# THE GRAVITY MODEL AS A TOOL FOR DECISION MAKING. SOME HIGHLIGHTS FOR INDIAN ROADS

**Oscar Martínez-Alvaro**
Managing Partner, Tool Alfa S.L, Spain; Ph. D. Civil Engineering. M. Sc. Economics
**Jose-Manuel García**
Project Manager, Tool Alfa, S.L, Spain
**Narender Kumar**
Country Manager, Tool Alfa Consultants p.l.c, India

## ABSTRACT

For many decades, there have been plenty of analyses all over the world about the relationship between socio-economic attributes and transport flows. One of the most fruitful tools is the gravity model, in the beginning used for road transport, but recently widely used for air transport and international trade.

India is an outstanding example of complexity, with a mixture of megapolises and vast rural areas. Its road network shows plenty of six and four lane expressways spanning hundreds of kilometers, complemented by a dense web of State and local secondary and tertiary links.

In the last decades, National and State Governments have improved vast tracts of roads, but there is still a huge gap. Investment priorities are usually decided on the ground of existing congestion or strategic issues, but not much on demand analyses.

For ascertaining whether in India socio-economic structure and transport flows follow a common pattern, complete corridor OD matrices were calibrated from partial screen matrices for a sample of long-distance corridors (NH-1, NH-6, NH-8, NH-58, NH-73). These matrices were later analyzed by means of gravity models that included parameters such as population or GDP per district (as zone attributes) and road distance among district centroids (as friction factors). Several formulae were tested, and the best fit was selected.

Results for main corridors are rather homogeneous, and rather consistent with research carried out in other countries. Simple formulae have a high explanatory capacity, even if the huge mega-cities of Delhi and Mumbai are included in the analysis. But results for rural corridors are much less consistent, probably due to a less mature structure in terms of spatial distribution and transport relationships.

## 1. BACKGROUND AND OBJECTIVES

India is one of the largest countries in the world, with well above 1.2 billion inhabitants and a road network of around 3.3 million kilometers. This size in itself entails a high complexity, with huge long distance corridors spanning hundreds of kilometers linking very large cities, surrounded by a web of secondary links, many of which have  local importance. Although the road network has greatly improved in the recent past, it still needs important investments for implementing further improvements in terms of quantity and quality. Owing to the importance of the investment needs, priorities should be clearly supported by analytical tools. But data are scarce and scattered and, accordingly, it is not easy to have a clear, simple vision of traffic flows beyond the usual, *ad hoc* vehicle traffic counts.

This paper's objective is to find out whether in India, as in other countries, there is a simple explanation for the traffic flow volume between different cities and, in particular:

- Taking into account the vast extension of the country and its huge diversity, do traffic flows in India follow a common, simple pattern on all long distance corridors? Are there groups of different corridors?
- Since the city size is highly diversified, does the size of cities on different corridors affect their traffic flow structure? Do huge metropolitan areas behave differently to the rest of cities?
- In spite of the general scarcity of data, are there any independent variables that can explain the aforementioned patterns?
- If any common flow pattern does exist, can it be used for supporting the investment decision-making process?

## 2. THE GRAVITY MODEL

Gravity models have been widely used in many fields of human activity, from phone calls to commuting trips, to describe behaviors that follow patterns that are somehow similar to the gravitational interaction as described by the famous Newton's law of gravity. There is some dispute as to who introduced this idea in the analysis of land use and transport. In the 1860s H. Carey first applied Newtonian Physics to the study of human behavior by means of the so-called "gravity equation" but it seems that it was H.N. Pallin, a Swedish researcher, who in 1930 (Schmidt & Campell, 1956), came up with the first gravity formula for metropolitan traffic.

For some time this gravity model got wide and deep attention, not just because of the mentioned similarity to Newton's law, but because it was found to have the potential of a higher level of generality (Wilson, 1970).

With this broader approach, the flows between any pair of zones ($T_{ij}$) are the “dependent” variables and have to be explained by some “independent” variables ($x_i$, $x_j$, $f_{ij}$).

$$T_{ij} = f(x_i\ , x_j, f_{ij}) \tag{1}$$

where:

- $x_i$ and $x_j$ are attributes of the origin and destination, such as population, domestic product, etc.
- $f_{ij}$ are “friction factors”, such a trip time or distance between zones.

Lately the gravity model has been used in many studies, to predict both freight and passenger flows, as well as international trade (Kepaptsoglou, Karlaftis & Tsamboulas, 2010). Its usefulness has been demonstrated for other transport modes, such as air transport. (Arvis & Shepherd, 2011).

In India, the gravity model has been used for freight flows at national level within a comprehensive several-steps model: it provided the distribution of generated and attracted flows (Dalvi & Das, 1983). It has also been widely used for India’s international trade analysis (Batra, 2004; Bhattacharyya & Banerjee, 2006) and, to a large extent, for urban transport. (Jaiswal & Sharma, 2012).

## 3. METHOD

The analysis of long-distance flows requires information on origin and destination of trips that is not easily available in India. Although there are some databases with some information on traffic, there is a general lack of structured information and most researches and consultancy studies on road traffic are carried out with information collected on an ad-hoc basis, by means of vehicle traffic counts and screen origin-destination surveys. In fact, there are several procedures clearly defined by the codes published by the Indian Road Congress (The Indian Road Congress, 1988 & 2001).

This usual method provides information that describes, with some detail, the attributes of users crossing the screen where the survey in question is being done. But, owing to the physical impossibility of obtaining statistically balanced and representative samples for all flows, some flows may be under-represented, while others may be over-represented.

Besides, some flows may cross several screens while others may cross only one. In order to solve this problem, the most used method (Ortuzar & Wilumsen, 1995) is a procedure of maximum likelihood calibration, by means of which a “seed matrix”, obtained as a combination of matrices obtained in the survey, is assigned to the network and the traffic volumes estimated by the model are compared to the vehicle traffic counts.

Any mismatches are adjusted on an iterative process to make sure the final traffic flows reproduce the traffic counts with enough accuracy.

The next step is to try to explain the matrices obtained. After an analysis of available socio-economic data, the ones that seemed best suited as independent variables were the following:

- "Attraction factors" ($x_i$): Population per district or, alternatively, domestic product per district. Also some other attempts were made with other variables that could be related to traffic: the most obvious candidate was the number of vehicles per district.
- "Friction factors" ($f_{ij}$): distance between districts' centers of gravity.

A further problem was this of the heterogeneous time reference, since for some variables there were reliable data for some years, but for others the time reference was different. Some simple algorithms to overcome this problem, such as making the assumption that GDP or population at District level has evolved at the same pace as the corresponding whole State, just introduces more hypothesis. Therefore, it was decided not to make any changes to the official data, except in the case when for the same variable there were different horizons (notably in the case of vehicle data). As a consequence, all values for each variable were referred to the same year, but different variables could be referred to different years, which changes formula as follows:

$$T_{ij\,y'} = k\,\frac{x_i^{\alpha\,y''}\,x_j^{\alpha\,y''}}{f_{ij}^{\beta\,y'''}} \tag{2}$$

Where $y'$, $y''$ and $y'''$ are the reference years for each variable.

Formula (2) can be calibrated easily thanks to the use of logarithms that transform the expression into a linear one, which allows for a very simple regression calibration. The software used was Microsoft's Excel.

## 4. CASES AND DATA AVAILABLE

In order to have a proper picture of the situation in India, a broad sample was needed: covering different parts of the country as well as different corridor types. According to the information available to the researchers, the following corridors were chosen:

- **NH-1**: Panipat - Jalandhar. 285 km. NH1 is, as expected from its code name, one of the main corridors in India. It connects the Indo-Pakistan border to Delhi. Major towns are Panipat, Jalandhar, Ambala, Ludhiana, Chandigarh and Delhi. Panipat oil refinery and industries at Ambala, Jalandhar and Ludhiana are major centers of activity.

Ludhiana is home to 90% of all woolen garments in India and Jalandhar is known for its sports good industry. There are many short distance and long distance routes competing with this corridor.

- **NH-6**: Gujarat/Maharashtra border - Amravati. 484 km. NH6 is the major national highway connecting East India to West India. Main cities around the corridor are Amravati, Akola, Jalgaon, Dhule, Nagpur, Surat and Mumbai. The project corridor mainly lies in the most backward economic regions of Maharashtra state with agriculture as the main occupation. The corridor is full of many small local bypasses and long alternate routes.
- **NH-8**: Kishangarh – Udaipur - Ahmedabad. 555km. This corridor runs through some feebly populated areas of Rajasthan, but it connects Delhi to Mumbai via many important economic hub cities like Jaipur, Kishangarh, Ahmedabad, Vadodara and Surat. NH3 could potentially compete with NH8, but its poor current condition prevents it from being a real alternative.
- **NH-58**: Meerut - Muzaffarnagar. 79 km. This corridor runs through the sugar belt of India wherein sugarcane production and processing is the main industry. Meerut, Ghaziabad and Muzaffarnagar are known for many industrial and agro-based activities. Sacred sites of Haridwar and Rishikesh attract thousands of pilgrims all year round. SH57 is a major competing route for NH58, though it is in a very poor condition at present.
- **NH-73**: Yamunanagar - Panchkula. 108 km. Chandigarh and Panchkula are major cities near or along the corridor. Yamunanagar, Ambala and Saharanpur are industrial towns around the corridor. Baddi, an industrial township near Panchkula, is being developed as pharmaceuticals industry capital of India. NH1 and a combination of some State highways pose choices to users of NH73 for both short and long distance routes.

For each of the aforementioned corridors, data on volume of traffic and on origin-destination flows on the very road and on competing ones were available on many screens, as shown in Table 1. On the other hand, independent variables were available from official sources at different years. The distance between District capitals was easily obtained by means of TransCAD on the corresponding network graphs.

| Corridor | Section | Year | Number of traffic count screens | Number of OD screens | Number of OD surveys |
|---|---|---|---|---|---|
| NH-1 | Panipat-Jalandhar | 2010 | 34 | 14 | 15,675 |
| NH-6 | Gujarat/Maharashtra border-Amravati | 2012 | 10 | 10 | 18,376 |
| NH-8 | Kishangarh-Udaipur-Ahmedabad | 2011 | 17 | 14 | (*) |
| NH-58 | Meerut-Muzaffarnagar | 2013 | 2 | 2 | 2,423 |
| NH-73 | Yamunanagar to Panchkula | 2011 | 11 | 2 | 6,035 |

(*) The precise figure was not available, but the sample was between 15,000 and 20,000 vehicles.

**Table 1 – Year of available traffic data and number of screens per analyzed corridor**

## 5. INTERMEDIATE RESULTS: OD MATRICES

Not all previous studies available had the same vehicle type structure, because they had been merged according to particular needs.

Therefore, the first step was a calibration of OD matrices following the existing vehicle types.

The goodness of fit between OD matrix and traffic counts was high in all cases, as shown in Table 2.

| Corridor | Car/Van/Jeep | LCV/Minibus | 2AT/3AT/Bus | 2AT/Bus | 3AT | MAV |
|---|---|---|---|---|---|---|
| NH1 | 0.992 | 0.992 | 0.974 | n.a. | n.a. | 0.982 |
| NH6 | 0.997 | 0.996 | n.a. | 0.943 | 0.870 | 0.932 |
| NH8 | 0.992 | 0.992 | n.a. | 0.969 | 0.966 | 0.976 |
| NH58 | 1.000 | 1.000 | 1.000 | n.a. | n.a. | 1.000 |
| NH73 | 0.997 | 0.995 | n.a. | 0.965 | 0.763 | 0.998 |

LCV = Light Cargo Vehicle; 2AT = Two Axle Truck; 3AT = Three Axle Truck; MAV = Multiaxle vehicle (> 3 axles); n.a. = not available.

**Table 2 – Coefficient of determination (R2) of regression analysis between assigned traffic values and traffic counts**

This somehow heterogeneous classification of vehicles prevented from a detailed, homogeneous analysis of vehicle types, but this was not a big hurdle, as the lack of proper independent variables made it useless to try a very detailed analysis of dependent variables.

Besides, the classical differentiation between passenger vehicles and goods vehicles could not be done in detail due to the mixture of types within some matrices (e.g. LCV and minibuses, trucks and buses).

The decision made was to distinguish between light and heavy vehicles for two reasons:

- This aggregation is fully compatible with existing data.
- This definition is based on the nature of the presumed distance for which each vehicle is best suited, which is coherent with the distance as the friction factor.

Therefore, two matrices were obtained for each corridor:

- One matrix for light vehicles, which in practice was almost reduced to cars and small cargo vehicles.
- One matrix for heavy vehicles, that is, buses and trucks of any size, which is in practice mostly trucks of any size.

## 6. FINAL RESULTS: GRAVITATION MODEL CALIBRATION

In order to have a broader analysis, the parameter k in the aforementioned formula was considered under two hypothesis: fixed (k=1) and calibrated. Thanks to the data available, for all large corridors (NH1, NH6, NH8) it was possible to carry out the regression analysis taking into account population, GDP or number of vehicles. For second-rank corridors (NH58 and NH73) it was possible only with population and GDP. In all cases, the friction factor was the distance.

When it was assumed that $k = 1$, results showed a high coefficient of determination, with values of 0.9 or above for all calibrations.

| Corridor | Attraction factor | Light vehicles | | | | Heavy vehicles | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | R2 | | F significance | | R2 | | F significance | |
| | | (1) | (2) | (1) | (2) | (1) | (2) | (1) | (2) |
| NH1 | Population | 0.92 | 0.92 | 0.0% | 0.0% | 0.92 | 0.92 | 0.0% | 0.0% |
| | GDP | 0.92 | 0.92 | 0.0% | 0.0% | 0.92 | 0.92 | 0.0% | 0.0% |
| | N of vehicles | 0.92 | 0.92 | 0.0% | 0.0% | 0.89 | 0.90 | 0.0% | 0.0% |
| NH6 | Population | 0.99 | 0.99 | 0.0% | 0.0% | 0.99 | 1.00 | 0.0% | 0.0% |
| | GDP | 0.99 | 0.99 | 0.0% | 0.0% | 0.99 | 1.00 | 0.0% | 0.0% |
| | N of vehicles | 0.99 | 0.99 | 0.0% | 0.0% | 0.99 | 1.00 | 0.0% | 0.0% |
| NH8 | Population | 0.96 | 0.97 | 0.0% | 0.0% | 0.96 | 0.97 | 0.0% | 0.0% |
| | GDP | 0.96 | 0.97 | 0.0% | 0.0% | 0.96 | 0.97 | 0.0% | 0.0% |
| | N of vehicles | 0.95 | 0.97 | 0.0% | 0.0% | 0.94 | 0.98 | 0.0% | 0.0% |
| NH58 | Population | 0.98 | 0.98 | 0.0% | 0.0% | 0.94 | 0.94 | 0.0% | 0.1% |
| | GDP | 0.98 | 0.98 | 0.0% | 0.0% | 0.94 | 0.94 | 0.0% | 0.1% |
| | N of vehicles | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. |
| NH73 | Population | 0.94 | 0.94 | 0.0% | 0.0% | 0.95 | 0.95 | 0.0% | 0.0% |
| | GDP | 0.94 | 0.94 | 0.0% | 0.0% | 0.95 | 0.95 | 0.0% | 0.0% |
| | N of vehicles | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. |

(1) With Delhi and Mumbai; (2) Excluding Delhi and Mumbai; n.a. = data not available

**Table 3 – Values of coefficient of determination by corridor, type of city and independent variable**

Focusing on the first-rank corridors (NH1, NH6, NH8), the sample excluding Delhi and Mumbai showed a slightly better coefficient of determination, but again in negligible amounts. In all cases, population, GDP and number of vehicles were good independent variables, not much different to one another in statistical significance. The values obtained for the calibrated parameters are presented in Table 4.

| Corridor | Attraction factor | Light vehicles | | | | Heavy vehicles | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | α | | β | | α | | β | |
| | | (1) | (2) | (1) | (2) | (1) | (2) | (1) | (2) |
| NH1 | Population | 0.44 | 0.50 | -1.58 | -1.91 | 0.45 | 0.48 | -1.75 | -1.92 |
| | GDP | 0.47 | 0.52 | -1.61 | -1.90 | 0.48 | 0.51 | -1.77 | -1.92 |
| | N of vehicles | 0.49 | 0.55 | -1.56 | -1.92 | 0.51 | 0.55 | -1.30 | -1.46 |
| NH6 | Population | 0.39 | 0.43 | -1.36 | -1.57 | 0.29 | 0.32 | -0.80 | -0.98 |
| | GDP | 0.47 | 0.49 | -1.60 | -1.70 | 0.37 | 0.38 | -1.06 | -1.11 |
| | N of vehicles | 0.52 | 0.58 | -1.64 | -1.93 | 0.45 | 0.46 | -0.94 | -0.98 |
| NH8 | Population | 0.46 | 0.63 | -1.60 | -2.55 | 0.36 | 0.56 | -1.10 | -2.22 |
| | GDP | 0.59 | 0.78 | -2.04 | -2.97 | 0.48 | 0.71 | -1.51 | -2.67 |
| | N of vehicles | 0.51 | 0.87 | -1.57 | -3.29 | 0.36 | 1.05 | -0.53 | -3.13 |
| NH58 | Population | 0.26 | 0.21 | -0.45 | -0.14 | 0.17 | 0.12 | -0.19 | 0.13 |
| | GDP | 0.26 | 0.23 | -0.34 | -0.09 | 0.18 | 0.13 | -0.14 | 0.15 |
| | N of vehicles | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. |
| NH73 | Population | 0.75 | 0.75 | -3.51 | -3.51 | 0.64 | 0.64 | -2.92 | -2.92 |
| | GDP | 0.72 | 0.72 | -3.08 | -3.08 | 0.62 | 0.62 | -2.61 | -2.61 |
| | N of vehicles | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. | n.a. |

α = power of attraction factors; β = power of friction factor (distance); (1) With Delhi and Mumbai; (2) Excluding Delhi and Mumbai; n.a. = data not available.

**Table 4 – Values of parameters calibrated by corridor, type of city and independent variable**

For all first-rank corridors (NH1, NH6, NH8), the sign of the powers of the attraction variables (population, GDP, number of vehicles) and friction (distance) were correct, while the two lesser corridors showed problems of sign consistency (NH58) or of statistical significance (NH73).

According to values in Table 4, the introduction or exclusion of the largest cities meant interesting variations in the values of the calibrated parameters: for corridors with only one of these large metropolitan areas in the surroundings (Delhi for NH1, Mumbai for NH6) the variation in the values was relatively low (from 3% to 23%, depending on the cases), but when both cities were relevant (NH8) the variation was much larger (above 32% and up to 491%).

A deeper analysis of values obtained for the power (β) of the distance is shown in Table 5, where minimum, maximum and non-weighted average values are presented by type of vehicle and independent variable.

| Attraction factor | Light vehicles | | | | | | Heavy vehicles | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | With Delhi & Mumbai | | | Excl. Delhi & Mumbai | | | With Delhi & Mumbai | | | Excl. Delhi & Mumbai | | |
| | Min | Max | Av | Min | Max | Av | Min | Max | Av | Min | Max | Av |
| Population | -1,60 | -1,36 | -1,51 | -2,55 | -1,57 | -2,01 | -1,75 | -0,80 | -1,22 | -2,22 | -0,98 | -1,71 |
| GDP | -2,04 | -1,60 | -1,75 | -2,97 | -1,70 | -2,19 | -1,77 | -1,06 | -1,45 | -2,67 | -1,11 | -1,90 |
| N of vehicles | -1,64 | -1,56 | -1,59 | -3,29 | -1,92 | -2,38 | -1,30 | -0,53 | -0,92 | -3,13 | -0,98 | -1,86 |

Min = minimum; Max = Maximum; Av = non-weighted average.

**Table 5 – Values of power (β) of the distance as friction factor**

According to Table 5, average values of this β parameter are quite homogeneous:

- In absolute terms, taking into account Delhi and Mumbai, they are around 0.9-1.5 for heavy vehicles and around 1.5-1.8 for light vehicles.
- Without these metropolitan cities, values are around 1.7-1.9 for heavy vehicles and around 2.0-2.4 for light vehicles.

It is noteworthy that for NH58 the values of this β parameter are much lower (in absolute terms, from 0.2 to 0.5), this showing again the difference between large corridors of National relevance and corridors of mere local importance.

The introduction of the parameter k reduced the accuracy of the calibration, which showed low values of the coefficient of determination.

## 7. CONCLUSIONS

The analysis summarized in this paper clearly shows the following:

- In India, as in all other developing countries, the lack of data introduces some degree of inaccuracy in traffic analyses. But results described in this paper are so consistent that can be considered, at least, good orders of magnitude.
- In large corridors, volumes of vehicles can be explained by the classical gravity model. Population, GDP and number of vehicles at District level have a noticeable explanatory capacity. Parameter values are remarkably homogeneous: as a rule, the friction due to distance has exponent values between 1.0 and 2.0 if all cities are included and between 1.5 and 3.0 if the largest cities are excluded. These values are not very much different from the ones obtained in more developed countries. Probably people tend to behave similarly in trips between large cities all over the world.

- In corridors in rural areas, the parameters are quite different from one area to another. There are no obvious patterns and each corridor has its own characteristics. It remains to be seen whether these rural corridors converge to the pattern found in the largest ones when GDP grows or whether their peculiarities are intrinsic.

## ACKNOWLEDGEMENTS

The research work described in this paper has been funded by TOOL ALFA, S.L. (Spain), parent company of TOOL ALFA CONSULTANTS, Pvt. Ltd. (India.)

## REFERENCES

ARVIS, J.F. & SHEPHERD, B. (2011). The Air Connectivity Index. Measuring Integration in the Global Air Transport Network. The World Bank. Policy Research Working Paper 5722.

BATRA, A. (2004). India’s global trade potential: the gravity model approach. Indian Council for Research on International Economic Relations. Working paper 151.

BHATTACHARYYA, R. AND BANERJEE, T. (2006). Does the Gravity Model Explain India’s Direction of Trade? A Panel Data Approach. Indian Institute of Management Ahmedabad. Working paper 2006-09-01.

DALVI M. Q. & DAS T. K. (1983). A multisectoral and multiregional transport model for India - data base and calibration techniques. Environment and Planning A 15(3), pp. 391 – 403.

JAISWAL, A. AND SHARMA, A. (2012). Optimization of Public Transport Demand: A Case Study of Bhopal. International Journal of Scientific and Research Publications, Volume 2, Issue 7.

KEPAPTSOGLOU, K., KARLAFTIS, M.G & TSAMBOULAS D. (2010). The Gravity Model Specification for Modeling International Trade Flows and Free Trade Agreement Effects: A 10-Year Review of Empirical Studies. The Open Economics Journal, 2010, 3, pp. 1-13.

ORTUZAR, J. D. AND WILUMSEN, L.G. (1995). Modelling transport. John Wiley & Sons, Ltd, New York.

PAPACOSTAS, C.S. (1987). Fundamentals of transportation engineering. Prentice Hall, New Jersey.

SCHMIDT, R.E. & CAMPELL, M.E (1956). Highway traffic estimation. Eno Foundation Publications, Connecticut.

THE INDIAN ROADS CONGRESS (1988). Traffic studies for planning bypasses around towns (IRC: 102 – 1988). IRC Publication, New Delhi.

THE INDIAN ROADS CONGRESS (2001). Manual for survey, investigation and preparation of road projects (IRC: SP 19-2001). IRC Publication, New Delhi.

WILSON, A.G. (1970). Entropy in urban and regional modelling. Pion Limited, London.