An extended Huff-model for robustly benchmarking and predicting retail network performance

Matthias De Beule[a,*], Dirk Van den Poel[a], Nico Van de Weghe[b]

[a] *Faculty of Economics and Business Administration, Ghent University ,Tweekerkenstraat 2, B-9000 Ghent, Belgium*
[b] *Department of Geography, Ghent University, Krijgslaan 281 (S8), B-9000 Ghent, Belgium*

Abstract

This study proposes a modified Huff model that takes directly into account spatial competition between stores of the same brand, brand attraction based on actual brand performance and spatially variable substitution. The model uses only publicly available or easily acquirable data as input, whereas model output is extensively validated on various levels. These levels include comparison of modeled and real market shares on block, store and brand level for the Belgian food market. Results show that multi-objective optimization of model parameters yields comparable results on block level to other models in the literature but improved results on store and brand levels, thereby ensuring model robustness. This robustness also enables the application of the model for various business purposes as store location determination, leaflet distribution optimization, store and store concept benchmarking, without loss of spatial generality.

Keywords:
Huff model, retail management, spatial competition, multi-objective optimization, store benchmarking, turnover prediction

1. Introduction

To monitor operational performance, retailers rely more and more on objective store benchmarks. Benchmarks are objective in a way that they quantify internal and external influences on store performance (store size, brand, competition, geodemographic characteristics of consumers, etc.) to obtain a measure indicating the performance of the management. The more fine-grained such store benchmark is, based on for instance loyalty card information, the more targeted improvement actions can be defined. A store benchmark on a fine-grained block level is therefore more valuable than a benchmark on an aggregate store level for defining and monitoring the impact of marketing actions such as door-by-door leaflet drops. In expansion strategy, accurately predicting turnover for a new outlet is also of primary importance for today's retailers. An accurate turnover

*Technologiepark 3, B-9052 Zwijnaarde, Belgium. Tel.: +32(0)92415635 Fax.: +32(0)92415656.
Email address: `Matthias.DeBeule@UGent.be` (Matthias De Beule)

prediction can quickly indicate whether it is still worthwhile to pursue a scarce city center development opportunity or to accurately assess the opportunity cost on the future network of opening a new store outside the city center, where supply of potential location alternatives is still more abundant.

In the next chapters, we propose a Huff-model that provides both a robust benchmark for current stores and an accurate turnover prediction for new stores, applied to the Belgian food market. In chapter 2, we explain in what ways our new approach extends the current state-of-art on store benchmarking and prediction techniques. Chapter 3 covers the development of the new model. In chapters 4 and 5, we explain what data we use as input and validation data and how model performance is measured. In chapter 6 we discuss the performance of our model after optimization, both in comparison with other Huff-models and of the individual contribution to overall effectiveness of the model of the different model building blocks. Finally, in chapter 7, the results of this study are summarized and managerial implications and limitations for using this model in practice are discussed.

2. Literature and own approach

Many approaches to benchmarking and predicting turnover exist, ranging from simple methods as experience and analogs, over regression analyses to more complex methods as spatial interaction modeling and neural networks [(1)].

Already in 1964, Huff showed that gravity modeling techniques can have a significant contribution to solving these retail network management issues [(2)]. By calculating customer's probabilities for store patronage, the Huff model embodied an important milestone in scientifically assessing store trade areas. The model states that the market share of a store in a given region is proportional to the utility for consumers in this region generated by this store to the total utility generated by all stores in the neighborhood of this region.

Ever since the formulation of the basic model in 1964, many extensions have been proposed to improve the predictive accuracy of this type of gravity model. Lakshmanan and Hansen [(3)] argued that a non-linear relationship between attraction and store size increases patronage prediction accuracy because the utility trade-off between store size and travel distance was now more flexible. Nakanishi and Cooper [(4)] proposed a strategy to estimate model parameters using ordinary least square estimations when a log transformation is applied to the different drivers of store attraction. Stanley and Sewall [(5)] added brand image to the attractiveness drivers of a store. Ghosh [(6)] was the first to account for spatial non-stationarity of the parameters used in a gravity model, because the relevance and impact of different drivers of store attractiveness can vary across geographic regions. Orpana and Lampinen [(7)] introduced different store concepts in the gravity model based on the size of grocery stores. A separate set of parameters for each store concept was estimated to model the varying impact of store attractiveness drivers on each store concept as they serve a different shopping purpose.

Next to finding the right drivers and estimation procedures, many applications of the Huff model have been proposed and tested in literature. These applications include university campus selection [(8)], store selection in the furniture market [(9)], the choice of movie theater [(10)], and the analysis of spatial access to health services [(11,12)]. The most common application in both literature and practice however, is found in the grocery

market, since it is one of the most saturated markets, for which benchmarking and a predictive model is most valuable.

We argue that in current approaches proposed in the literature several shortcomings can be found. Firstly, very few research has looked into the impact of the spatial configuration of the store networks and more specifically how the presence of multiple stores of the same retail chain in a customer's choice set can influence store results in that area. Secondly, we noticed a lack of variety of information used to validate the proposed models. This is mainly due to the fact that most, if not all, papers focus solely on answering one management issue. For example, Orpana and Lampinen [7], Yingru and Liu [13], and Sandikcioglu et al. [14] focus solely on the prediction accuracy for retail locations. For this purpose they only use information on a store level, which yielded good results for their purpose. Less research has been conducted on block level, based on questionnaires or loyalty card information. Gauri et al. [15] use such block level information and gravity modeling techniques for a store performance benchmark exercise. Although the results on block level for the performance benchmark were good, the results on a more aggregate store level were less satisfactory. None of the existing work on gravity modeling has incorporated results on a higher level, the food retail chain, despite being readily available in a nation's database of financial statements. A final shortcoming can be found in the type of input data used in existing gravity models. Collecting a wide variety of input data to capture more influencing factors [16] can be extremely time consuming or very costly when bought. Retailers are therefore often reluctant to acquire these data because the marginal benefit of incorporating these data in practice has become questionable. In this paper, we show how easily available information can be used for maximum applicability and results in practice, ensuring high return on investment.

This paper aims at constructing a robust gravity model for the whole Belgian grocery market, using an extensive set of easy-to-gather input and validation data. In doing so, we address the three aforementioned shortcomings. Firstly, the state of art of the Huff-model is extended by incorporating more spatially influencing factors, such as brand recognition and internal cannibalization of sales between stores of the retail chain. The inclusion of such factors can provide valuable insights in a retail chain's network expansion strategy. Secondly, block level information drawn from a grocery retailer's Customer Relationship Management database is used in addition to annual store turnovers from the same grocery retailer and annually reported group turnovers for all competitors as reported in their financial statements. Validation on these three levels is applied for an improved robustness of the proposed model. Lastly, in our approach, only easy-to-gather input data on a national scale is used. Therefore, we limit our model to the store surface and the store brand as a measure of store attractiveness. Addresses and brands of stores can easily be acquired using company websites and common knowledge of the competitive landscape. While calculating surfaces on a large scale can be time consuming, the spread of freely accessible aerial photographs (Google Earth, Bing Maps) [13] and more detailed socio-economic permits have sped up its calculation considerably.

3. Model Development

Starting from the basic Huff model, this section explains the extensions that seek to improve predictive and benchmarking accuracy on block, store and chain level.

Basic Huff model

As a starting point for our model we use the Huff model as proposed in 1964. It states that the patronage probability P_{ij} of a store j for inhabitants and workers in a given region i (henceforth named 'residents of block i') is equal to the proportional utility of this store (U_{ij}) compared to the total utility generated by all N stores in the neighborhood of this region:

$$P_{ij} = \frac{U_{ij}}{\sum_{q=1}^{N} U_{iq}} \tag{1}$$

The utility generated by grocery store j for residents of block i is calculated as:

$$U_{ij} = \frac{A_j}{{D_{ij}}^{\beta}} \tag{2}$$

The value A_j represents the aspatial attractiveness component for store j. In the basic Huff model, store size is used for A_j. As mentioned in section 2, it is however possible to incorporate more drivers for aspatial store attractiveness by averaging or multiplying different drivers. D_{ij} is the distance between store j and the centroid of block i. In most research, Euclidian distance based drive times are used. However, with recent technology advances, the calculation of fastest route drive times has become feasible, even for large scale projects. The parameter β shows the relationship between distance and attractiveness of the store.

To translate probabilities from formula 1 into monetary allocations, it is assumed that the total spending potential of a block is divided evenly according to the store visit probabilities P_{ij} for all stores j in close proximity.

$$F_{ij} = P_{ij} * SP_i \tag{3}$$

Where F_{ij} equals the monetary flow between store j, and block i and SP_i is the total spending potential on groceries of all residents of block i.

Extending the Huff model

Taking the above basic formulation as a starting point, we now further develop this model to incorporate more influencing factors on store choice probabilities. The development of the model is explained in three phases. In the first phase, an Unrelated Total Attraction (UTA_{ij}) for every block i in regard to store j is calculated. In the next phase, UTA_{ij} is modified to account for weakening and fortifying effects of regional brand presence, resulting in a Related Total Attraction (RTA_{ij}). Finally, after incorporating substitution for grocery spending in grocery stores in the model, store visit probabilities are calculated using the Related Total Attraction. The resulting monetary allocations then can be validated with real sales information.

Phase 1: the construction of UTA_{ij}

S_j **- Store size**

Larger stores carry a more complete and voluminous range of grocery products. More choice options and a better product availability tends to be more attractive to consumers.

BA_{b_j} **- Brand Attraction**

Another important influencing factor on store choice is the brand each grocery store belongs to, as each grocery store chain has its own store format. Incorporating a brand related attraction value in the model thereby reflects two influencing factors: shelve density and attraction of the brand format to consumers. Although store size is an adequate proxy for the range of products carried, the different store formats have varying shelve densities, resulting in fluctuating sales per square meter. The incorporation of such a brand attraction measure can then refine the impact of store size on store attractiveness. Also, due to pricing and/or product strategy differences, some grocery store chains are more attractive to consumers than others. Using the global turnover results of each grocery store chain and the total surface of their stores in Belgium, an average annual turnover per square meter, BA_{b_j}, can be calculated, which is a good relative approximation of the attractiveness of the brand concept b_j, independent of the store j's size. For a market entrant the application of this approach is difficult, as they haven't realized any turnover yet. This can however be overcome by using the same BA as an existing firm following a similar strategy.

LB_{ij} **- Language Borders**

Belgium is characterized by its division in three major geodemographic areas: Flanders, Brussels and Wallonia. In Flanders the mother tongue is Dutch, while the native language in Wallonia is French. Finally, Brussels is characterized by both Dutch and French speakers. Due to these language borders, there is a preference for most people to shop only in their own geodemographic area. To model these geodemographic borders, penalties for cross-border utility calculations are calculated, according to which specific geodemographic border is crossed (Figure 1). These penalty values have been estimated based on expert interviews. A penalty of 0.1 corresponds for example with a 90% reduction of the store attractiveness. Moreover, since the majority of the focal brand's stores are located in the southern part of Belgium, we also took French grocery stores close to the Belgian border into account. These cross-nation allocations are also subject to a penalty according to the language of the resident of a block and the area in which the store is located.

K_j **- Grocery Store Concepts**

Different store concepts have also spatial differences in attraction. Hypermarkets are characterized by the largest store surfaces in the grocery market and usually have the largest parking spaces. From a spatial point of view, it significantly increases the fixed time cost of visiting this type of store concept. From an aspatial point of view, they also carry the most complete range of grocery products, as covered in brand attraction and store surface. This store configuration tends to be more attractive to consumers from distant areas, who prefer large quantity one-stop shopping trips, thereby reducing the relative impact of the larger fixed time costs on the total time cost of their shopping trips. For residents at closer distances however, the impact of the higher fixed time costs is often too high for top-up shopping trips, which reduces the relative attractiveness of these hypermarkets for consumers at closer distances. Local shops are characterized

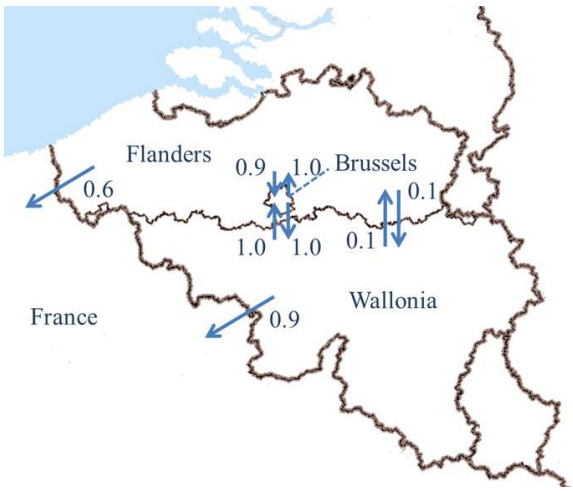

Figure 1: Graphical depiction of the attraction multipliers across geodemographic borders.

Sales surface	Store concept
$< 400\,m^2$	Local grocery store
$\geq 400\,m^2$ and $< 2,500\,m^2$	Supermarket
$\geq 2,500\,m^2$	Hypermarket

Table 1: Different store concepts used.

by the inverse relative attractiveness. They are very attractive for local residents for quick top-up shopping, while being less attractive to more distant residents as their limited range of products prevents a time-equitable one-stop shopping trip. To model these spatial differences in attractiveness between different store concept, we divide the grocery stores in scope into three categories: local grocery stores, supermarkets and hypermarkets. For each of these grocery store concepts, separate travel-time dependent parameters are introduced. Table 1 presents the classification as proposed by Orpana and Lampinen[7], which is also used in this study.

Also, a fourth store concept is introduced for the retailer who provided the sales data, both loyalty card information and store turnovers. This choice is motivated by the possibility these sales data offer to model their specific market dynamics more accurately than brands for whom we only have sales data on brand level, while avoiding overfitting for these other brands. When using this model for another retailer, it also means the model has to be re-estimated using their specific data.

To accurately model these differences in spatial attractiveness, we introduce both a global attractiveness parameter SC and a distance related parameter DP for every store concept k_j:

SC_{k_j} **- Global impact of store concepts**

The typology of store concept has a fixed influence on the incurred time cost. Other researchers have also implemented these ideas, either implicitly or explicitly: Pauler et al.[17] and Gauri et al.[15] augment the Euclidian distance between consumers and grocery stores by a fixed increment, thereby implicitly accounting for a fixed time cost. Orpana and Lampinen[7] also add a fixed time increment to the distance function. We propose a similar modification in the distance function specification which also accounts for an incurred fix time cost. Section A of Figure 2 shows such a classic Huff distance-attraction decay with a fixed time penalty. In literature, many other forms of distance-attraction decay have been proposed[18]. In this study, an exponential relationship is used: $SC_{k_j}/exp(D_{ij} * DP_{k_j})$. For every store concept k_j, parameter SC_{k_j} will indicate the relative fixed time cost increment, as shown in section B of Figure 2.

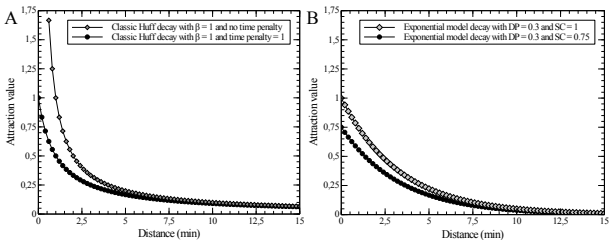

Figure 2: Comparison between classic Huff decay with time penalty and the Huff decay proposed in this study.

DP_{k_j} - The impact of distance

We measured the distance between customers and stores as the average between Euclidian distance based travel time and fastest route travel time, since customers not only judge the spatial attractiveness of a store on the travel time of the fastest route but on geographical proximity as well. We refer to section 6 for a proof of the contribution of this approach to the overall effectiveness of the model.

Parameter DP_{k_j}, combined with the fixed time cost parameter SC_{k_j}, determine the time-dependent attraction of each store concept. Figure 3 shows a distance-attraction relation for each store concept. Independent of their surface, a local grocery store has greater local attraction then any of the other store concepts, while a hypermarket has greater attraction on longer distances.

Combining the previous drivers of store attractiveness, we can now calculate the Unrelated Total Attraction of every grocery store j close to block i:

$$UTA_{ij} = \frac{S_j * BA_{b_j} * LB_{ij} * SC_{k_j}}{e^{D_{ij} * DP_{k_j}}} \tag{4}$$

The Unrelated Total Attraction of every grocery store j for block i is thus directly proportional to the average turnover of a store from brand b_j with surface S_j weighted by language border penalties LB_{ij} and the store concept impact SC_{k_j} and inversely proportional to an increasing function of the distance to the store D_{ij}.

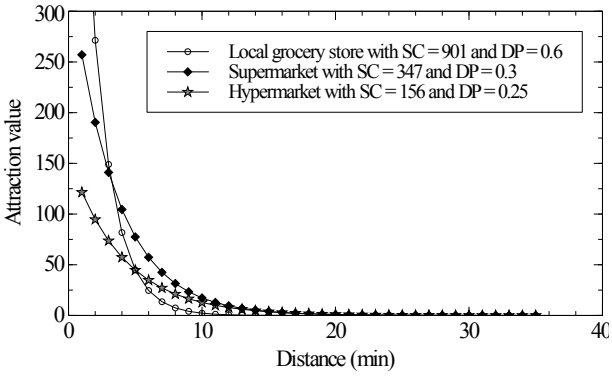

Figure 3: The distance-attraction relationship for each store concept. The values for SC and DP were chosen based on the optimal solution obtained in the results section.

Phase 2: the calculation of RTA_{ij}

The presence of stores of the same brand in a region can have both fortifying and weakening effects on the attractiveness of a grocery store. First of all, the biggest competitors of a grocery store that is part of a chain are neighboring stores of the same chain. While the Huff model takes competition between stores of different brands directly into account in the utility values, it is not as accurately accommodated to take competition within a brand into account. If, for example, for a certain geographic area, two grocery stores of different brands are in scope, customers will divide their purchases according to the stores' respective attractiveness values. This division is however much more unlikely if the stores belong to the same brand. In this situation, the store with the highest attractiveness is likely to attract more than its share attributed by a classic Huff model, because both stores are almost perfect substitutes and rational consumers will virtually only visit the store providing them with the highest utility. Therefore, we attribute a penalty to all but the most attractive stores per brand in the eyes of the residents of every region.

This penalty is calculated as follows:

$$CF_{k_j}^{\left(\frac{\sum_{q|b_q=b_j, UTA_{iq}>UTA_{ij}} UTA_{iq}}{UTA_{ij}}\right)} \tag{5}$$

Where CF_{k_j} $(0 \leq CF_{k_j} \leq 1)$ is a cannibalization penalty factor and is a parameter that will be estimated per store concept k. The power to which CF_{k_j} is raised depends on the ratio between the Unrelated Total Attraction of more attractive stores of the same brand for residents of region i and store j's Unrelated Total Attraction. A similar approach was used by Kaufmann and Rangan [19], who developed a model for site location for a franchise company. In this model, they argue that customers choose the franchisee that provides them the highest utility among all other available franchisees. Such an approach can be achieved in our model when CF_{k_j} approaches zero. When CF_{k_j} is 1, the classic Huff model is attained. A similar notion is used by Wan et al. [11] for correctly determining the demand for health services. In the proposed three-step floating catchment area (3SFCA) method, the demand for health services provided by a medical

facility is also cannibalized by the presence of other facilities in closer proximity to a block.

At the same time, the presence of multiple stores of the same brand in close proximity has a reinforcing effect on the attractiveness of all of these stores. Naert and Bultez [20] argued that a logistic 'S' relationship exists between market share per store and the number of stores of the same brand in geographic proximity. When opening a first store in a region, consumers are not yet familiar with the format of the chain. The more stores of the brand that have opened in the region, the more familiar consumers become with the concept, hence the increased market share per store. Naturally, with an even larger increase in numbers, the marginal effects of an additional store start to decrease. The brand presence BP_{ib_j} of a brand b_j for block i is calculated as follows:

$$BP_{ib_j} = 1 + BPF * BPS_{ib_j} \tag{6}$$

where BPS_{ib_j} is defined as the relative share of grocery stores of brand b_j for every geographic block i and BPF is a parameter optimizing the impact of the brand presence. The relative share of grocery stores of brand b_j for block i is calculated as the number of stores of brand b_j within a 20 minute drive time radius on the total number of stores within the same time radius. As figure 4 indicates, the BPS factor for the focal retailer is zero for the majority of blocks, since its network contains only 61 stores. Furthermore, the maximum BPS of 25% -meaning that 1 in 4 grocery markets within 20 minutes of these blocks belong to the focal retailer- indicates high local concentrations of focal stores.

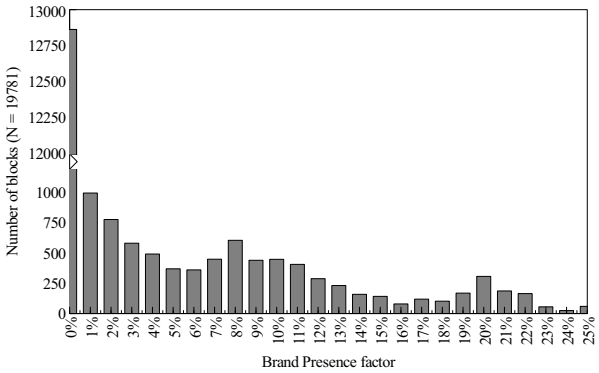

Figure 4: The spread of Brand Presence factors for the focal retailer for all geodemographic blocks.

When comparing figure 5A with 5B, it is clear that brand presence has reinforced the individual attraction of each of both stores in close proximity to the store, while for the zones in between both stores, where the internal cannibalization is the strongest, the clear preference for one of both stores has weakened the aggregated attraction of both stores.

By taking these factors into account, the Indepency of Irrelevant Alternatives (IIA) property from which a classic Huff model suffers, is also partially addressed. The IIA property states that the ratio of the probabilities of an individual selecting two alternative

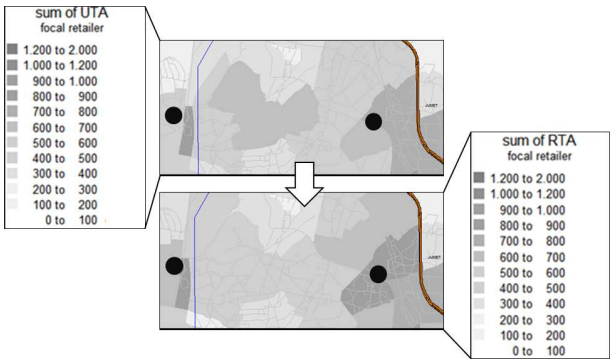

Figure 5: The aggregated Unrelated and Related Total Attraction values per geographic block for 2 stores of the focal retailer.

stores is unaffected by the addition of a third alternative store[21]. In our model, the introduction of a new alternative effectively influences the relative preferences of existing choice options when taking brand into account, as fortifying and weakening effects of brand presence will also influence the attractiveness of existing store options.

With these fortifying and weakening effects of brand presence, we can now calculate the Related Total Attraction of every store j close to block i:

$$RTA_{ij} = BP_{ib_j} * UTA_{ij} * CF_{k_j}{}^{\left(\frac{\sum_{q|b_q=b_j, UTA_{iq}>UTA_{ij}} UTA_{iq}}{UTA_{ij}}\right)} \tag{7}$$

Phase 3: The calculation of the store visit probabilities.

In this final phase, we transform the Related Total Attraction values to store visit probabilities and finally to monetary allocations.

It is however highly unlikely that all of the grocery budget within a family will be allocated to the grocery stores in our database. Substitution is often triggered by the absence of close grocery stores or by servitized alternatives like restaurants. Therefore, we incorporate two parameters FS and RS that model these two substitution possibilities when calculating store visit probabilities P_{ij} (see Equation 8). FS reflects a fixed attraction to substitutes regardless of any region specific characteristics. If there are abundant grocery stores in close proximity, i.e. large aggregated RTA values, much of the potential demand will be triggered. This is reflected in the fact that FS will be relatively small compared to $\sum_{q=1}^{N} RTA_{iq}$ and substitution thus will be minimal. Servitized substitution alternatives are more likely to be located in densely populated areas. Therefore, we multiply the population density PD_i in region i with a parameter RS to obtain a measure of servitized substitution.

$$P_{ij} = \frac{RTA_{ij}}{(\sum_{q=1}^{N} RTA_{iq}) + FS + RS * PD_i} \tag{8}$$

Finally monetary allocations can be calculated using equation 3:

$$F_{ij} = P_{ij} * SP_i * PM \tag{9}$$

Where PM is a Potential Multiplier used to fine-tune the spending potential figures SP_i we pre-calculated.
After obtaining all allocations, aggregations can be made to obtain results on store and brand level. Figure 6 shows such an aggregation for the focal retailer. Comparison with true allocations on block level or turnovers on store level is then the basis of model optimization.

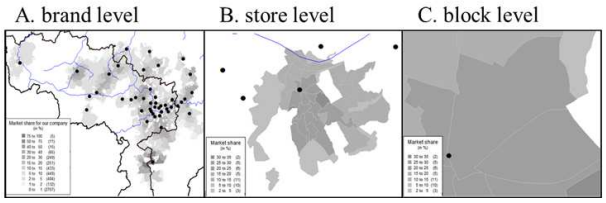

Figure 6: Graphical depiction of the three benchmarking levels.

4. Test Design

Solution procedure

Due to the non-linear nature of the proposed model, linear regression techniques for parameter optimization cannot be applied. Optimization techniques that are capable of dealing with such highly complex non-linear optimization problems, are for example meta-heuristics. They however cannot ensure an optimal solution. We opted for a simulated annealing (SA) solution procedure, which is part of the descent family of meta-heuristics. Simulated Annealing was introduced in 1983 by Kirkpatrick, Gelatt and Vecchi as a probabilistic solution method capable of finding very good results in limited computing time. It is also commonly used as a multiobjective optimization strategy [(23)]. The use of validation data on different levels allows for a multiobjective optimization and provides a robust model. When optimizing a multiobjective optimization problem, a set of pareto-optimal solutions is obtained. Pareto-optimal solutions are solutions for which there exists no feasible solution that equals or outperforms this solution on all criteria of the multiobjective optimization problem, in this on case block, store and brand level. However, it needs to be pointed out that an intelligent steering of the SA procedure for this problem formulation is very difficult. The highly complex definition of the problem makes a neighborhood definition around an accepted solution very difficult to define. In order to use the benefits of the SA intelligence to its best, the control of the deterioration acceptance for temporary solutions was controlled on store level, having stabler neighborhoods than the more fine-grained block level, while allowing for better parameter fitting than the more aggregated brand level. To ensure good results within reasonable time however, we opted for a standard multiobjective simulated annealing (MOSA) procedure with slow temperature decrease and allowed the search procedure 10.000 iterations to calibrate the optimization parameters.

Performance benchmarks

The quality of the proposed model is evaluated in comparison to a classic Huff model and an extended Huff model found in literature that was applicable to our dataset. Benchmarking with other models in the literature has indeed proven to be very difficult, since most papers work with a very broad range of input data, which are often not available or not relevant outside their test environment. A first benchmark is made with the basic Huff model as proposed by Huff (see equation 2). Model M1 of the extended Huff model proposed by Orpana and Lampinen [7] is also tested:

$$U_{ij} = SC_{k_j} * {S_j}^{\alpha_{k_j}} * D_{ij}^{\beta_{k_j}} \tag{10}$$

where the number of store concepts j is only 3, compared to 4 in our model.

As a performance measure on all 3 levels, the Mean Average Percentage Error (MAPE) is calculated:

$$MAPE = \frac{1}{N} \sum\nolimits_{n=1}^{N} \frac{|O_n - M_n|}{O_n} \tag{11}$$

Where N is the number of observations depending on the level of the validation data (block, store or brand level). O_n is the true observed result for area n, while M_n is the modeled result for area n (if on block level, $M_n = F_{ij}$). For more comparative results, the pseudo R^2 fit measure, used by Gauri et al. [15], is also reported:

$$PseudoR^2 = \frac{var(O(MS_n)) - var(\epsilon_n)}{var(O(MS_n))} \tag{12}$$

Where $O(MS_n)$ is the observed market share for area n (block, store or brand level) and $\epsilon_n = O(MS_n) - E(MS_n)$ where $E(MS_n)$ is the expected market share in area n according to the model. This fit measure thus indicates how much of the observed variance is explained by the model. However, since we are unable to use regression techniques and have opted for a meta-heuristic, $E(\epsilon_n) \neq 0$. If optimized towards this performance measure, robustness of the solution cannot be guaranteed as skewness in the results cannot be prevented. To still ensure comparable results, we added a 6% deviation constraint to the mean percentage deviation of the results on store level. Due to the limited number of observations on brand level, the pseudo R^2 will be reported only on block and store level. To avoid overfitting, we subdivided the allocations on block level into a 2/3 training and a 1/3 validation set. We did not opt for a test and validation set on store and brand level, given the limited number of observations.

Calculation performance improvements

Because of the scale of this research, calculation time per iteration can be quite long. To reduce the calculative burden, we added constraints on how many stores are evaluated per block. Indeed, it can be assumed that from a certain number of stores on, the true allocations of spending potential become negligible. We fixed this number on the 18 closest local grocery stores and supermarkets and the 2 closest hypermarkets. Preliminary evidence showed that this number yielded sufficiently accurate results while maintaining acceptable calculation times.

5. Data

The proposed model was tested on a national scale. For this purpose, an inventory of grocery stores in Belgium was fetched. For ease of gathering this information, only grocery stores belonging to a food chain were added, provided that the food chain had at least seven grocery stores in Belgium. This process yielded a database with 3,420 grocery stores, belonging to 34 food chains. To complete this database, we added net sales surfaces of these grocery stores. Due to the availability of high resolution aerial photographs it is much easier to accurately estimate sales surfaces of stores. This process is for example used by Yingru and Liu[13]. Next, we obtained annual sales data from a major food retailer in Belgium for the year 2010. For 61 (out of 63) supermarkets (with surfaces between $600m$ and $2,400m$) we obtained loyalty card information. After geocoding the addresses, the annual sales quantities were allocated to the different blocks in Belgium. This aggregation resulted in 27,143 monetary allocations from geographic blocks in Belgium to the 61 stores for which loyalty card information was available. Since not every transaction is logged with a loyalty card, the current spread of sales registered with loyalty cards is corrected to obtain the complete annual turnover for every store. From the National Institute for Statistics, we obtained geodemographic information on the 19,781 geographic blocks in Belgium. Taking the number of families and the average revenue of each block into account, a partial expenditure potential on groceries was calculated. This potential was further augmented with an expenditure potential of the total workforce active in each block, since they too are prone to buy groceries before, after or during their stay at their workplace. This resulted in a total expenditure potential per block i, SP_i. Since the expenditure potentials are indications, parameter PM was introduced in the model to fine-tune the global expenditure potential. Fastest route driving times were calculated from the center of the geographic block to the exact location of the store using Microsoft MapPoint Europe 2011.

6. Results

In this section, the results of a 10,000 iteration optimization run are presented. In the first section we discuss the optimized parameters and compare the results with the performance of other gravity models from literature on our dataset. In the second section, we perform iterative sensitivity analyses to measure the contribution of several proposed drivers of store attractiveness to the overall performance of the model.

Comparative results

Table 2 shows the optimized parameters for each tested model after a 10,000 iterations optimization run. Due to the fact that substitution was explicitly taken into account in our model, the potential multiplier PM is higher compared to the other models as part of the market is not allocated to stores in our database. The store concept multipliers SC are both used in our model and the M1 model of Orpana and Lampinen. The difference in absolute magnitude of the parameter values between both models is not important as store visit probabilities are calculated, which involves a relative weighing of attraction scores (see equation 1). The relative difference in parameter value between store concepts however, is much more important and is relatively comparable between both models, indicating the same capture of store concept dynamics. This finding is

also confirmed when looking to the distance related parameters DP for our model and β for the other 2 models: the larger the store concept, the lower the impact of distance becomes. Moreover, in our model the impact of distance for our focal retailer is larger than for a comparable supermarket, indicating the somewhat more 'local store' image of the focal retailer. To correctly interpret the resulting cannibalization factors CF, it is to take the varying impact of distance for different store concepts into account. For local stores, having very small trade areas, it is much more difficult to accurately assess internal cannibalization as it is not so common that their trade areas converge. The bigger the stores become however, the more the trade areas of stores of the same branch are likely to converge and internal cannibalization becomes more important, hence the increasing penalty values for bigger store concepts.

Parameter		Our model	Basic Huff	M1
Name	Store concept			
FS	-	570	-	-
SF	-	0.0798	-	-
PM	-	0.94	0.6	0.6
BPF	-	2.1	-	-
SC	1	901	-	8
	2	347	-	4
	3	156	-	2
	4	347	-	-
DP/β	1	0.6	2	3.2
	2	0.3	-	3
	3	0.25	-	2
	4	0.4	-	-
CF	1	0.6	-	-
	2	0.35	-	-
	3	0.2	-	-
	4	0.6	-	-
α	1	-	-	0.7
	2	-	-	1
	3	-	-	1

Table 2: Parameters of the best solutions found.

Results from table 3 show that our proposed model outperforms the basic Huff model significantly on all performance measures, although our model only uses global brand results as additional input data. This also indicates that the complex non-linear relationships in our model result in significant improvements in overall accuracy. More specifically, a 66.4% mean average deviation was found on the test set on block level, whereas the validation set confirmed this result with a MAPE of 62.99%. The performance measures on block level result in relatively high mean average percentage errors, especially compared to the MAPE on store level. Since we did not trim the observed allocations, every observed allocation that did not have a modeled counterpart (or vice versa) resulted in a 100% deviation for that block. When taking only the 500 biggest allocations into account, we see a remarkable decrease in MAPE on block level to 37%. Since

these minor allocations have a relatively small impact on store level, the mean average percentage error on store level drops significantly to 22.34%. On brand level a MAPE of 22.28% was found, which is very satisfactory given the fixed nature of the Brand Attractions. Furthermore, a nearly 50% increase in explanatory power of the model compared to the basic Huff model is found when looking at the pseudo R^2. Our model explains 76% of the variance in market share on block level, which is in line with the results of Gauri et al.[15]. When looking at the M1 model of Orpana and Lampinen, it also outperforms the basic Huff model on all levels, thanks to the addition of parameters on store concept level. Compared to our model, results are comparable on block level, while our model outperforms the M1 model on store level, indicating that the addition of extra spatial or brand related attractiveness drivers in our model have improved accuracy on store level. Finally, the addition of Brand Attraction indicators for each brand in our model clearly benefits the result on brand level, thereby also drastically improving model robustness when for instance testing a potential store location in this competitive landscape.

Level	Block			Store	Brand
Performance measure	MAPE		R^2	MAPE	MAPE
Set	Test	Validation	Complete	Complete	Complete
Our model	66.40%	62.99%	76.11%	22.34%	22.28%
Basic Huff	117.48%	115.66%	26.47%	35.17%	57.36%
M1 Orpana and Lampinen[7]	66.19%	58.23%	71.56%	26.78%	47.78%

Table 3: Comparative results

Figure 7 shows the distribution of the 63 focal stores according to their percentage error. The maximum absolute error for one store was 66%. For all other stores the absolute percentage error was contained within 50%. Compared with other results in literature, this is a very solid outcome. Although comparison between two different geographic regions is difficult, our model returns for instance a modeled store turnover for 98.41% of the focal stores within a 50% deviation from their real turnover while the Store Performance Index presented by Gauri et al.[15] yields modeled store turnovers of approximately 85% of the focal stores within a 50% deviation from their real turnover. From a practical point of view, these deviations can be discussed with store management to improve performance or learn best practices. For simulation purposes, these figures however indicate that individual on-the-field insight in each case is necessary for an even more accurate prediction.

Sensitivity analysis

Next to comparative results, we also measured the impact of the different newly proposed attractiveness drivers on the total performance of the model. We iteratively dropped or changed one of the model building blocks to measure the drop in model accuracy. Table 4 shows the optimization results after 10,000 iterations. For generality purposes we did not test the impact of language border penalties since it is a specific Belgian characteristic. On block level, we noticed no clear evidence with the MAPE measure of contribution to overall effectiveness in all sensitivity tests. However, the R^2

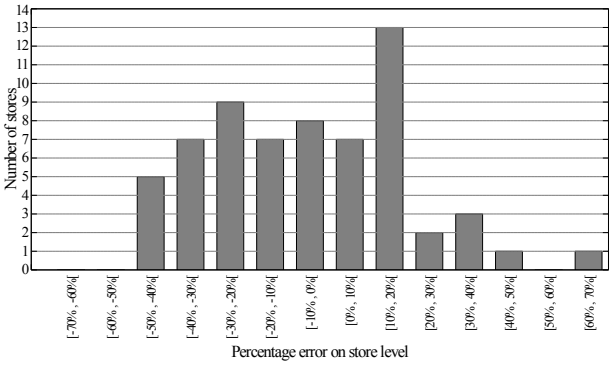

Figure 7: Distribution of percentage errors on store level

measure indicates that the contribution of brand attraction, internal cannibalization and the combination of real and Euclidian distance based drive times are significant. The improvement with the incorporation of brand presence on block level is rather small, as was also noticed when comparing our model to the M1 model of Orpana and Lampinen [7], but can be explained recalling 4, as it has a significant impact on few blocks with high brand presence.

Looking at store level results, we noticed a decrease in error terms in all six test cases. Small deteriorations are found when dropping brand attraction and brand presence drivers. The limited effect on brand attraction is due to the fact that the focal retailer has its own store concept, for which SC_{k_j} is now optimized to act as a brand attraction driver for the focal retailer. Brand presence in turn, has only a small spatial impact due to the focal retailer's specific network configuration, as shown in Figure 4, but has a strong local impact in the few areas with high brand presence. The most significant decrease in predictive error is found in the use of hybrid drive times. Figure 8 shows the accumulated turnover for blocks located at a certain drive time (x-axis) to the stores for whom we have loyalty card information. From these graphs it is clear that that our proposed model closely matches the real sales from a spatial point of view. Using solely Euclidian distance based drive times however, is clearly not as capable to capture the spatial dynamics in the market, specifically at shorter distances. Using fastest route drive times, on the other hand, enables accurate modeling at shorter distances while it fails to do so at longer distances. Using hybrid drive times significantly mitigates the shortcomings of both approaches.

Finally, on brand level, the deterioration of dropping Brand Attraction is significantly, as expected. Dropping other drivers has only marginal effects on brand level, except for the moderate influence of Internal Cannibalization, which indicates that it is an important concept to take into account when modeling a whole market segment.

7. Managerial implications, limitations and suggestions for future research

We showed that starting from a limited variety of input data that are easy to acquire, a robust multi-purpose gravity model with high accuracy can be formulated. Moreover,

Level	Block			Store	Brand
Performance measure	MAPE		R^2	MAPE	MAPE
Set	Test	Validation	Complete	Complete	Complete
Our model	66.40%	62.99%	76.11%	22.34%	22.28%
Dropping Brand Attraction	67.68%	64.72%	71.90%	22.96%	41.47%
Using Euclidian distance based drive times	65.52%	71.90%	60.20%	25.91%	23.56%
Using Fastest route drive times	63.29%	61.52%	75.45%	23.12%	25.89%
Dropping Internal Cannibalization	60.68%	65.70%	72.87%	23.51%	28.34%
Dropping Brand Presence	61.27%	65.12%	76.05%	23.01%	22.27%

Table 4: Sensitivity analysis on model building blocks

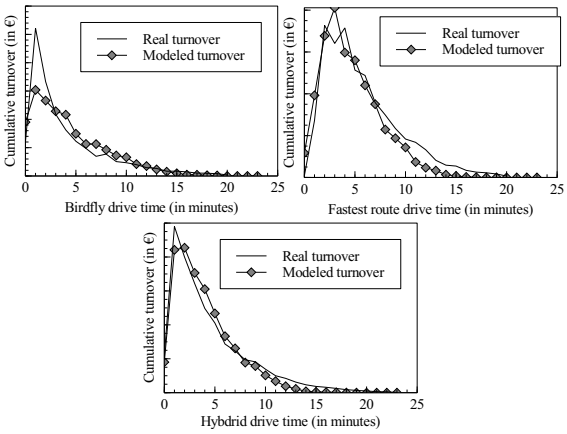

Figure 8: Comparison between modeled and true turnover for every minute driving time using different driving time calculations.

for the first time a gravity model has been validated on three different levels: block, store and brand level. On all three levels the results clearly indicate the benefit of our proposed model compared to a standard gravity model and models previously proposed in literature. More precisely, we showed that incorporating both spatial and aspatial brand related drivers of store attractiveness have a significant positive impact on predictive accuracy for a focal retailer.

The model can be used for multiple purposes in practice. The deviations on block level can be discussed with the management, and targeted actions can be defined. On a more aggregate leve, the store level, we see that the predictive accuracy is very satisfactory. Such predictive accuracy can be used for predicting turnover of a new location and, especially for our model, for accurately predicting the impact on existing stores. Although a gravity model can rapidly indicate potential turnovers and impacts on current networks, it still must be used with caution. Although we believe we have captured the most important drivers of store success in the model - except for difficult to capture drivers as store management- many more drivers have an influence on store success. Therefore, a model can never replace a visit on site as it will provide many more insights in the choice behavior of local consumers [(1)].

The aim of a model should then not be to act as a final predictor, but as an effective funneling instrument to filter sets of potential locations, as shown in figure 9.

This paper therefore aims to provide a valuable, robust starting point for retailers in their attempt to formulate a good predictive and benchmarking model. Augmenting the model with more elaborate and relevant data will virtually always contribute to an increased model accuracy and should thus be encouraged. In order to deepen the impact of brands on individual store results even further, it is for instance worthwhile looking into geodemographic segmentation of the population to model the targeted population groups of the different retailers [(25)]. Also, a more extensive validation based on detailed

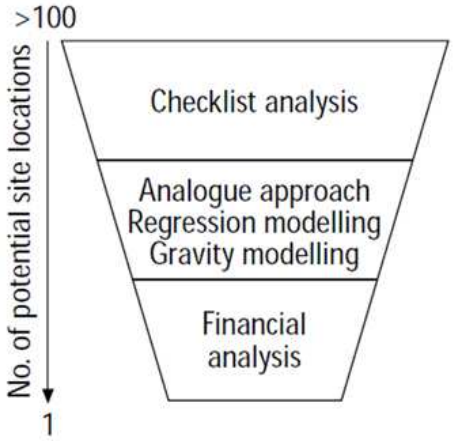

Figure 9: The process of retail location assessment [24]

results on block and store level from multiple retailers across multiple store concepts can increase model robustness and generalization. Finally, as pointed out in section 4, an intelligent optimization procedure is very difficult to configure for the highly complex formulation of the model. Limited intelligence was introduced in our optimization due to very difficult neighborhood definitions. Although the used procedure yielded satisfactory results on all levels, it cannot be assured the optimal solution was found, and a better solution still can be found using an improved optimization methodology.

8. Acknowledgment

This study was carried out within the framework of a Baekeland Project (A11/TT/1037) sponsored by the Flemish Institute for the Promotion of Scientific and Technological Research in Industry (IWT) and in cooperation with Geo Intelligence BVBA.

References

1. Wood S, Tasker A. The importance of context in store forecasting: The site visit in retail location decision-making. *Journal of Targeting, Measurement and Analysis for Marketing* 2008;:139–155.
2. Huff DL. Defining and estimating a trade area. *Journal of Marketing* 1964;:34–38.
3. Lakshmanan JR, Hansen WG. A retail market potential model. *Journal of the American Institute of Planners* 1965;:134–143.
4. Nakanishi M, Cooper LG. Parameter estimation for a multiplicative competitive interaction model - least squares approach. *Journal of Marketing Research* 1974;:303–311.
5. Stanley TJ, Sewall MA. Image inputs to a probabilistic model: Predicting retail potential. *Journal of Marketing* 1976;:48–53.
6. Ghosh A. Parameter nonstationarity in retail choice models. *Journal of Business Research* 1984;:425–436.
7. Orpana T, Lampinen J. Building spatial choice models from aggregate data. *Journal of Regional Science* 2003;:319–347.
8. Bruno G, Improta G. Using gravity models for the evaluation of new university site locations: A case study. *Journal of computers and Operations Research* 2008;:436–444.

9. Cliquet G. Implementing a subjective mci model: An application to the furniture market. *European Journal of Operational Research* 1995;:279–291.
10. Davis P. Spatial competition in retail markets: Movie theaters. *The RAND Journal of Economics* 2006;:964–982.
11. Wan N, Zou B, Sternberg T. A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science* 2012;26(6):1073–1089.
12. Wan N, Zhan FB, Zou B, Chow E. A relative spatial access assessment approach for analyzing potential spatial access to colorectal cancer services in texas. *Applied Geography* 2012;32(2):291 – 299.
13. Yingru L, Liu L. Assessing the impact of retail location on store performance: A comparison of wal-mart and kmart stores in cincinnati. *Applied Geography* 2012;:591–600.
14. Sandikcioglu M, zden GA, Sayin S. A new gravity model with variable distance decay. Neringa, Lithuania: eurOPT; 2008:376–380.
15. Gauri DK, Pauler JG, Trivedi M. Benchmarking performance in retail chains: an integrated approach. *Marketing Science* 2009;:502–515.
16. Jones K, Simmons J. Location Location Location. Canada: Nelson; 1993:345.
17. Pauler G, Trivedi M, Gauri DK. Assessing store performance models. *European Journal of Operational Research* 2009;:349–359.
18. Kwan MP. Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis* 1998;30:191–216.
19. Kaufmann PJ, Rangan VK. A model for managingsystem conflict during franchise expansion. *Journal of Retailing* 1990;:155–173.
20. Naert PA, Bultez AV. A model of a distribution network aggregate performance. *Management Science* 1975;:1102–1112.
21. Yrigoyen CC, Otero JV. Spatial interaction models applied to the design of retail trade areas. ERSA conference papers; European Regional Science Association; 1998.
22. Kirkpatrick S, Gelatt CD, Vecchi M. Optimization by simulated annealing. *Science* 1983;:671–680.
23. Suman B, Kumar P. A survey of simulated annealing as a tool for single and multiobjective optimization. *The Journal of the Operational Research Society* 2006;:1143–1160.
24. Clarkson RM, Clarke-Hill CM, Robinson T. Uk supermarket location assessment. *International Journal of Retail & Distribution Management* 1996;:22–33.
25. Gonzalez-Benito O, Gonzalez-Benito J. The role of geodemographic segmentation in retail location strategy. *International Journal of Market Research* 2005;:295–316.