Trip distribution for limited destinations:
A case study for grocery shopping trips in the Netherlands

S.A. Veenstra (corresponding author): Centre of Transport Studies, University of Twente, p.o.box 217, 7500 AE Enschede, The Netherlands, tel: +31 53 489 4704, fax: +31 53 489 4040, s.a.veenstra@utwente.nl

T. Thomas: Centre of Transport Studies, University of Twente, p.o.box 217, 7500 AE Enschede, The Netherlands, tel: +31 53 489 3821, fax: +31 53 489 4040, t.thomas@utwente.nl

S.I.A. Tutert: Centre of Transport Studies, University of Twente, p.o.box 217, 7500 AE Enschede, The Netherlands, tel: +31 53 489 4517, fax: +31 53 489 4040, s.i.a.tutert@utwente.nl

DATE: November 15, 2009

ABSTRACT
In this paper, we introduce a new trip distribution model for destinations that are not homogeneously distributed. The model is a gravity model in which the spatial configuration of destinations is incorporated in the modeling process. The performance was tested on a survey with reported grocery shopping trips in the Dutch city of Almelo. The results show that the new model outperforms the traditional gravity model. It is also superior to the intervening opportunities model, because the distribution can be described as a function of travel costs, without increasing the computational time. In this study, the distribution was described by a simple function of Euclidean distance, which provides a good fit to the survey data. The slope of the distribution is quite steep. This shows that most trips are made to nearby supermarkets. However, a significant fraction of trips, mainly made by car, still goes to supermarkets further away. We argue that modeling of these trips by the new method will improve traffic flow predictions.

KEY WORDS: gravity model, intervening opportunities, supermarket, aggregated approach

WORD COUNT: 6,581 words, 5 figures and 0 tables

1.   INTRODUCTION

The understanding of traffic flows in an urban environment is an important issue with reference to policy making. Problems with traffic nuisance and air quality are increasingly seen as being a threat to the livability in Dutch cities. Pollution of small particles by cars decreases the air quality in urban areas (*1*), while over forty percent of the population in Dutch cities also indicated that they experience traffic nuisance as an impediment to the livability of residential areas (*2*).

A better insight to traffic flows will help local policy makers to implement policies regarding urban traffic. This requires reliable data on the different trip purposes. According to the Dutch National Travel Survey (*3*), on average more trips are made to grocery shops than to work. A good description of trip generation and distribution for these trips is therefore important. In this paper, we use high resolution survey data from the Dutch city of Almelo to model the distribution of grocery shopping trips.

There are several methods to model trip distribution. Among others, Ibrahim (*4*) and Jang (*5*) used disaggregated models to describe the generation and distribution of shopping trips. These methods model individual choices of travelers. By including many relevant attributes, they can model several aspects of travel behavior and policy interventions simultaneously. However, due to many unknown factors, it is almost impossible to model how each individual values an alternative, and the evaluation of the results is also not straightforward. Most practitioners in the Netherlands therefore still use aggregated methods to model trip distribution, especially in the case of shopping trips, e.g. (*6*). In this study, aggregated observations are parameterized by a simple empirical model, for which the results can be inspected visually. Although this approach is not often used in traffic engineering, we will show that it provides reliable results that can be interpreted in a straightforward way.

The traditional gravity model, e.g. (*7*), is probably the most popular aggregated method. Within gravity models a 'deterrence function' is used to describe the propensity to travel at increasing generalized costs. This function is often presented in the form of a power law, an exponential function or a combination of both (*7*). A different approach to estimate the aggregated trip distribution is to use the concept of intervening opportunities, which was proposed by Stouffer (*8*). In this method, the number of persons going to a particular destination is inversely proportional to the number of intervening opportunities between the origin and destination.

In the past, a few comparisons were made between the traditional gravity (TG) model and the intervening opportunities (IO) model. In general, both models performed comparably (*9*). Eash (*10*) stated that the TG model and the IO model are fundamentally the same as they both can be derived from the entropy maximization approach (*11*). The difference between the models arises from the approach to determine the disutility. Within the TG model the disutility is described by travel costs, while in the IO model the disutility depends on the number of intervening opportunities (*12*), (*13*). IO models are especially useful for trip purposes, in which the opportunities are not homogeneously distributed, but form discrete attraction points in the urban environment. In an effort to make a unification of both aggregate approaches, the gravity-opportunity model was developed. This model can be seen as an IO model in which deterrence as a function of travel costs is included (*12*). However, for this model the computational complexity is substantial (*9*).

In this paper, we introduce an aggregated method to model trip distribution for the shopping purpose. The model is based on the gravity method, but, like IO models, it takes the spatial configuration of supermarkets into account. However, this is done without introducing extra computational time. In section 2, we describe the data. In section 3, we introduce the new model and we shortly summarize how we applied the TG, the IO and the new model to our sample. In section 4, we compare the results of the three different models. Section 5 provides conclusions.

2.   DATA

In the Netherlands and in many other countries, data from National Travel Surveys (NTS) are used to describe traffic flows. The Dutch NTS (MON) is a large survey which is carried out each year. In the questionnaire of the MON respondents are asked to fill in which trips they have made at a given day. The MON survey, however, is not very suitable for estimating the distribution of grocery shopping trips. First,

short non-commuting trips are underreported, because these trips are more easily forgotten by respondents (*14*), (*15*). Since underreporting is distance dependent, the distribution is also affected by this bias. Secondly, the spatial resolution of MON is relatively poor and precise distance estimates are lacking.

For these reasons, we used data from a local survey, Omnibus, which was conducted in the Dutch city of Almelo for many subsequent years. In general, the main objective of this questionnaire was to acquire information on the attitudes and behavior of its inhabitants. One person per household is asked which two supermarkets are visited most frequently, what their corresponding frequency rates are, and which mode of transport is used. Both the location of the household and the supermarkets are known on a Dutch postal 6 zone level. An average postal 6 zone in the municipality of Almelo is 0.04 km^2. This resolution is high enough for a reliable estimate of the distribution for grocery shopping trips. Figure 1 shows the distribution of the population and the locations of the supermarkets in Almelo. The cross-section of the figure is about 5 km.

The Euclidean distance between the centroids of the postal 6 zones was taken as an indicator for the deterrence. For internal trips (when the residents and supermarket are in the same zone), the Euclidean distance was calculated by $0.5r$ with r being the radius of the postal 6 zone. Although Euclidean distance is strongly correlated with network distance, network distance and travel time may be better indicators for the deterrence. However, these variables are difficult to obtain, because estimates are often biased for short distances e.g. (*16*), (*17*), (*18*). We lacked reliable information on network distance and travel time.
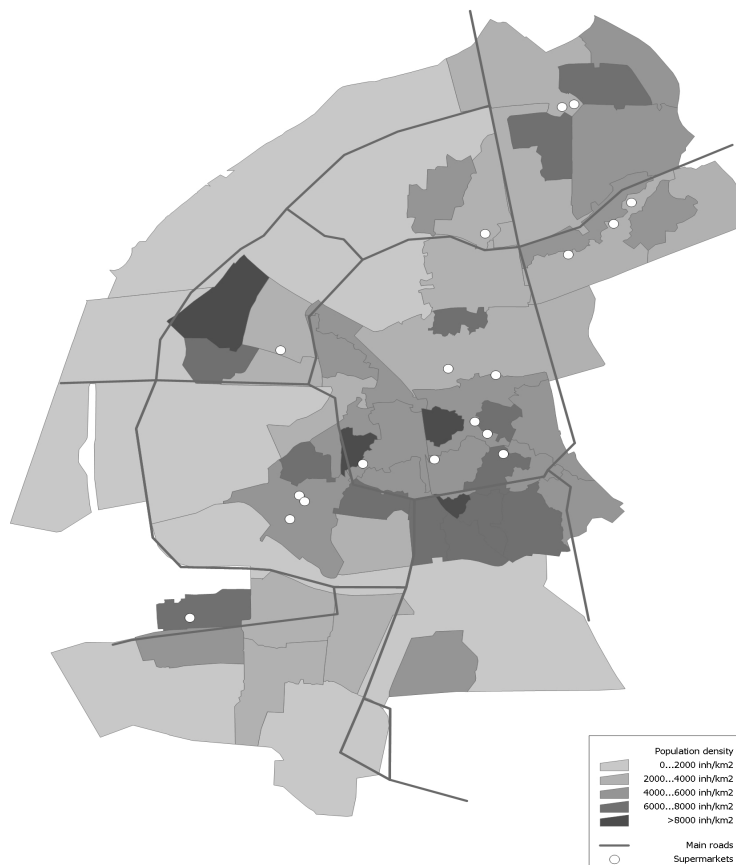
FIGURE 1 Population and location of supermarkets in the neighborhoods of Almelo

Despite the comprehensive information on travel behavior, the Omnibus questionnaire has its drawbacks. Although the sample is representative for the types and locations of the households in Almelo, it is not representative for the entire population of Almelo, because it contains more women and relatively fewer workers. We, however, assume that the trip distribution of the respondents is comparable to those of other household members. Also, respondents can only report on two supermarkets, while they might visit a third or fourth supermarket. Although it is hard to estimate the effects of this bias, we think that it has little effect on the distribution as function of distance. Finally, only supermarkets within Almelo are considered. Supermarkets in neighboring villages are not taken into account. To avoid biases in the choice set, households in the border zones of Almelo were excluded from the sample.

By combining the databases from 2001 to 2007, we obtained a data set with 8700 cases. For this sample, it is possible to compare the performance of different distribution models. A major limitation, however, is that the distribution for just one city is estimated. Although it is quite likely that this distribution is valid for other cities in the Netherlands as well, no data were available to test this.

3.   METHOD

In this paper, we distinguish three distribution models. First, the distribution of grocery shopping trips is estimated by the traditional gravity (TG) model. In this estimate, the spatial configuration of supermarkets is not included in the modeling process. This will appear to be a disadvantage. Secondly, the distribution is estimated by the intervening opportunities (IO) model, which takes the spatial configuration of supermarkets into account. However, in the traditional IO model the distribution is not described as function of other attributes, like distance. The advantages of both the TG and IO model are combined in a new gravity model, which is called the Limited Destinations (LD) model. In the LD model, the distribution is described as function of distance, while the spatial configuration of supermarkets is also included in the modeling process. In the sections 3.1 – 3.3, we describe the three models in more detail. In section 3.4, we explain how we estimated the quality of the models by a likelihood measure.

For the three models, the distribution function is estimated from the survey data. The number of residential postal zones, I, is 1956. The number of supermarket zones, J, is 18. (20 supermarkets situated in 18 postal 6 zones). The Origin Destination matrix, T_{ij}, is thus an I by J matrix of 1956 rows by 18 columns. The observed OD matrix from the survey, T_{ij}^{obs}, gives the total of reported trips from i to j. Thus, each OD pair contains the sum of the number of trips of each respondent. Both the first and second most frequently visited supermarket were taken into account, by giving them weights according to their reported trip frequencies.

In the OD matrix, the total of row i is equal to the number of trips made from origin i, and the total of column j is equal to the number of trips made to destination j. The observed row totals are equal to the sample production rates per zone, O_i^{obs}. For commuting, the observed column totals are equal to the sample attraction rates per zone, because in a saturated labor market, the number of trips to a working zone is equal to the number of opportunities, i.e. jobs, in that zone. However, this is not the case for shopping trips. The number of trips to a supermarket is proportional to the intrinsic attraction (e.g. characteristics of the supermarket), but also depends on the location of the supermarket. If a supermarket is located in the proximity of many residents, and at the same time near few other supermarkets, it will attract many shoppers. This so called attraction is in fact the result of the way shoppers and supermarkets are distributed throughout the city.

It is not trivial to disentangle the intrinsic attraction from this distribution effect. Many authors (6), (19), (20) use the amount of floor space of a supermarket as a measure for the intrinsic attraction. The data of shopping behavior in Almelo show the same pattern. A supermarket with twice the size of an average supermarket will attract roughly twice the number of customers. The attraction is therefore assumed to be directly proportional to the floor space of a supermarket. The observed intrinsic trip attraction of zone j, D_j^{obs}, can thus be defined as the total floor space of supermarkets in zone j times a normalization factor, which is the ratio between the total number of trips in the survey and the total floor space of all supermarkets in the survey. Note however that large supermarkets are often located in large residential areas, which means that the size of a supermarket may be the result rather than the cause of

attractiveness. Also, other intrinsic characteristics like operating hours, economic activity in the environment, access to major roads, and transit stops, may influence the attractiveness of a supermarket.

It is difficult to estimate which attraction characteristics should be used. Moreover, it will sometimes be difficult to quantify certain characteristics. To get a feeling for the sensitivity of attraction measures, we also estimated the distribution in which all supermarkets have an equal intrinsic attraction. In that case, the observed intrinsic trip attraction of zone j, D_j^{obs}, is defined as the number of supermarkets in zone j times a normalization factor, which is the ratio between the total number of trips and total number of supermarkets in the survey. We find that the results are not significantly different for both attraction measures. In this paper, we only show the results for the former attraction measure, because this is in accordance with the literature.

3.1   Traditional Gravity model

In the TG model the number of trips between i and j is estimated by multiplying the production O_i (total number of trips generated in origin i) and attraction D_j (total number of trips attracted by destination j) with the distribution values f_{ij}. We describe the TG model as follows.

$$T_{ij} = f_{ij} * H_{ij} \tag{1}$$

$$H_{ij} = \frac{O_i * D_j}{\sum_i O_i} \tag{2}$$

T_{ij} is the estimated OD matrix and H_{ij} is the production attraction matrix. The latter one provides the number of trips between i and j according to the production in i and attraction in j. The distribution values can be represented by a distribution function f. The function can be interpreted as a measure for the attractive force between origin and destination per unit of production (in the origin) and unit of attraction (in the destination). When the distribution function is independent of any attribute, i.e. $f \equiv 1$, the OD matrix is equal to the production attraction matrix.

The distribution can be modeled in different ways. The traditional approach is to adopt a function form and to calibrate the parameters subsequently. Thomas and Tutert (*18*) adopted a different approach. They searched for the most simple function form that follows directly from the survey data. We followed the same approach. For the observed OD matrix, T_{ij}^{obs}, and the observed production attraction matrix, H_{ij}^{obs}, trips with similar Euclidian distances were aggregated in nine Euclidian distance bins: 0 – 250, 250 – 500, 500 – 750, 750 – 1000, 1000 – 1500, 1500 – 2000, 2000 – 2500, 2500 – 3500 and 3500 – 5000 meters. These bins were chosen to ensure similar trip frequencies in each bin. Distances over 5 kilometers were not taken into account because of the lack of supermarkets in the sample at those distances. For the bins, the observed distribution values f^{obs} were calculated by:

$$f^{obs}(d) = T^{obs}(d) / H^{obs}(d) \tag{3}$$

With T^{obs} being the observed trip length distribution, H^{obs} the observed production attraction distribution and d the average (weighted by T^{obs}) Euclidean distance in each bin. In section 4.1, the distribution values and function fit are shown and the results are discussed.

3.2   Intervening Opportunities model

In the previous subsection, the TG model is used to estimate the distribution. Given the distribution function, the TG model can also predict the probability, π_{ij}, that someone from zone i makes a trip to a supermarket in zone j. The number of trips from i to j is equal to this probability times the production in i:

$$T_{ij} = O_i * \pi_{ij} \tag{4}$$

With the restriction that $\sum_j \pi_{ij} = 1$, which makes the OD matrix single-constrained.

In the TG model (equations 1 and 2), the probability of visiting a supermarket in j depends on the distance between i and j (according to the distribution function) and on the intrinsic attraction of zone j, D_j. In the IO model, the probability π_{ij} does not depend on the distance between i and j, but on the amount of intervening opportunities. This amount is defined as the total or cumulative intrinsic attraction of supermarkets between i and j. In the IO method, orientation or direction does not play a role. If the distance between i and zone k is smaller than between i and j, supermarkets in k are considered to be intervening opportunities, even if zone k lies in the opposite direction.

The IO model consists of the following steps. For each origin i, the destination zones j are ranked according to their distance to i, i.e. $j = 1$ for the nearest zone, $j = 2$ for the second nearest zone, etc., and $j = J$ for the zone furthest away. The probability π_{ij} that someone from zone i makes a trip to a supermarket in zone j is then estimated by (*11*):

$$\pi_{ij} = \frac{\exp(-\alpha D_{j-1}^{cum}) - \exp(-\alpha D_j^{cum})}{1 - \exp(-\alpha D_J^{cum})} \tag{5}$$

Where D_j^{cum} is the cumulative attraction of intervening opportunities between origin i and destination zone j, including those in j, and α is a positive scale parameter. The cumulative attraction can be calculated by iteration: $D_j^{cum} = D_{j-1}^{cum} + D_j$.

For the nearest zone, $j = 1$, the cumulative attraction in j is equal to the attraction in j, D_j, and the attraction in $j - 1$ is equal to 0. The numerator is then equal to $1 - \exp(-\alpha D_j)$. From this, it follows that the denominator should be included to satisfy $\sum_j \pi_{ij} = 1$.

The parameter α is calibrated by the survey data. A small value of α corresponds with a distribution that slowly decays. In this case, not all travelers go to intervening opportunities, but some are still going to opportunities far away. A large value of α corresponds with a steep decay of the distribution in which almost all travelers go to the nearest opportunities.

The parameter α is actually origin specific. For our survey, it is time consuming and difficult with the available software to calibrate the α of about 2000 origins and let them converge to a steady state. Also, a proper calibration of a location specific α is complicated by the fact that for most origin zones, only a small number of trips was observed. Therefore, we estimate one α value for all origins, which is probably the only way the IO model can be applied in practice. In section 4.2, the distribution for the IO model is shown and the results are discussed.

3.3   Limited Destinations model

Although the TG and IO model appear to be different, they are fundamentally the same (*10*). The disutility in the TG model also depends on the number of intervening opportunities. This is explained in the next example. Suppose that residents in i can shop in either a supermarket in zone j or a supermarket in k. The supermarkets have the same size, but the distance from i to k is smaller than from to i to j. Thus, $\pi_{ik} > \pi_{ij}$ (and $\pi_{ij} + \pi_{ik} = 1$). Suppose that the supermarket in k is closed. The intervening opportunity in k has disappeared and the respondents from i only have one alternative to do their shopping. Hence, $\pi_{ij} = 1$. This probability has changed, because it also depends on the number of intervening opportunities in k.

The problem of the TG model is not that the distribution is a function of distance (or travel costs), but that the spatial configuration of supermarkets is not included in the modeling process. For commuting trips, this is not a problem, because the activities, i.e. jobs, are homogeneously distributed throughout the city. Each postal zone contains jobs. However, when activities are not homogeneously distributed, but

form discrete attraction points in the urban environment (see Figure 1), the configuration of opportunities should be taken into account. Consider an origin zone for which there are only opportunities far away. In the TG model, trips from this origin are aggregated in bins that also contain long distance trips from other origin zones. These other origins, however, could very well have nearby opportunities. The mixing of origins that have different configurations of opportunities leads to a bias in the distribution estimate. All trips from the origin without nearby opportunities have long distances. This origin therefore disproportionately contributes to the long distance trips. The distribution for long distances will thus be overestimated, and the slope of the estimated distribution will be shallower than the true slope.

To prevent this bias, a new model was introduced. The method is a gravity model, like the traditional one. Thus, the OD matrix is estimated according to formulas (1) and (2). However, the distribution function, f, in formula (1) is estimated in a fundamentally different way. In the new method, we implicitly distinguish between origins that have different supermarket configurations. The method estimates the distribution ratio of any two distance bins. This ratio can be determined by aggregating all origins that have opportunities in both distance bins. Origins that do not have opportunities in either distance bin are excluded from the estimate. This is the essential difference with the traditional method in which all origins are aggregated. According to formula (1), the number of shopping trips from i to j, T_{ij}, is proportional to $f_{ij}O_iD_j$, or the distribution value, f_{ij}, is proportional to $T_{ij} / (O_iD_k)$. Thus, viewed from the origin i, the ratio between the distribution values for, let us say, destination zones j and m, f_{ij} / f_{im}, is equal to $(T_{ij}/D_j) / (T_{im}/D_m)$. Because the distribution value is proportional to the observed frequency divided by the intrinsic attraction, the distribution ratio for two arbitrary distance bins k and l could be estimated in the following way.

$$\frac{f^{obs}(d_k)}{f^{obs}(d_l)} = \frac{\sum\limits_{i \in I_{kl}} \left(T_{ik}^{obs} / D_{ik}^{obs} \right)}{\sum\limits_{i \in I_{kl}} \left(T_{il}^{obs} / D_{il}^{obs} \right)}, \text{ with } i \in I_{kl} \text{ if } D_{ik}^{obs} > 0 \text{ and } D_{il}^{obs} > 0 \qquad (6)$$

Where d_k and d_l are the (average) distances in bins k and l, T_{ik}^{obs} and T_{il}^{obs} are the total number of observed trips from origin i to all zones within distance bin k and l respectively, D_{ik}^{obs} and D_{il}^{obs} are the total observed intrinsic attractions in those bins, viewed from origin i, and I_{kl} is the set of origins for which there are supermarkets in distance bins k and l. Note that we first aggregated the numerator and denominator before dividing them by each other. In this way, the result is less sensitive to the variation between different origins.

We use the same nine distance bins as for the TG model. There are in total 36 (= 9*8 / 2) pairs of distance bins. By setting the distribution value of the first bin, $f^{obs}(d_1)$, equal to 1, all distribution values can be calculated. Because there are more distribution ratio's than distance bins, distribution values can be calculated in different ways. The distribution value in distance bin 3, for example, follows from the ratio $f^{obs}(d_3) / f^{obs}(d_1)$, but also from the product of the ratio's $f^{obs}(d_2) / f^{obs}(d_1)$ and $f^{obs}(d_3) / f^{obs}(d_2)$. These estimates are not necessarily similar. For each distance bin, we therefore took the averages of all possible combinations. It appears that this estimate is very comparable with the estimate for which the ratios between successive bins are used. The average estimate, however, can be quite different from the estimate for which the ratio with respect to the first distance bin is taken. The latter estimate is less reliable in our opinion, because the distribution would then be completely based on the first distance bin. In section 4.3, the distribution values and function fit are shown and the results are discussed.

3.4   Goodness of fit

To test the quality of each model, the log likelihood of the OD-matrix was calculated. This value shows the probability to obtain the observed OD-matrix given the assumed model parameters. The likelihood was calculated by multiplying the model probabilities of the individual observed trips from all origins to all destinations. Due to the large number of trips, the overall probability is very small. For a manageable

comparison, the logarithm (log$_{10}$) of the likelihood was taken. The higher the (log) likelihood value, the better the model performs. Because all individual trips from i to j, with observed trip frequency T_{ij}^{obs}, have the same probability π_{ij}, the log likelihood can be expressed in the following aggregated form:

$$\log_{10} L = \sum_i \sum_j T_{ij}^{obs} \ln(\pi_{ij})$$ (7)

4. RESULTS

In this section, the results for the three different models are shown. The distribution functions are shown, and the quality is expressed by the log likelihood. In section 4.1, the results of the traditional gravity model are shown. In section 4.2 and section 4.3 respectively, the results are shown for the intervening opportunities model and the limited destinations model. In section 4.4, the results are validated. By comparing model and observed trip length characteristics, we show what the implications of the model results are.

4.1 Traditional Gravity model

We investigated which distribution function best describes the data. We find that the distribution can be adequately described by the following simple function form

$$\ln f(d) = a + b * d^{\beta}$$ (8)

First, the best value for the power β was estimated. In Figure 2 we plotted the natural logarithm of the observed distribution values f^{obs} versus the average Euclidean distance to the power 0.3, $d^{0.3}$. The figure shows that, except for the first bin, the observations are fitted rather well by a linear function (solid line). The result for the first bin is not unexpected, because the relation between travel time or distance and Euclidean distance is highly uncertain for this bin. In fact, it is likely that the travel time or distance is relatively high in the first bin (compared to the Euclidean distance), and that these parameters therefore would provide a better fit in theory. Excluding the first bin, a good fit is obtained when $\beta = 0.3$. According to the least-square method, $\beta = 0.3$ provides the best linear fit. Given this value for β, the best fit was obtained with $a = 5.86 \pm 0.11$ and $b = -5.48 \pm 0.21$.
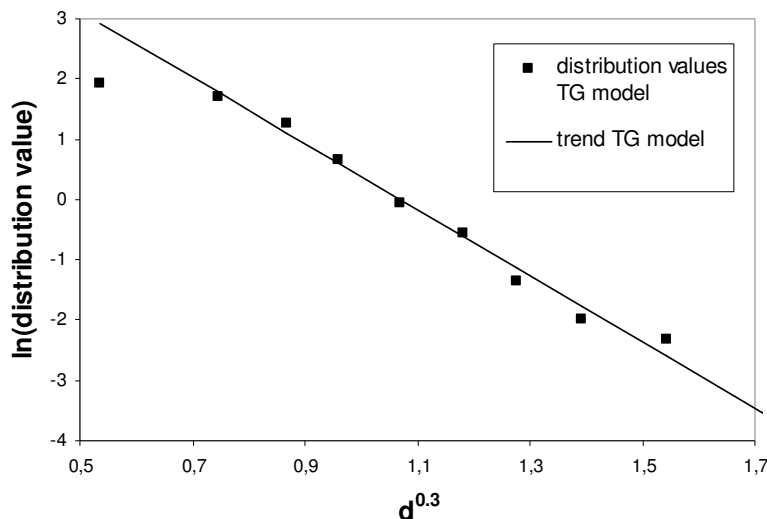

FIGURE 2 Distribution function for the traditional gravity model

The coefficient *a* is a scaling factor. In equation 8, *a* is an average for the whole sample. Normally, there are scaling factors for each row and / or column, which are determined by applying the method of Furness (*21*). These scaling factors constrain the OD matrix in such way that the row and /or column totals are equal for observations and model. However, for the interpretation of the results, they are less relevant. The coefficient *b* is the slope. The slope is highly relevant. A shallow slope indicates that opportunities far away are still an alternative. A steep slope indicates that almost all trips will go to nearby opportunities if available. According to the bias as described in section 3.3, it is expected that the slope is steeper in reality.

The log likelihood for this method is -6244. As mentioned, the log likelihood values are small due to the large OD-matrix (i.e. 1956 origins and 18 destination zones).

4.2   Intervening Opportunities model

For the IO model, the value of the scale parameter α in equation 5 needs to be estimated. This was done by maximizing the log likelihood. We found that the likelihood is maximized for $\alpha = 0.00085$. As mentioned before, calibration of the model was done assuming all origins have the same value for α. This assumption might have resulted in a sub-optimal fit. However, the computational complexity is reduced considerably.

The (maximal) log likelihood for the IO method was -6593. This is significantly lower than that of the TG model. We therefore conclude that the estimated distribution function of the IO method is inferior to that of the estimate from the TG model.

4.3   Limited Destinations model

For the LD model the same approach was followed as for the traditional estimate (section 4.1). The best linear fit (equation 8) was obtained for $\beta = 0.3$. This is similar to the estimate in the TG model. In Figure 3, the natural logarithm of the observed distribution values f^{obs} is plotted against the average Euclidian distance to the power 0.3, $d^{0.3}$. Given this value for β, the best fit (solid line in figure 3) was obtained with $a = 4.57 \pm 0.12$ and $b = -6.71 \pm 0.20$. The fit is quite good, although again the first bin is quite far off the trend line. As a reference the trend of the TG model is shown as well (dashed line in figure 3).
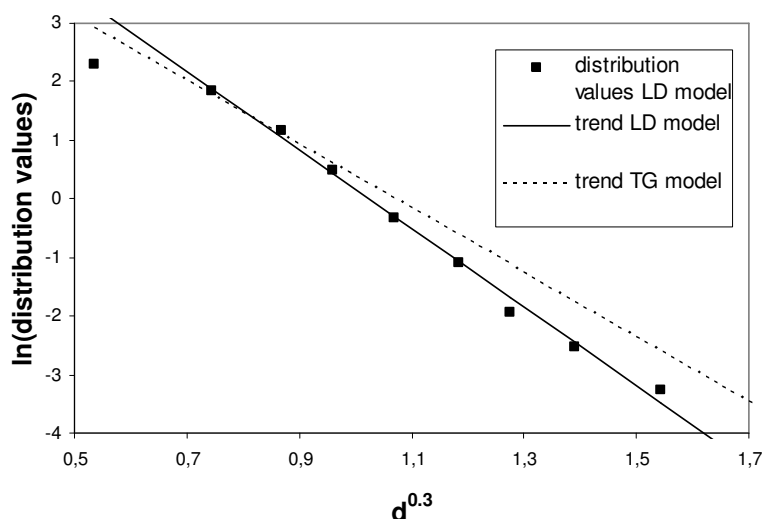

FIGURE 3 Distribution function for the limited destinations model compared to the TG model

As expected, the slope *b* is significantly steeper for this model: -6.71 compared to -5.48 in the TG model. It is assumed that there is no bias left, and that the LD model provides a better slope for the distribution function.

This result is also retrieved in the log likelihood estimate, which is -6166 for the LD model. This is significantly higher compared to the log likelihood of -6244 in the TG model. It is also significantly higher than the log likelihood of the IO model, which was -6593. One can therefore conclude that the LD yields the best distribution function.

4.4   Validation

In the previous subsections, we compared the different distribution models. We concluded that LD is the best distribution model. However, the implications of this result are not clear yet. The likelihood measure is a rather theoretical measure after all. Trip length characteristics may provide better indicators to validate the model outcomes.

In Figure 4, we show the observed trip length frequencies. The figure illustrates that shoppers mainly choose nearby opportunities, because the frequency peaks at a distance between 250 and 500 meters. However, a significant fraction of car trips still goes to supermarkets further away. This may have implications for flows on the urban network, because these trips are served by transit roads. Since the number of shopping trips is significant, the modeling of these trips should not be neglected. If trip length frequencies, for example, show that only the LD model is comparable with the observations, this would be enough reason to be concerned about traditional models.

The average trip length is perhaps the most simple validation measure. According to the survey, the average trip length is 888 meters. This is very comparable with the average trip length from the LD model, which is 884 meters. The TG model, however, yields a larger average trip length of 1033 meters. This is an overestimation of about 16%, which is quite significant. The result is not unexpected. Because the slope of the distribution function is too shallow, longer trips are significantly overestimated in the TG model.
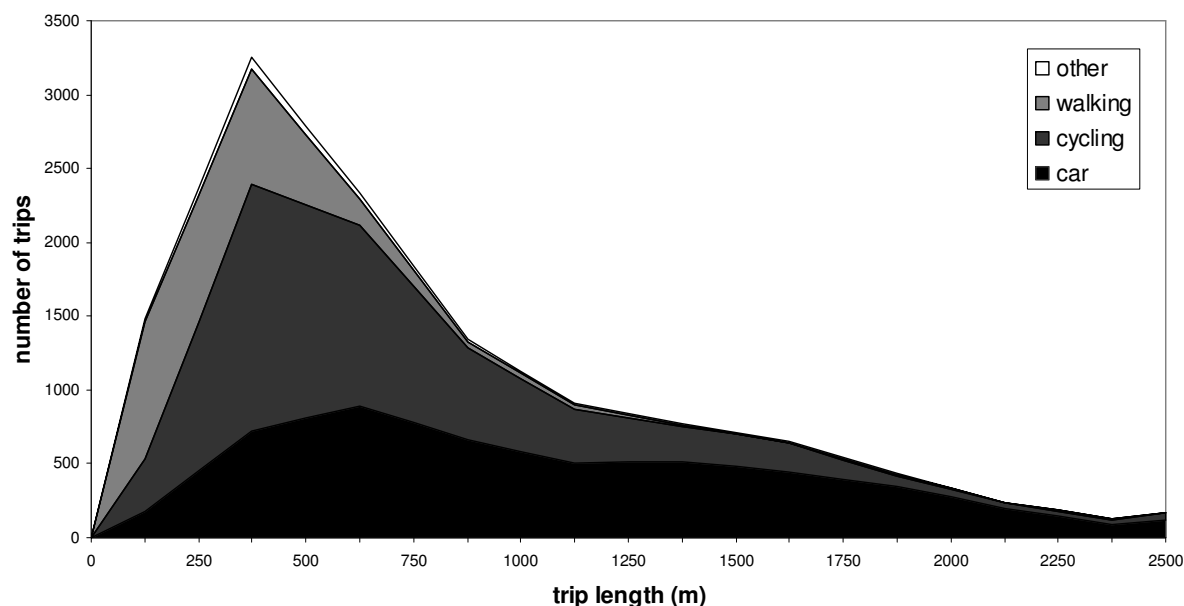
FIGURE 4 Observed trip length distribution in which the shares of the different modes are shown

The IO model yields an average trip length of 911 meters, which is quite comparable with the observed average trip length. This seems to be contradictory with the likelihood measure, which was

worst for the IO model. However, in the IO model, the spatial rank of supermarkets, rather than the distances between consecutive supermarkets, is assumed to be decisive. Dependent on the spatial configuration of supermarkets (e.g. whether supermarkets are clustered or not), this assumption may lead to a wrong assignment of the number of trips to consecutive supermarkets. This is shown in Figure 5, in which the percentages of trips to the seven closest supermarkets are plotted. For each origin, the seven closest supermarkets are ranked, i.e. the supermarket with rank one is nearest to that specific origin, supermarket with rank two, the second nearest, etc. The number of trips to each ranked supermarket is then added to the total according to its rank.
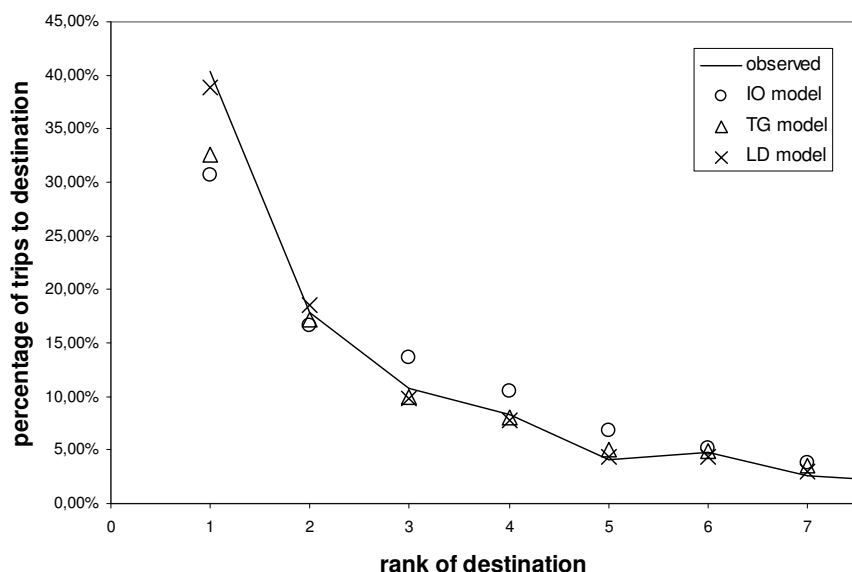
FIGURE 5 Comparison between the predicted number of trips by the models and the observed number of trips to the seven closest supermarkets. The supermarket with rank one is nearest to its origin, the one with rank two, the second nearest, etc.

The figure shows that the IO model underestimates the number of trips to the nearest supermarket, while it overestimates the number of trips to supermarkets with intermediate ranks (3, 4 and 5). The TG model also underestimates the number of trips to the nearest supermarket, but does not overestimate the number of trips to supermarkets with intermediate ranks. In fact, it overestimates the trip rates to higher ranked supermarkets further away. To keep the figure orderly, we do not show the percentages of the higher ranks.

Figure 5 shows that the LD percentages match the observed percentages quite well. Thus, from a comparison of likelihoods, average trip lengths and number of trips to the closest supermarket, we can conclude that the LD model is superior to the two other models.

5.   CONCLUSIONS

We introduced a new trip distribution model, the Limited Destinations (LD) model, for destinations that are not homogeneously distributed. Its performance was tested on a survey with reported grocery shopping trips in the Dutch city of Almelo. The results show that the LD model is reliable and that it outperforms the Traditional Gravity (TG) and Intervening Opportunities (IO) model.

The TG model implicitly assumes a homogeneous distribution of opportunities, and therefore does not pay special attention to the specific spatial configuration of destinations. As a result, the TG model overestimates the number of relatively long trips. The traditional IO model does not take the travel costs into account. As a result, it does not assign the right number of trips to consecutive destinations.

The LD model is a gravity model that combines the advantages of a TG and IO model. It describes the distribution as a simple function of Euclidean distance, which is a measure of travel costs. At the same time, it explicitly takes the spatial configuration of opportunities into account, without introducing a lot of extra computational time. As a result, the model has the highest likelihood to reproduce the observations. Compared to the observations, it also shows similar trip length characteristics.

The new distribution model was tested for grocery shopping trips, but can also be used for other trip purposes. In general, the model is suitable for distributing trips from a large number of origins over a small number of destinations. The distribution of recreational trips on a local scale (e.g. parks), on a regional scale (e.g. nature reserves) or on a national scale (e.g. amusement parks) can probably also better be estimated with the LD model. This may be tested in future studies.

A general problem is how to determine the deterrence. In this case the deterrence is assumed to be directly related to the Euclidean distance. Network distances or actual travel times would probably be better indicators for the deterrence, but these variables are more difficult to obtain. It might be possible that the fit will improve if the distribution is described as a function of network distance or travel time. We assume, however, that uncertainties in these estimates would yield less reliable results in the end. Besides travel time or distance, the deterrence may also depend on other spatial factors. Trip chaining may influence the distribution, i.e. the choice of a supermarket does not only depend on its location with respect to the residence, but also on the locations of other activities, e.g., (22). The Omnibus survey only provides information of home-bound trips. However, we think that the results from this study are still very useful, because about 80% of shopping trips in the Netherlands are home-bound according to MON.

The observed trip length frequencies show that a significant fraction of grocery shoppers does not shop at the nearest supermarket. As grocery shopping trips form a large fraction of car trips within cities, we suggest that the distribution model of grocery shopping is quite relevant for the prediction of urban traffic flows.

ACKNOWLEDGEMENTS

REFERENCES
(1)   MNP. *Fijn stof nader bekeken; de stand van zaken in het dossier fijn stof.* Milieu- en Natuurplanbureau en de sector Milieu en Veiligheid van het Rijksinstituut voor Volksgezondheid en Milieu, 2005, Bilthoven

(2)   Central Bureau for Statistics. *Vier op tien Nederlanders heeft last van verkeer;* March 10 2008. http://www.cbs.nl/nl-NL/menu/themas/veiligheid-recht/publicaties/artikelen/archief/2008/2008-2409-wm.htm. Accessed, February 4, 2009

(3)   MON. *Mobiliteitsonderzoek Nederland 2006; tabellenboek.* Ministerie van Verkeer en Watertstaat, Rijkswaterstaat, 2007

(4)   Ibrahim M.F. Disaggregating the travel components in shopping centre choice. An agenda for valuation practices. *Journal of Property Investment & Finance;* Vol 20, No. 3, 2002, pp. 277-294

(5)   Jang T.Y. Count data models for trip generation. *Journal of Transportation Engineering,* Vol. 131, No. 6, 2005, pp. 444-450

(6)   Simma, A., P. Cattaneo, M. Baumeler, and K.W. Axhausen. *Factors influencing the individual shopping behavior: The case of Switzerland;* ARE - Swiss Federal Office for Spatial Development, 2004

(7)   Ortuzar J. de Dios, and L.G. Willumsen. *Modelling Transport; third edition.* Wiley & Sons, New York, 2001

(8)   Stouffer S.A. Intervening opportunities: A theory relating mobility and distance. *American Sociological Review 5,* 1940, pp. 295-303

(9)   Chen B. *Modeling destination choice in hurricane evaluation with an intervening opportunities model;* Louisiana State University, Tsinghua University, May 2005 http://etd.lsu.edu/docs/available/etd-04132005-173111/unrestricted/Chen_thesis.pdf. Accessed, July 30, 2009

(10)   Eash R. *Development of a Doubly Constrained Intervening Opportunities Model for Trip Distribution;* Chicago Area Transportation Study (CATS) working paper 84-7, 1984

(11)   Wilson, A.G. *Urban and regional models in geography and planning,* Wiley & Sons, United Kingdom, 1974

(12)   Cascetta E., F. Pagliara, and A. Papola. Alternative approaches to trip distribution modelling: A retrospective review and suggestions for combining different approaches. *Papers in regional Science 86/4,* 2007, pp. 597-620

(13)   Akwawua S., and J. Pooler. The development of an intervening opportunities model with spatial dominance effects. *Journal of Geographical Systems 3,* 2001, pp. 69-86

(14)   Clarke M., M. Dix, and P. Jones. Error and uncertainty in travel surveys. *Transportation 10,* 1981, pp. 105-126

(15)   Stopher P.R., and S.P. Greaves. Household travel surveys: Where are we going? The Institute of Transport and Logistics Studies, University of Sydney, NSW 2006, Australia. *Transportation Research Part A 41,* 2007, pp. 367–381

(16)   Chalasani, V.S., J.M. Denstali, O. Engebretsen, and K.W. Axhausen. *Precision of geocoded locations and network distance estimates.* Arbeidsbericht Verkehrs- und Raumplanung, 256, IVT. ETH Zürich, 2004

(17)   Witlox F. Evaluating the reliability of reported distance data in urban travel behavior analysis. *Journal of transport geography 15,* 2007, pp. 172-183

(18)   Thomas T., Tutert S.I.A. Parameterization of distribution using survey data: The influence of spatial factors on commuting trips in The Netherlands. In *Proceedings of the International Workshop on Traffic Data Collection and its Standardization*, Barcelona, September 2008

(19)   Riet H. van, and P. Hospers. Altijd plaats bij de supermarkt? Nieuwe normen voor verkeersproductie en parkeerbehoefte supermarkten; adviesbureau RBOI. *Verkeerskunde, edition 2003* no. 5 pp. 46-52

(20)   CROW. *Verkeersgeneratie woon- en werkgebieden, vuistregels en kengetallen gemotoriseerd verkeer;* CROW, Ede, 2007

(21)   Furness K.P. Time Function Interaction. *Traffic Engineering and Control,* Vol. 7, No. 7, 1970, pp. 19-36

(22)   Bernardin V.L., Koppelman F., and D. Boyce. Enhanced destination choice models Incorporating agglomeration related to trip chaining while controlling for spatial competition. In *Proceedings of the TRB congress*, Washington, 2008