NEW UNCONVENTIONAL SOURCES OF INFORMATION FOR TRANSPORT ANALYSIS. THE CASE OF A MODEL FOR PUBLIC TRANSPORT SERVICES IN MAJORCA

Oscar Martínez-Alvaro
Transport Professor (UPM). Transport Researcher (Transyt); Managing Partner (Tool Alfa)
Iván Hernández Santana
Junior Consultant (Tool Alfa)

ABSTRACT

Most transport modeling requires the use of data available in official sources plus complementary information that helps to fill gaps. This later information is mostly obtained with the help of private data and field data collection. But sometimes this field data collection is not possible for several reasons. On the other hand, there are more and more sources of information readily available, which may provide indirect clues about users' behavior.

Within the framework of a bidding process, a modelling of Majorca's public transport was carried out in order to ascertain elasticities of demand to several parameters of supply that were expected to change. For that purpose, there appeared a need of getting detailed information on tourist activities, which are a huge share of total mobility in summertime. Official information on population is rather detailed, but tourism data are available only on an aggregated basis.

Some indirect sources of information were analyzed. The easiest and most useful ones were *TripAdvisor* and *Booking*. The former is sort of a macro-survey of popularity and it seems logical that those nuclei with more opinions are those most visited, thus acting as an indicator of attraction. The latter provides information on location of accommodation offered, thus providing a tool for the geographical distribution of trip generation on a very detailed basis. In total, 933,934 opinions for 198 nuclei from *TripAdvisor*, and 2,114 accommodation assets for 51 Municipalities from *Booking* were used. With this information it was possible to calibrate different models for the whole Island, this opening the door to further analyses in the field of transportation and mobility.

The Transport Consortium of Majorca (CTM), entity in charge of planning and management of scheduled public transport services on that island, published in March 2019 the call for tenders for all intercity public passenger transport services by bus. It opened a scenario of important changes (CTM 2017a and 2017b) that sought to increase the total demand of the intercity bus network by 25% between 2015 and 2028, by means of a substantial variation in the supply side (travel times, timetables, connectivity between lines, etc.). For example, an increase in supply between 35.8% and 73.1% of vehicle-km was planned, with different

figures for each of the three packs into which the tender was divided. In addition, a new fare system was introduced, with a very complex structure that imposed a heavy surcharge when purchasing the ticket on the bus. However, no technical study with more information on the expected demand figures was released.

Under these circumstances, any projection of the historical data was irrelevant and one of the bidding companies, which also operated a good part of the existing services, felt the need of calculating the expected demand with scientific tools.

The analysis of the problem in question faced, fundamentally, an important lack of detail of the available information:

- As for transport supply or demand, the official information was limited to total figures by line (CTM, 2017 and “Open Data” site).
- In terms of demand factors, the greatest disaggregation available in official sources (ibestat) was at municipal level.

The following sections focus on the cross-section analysis of the general services, without dealing with other issues analyzed at the time, such as the new lines to the airport and the time-series modeling.

1. THE CONTEXT

Majorca has about 3,600 km² and 868,700 inhabitants (2017). It has 53 municipalities, overshadowed by the capital, Palma, with 406,500 inhabitants (47% of the total). It is a top-level tourist destination: in 2017 Majorca received 11.6 million tourists.

The most important tourist areas are Calvià and Palma, as well as the east coast, while the municipalities in the center of the island have a much lower tourist activity. Tourism presents strong seasonality, with the maximum in the summer months and small volumes in winter.

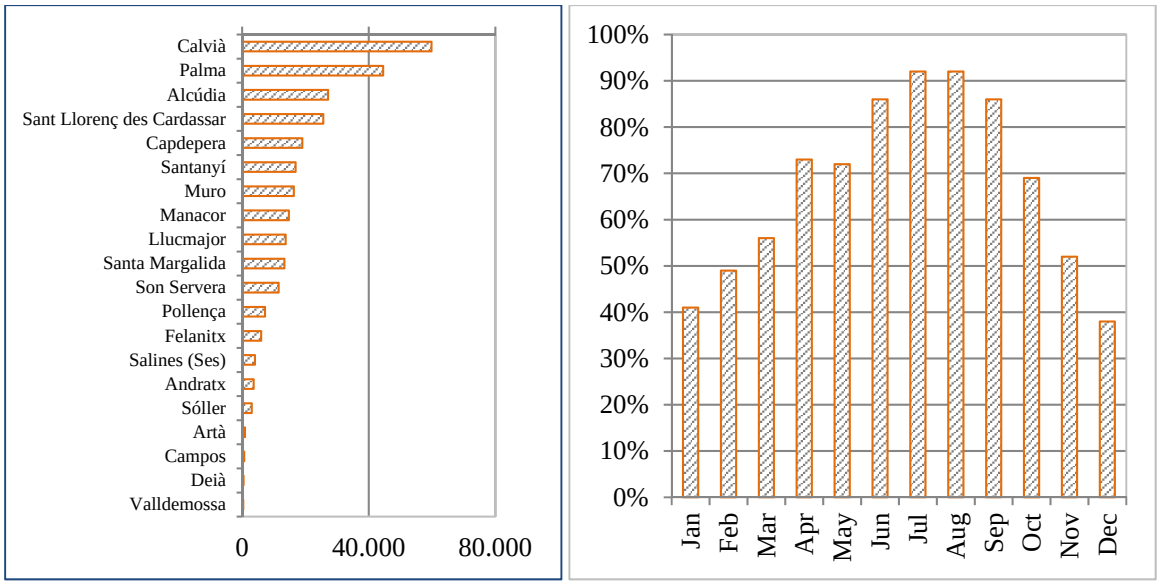

Fig. 1 - Bed capacity of the 20 municipalities with the highest volume (left) and hotel occupancy of the whole island per month (right). (2017). Source: Own elaboration with IBESTAD data

The number of vehicles is very high, due to the presence of an important rent-a-car activity. Vehicles are distributed territorially by fiscal considerations: the most extreme case is this of Escorca, where the car ownership rate is 22,000 vehicles per 1,000 inhabitants.

The island has two railway lines and one subway line, but its territorial coverage is feeble, particularly in tourist areas, with the exception of the Sóller railway, a tourist attraction in itself. The structure of the intercity bus network is essentially this of spokes directed towards Palma, with great concentration on the east coast.

The demand for the rail is not subject to the same seasonality of the intercity bus, whose demand is closely linked to tourism. The correlation between the airport demand and the bus demand is remarkable, while the huge share of the single ticket for bus users is another proof of the sporadic nature of its patronage.

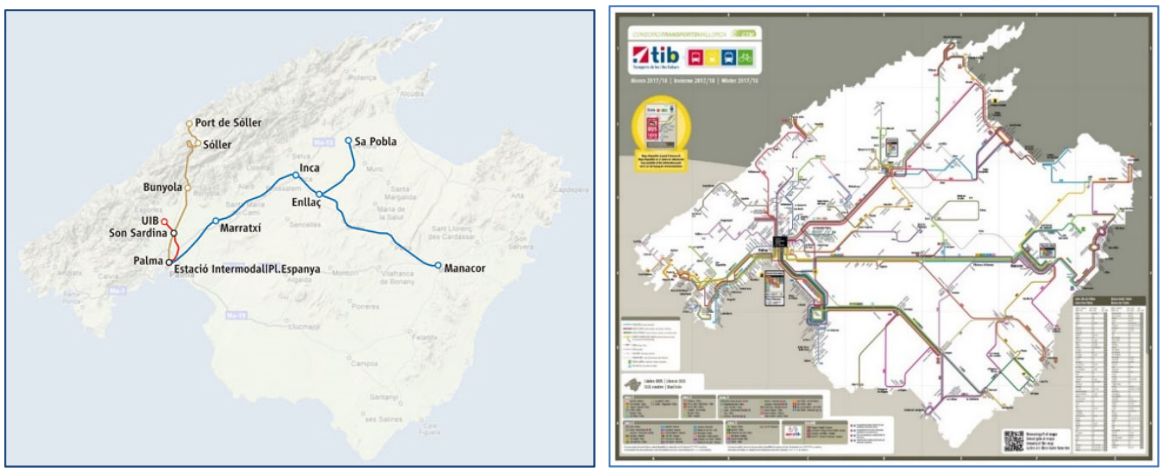

Fig.2 - Railway network (left) and intercity bus (right). 2017. Source: trenscat.com and CTM (2018)

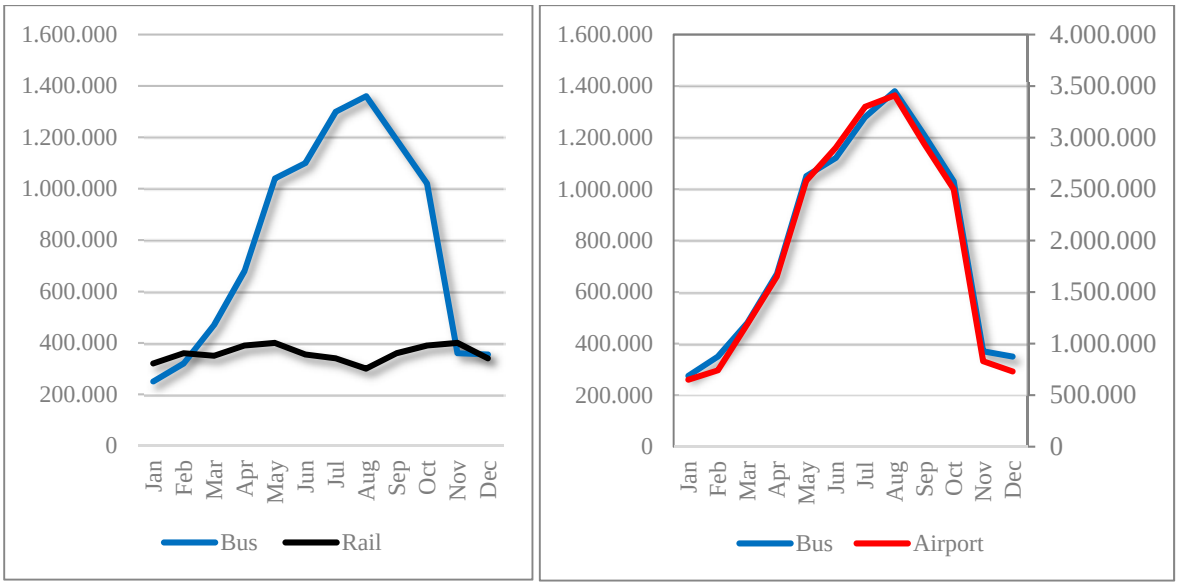

Fig. 3 - Monthly demand by transport mode: Bus and rail (left) and bus and plane (right). 2015. Source: CTM (2017 c).

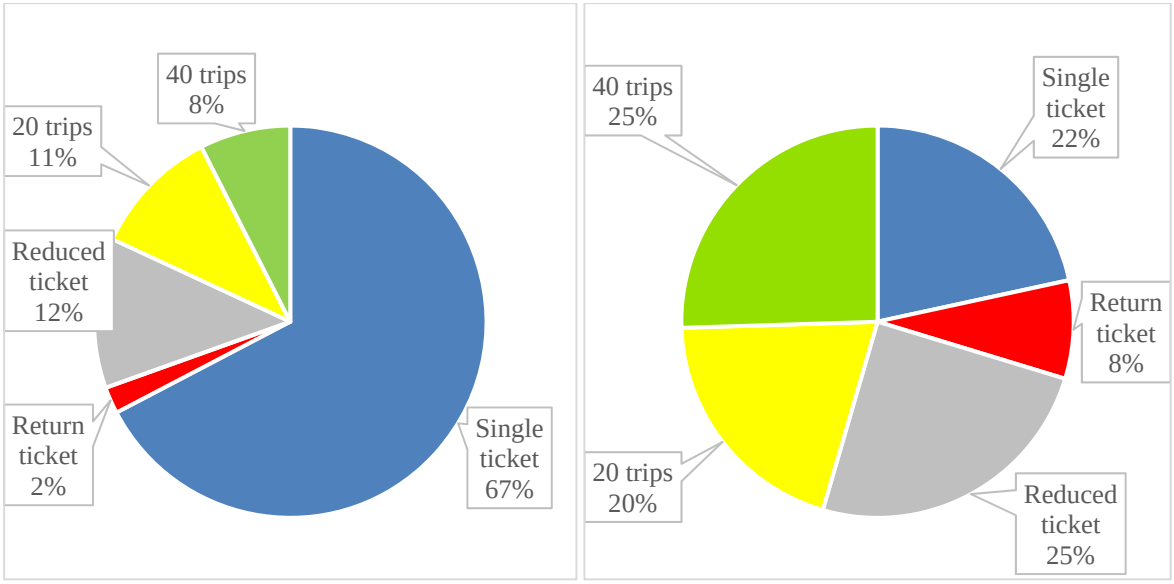

Fig.4 - Share of transport tickets: Bus (left) and Rail (right). 2016. Source: CTM (2017 c).

2 THE INPUTS

2.1Conventional sources

The information available for calibration was initially of two types:

- The one from official sources, with disaggregation, at best, at municipal levela or at line level.
- The one available by the operator itself, of a strictly internal nature, on a line basis lines (lines 100, 102, 104, 105, 106, 107, 111, 500, 501, 502, 503, 505, 507, 515, 520, 525, 530 y 811). It included demand data for 271 Origin-Destination pairs in nine tariff groups and other data (pax-Km, total revenues per line, etc.).

Given the complexity of the variation of the new tariff structure (based on number of areas crossed and accumulation of travel points), a face-to-face survey of revealed preferences was carried out among almost one thousand users, which shed enough light on that particular problem.

2.2 Possible alternatives

As the problem of lack of enough granularity to characterize the demand factors persisted, a research of alternative means was carried out. Many references to the potential of big data were found (Anda, 2017) and a large sample of analyses based on mobile phone data were of public domain (Alexander, 2015; Bonnel, 2015), besides many other related to different innovative sources of information, such as smart cards. But no solution that could identify trips on a detailed basis was available. In particular, mobile phone data could not provide transport mode information with the required reliability.

Some warnings were also found on the topic of potential problems of trying to use data not oriented to mobility analyses, mainly in terms of data volume availability and owners' willingness to allow it to be used by third parties (Milne y Watling, 2019). But something had to be done and two highly atypical sources of information showed some interest:

- *TripAdvisor* as an indicator of attraction.
- *Booking* as an indicator of tourist residence location.

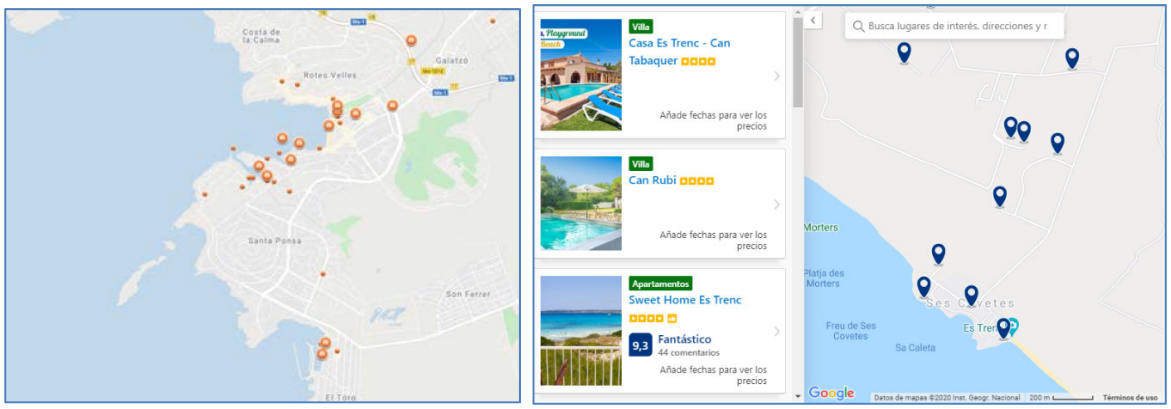

Fig. 5 - Screenshots of TripAdvisor (left) and Booking (right) websites. Source: TripAdvisor.es and Booking.com.

2.3 TripAdvisor

TripAdvisor.com is a website owned by Tripadvisor, Inc. that provides reservations for travel-related activities, with the particularity of collecting user reviews. In that sense, *TripAdvisor* acts as a popularity survey. The basis of using this source is the hypothesis that the number of visits to a certain site of tourist interest is strongly correlated with the number of opinions recorded on *TripAdvisor*, irrespective of their attitude (positive, neutral or negative).

In total, 198 locations were analyzed manually on which 933,934 opinions were found.

The information obtained was as follows:

- Location of the tourist site.
- Type of tourist interest (general, leisure, accommodation, catering).
- Number of opinions collected.

2.4 Booking

In Majorca there is official information on the number of accommodation places per municipality, but without further disaggregation. As there is a notable dispersion and there is a great diversity of establishments of very different sizes, it was hypothesized that the territorial distribution of beds within each municipality could be significantly proportional to the location of establishments.

The source of information in this case was Booking, a virtual agency specialized in hotel reservations, owned by Booking Holdings Inc., a holding company listed on the American Nasdaq that includes other tourism-related companies such as Kayak or Momondo.

Booking offers reservations for establishments whose location can be viewed in an on line map. As in TripAdvisor, Booking has an option for user valuation, but this feature was not considered relevant and the analysis consisted of merely identifying the different locations according to a predefined zoning, related to the bus stops, simply using the "map view" available in the search engine. A total of 2,114 establishments were identified in a total of 203 different zones.

3. THE FORMULATION

In this case, the objective was to calculate the bus demand, so the analysis was focused only in this transport mode, leaving aside the railway (both because the corridors mostly do not coincide, and because of its very different seasonality) and the car (with a much larger market share).

A generalization of the classic gravity model was used, as follows:

$$T_{ij} = \alpha\ \pi(G_{rij}^{g})\ \pi(A_{sij}^{a})\ c_{ij}^{k} \quad (1)$$

Where the notation is as follows:

- T_{ij} = volume of demand between zones "i" and "j".
- α = parameter.
- π = multiplication operator.
- r = ordinal index of the different possible indicators of trip generation.
- s = ordinal index of the different possible indicators of trip attraction.
- G_{ij}= Generating potential of zone "i" and zone "j".
- g = parameter of the generation factors, equal to the corresponding elasticity of demand.
- A_{ij}= Attractor potential of zone "i" and zone "j".
- a = parameter of the attraction factors, equal to the corresponding elasticity of demand.
- c_{ij}= friction factor between both zones.
- k = parameter of the friction factor, equal to the elasticity of demand to the friction factor.

The selection of the different model formulations and the selection of the corresponding variables followed a trial and error process, through regression analysis between the 2017

demand data for the 271 origin-destination pairs and various independent variables (all of them oriented to the forecast in the following phase). The best explanatory capacity was found with the following variables:

- Sum of population of origin and destination.
- Sum of indicators of tourist attraction (*TripAdvisor*) of origin and destination.
- Sum of tourist places by zone (*Booking*) of origin and destination.
- Fare distance (km). It is the nominal distance between origin and destination used for pricing. It is almost identical to actual road distance, with minor differences.
- Fare rate (€/pax-km). It is the price per pax-km that multiplied by the fare distance gives the tariff.

The number of services per line was discarded out, although it is an excellent indicator of quality from the point of view of users. The reason was the high correlation between number of services and demand, as expected in any sensible operation: the lesser the demand volume, the less services the operator provides.

3.1 The results

The results can be synthesized as follows, expressed through the coefficients of the corresponding variables (which are the corresponding elasticity) and the regression statistics.

Variable/Indicator		**Tourist generation only**	**Tourist attraction only**	**Generation and tourist attraction**
Elas-ticity	Population	0,72	0,64	0,56
	Tourist attraction	-	0,70	0,51
	Tourist generation	0,65	-	0,37
	Fare distance	-2,27	-2,37	-2,41
	Fare rate	-2,51	-2,53	-2,62
Statistic s	Multiple correlation coefficient	0,66	0,67	0,68
	R2 adjusted	0,42	0,44	0,45

Table 1 - Elasticities and indicators obtained. Source: Own elaboration.

Although the quality of the adjustment is not high, the values of the exponents are very reasonable. From the point of view of forecasting, the similarity of exponents of the fare distance and the fare rate were interesting, corroborating the fact that the trip friction is given, simply, by the trip price.

4. CONCLUSIONS AND FINAL REFLECTIONS

An adequate understanding of mobility in a specific area requires data of according quality, which usually official sources do not provide, for lack of detail. Sometimes these data can be obtained through surveys or other measurements, but often there are important restrictions due, for example, to seasonality.

Many studies have a strict completion period, especially if they are related to tendering processes, and then it is not reasonable to waste any effort that is expected to be fruitless.

In today's world there are numerous sources of information that can act as explanatory variables of demand or as *proxies.* In the study described above, even with a very simple approach, it was possible to find some clues about the trip attraction and generation, in particular thanks to two portals with a purely commercial objective: *TripAdvisor* and *Booking*.

The results obtained, have not only been useful for the study, but also point in the direction of the potential value of the huge amounts of information currently available, which are usually a mere by-product of commercial businesses.

The data obtained in the study were manually obtained, but this did not entail excessive consumption of resources.

Although it is conceivable to obtain that information massively, thanks to some automation, the data is not designed for that purpose and the automation would foreseeably face strong protections.

But it is not unlikely that the companies that own this type of data can sell them to third parties, as some mobile phone operators are already doing, provided privacy and other legal issues are respected. For such companies, this is not their core business and it will never be a source of large-scale revenue.

But it seems that there is a certain market and the companies in question have experience in diversifying very intelligently. Everybody would benefit from it, not least transport planners and transport economists.

REFERENCES

ALEXANDER, L., JIANG S., MURGA M., GONZÁLEZ, M. (2015) Origin–destination trips by purpose and time of day inferred from mobile phone data. Transportation Research Part C: Emerging Technologies. Volume 58, Part B, September 2015, Pages 240-250. Transportation Research Part C: Emerging Technologies.

ANDA, C., ERATH, A., FOURIE, P. (2017). Transport modelling in the age of big data. International Journal of Urban Sciences. 21. 1-24.

BONNEL, P., HOMBOURGER E., OLTEANU-RAIMOND, A-M, SMOREDA, Z. (2015). Passive mobile phone dataset to construct origin-destination matrix: potentials and limitations. 10th International Conference on Transport Survey Methods. Transportation Research Procedia 11 (2015) 381 – 398.

CONSORCIO DE TRANSPORTES DE MALLORCA (2017 a) Anteproyecto de servicios de transporte regular y de uso general de viajeros por carretera de Mallorca.

CONSORCIO DE TRANSPORTES DE MALLORCA (2017 b) Plan insular de servicios de transporte regular de viajeros por carretera de Mallorca.

CONSORCIO DE TRANSPORTES DE MALLORCA (2017 c) Memoria 2016.

CONSORCIO DE TRANSPORTES DE MALLORCA (2018) Memoria 2017.

CONSORCIO DE TRANSPORTES DE MALLORCA “Open Data” Consorcio de Transportes de Mallorca. https://www.tib.org/es/web/ctm/open-data

MILNE, D., WATLING, D. (2019). Big data and understanding change in the context of planning transport systems. Journal of Transport Geography. Volume 76, April 2019, Pages 235-244.

INSTITUT D'ESTADÍSTICA DE LES ILLES BALEARS. https://ibestat.caib.es/ibestat/inici. Accessed in 2019.

TRENSCAT. http://www.trenscat.com/balears/sfm_ct.html. Accessed in 2019.