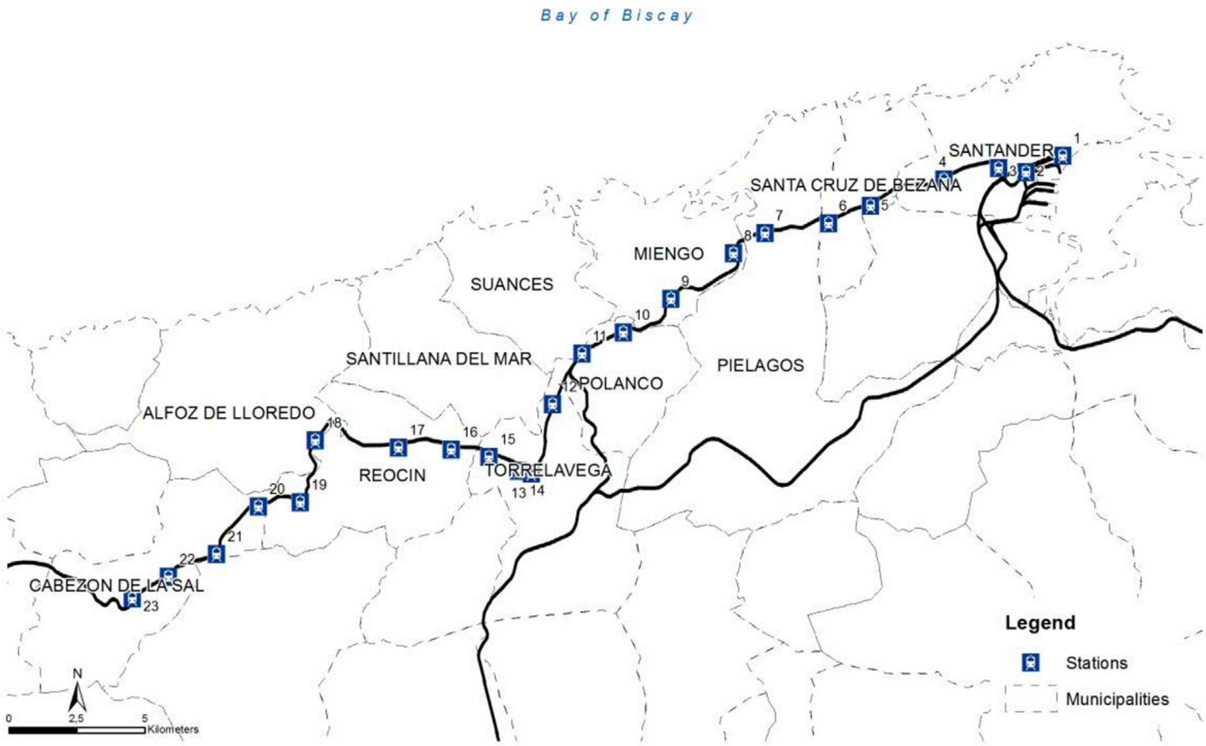

Fig. 1. Stations on the narrow gauge line Santander – Cabezón de la Sal.

Table 1
Descriptive statistics of the variables contained in the database.

Variable	Description	Units	Average	Standard Deviation	Minimum	Maximum
V_{ij}	Trips between origin i and destination j	No. Trips	52.47	255.40	1	2,900
O_i	Trips produced by origin i	No. Trips	1,146.57	2,083.57	43	8,842
D_j	Trips attracted by destination j	No. Trips	1,146.57	2,066.93	43	8,913
C_{ij}	Generalised cost between i and j	Euros	8.59	4.80	1.90	21.05
Cont	Dummy variable if the stations are contiguous	1/0	0.09	0.28	0	1
SantTorre	Dummy variable if the O-D pair ij corresponds to Santander and Torrelavega	1/0	0	0.06	0	1

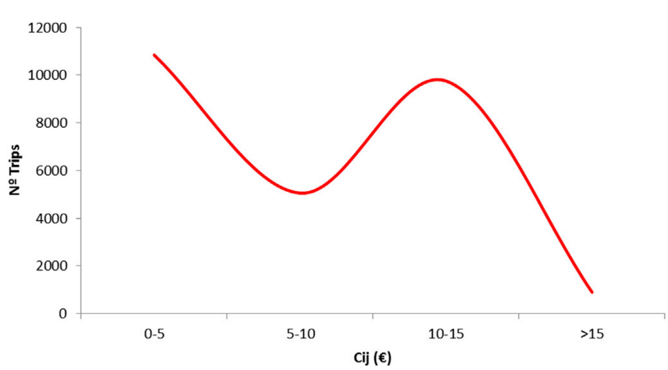

Fig. 2. Histogram of journeys according to generalised cost.

negative sign. This sign provides evidence that, if a greater number of journeys are made between points with low generalised costs (see Fig. 2), these are not normally made between adjacent stations given that the parameter implies a reduction of around 72% in the number of journeys. The P-2 model had a slightly better fit than P-1 according to the AIC index as well as a superior R^2 comparing the estimated with the observed journeys. The Poisson P-3 model included an additional dummy variable corresponding to whether the O-D pair was Santander – Torrelavega or Torrelavega – Santander. The sign of the parameter was negative with a reduction of 10% in the number of expected trips which is almost certainly due to the fact that the O and D factors overestimate the potential for interaction between the two locations. This model had a slightly better fit than P-2 with all the estimated parameters being clearly significantly different from 0. The specification of P-3 is therefore:

$$\ln(u_{ij}) = \beta_1 + \beta_2 O_i + \beta_3 D_j + \beta_4 C_{ij} + \beta_5 Cont_{ij} + \beta_6 SantTorre_{ij} + \varepsilon_{ij} \quad (5)$$

Finally, the P-4 model was estimated using the Spatial Filtering technique to eliminate the possible presence of spatial correlation in the origins and destinations (Griffith, 2007). All the pairs with identical origins or identical destinations were considered to have neighbourhood relationships. The spatial filtering selected two eigenvectors, one at origins (EvO) and another at destinations (EvD), which were introduced into the Poisson regression. The fit of the model increased and reduced the AIC to 16,880, the Root Mean Squared Error (RMSE) to 76.59 and the Standardized Root Mean Squared Error (SRMSE) to 0.96. This latter indicator is recommended by Fotheringham and Knudsen (1987) because it is less sensitive to the magnitude of the data.

The Wilson gravity type models are summarised in columns W-1 to W-3 in Table 2. Rows A_i and B_j show the average of the 23 balancing factors estimated for origins and destinations respectively, using the Furness method with a stop criterion of a 0.1% maximum change in the factors between one iteration and the next. The rest of the parameters are the same as those specified in the Poisson type models, having been estimated using a log-linear model which also allows their statistical

Table 2
Estimated Distribution Models (in brackets the p – value with the statistical significance of the parameters).

Variable	P-1	P-2	P-3	P-4	W-1	W-2	W-3
(Intercept)	1.9490 (.000)	2.2190 (.000)	2.1460 (.000)	1.7180 (.000)	–	–	–
O/A_i	0.0005 (.000)	0.0005 (.000)	0.0005 (.000)	0.0005 (.000)	0.0001	0.0002	0.0002
D/B_j	0.0005 (.000)	0.0005 (.000)	0.0005 (.000)	0.0006 (.000)	0.8846	0.9690	0.9307
C_{ij}	−0.0969 (.000)	−0.1187 (.000)	−0.1156 (.000)	−0.1092 (.000)	−0.1102 (.000)	−0.1690 (.000)	−0.1652 (.000)
Cont	–	−1.2890 (.000)	−1.2800 (.000)	−1.2460 (.000)	–	−2.3657 (.000)	−2.4135 (.000)
SantTorre	–	–	−0.1097 (.000)	−0.5648 (.000)	–	–	−0.2789 (.000)
EvO	–	–	–	−8.9470 (.000)	–	–	–
EvD	–	–	–	4.4020 (.000)	–	–	–
AIC	20,279	18,614	18,591	16,880	10,295	6,998	6,910
R^2	0.85	0.88	0.89	0.91	0.94	0.98	0.99
RMSE	98.84	85.64	83.78	76.59	55.12	32.56	28.41
SRMSE	1.24	1.07	1.05	0.96	0.69	0.41	0.36
Residual Deviation	18,531	16,864	16,839	15,124	8,463	5,164	5,074

significance to be estimated. The fit of the constrained gravity models was better than that of the Poisson regression models with R^2 superior to 0.9 in all cases, up to a fit of 0.99 for the observed data in model W-3 considering the contiguity of the stations and the specific interaction between Santander and Torrelavega. The W-3 model was specified as:

$$V_{ij} = A_i O_i B_j D_j \exp\left(\beta_4 C_{ij} + \beta_5 Cont_{ij} + \beta_6 SantTorre_{ij}\right) \tag{6}$$

The parameter of the W-3 travel cost variable highlights a journey reduction of 15% for each additional euro in cost. The Cont (β_5) parameter clearly had one magnitude greater than it did in the Poisson models, estimating a reduction of 91% in the journeys between adjacent stations compared with what would be obtained by only considering the row, column and travel cost factors. The SantTorre (β_6) parameter also had a greater magnitude than in the P-3 model estimating a reduction, ceteris paribus, of 24% in the number of journeys.

If an LR test is conducted between the gravity and Poisson regression models, the former show a test value which is clearly superior to the critical value even considering the greater number of parameters used by the constraints on the origins and destinations. This is the case, for example, with the W-3 model compared with P-3, where the test presented a value greater than 11,000 for a critical value of 95% of the confidence level of 55.8.

If the residual deviation between the estimated models and the completely saturated model is considered, the test value was always superior to the critical value of the distribution, although the Wilson type models clearly got closer to the maximum fit provided by the saturated model. The RMSE and the SRMSE between the modelled and observed values showed similar results with the best fit in the W-3 model.

An examination of the fit of the models with respect to the observed data by cost ranges (see Fig. 3 and Fig. 4) shows how the Poisson models have a worse fit for the intermediate cost ranges (5–10 and 10–15 euros). On the other hand, the gravity models and especially the W-2 and W-3 models with dummy variables considering spatial effects showed a better fit over all the cost ranges. The fit provided by these models was significantly better than that of the W–1 model using the LR test with one (W–2) or two degrees of liberty (W–3).

An analysis of the highest residuals of the P-1 model shows how (see Fig. 5) the W-1 and W-3 models reduced both the positive errors and the negative errors thanks to the constraints on the origins and destinations. The P-4 and W-3 models also enabled a reduction in the negative residuals caused by the over prediction of trips between contiguous stations (see the 1–2 and 2-1 pairs in Fig. 5) which represented 11% of the overall errors in the P-1 model. So, while the P-1 model showed a quotient of 0.33 between the observed trips and the modelled trips between contiguous stations, the P-4 model's quotient was 0.98. Similar values were also found with the W-1 (0.32) and W-3 (0.98) models, respectively. In the cases of the Santander – Torrelavega (1–13), Torrelavega – Santander (13-1) pairs, the errors of the model represented 2% of the total errors and the quotient between the observed trips and the modelled trips changed from 0.94 in the P-1 model to 0.99 in the W-3 model.

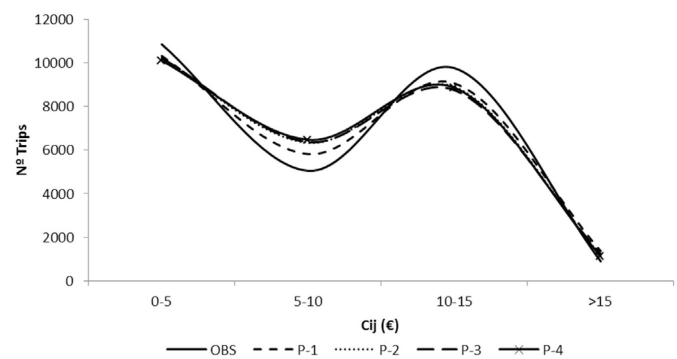

Fig. 3. Histogram of the observed trips compared to estimated trips for the unrestricted Poisson regression models.

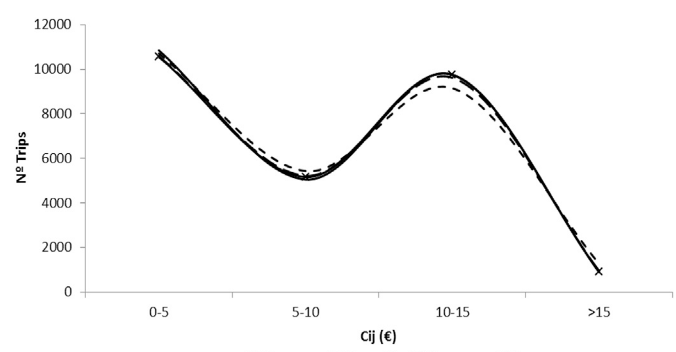

Fig. 4. Histogram of the observed trips compared with the estimated trips for the gravity models.

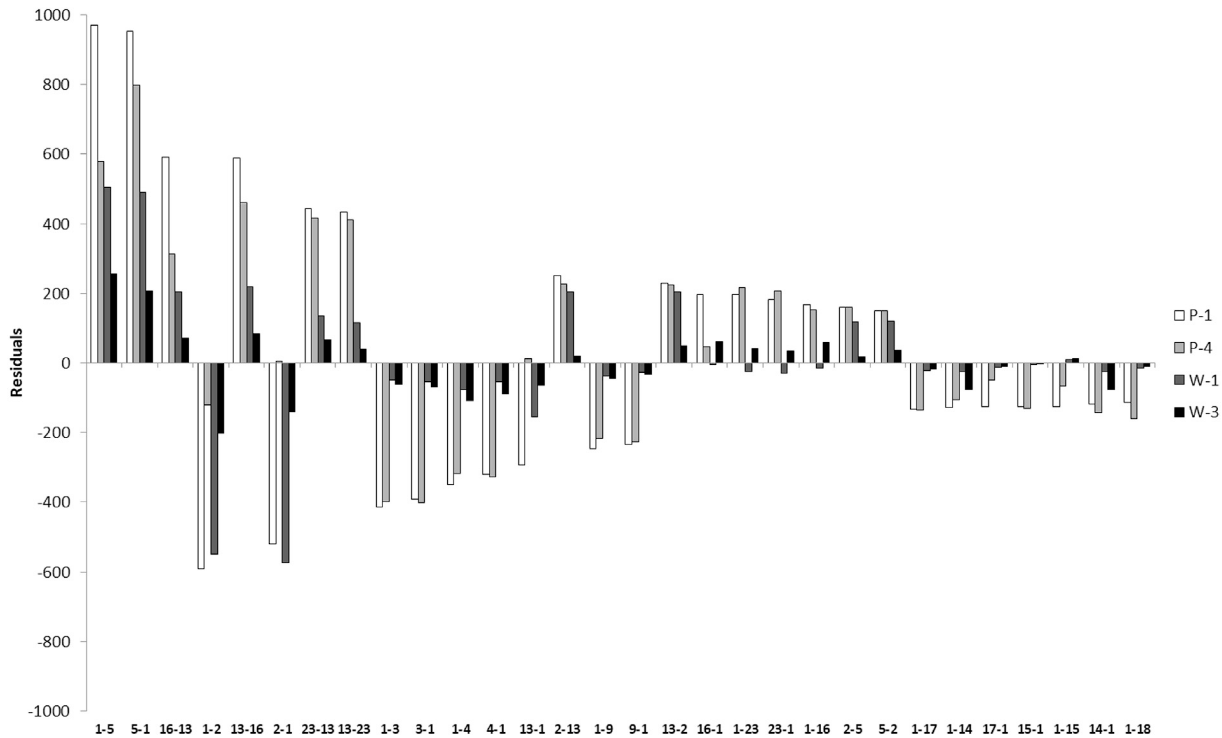

Fig. 5. Residuals of the P-1, P-4, W-1 and W-3 models in the 30 origin – destination pairs with the highest absolute residuals in the P-1 model.

4.3. Application of the distribution models to the prediction of journey patterns

The simpler P-1 and W-1 models will be applied along with those models estimated with more parameters, P-4 and W-3, for predicting journey patterns resulting from changes in the production and attraction of trips from stations on the studied railway line. These models could also be applied, as suggested by Raymer (2007), to the estimation of new origin-destination trip pairs (e.g. a new train station) by applying the ratio between the observed and the predicted trips between existing origin-destination pairs. In this example, however, only changes in the number of trips produced and attracted at stations along the current line will be simulated.

It is assumed that stations 1, 13 and 23 along the studied line experience increased trip production and attraction, as shown in Table 3. These increases could be obtained from a trip generation model and be due to various factors like population or economic growth in the catchment areas or an increase in the number of available services, among others.

Fig. 6 describes trips estimated by the P-1, P-4, W-1 and W-3 models for the 30 Origin-Destination pairs with the greatest average journey demand in all the models. The journeys produced in the current situation have also been represented (without increases in the produced/attracted journeys) to provide a reference pattern. It can be seen how the P-1 and P-4 models estimated much more trips than the W-1 and W-3 models or the current situation for the pairs where the origin or the destination is the station at Santander (1) and particularly for the Santander – Torrelavega (1–13) or the Torrelavega – Santander (13-1) pairs. This is because the P-1 and P-4 models had no constraints on the origins and destinations allowing them to overestimate the number of journeys produced between origin-destination pairs with high overall rates of trip production and attraction, as is the case of the station at Santander. However, on the contrary, between stations with a lower capacity for producing and attracting journeys, such as is the case of Cabezón de la Sal (23), the effect is the opposite and the W-1 and W-3 models estimate more journeys than the P-1 and P-4 models. It is also worth noting how between adjacent stations (pairs 1–2 and 2-1) the P-4 and W-3 models, which consider this spatial effect, estimate a considerably lower number of journeys than the P-1 and W-1 models, at closer magnitudes to the current situation.

Table 3
Changing trip production and attraction scenario at stations.

Station	Growth in trip production	Growth in trip attraction
1-Santander	+5%	+4%
13-Torrelavega	+3%	+4%
23-Cabezón de la Sal	+1%	+4%

The journey patterns obtained by the W-1 model and above all by the W-3 model, therefore seem to be more realistic and consistent with an increase in the capacity of certain stations to produce and attract journeys, when compared with the current situation.

5. Conclusions

This article has presented the estimation of trip distribution models using two methods: nonlinear Poisson regression and gravity models with constraints on origins and destinations. The goal was to assess whether or not the gravity models fit to the data significantly better considering they require a greater number of parameters. Additional variables have also been introduced to account for the spatial effect of contiguity between stations controlled by the effect of spatial correlation which may be present in the trip distribution data. The estimated models could be useful tools for simulating changes that passengers make in their choice of destination as a result of new policies such as the opening and closing of stations or changes in the service conditions.

The results confirm that the gravity model with constraints on origins and destinations had a significantly better fit to the data, according to the LR test, than the Poisson regression models without constraints. This fact was true even considering that the gravity models were estimated with 40–42 more parameters and that in a Poisson model the presence of spatial correlation was controlled. The models that considered contiguity between stations and the specific effects of interaction also showed a

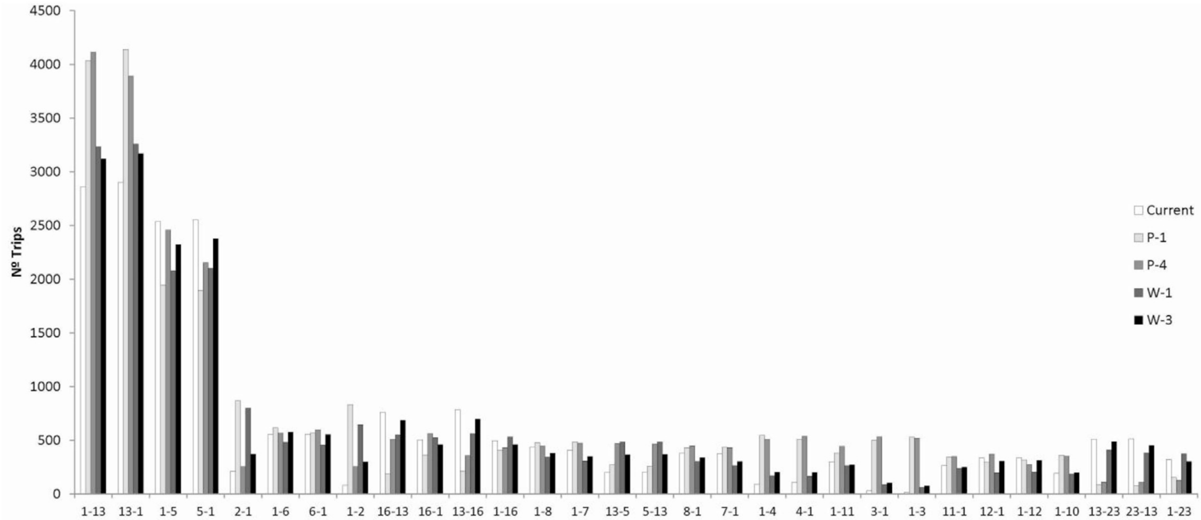

Fig. 6. Number of journeys in the current and simulated situations for the 30 origin – destination pairs with the highest average number of journeys.

significantly better fit with only one or two more parameters than the models that did not consider these effects. It would therefore seem recommendable to estimate gravity models constrained by production and attraction data obtained from a trip generation model when creating a trip distribution model. Even more so when to estimate a gravity model using a log-linear model does not imply any additional costs other than those involved in the iterations needed to obtain the balancing factors. The possibility of specifying addition spatial variables also gives the model an extra capacity of adaptation to the study area.

The prediction exercise performed on a future scenario also showed how the constrained gravity models and in particular the more complex model considering contiguity between stations produced a more realistic journey pattern in accordance with the current situation. This kind of distribution model combined with the trip generation models could therefore be easily applied to the simulation of different events and policies occurring to the railways, such as the changing demand at stations, the creation of new stations and changes in journey travel cost, among others. A realistic journey pattern between stations for future scenarios would be a very useful support tool for operators in order to correctly plan the supply of railway services.

A future line of research would be the estimation of gravity models which consider the presence of spatial autocorrelation at origins, destinations and at points of interaction between O-D pairs. The estimation of this type of model currently requires considerable computing power which makes necessary additional research (Griffith, 2009).

Acknowledgements

This research was made possible thanks to financing from the Horizon 2020 Framework Programme, Shift2Rail Joint Undertaking 2015 call through the "NEAR2050: future challenges for the rail sector" project (Grant Number: 730838). The authors would also like to thank ADIF – RAM for having provided the data used.

References

Agresti, A., Kateri, M., 2011. Categorical Data Analysis. Springer.
Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete Choice Analysis : Theory and Application to Travel Demand. MIT Press, Cambridge, Mass.
Cascetta, E., 2009. Transportation Systems Analysis : Models and Applications, second ed. Springer, New York.
Cascetta, E., Pagliara, F., Papola, A., 2007. Alternative approaches to trip distribution modelling: a retrospective review and suggestions for combining different approaches. Pap. Reg. Sci. 86 (4), 597–620.
Cochrane, R., 1975. A possible economic basis for the gravity model. J. Transport Econ. Pol. 34–49.
de Grange, L., Ibeas, A., González, F., 2011. A hierarchical gravity model with spatial correlation: mathematical formulation and parameter estimation. Network. Spatial Econ. 11 (3), 439–463.
Dennett, A., 2012. Estimating Flows between Geographical Locations:'get Me Started in'spatial Interaction Modelling. UCL working paper series.
European Commission, 2011. Roadmap to a Single European Transport Area — towards a Competitive and Resource-efficient Transport System. European Commission.
Flowerdew, R., 2010. Modelling migration with Poisson regression. Technol. Migrat. Commuting Anal. 261–279.
Flowerdew, R., Aitkin, M., 1982. A method of fitting the gravity model based on the Poisson distribution. J. Reg. Sci. 22 (2), 191–202.
Flowerdew, R., Lovett, A., 1988. Fitting constrained Poisson regression models to interurban migration flows. Geogr. Anal. 20 (4), 297–307.
Fotheringham, A.S., Brunsdon, C., Charlton, M., 2000. Quantitative Geography: Perspectives on Spatial Data Analysis. Sage Publications, London; Thousand Oaks, Calif.
Fotheringham, A.S., Knudsen, D.C., 1987. Goodness-of-fit Statistics. Geo Books.
Furness, K., 1965. Time function iteration. Traffic Eng. Contr. 7 (7), 458–460.
Green, P.J., 1984. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. J. R. Stat. Soc. Series B (Methodological) 149–192.
Griffith, D.A., 2007. Spatial structure and spatial interaction: 25 years later. Rev. Reg. Stud. 37 (1), 28.
Griffith, D.A., 2009. Modeling spatial autocorrelation in spatial interaction data: empirical evidence from 2002 Germany journey-to-work flows. J. Geogr. Syst. 11 (2), 117–140.
Grupo de Investigación de Sistemas de Transporte, 2008. Optimización de la Coordinación Intermodal mediante la Modelización del Comportamiento del Usuario - INTERCOR. Ministerio de Fomento, Grupo de Investigación de Sistemas de Transporte, Ministerio de Fomento, Gobierno de España.
Gujarati, D.N., Porter, D.C., 2009. Basic Econometrics, fifth ed. McGraw-Hill Irwin, Boston.
Hall, R., 2012. Handbook of Transportation Science. Springer Science & Business Media.
Huff, D.L., 1959. University of Washington Business Review. Geographical Aspects of Consumer Behavior, vol. 18, pp. 27–37.
Hyman, G., 1969. The calibration of trip distribution models. Environ. Plann. 1 (3), 105–112.
Lythgoe, W., Wardman, M., 2002. Demand for rail travel to and from airports. Transportation 29 (2), 125–143.
Ortúzar, J.d.D., Willumsen, L.G., 2011. Modelling Transport. John Wiley & Sons.
Ravenstein, E.G., 1885. The laws of migration. J. Roy. Stat. Soc. 48, 68.
Raymer, J., 2007. The estimation of international migration flows: a general technique focused on the origin-destination association structure. Environ. Plann. 39 (4), 985–995.
Sen, A., Smith, T., 2012. Gravity Models of Spatial Interaction Behavior. Springer Science & Business Media.
Thill, J.-C., Kim, M., 2005. Trip making, induced travel demand, and accessibility. J. Geogr. Syst. 7 (2), 229–248.
Tiefelsdorf, M., 2003. Misspecifications in interaction model distance decay relations: a spatial structure effect. J. Geogr. Syst. 5 (1), 25–50.
Tiefelsdorf, M., Griffith, D.A., 2007. Semiparametric filtering of spatial autocorrelation: the eigenvector approach. Environ. Plann. 39 (5), 1193–1221.
Wang, S., Sun, L., Rong, J., Hao, S., Luo, W., 2016. Transit trip distribution model considering land use differences between catchment areas. J. Adv. Transport 50 (8), 1820–1830.
Wardman, M., 1998. The value of travel time: a review of British evidence. J. Transport Econ. Pol. 285–316.

Wardman, M., 2006. Demand for rail travel and the effects of external factors. Transport. Res. Part E Logist. Transport. Rev. 42 (3), 129–148.
Wilson, A.G., 1970. Entropy in Urban and Regional Modelling. Pion, London.
Wilson, A.G., Bennett, R.J., 1985. Mathematical Methods in Human Geography and Planning. Wiley, Chichester, [West Sussex]; New York.
Willekens, F., 1983. Log - linear modelling of spatial interaction. Pap. Reg. Sci. 52 (1), 187–205.
Winkelmann, R., Zimmermann, K.F., 1995. Recent developments in count data modelling: theory and application. J. Econ. Surv. 9 (1), 1–24.