

Locating a shopping centre by considering demand disaggregated by categories

RAFAEL SUÁREZ-VEGA[†]

*Departamento de Métodos Cuantitativos en Economía y Gestión and
Instituto Universitario de Turismo y Desarrollo Económico Sostenible (Tides),
Universidad de Las Palmas de Gran Canaria, Spain*

[†]Corresponding author. Email: rafael.suarez@ulpgc.es

JOSÉ LUIS GUTIÉRREZ-ACUÑA

*Departamento de Economía y Dirección de Empresas, Universidad de Las Palmas
de Gran Canaria, Spain*

AND

MANUEL RODRÍGUEZ-DÍAZ

*Instituto Universitario de Turismo y Desarrollo Económico Sostenible (Tides) and
Departamento de Economía y Dirección de Empresas, Universidad de Las Palmas
de Gran Canaria, Spain*

[Received on 5 October 2016; accepted on 7 June 2017]

We model a shopping centre. The demand for goods and services in shopping centres is classified in four different categories: food, leisure, household goods and clothing. As some of these sectors do not provide essential goods and services, a Huff customer-choice model is applied that sets a parameter absorbing any lost demand when there is a shortfall in customer attraction. For each category, the parameters for the Huff model are estimated both globally (by means of ordinary least squares, assuming the same effect for the parameters throughout the entire market), and locally (using geographically weighted regression, considering that parameters depend on the customers' location). The proposed model was applied to a real data case on the island of Gran Canaria (Spain) to determine the best location for a shopping centre selling all four categories of goods. Finally, a study is conducted to determine how robust the solution is with respect to the lost demand parameter, and a comparison is made between the solutions obtained, using both global and local calibration methods.

Keywords: Huff model; spatial non-stationarity; geographically weighted regression; competitive location; disaggregate demand.

1. Introduction

The growth of the urban population has meant the appearance of different types of commercial facilities, and shopping centres have been one of the more recent retail options. Following [The International Council of Shopping Centers \(2004\)](#), a shopping centre can be defined as a group of retail and other commercial establishments that is planned, developed, owned and managed as a single property, with on-site parking provided. This type of retail structure has proliferated in the last two decades because it provides customers with the possibility of making multipurpose shopping trips, reducing transportation/time costs and increases the probability of finding what they are looking for ([Arentze *et al.*, 2005](#)).

Moreover, the selection of the location for a new shopping centre is one of the most important decisions that managers have to face when the long-term capital investment it involves has to be evaluated. Although this topic has been the subject of in-depth scientific research for over a century, real operational solutions have not been found until only recently. The competition in the retail sector is forever increasing, thus leading to a decline in margins and economic profitability. Any competitive element that can make a difference in this context has a very high business value (Clarke, 1998). The search for an optimal location strategy has the potential of being that differentiator; subsequently, further research into strategies of retail location sites is essential to achieve more efficient business planning. Although traditional models are focused on determining an optimal location, there are other managerial decisions that require the information to be disaggregated in categories in order to design the layout and the best distribution of the sales space for each category. This decision, in itself, is as important as the location where the shopping centre is built. One of the aims of this study is to develop a practical location model that includes the disaggregation of the demand to deal with this limitation that exists in traditional models.

The location of retail sites and their implications on consumer behaviour have been studied in different areas such as marketing, urban sciences, geo-marketing and so on. Gravity models, which assume that the utility perceived by individuals using the facilities is negative related to the distance between them and positive related to the facilities' attractiveness, have been widely employed in the field of retail distribution because of their easy interpretation. Reilly (1931) and Converse (1949) were the first to apply gravity models for estimating trading areas. Later, Huff (1964) defined a customer-choice gravity model where customer attraction was defined by taking both the size of the facility and the distance into account. The multiplicative competitive interaction models (MCI) are the generalization of gravity models, but consider other variables, in addition to the distance, so as to determine customer attraction. Nakanishi & Cooper (1974) showed that MCI models can be easily calibrated by means of ordinary least squares (OLS). In this article, we use geographic information system (GIS) procedures to resolve a competitive location issue based on the Huff customer-choice model. Other works combining location models and GIS can be found in Spaulding & Cromley (2007), Suárez-Vega & Santos-Peñate (2014), Suárez-Vega *et al.* (2011, 2012, 2015), Roig-Tierno *et al.* (2013a) as well as in references cited by Church (2002) and Murray (2010).

A shopping centre is a concentration of stores providing or selling goods belonging to different categories. Therefore, it may seem quite realistic to assume that customers visiting the shopping centre will probably purchase goods belonging to different categories. Several articles concerning the multipurpose nature of shopping trips to shopping centres can be found with regard to retailing (see for example Ghosh (1986) and Arentze *et al.* (2005)). However, most of them deal with this issue by making their analysis from an aggregate perspective, considering the shopping centre as a unit, without taking into account the differences that may exist between the structures of the shopping centres (Arentze *et al.*, 2005). As an alternative to traditional models, Arentze *et al.* (2005) analysed the probabilities of customers making trips to a shopping centre bearing in mind the multipurpose-trip element, while disaggregating the sales surface into three categories. In their article, they concluded that the different distribution in size for the categories in the shopping centre influences the location choice, including the cases in which only single trips were observed.

The problem of locating a shopping centre has been previously studied (see, e.g. Cheng *et al.*, 2007), but where the aggregate demand is always examined as a whole. In these cases, the probabilities of customers visiting the shopping centre are usually estimated. To be able to do this, different characteristics of the shopping centre, for example, the sizes for the different categories, can be considered. On employing an aggregate model, the probabilities of customers visiting the shopping centre can be estimated, however, the purchase made in each category cannot be identified. Given that the expected expenditure is different for each category, and the design and management of the shopping centre require more detailed information,

the aggregate model is deficient when the expected income of the shopping centre has to be estimated. Therefore, in this article we have assumed that the demand is disaggregated in four different categories (food, household goods, clothing and leisure) to be able to obtain the estimated income for each one; subsequently, these partial estimations are added together to obtain the total income for the shopping centre.

The traditional Huff model can be globally calibrated by means of OLS. This implies that customer perception of the distance and the size of the facilities is the same throughout the market. [Suárez-Vega et al. \(2015\)](#) employed a locally calibrated Huff model to illustrate that customer perception may vary across the study area. More specifically, geographically weighted regression (GWR) was used to calibrate the Huff model with a view to determining the site for a new hypermarket. GWR allows estimating the parameters not only at the sample points but at the considered demand points as well (see [Fotheringham et al. \(2002\)](#) for a complete summary of this technique). Hence, with a spatially distributed sample, Huff parameters can be estimated for each demand point in order to obtain the estimated purchase probabilities. In the current article, a local calibration has been applied, not only to the food category, as was the case of [Suárez-Vega et al. \(2015\)](#), but also to the clothing, household goods and leisure categories in order to cover all the goods and services provided by a shopping centre.

Research on competitive location has usually dealt with locating food distribution facilities (i.e. groceries or supermarkets), where the products sold are essential items. The demand for essential goods is considered to be inelastic, i.e. consumers satisfy their demand independently of the price and place of purchase. However, in shopping centres, in addition to food products, non-essential products and services (for example, leisure, clothing or household goods) are available. Therefore, the model should take the spending on non-essential items into account. When goods or services are not essential, the use of an elastic type demand, which depends on the final utility perceived by the customers, is more appropriate. Location problems considering elastic demand have been proposed by [Berman & Krass \(2002\)](#), [Aboolian et al. \(2007a,b\)](#), [Redondo et al. \(2012\)](#) and [Drezner & Drezner \(2012\)](#). The first two articles inferred customer demand as a concave non-decreasing function of the utility. [Redondo et al. \(2012\)](#) presented simulations showing that significant differences occurred when a location model is solved considering inelastic demand instead of elastic demand. [Drezner & Drezner \(2012\)](#) incorporated a parameter in the Huff model simulating the existence of a “dummy” facility that absorbs the lost demand caused by the reduction of the attraction perceived by customers. This idea has also been implemented by [Zhang et al. \(2012\)](#) who used a multinomial logic model to design a preventive healthcare facility network. In our article, for those categories offering non-essential services and goods, we have incorporated a lost demand parameter in order to simulate elastic demand. As the real value of the lost demand parameter was unknown, a simulation was carried out in order to analyse the effect of this parameter on the selection of the new shopping centre’s location. To our knowledge, this procedure has yet to be used in a real application in competitive locations, and moreover, to resolve the issue of locating a new shopping centre.

Most of the articles dealing with the issue of finding a suitable location for a new shopping centre only take into account the purchases made by customers at shopping centres, and ignore the possible purchases that they are able to make at other types of stores. From the interviews, we made to calibrate our model, we inferred that customers living in the study area do not make all their purchases in shopping centres. This fact makes it necessary to include the offer of goods and services outside shopping centres for the different categories in order to distribute the expected purchases made by customers. In order to achieve this, purchases made in stores outside shopping centres have been modelled based on two possible customer choice models: one, in which customers opt to purchase at the nearest store providing goods or services in the corresponding category, and the other, where customer purchases are shared among the stores outside the shopping centre in a certain influence-area surrounding their homes. [Suárez-Vega et al. \(2015\)](#) used the influence-area, while [Vega et al. \(2015\)](#) used the closest facility to model what they

called the proximity purchase (that made out of the big supermarkets), with a view to calibrating the Huff model in the food category. Both customer choices have been included in our article in order to analyse the following two objectives, firstly, to identify which of them better reflects the customers' preferences for the purchases at stores which are found outside shopping centres for each category; and secondly, to analyse whether customer behaviour depends on the categories themselves or not.

Furthermore, most of the articles calibrating retail models used samples consisting of established clients instead of potential customers. In other words, the data provided by the firms were extracted from the information based on, for example, loyalty card schemes, or from the results after a survey carried out on clients using the facilities operating in the market. In this case, we have used a sample made up of possible clients; that is to say, we have randomly selected individuals living in the potential market area in order to reflect the preferences of both clients and non-clients of shopping centres.

The rest of this article is organized as follows. Section 2 contains a brief description of the methodology applied in the study, covering the GWR and how this technique is used to calibrate the Huff model. Next, in Section 3, the methods are applied to a real data example, where an entry firm seeks to determine the location of a new shopping centre in the northern part of the island of Gran Canaria (Spain). Results obtained by means of local and global parameters are compared. An analysis of how robust the solution is with respect to the lost demand parameters is also carried out. Finally, Section 4 contains our conclusions.

2. Methodology

The methods used in determining the location for a new shopping centre are described in this section. A shopping centre consists of stores belonging to four categories (Food, Leisure, Household Goods, and Clothing). Therefore, the problem has been solved estimating the income for each category individually and aggregating them in order to estimate the total income of the shopping centre. Each individual problem is considered as a Huff customer-choice model and it is calibrated both globally (using OLS as suggested by [Nakanishi & Cooper \(1974\)](#)) and locally (using GWR as proposed by [Suárez-Vega et al. \(2015\)](#)).

First of all, a short introduction to the GWR model is presented. Then, the approach considered in applying this procedure to locally calibrate the Huff model is described. Although the Huff model can be solved as a particular case of the MCI models, the proposed methodology can be generalized to MCI models.

2.1. Geographically weighted regression

Suppose there is a continuous random variable y that we want to know, but which cannot be measured directly. Also suppose that there are other continuous random variables x_1, \dots, x_{K-1} (independent variables) that can be measured and used to estimate y (dependent variable). In particular, assuming that y depends linearly on variables x_k , the linear regression model

$$y = \beta_0 + \sum_{k=1}^{K-1} \beta_k x_k + \varepsilon$$

is proposed, where β_k represents the coefficients to be estimated, and ε is a normal distributed error term with zero mean. To estimate the parameters of the linear model, a sample of n observations is obtained $(y_i; x_{i1}, \dots, x_{iK-1})$, $i = 1, \dots, n$, and a OLS put forward to the system

$$y_i = \beta_0 + \sum_{k=1}^{K-1} \beta_k x_{ik} + \varepsilon_i, \quad i = 1, \dots, n.$$

If X is the $n \times (K-1)$ matrix of observations of the independent variables and y the $(n \times 1)$ vector of observations of the dependent variable, the estimation of the $(K \times 1)$ vector of parameters for the OLS model is

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

In OLS, it is assumed that all the observations have the same influence on the estimations. The weighted least squares (WLS) model is the generalization of the OLS to the case for which each observation i has a specific weight w_i when parameters $\hat{\beta}$ are estimated. In this case, the expression of the estimated vector of parameter is

$$\hat{\beta} = (X^T W X)^{-1} X^T W y,$$

where $W = \text{diag}(w_1, \dots, w_n)$ is the weight matrix.

GWR is similar to WLS, except that coefficients are not global, but depend on the location where the data were obtained. This method was initially proposed by [Brunsdon et al. \(1996\)](#) and [Fotheringham et al. \(1996, 1997\)](#) and assumes that close elements tend to have similar values. From a business point of view, this means that close customers present similar preferences. Therefore, GWR can be written as

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{K-1} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \quad i = 1, \dots, n$$

where (u_i, v_i) represents the coordinates of the i th sample point in space, and $\beta_k(u_i, v_i)$ is the estimated parameter for variable x_k at point i . In this case, the vector of estimated parameters is given by

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y,$$

where

$$W(u_i, v_i) = \begin{pmatrix} w_1(u_i, v_i) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n(u_i, v_i) \end{pmatrix}$$

is a weight matrix. Each element $w_j(u_i, v_i)$ on the diagonal represents the weight of the observation j for estimating the parameters in location (u_i, v_i) .

The function used to calculate $w_i(u)$ is called a kernel and reflects that when estimating $\hat{\beta}(u_i, v_i)$, closer elements to (u_i, v_i) are more influential than points farther away. The kernels used in this article are the

$$w_i(u) = e^{-0.5\left(\frac{d_i(u)}{h}\right)^2}, \quad (\text{Gaussian kernel})$$

and

$$w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ji}}{h}\right)^2\right)^2, & \text{if } d_{ji} < h \\ 0, & \text{if } d_{ji} > h \end{cases}, \quad (\text{Bisquare kernel}),$$

where $d_i(u)$ is the Euclidean distance between location (u_i, v_i) and observation j , and h is the bandwidth (a quantity expressed in the same coordinates used in the dataset and that controls the rate of decay in the distance decay function considered to obtain the weights).

Note that, as the weight matrix W can be calculated at every point in the space, this procedure allows for the estimation of $\hat{\beta}$ at any point in space, not only for the sample points (Fotheringham *et al.*, 2002, pp. 53–54).

The estimated parameters depend on the bandwidth of the spatial kernel considered to calculate the different weights. While the specific kernel function does not have a significant influence, the selection of the bandwidth may produce considerable changes in the estimated parameters. When the sample elements are regularly distributed in the study area, a kernel with fixed bandwidth is recommended; otherwise an adaptive bandwidth is indicated. When the fixed form is selected, all the estimations are made using the same bandwidth. If an adaptive bandwidth is chosen, this parameter is determined for each observation, ensuring that all the subsamples used to calculate the weights contain the same number of observations (see Páez *et al.* (2002a,b) and Fotheringham *et al.* (2002, pp. 56–59)) for a more detailed discussion about the bandwidth selection).

The corrected Akaike information criterion (AICc) can be used for comparing the relative quality of the models. Fotheringham *et al.* (2002) proposed the following version of the AICc:

$$\text{AICc} = 2n \log_e(\hat{\sigma}) + n \log_e(2\pi) + n \left\{ \frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \right\} \quad (1)$$

where n is the number of observations, $\hat{\sigma}$ is the estimated standard deviation of the error term, and $\text{tr}(S)$ is the trace of the hat matrix S (which is the matrix verifying $\hat{y} = Sy$, with \hat{y} being the estimated values for y). As $\text{tr}(S)$ is a function of the bandwidth, the AICc can be used to select the optimal bandwidth by minimizing it using, for instance, the golden selection method (Fotheringham *et al.*, 2002, p. 212). The AICc can also be used to compare the performance of different models (including OLS and GWR) because it takes into consideration the effective number of parameters in the models. In general, models with lower AICc are preferred, and a difference of at least 3 is required to consider that there is a difference between models (Fotheringham *et al.*, 2002, p. 96).

The existence of collinearity among the covariates in a regression model may reduce the precision and the robustness of the estimated parameters. Local collinearity may appear, even if it does not exist in the global model, because of the effect of the weights, usually higher in the closest observations, and the normally lower sample sizes. For diagnosing collinearity in GWR models, Wheeler (2007) proposed the local versions of the variance inflation factors (VIF) and the condition numbers (local-CN). As a general rule, local regressions with VIFs larger than 10 and/or local condition numbers greater than 30 can be affected by collinearity (Belsley *et al.*, 1980; O'Brien, 2007).

The significance of the parameters in the local regressions is tested by means of a t -value. Since GWR uses almost the same sample for calibrating the adjoining local models, a certain degree of dependency between the models exists, which artificially increases the t -statistic for the local coefficients. To avoid this problem, da Silva & Fotheringham (2016) suggested the following Bonferroni style family-wise error correction for testing the significance of the GWR coefficients

$$\alpha = \frac{\xi_m}{\frac{p_e}{K}} \quad (2)$$

where ξ_m is the desired significance level for the estimations, and α is the corrected significance level to take into account the model dependences, p_e is the effective number of parameters ($p_e = 2\text{tr}(S) - \text{tr}(S'S)$), and K is the number of parameters in each model.

Different analysis of variance F -tests have been proposed to test the significance of the improvement obtained by GWR with regard to OLS. [Leung et al. \(2000\)](#), pp. 16–17) proposed what we call the FL1 and FL2 tests (F_1 and F_2 using their notation). In the FL1 test, the F -statistic is obtained comparing the residual sum of squares of the GWR and the OLS. In the FL2 test, the F -statistic compares the difference in the residual sum of squares of the OLS and GWR to the sum of residual squares of the OLS model. [Brunsdon et al. \(1999\)](#), pp. 502–503) proposed another test (FF1 test) to compare the difference in the residual sum of squares of the OLS and GWR to the sum of residual squares of the GWR model. Essentially, the null hypothesis for these tests is that the GWR model does not improve the OLS model.

In order to contrast the spatial variability of the coefficients associated with the regressors, [Leung et al. \(2000\)](#), pp. 21–22) proposed the FL3 test (F_3 in their notation) based on the estimation of the sample variance of the estimated values for each parameter. This test is evaluated for each parameter and reports an F value, being the null hypothesis the stationarity of said parameter.

Different authors have analysed GWR prediction capacity by comparing this technique with other geostatistical alternatives (see for instance, [Páez et al. \(2008\)](#), and [Harris et al. \(2010\)](#)). The accuracy of prediction was measured by means of the root mean squared prediction error (RMSPE) and the mean absolute prediction error (MAPE). The smaller these values are, the more accurate the model's predictions are. The uncertainty of the prediction can be evaluated using the mean and the standard deviation of the z -score data (MZS and SDZS, respectively), where the z -score is defined as

$$z\text{-score}(u, v) = \frac{y(u, v) - \hat{y}(u, v)}{\sigma_{pred}(u, v)} \quad (3)$$

being $y(u, v)$ and $\hat{y}(u, v)$ the sample value at location (u, v) and its prediction, respectively, and $\sigma_{pred}(u, v)$ the variance of the prediction of GWR ([Leung et al., 2000](#)). For unbiased prediction standard error, MZS and SDZS must tend to zero and unity.

2.2. Huff model calibration

Given a category c , according to the Huff model, the utility perceived by customers at demand point i from the services in category c at facility j can be expressed as

$$U_{ijc} = S_{jc}^{\alpha_c} d_{ij}^{\lambda_c} \quad (4)$$

where S_{jc} is the size (sales surface) that facility j dedicates to category c , α_c is the weight associated with this size, d_{ij} are the transportation costs (distance or travel time) from demand point i to facility j , and λ_c is the parameter that reflects the effects of this transportation cost. Thus, the probability that a customer at i buys goods or services of category c at a facility j is given by

$$p_{ijc} = \frac{S_{jc}^{\alpha_c} d_{ij}^{\lambda_c}}{\sum_{j=1}^m S_{jc}^{\alpha_c} d_{ij}^{\lambda_c}} \quad (5)$$

where m is the number of facilities operating in the market.

[Nakanishi & Cooper \(1974\)](#) proposed the following log-transformed-centred form to obtain least squares estimates of the parameters:

$$\ln(p_{ijc}/\tilde{p}_{ic}) = \alpha_c \ln(S_{jc}/\tilde{S}_c) + \lambda_c \ln(d_{ij}/\tilde{d}_i) \quad (6)$$

where p_{ijc} is the probability that a customer at location i purchases goods of category c at facility j , and $\tilde{p}_{ic}, \tilde{S}_c$ and \tilde{d}_i are the geometric means of p_{ijc}, S_{jc} and d_{ij} over j , respectively.

Parameters in (6) can be estimated by means of OLS. The application of the OLS method assumes that regressors affect homogeneously throughout the study area. This assumption may not hold when spatial data is considered in the analysis. In the case of the Huff model, this homogeneity assumption implies that all the customers in the study area have the same perception of the different variables that define the attraction of the stores, which is unlikely when socio-demographic differences exist in the market (Ghosh, 1984). While OLS assumes that all parameters associated with regressors are constant throughout the study area, GWR allows for the modelling of customers' behaviour taking into account possible variations in the estimated parameters along the market (Suárez-Vega et al., 2015).

The transportation cost and sales surface (i.e. area size) parameters in the Huff model may present spatial non-stationarity. This situation gives rise to the following reformulation of the utility perceived by customers at location i from facility at location j for a given category c :

$$U_{ijc} = S_{jc}^{\alpha_{ic}} d_{ij}^{\lambda_{ic}} \quad (7)$$

where, $\alpha_{ic} = \alpha_c(u_i, v_i)$ and $\lambda_{ic} = \lambda_c(u_i, v_i)$ are the estimated parameters reflecting the effect of the sales surface for category c and the transportation cost for customers located at demand node i at coordinates (u_i, v_i) . Consequently, probabilities p_{ijc} are expressed as:

$$p_{ijc} = \frac{S_{jc}^{\alpha_{ic}} d_{ij}^{\lambda_{ic}}}{\sum_{j=1}^m S_{jc}^{\alpha_{ic}} d_{ij}^{\lambda_{ic}}} \quad (8)$$

Following the Huff model, if a new facility of size S_c is located in the market, the probability that customers at i purchase goods of category c at this new store located at point $P = P(u, v)$ is given by:

$$p_{iPc} = \frac{S_c^{\alpha_{ic}} d_{iP}^{\lambda_{ic}}}{S_c^{\alpha_{ic}} d_{iP}^{\lambda_{ic}} + \sum_{j=1}^m S_{jc}^{\alpha_{ic}} d_{ij}^{\lambda_{ic}}} \quad (9)$$

where d_{iP} is the transportation cost for customers moving from demand point i to P . Therefore, the estimated income for the new store for category c is

$$MS_c(P) = \sum_{i=1}^D w_{ic} p_{iPc}, \quad (10)$$

being w_{ic} the buying power in category c of the demand node i , and D the number of demand nodes.

The parameters involved in formula (8) can be locally estimated by applying GWR to the Nakanishi and Cooper's transformation

$$\ln(p_{ijc}/\tilde{p}_{ic}) = \alpha_{ic} \ln(S_{jc}/\tilde{S}_c) + \lambda_{ic} \ln(d_{ij}/\tilde{d}_i) \quad (11)$$

where $\tilde{p}_{ic}, \tilde{S}_c$ and \tilde{d}_i are the geometric means of p_{ijc}, S_{jc} and d_{ij} over j , respectively.

2.3. Estimating the total income of the shopping centre

GWR allows parameters α_{ic} and λ_{ic} to be estimated not only at the location of the sample elements but also at every point $P = P(u, v)$ in the study area. This implies that probabilities p_{iPc} in (9) can be estimated and, therefore, also the estimation for the capture by category using (10). Finally, the total income for the new shopping centre can be estimated as

$$MS(P) = \sum_{c=1}^C MS_c(P) \quad (12)$$

with C representing the number of categories considered.

Formula (9), as it has been proposed, can only be applied when the good or service provided is essential, i.e. when consumers spend their entire budget w_{ic} . This may be the case of food, in which the consumed quantities do not depend on the location of the facility or its size. However, in a shopping centre, some goods/services can be non-essential (e.g. leisure), which means that providing them may not use the full budget that consumers had intended to spend, producing what [Drezner & Drezner \(2012\)](#) call lost demand. To model this lost demand, these aforementioned authors proposed the inclusion of a parameter in (5) in order to simulate a facility that captures the lost demand. The function to determine the probability of consumer in i buying goods of category c in the facility located at j is:

$$P_{iPc} = \frac{S_c^{\alpha_{ic}} d_{iP}^{\lambda_{ic}}}{S_c^{\alpha_{ic}} d_{iP}^{\lambda_{ic}} + \sum_{j=1}^n S_{jc}^{\alpha_{ic}} d_{ij}^{\lambda_{ic}} + LD_c} \quad (13)$$

where LD_c is the parameter associated with the lost demand in category c . In their work, [Drezner & Drezner \(2012\)](#) claimed that a correct choice of this parameter would be equivalent to considering a concave demand function, depending on the total utility perceived by the consumer. However, in the application presented in their article they used a value that was chosen-ad-hoc, without doing a previous study justifying the resulting demand function.

To be able to consider the lost demand when the goods are non-essential, the total estimated income given by equation (12) can be calculated using the proposed probabilities in (13). Needless to say that in the case of essential goods (e.g. food), the LD parameter is zero.

3. Application: locating a new shopping centre

In this section, the methodology previously described is applied to solve the problem of locating a new shopping centre in the north of the island of Gran Canaria, Spain. First, a sample containing the revealed probabilities for purchasing in each category at the different shopping centres available was collected. Then, the GWR model that best fits the sample data for each category was investigated. Using the kernel and the bandwidth that best performed in the GWR model, parameters α_{ic} and λ_{ic} were estimated for each demand point i and category c . Finally, these parameters were used to estimate the probabilities (9) and thus obtain the estimated income for the new shopping centre (12). This process is similar to the one proposed by [Suárez-Vega et al. \(2015\)](#), which locally calibrates the Huff model with the aim of searching for a suitable location for a new supermarket (food category).

TABLE 1 *Distribution of the population and commercial index in the study area*

Municipality	Population*	Surface (km ²)	Population density*	Commercial index**	Valid surveys
Agaete	5765	46	125.33	4	31
Artenara	1301	67	19.42	1	0
Arucas	35542	33	1077.03	43	125
Firgas	7424	16	464.00	4	28
Gáldar	23951	62	386.31	34	90
Moya	8010	32	250.31	7	30
San Nicolás de Tolentino	8403	124	67.77	10	32
Santa María de Guía	14207	43	330.40	16	79
Teror	12818	26	493.00	15	52
Valleseco	4121	22	187.32	3	9
Las Palmas de Gran Canaria	381114	101	3773.41	1336	248

*2008; **2012 data from the *Anuario Económico de España* 2013.

3.1. Data

Gran Canaria island belongs to the Canary Archipelago (Spain), just off the west coast of Africa (In Fig. 1, a map shows the geographical situation). For over 10 years, the need for a shopping centre in the northwestern part of the island has been upheld by both the private and public sectors. In order to achieve this aim, the problem of locating a new shopping centre in the northern part of the island is analysed. The study area includes the capital of the island (Las Palmas de Gran Canaria) and the 10 municipalities which constitute the Consortium of Northern Municipalities in Gran Canaria (see Fig. 1). This study area is characterized by the existence of significant socio-demographic differences between the capital of the island and the other municipalities encompassed in the analysis. This may suggest that there is some spatial variation concerning customer perception of the characteristics that define the amenities (Suárez-Vega *et al.*, 2015).

Table 1 shows the distribution of the population and the commercial index among the municipalities that make up the study area. Data regarding population were taken from the 2008 census, since this was the last year where compatible statistics and geographical data were available. The values of the commercial index reflect the relative weight (per hundred thousand) of the commercial activity in the area with respect to Spain as a whole. With 75.82% of the total population of the study area (502,656 inhabitants) living in Las Palmas de Gran Canaria, the city is the most important urban area in the Canary Islands. The other 10 municipalities are regarded as rural i.e. with a low population density and a very low commercial index, thus, the aim of this study is to determine the location of a new shopping centre within this rural area. The fact that the capital has great potential to attract customers in comparison to the demand in the rural areas means it must be included in the analysis.

For this approach, the utility perceived by customers for each category in shopping centres was calculated by considering the size (sales surface area) allocated to one particular category together with the travel time to the shopping centre using formula (7), and thereafter, formula (8) to define the probabilities p_{ijc} . The shopping centres as well as the stores in the study area belonging to the four categories were georeferenced using the addresses attained from the Economic Census for Gran Canaria in 2012 (*Censo Comercial de Gran Canaria*, 2012). In this database, the sizes of the stores are classified as follows: less

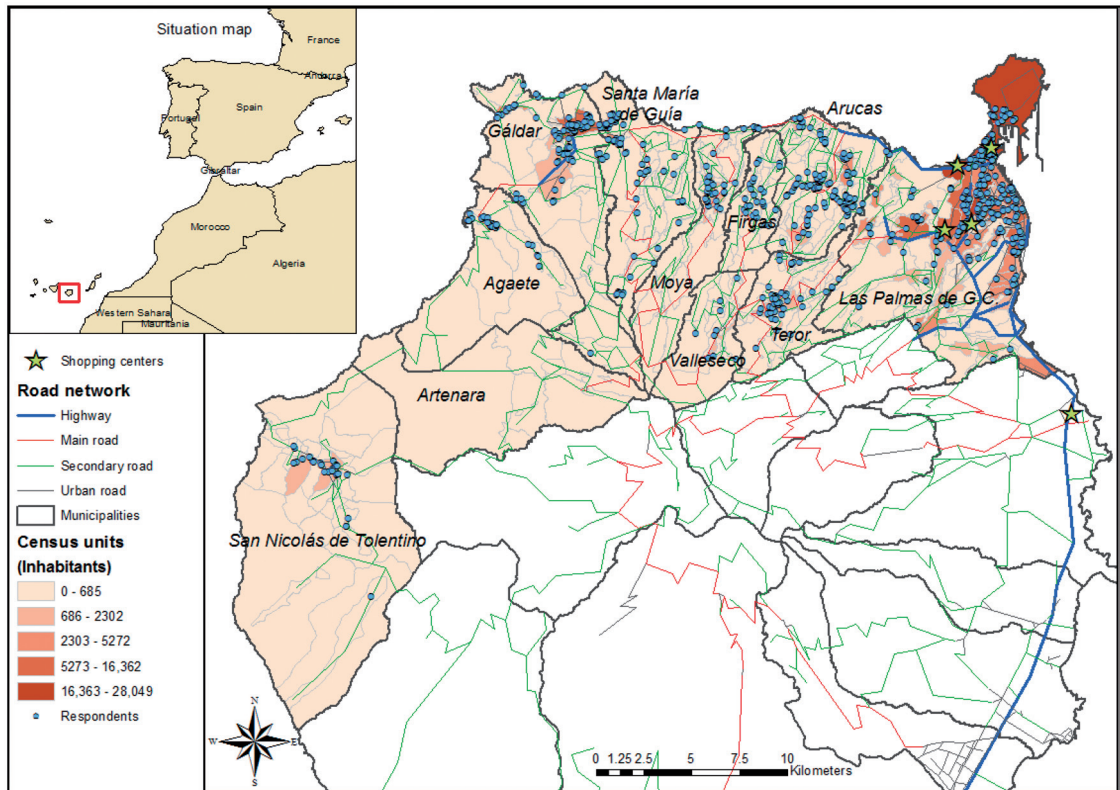


FIG. 1. Location map and description of the market.

than 120 m², from 120 m² to 399 m², from 400 m² to 999 m², from 1000 m² to 2499 m², from 2500 m² to 4999 m² and larger than 5000 m². In order to estimate the weights corresponding to the size of the store, sizes of 90, 260, 700, 1750, 3750 m² were allocated to the five intervals, respectively. The exact sizes of the biggest supermarkets were also taken into account (these sizes varied from 5000 m² to 13,387 m²). Figure 1 shows the locations of each of the five shopping centres undertaken in the study (although one of them cannot be found within the study area, it was accounted for due to its proximity to the analysed area), as well as the population distribution across the study area.

Table 2 shows the distribution by category and size of the non-shopping centre sales areas for each category in both the Consortium of Northern Municipalities and Las Palmas de Gran Canaria. These sizes (see Table 2) were calculated considering sales surface areas of 90, 260, 700, 1750 and 3750 m² which correspond to the five sizes intervals, respectively. The percentage which appears in parentheses corresponds to the surface area in the Consortium of Northern Municipalities with respect to the surface area in Las Palmas de Gran Canaria. Taking into account that the total population of the Consortium of Northern Municipalities is 31.42% of the population of Las Palmas de Gran Canaria, this figure could be used to compare the concentration of the different types of stores in the two areas. We were able to conclude that the percentages of sizes which are lower than the aforementioned figure mean that the sales surfaces allocated to a particular category, per person in the Consortium, amounts to less than the

TABLE 2 *The non-shopping centre sales surfaces dedicated to the different categories in the Consortium of Northern Municipalities and in Las Palmas de Gran Canaria*

Size (m ²)	Food		Clothing		Household goods		Leisure	
	Consortium	LPGC	Consortium	LPGC	Consortium	LPGC	Consortium	LPGC
<120	18,360 (39.23)	46,800	8640 (18.97)	45,540	10,980 (28.18)	38,970	8910 (20.97)	42,480
120-399	4320 (47.37)	9120	2080 (17.11)	12,160	4000 (23.15)	17,280	160 (3.13)	5120
400-999	8400 (52.17)	16,100	3500 (25.0)	14,000	1400 (6.45)	21,700	0 (0.0)	2800
1000-2499	3500 (10.0)	35,000	1750 (25.0)	7000	8750 (50.0)	17,500	0 (0.0)	1750
2500-4999	0 (0.0)	11,250	0	0	0	0	0	0
≥ 5000	0 (0.0)	10,000	0	0	0	0	0	0
Total	34,580 (26.96)	128,270	15,970 (20.29)	78,700	25,130 (26.33)	95,450	9070 (17.39)	52,150

sales-surface allocated in the capital. Table 1 shows that in the four categories the surface per person in the Consortium is less than in the capital, especially in the leisure and clothing category. In general, this tendency is observed in all types of stores, except in the food category and in stores with a surface area of between 1000 m² and 2499 m² in the household goods category. Nonetheless, in the food category, the Consortium presents a higher density of small stores than in the capital, as opposed to a reduced number of large stores.

Estimates of the transportation times between demand points and amenities were calculated using the transportation network made up of the island's main roads (see Fig. 1). The demand in the study area was distributed among 509 census units (Fig. 1 shows their distribution), and the population of each unit was allocated to the gravity centre of the housing units observed there.

In order to know the customers' purchase behaviour living in the study area, a survey containing 724 valid questionnaires was carried out (a more detailed explanation of the sampling process can be found in the Supplementary Appendix 1). The survey respondents were georeferenced and the fastest routes between them and the stores were calculated. The valid surveys for each municipality are shown in Table 1 and their location is drawn in Fig. 1. In each category in the survey, people were asked about the proportion of purchases they made in the five shopping centres in the study. The remaining purchase power (one minus the purchase proportion from the shopping centres) was considered as non-shopping centre purchases. As no information about how customers organized their non-shopping centre purchases was available, we proposed two different behaviours that are described in the next section.

3.2. *Modelling the non-shopping centre purchases*

The probabilities of purchasing at the shopping centres were revealed from the data sample. The information collected revealed that the non-shopping centre purchases are very important in the study area (43.46% in household goods, 45.18% in clothing, 59.15% in leisure and 71% in food); however, the customers' behaviour that best reflects these purchases is not obvious. In this work we propose two types of customer choices in order to model the non-shopping centre purchases.

When the influence area behaviour (IAB) is considered, we assume that the non-shopping centre purchases in a particular category were made at the stores within a specific influence area. Two types of influence area have been considered, the centred influence area and the central store (CS). In the former, an area of influence around each respondent is considered (500, 1000 and 3000 m were taken

as alternative radii for this influence area), and the non-shopping centre purchases per category were allocated to a dummy store with a size equal to the sum of all the sales surfaces of the stores in the same category observed in the influence area (excluding those located in the shopping centres), along with average travel times to the stores in the influence area. In the CS case, the stores for each category operating in the study area were clustered in zones, and the central store (the closest store to the gravity centre of the facilities belonging to the zone) for each zone was calculated. The non-shopping purchases were allocated to a store located at the CS with a size equal to the sum of all the sizes of the stores belonging to the zone.

In the closest facility behaviour (CFB), the non-shopping centre purchases in a specific category were allocated to the closest facility (the sales surfaces and the transportation times were calculated with respect to this facility). In this case, different scenarios were selected: the closest store (independently of its size), or the closest store belonging to a specific size group in which the stores were classified, each of these being different sizes (90, 260, 700, 1750, 3750 m² were allocated to the aforementioned groups).

Consequently, for each customer choice scenario and category, a sample containing 4344 observations p_{ijc} was obtained from this survey (724 valid questionnaires \times (non-shopping centre purchase plus the five shopping centre purchases)).

We inferred that customers would behave differently regarding the non-shopping centre purchases according to one category or another, so both customer-choices were analysed. The selection of the customer choice that best describes the non-shopping centre purchases in each category was made according to the goodness of fit for the OLS regression on (6). To be able to carry this out, the different possibilities for the IAB (changing the radius of the influence area and including the CS option) and the CFB (including all the stores or only those belonging to the same size group) were studied. Analysing the IAB customer choice, we can appreciate that the influence area for purchasing clothes (outside the shopping centres) was higher than that concerning food and household goods purchases (3000 m vs 1000 m). With regard to the leisure category, customers preferred to visit the facilities located at their cluster (zones with a high concentration of restaurants, bars and so on). If non-shopping centre purchases were made in the closest facility, the results reflected that customers preferred to visit small facilities (smaller than 120 m²) in the food and leisure categories, whereas preferences for household goods and clothing appear in the following group, which comprises larger surface areas (stores between 120 and 400 m²). Supplementary Appendix 2 contains details about this estimation process.

Important differences in R_a^2 regarding the two customer choices do not exist, but according to the AICc values, we can deduce that the IAB behaviour better fits the model in food, clothing and leisure, the CFB being better for modelling the household goods category. The poor adjustments to the models, mainly in the clothing and household goods categories, may be due to the fact that a sample consisting of potential customers (not only frequent clients) was used. For instance, using a similar sampling process to calibrate an MCI model for the grocery sector, González-Benito *et al.* (2000) obtained an R^2 of 0.1198. They needed to disaggregate the sample to reach an R^2 of 0.5899 in one of the segments. These low levels of goodness of fit may indicate low accuracy of the predictions when the estimated parameters are used to predict customers' probability of purchase and, therefore, the best location and design for the shopping centre. In order to improve this accuracy, the GWR is proposed to obtain a local estimation of the parameters.

Although the parameters' signs associated with the sizes and travel times in the IAB matched the traditional premises in the Huff model (i.e. positive for the size and negative for the time), this was not the case with the CFB. When the closest facility was selected for modelling the non-shopping centre purchases, negative signs were observed for the size parameters (except for clothing.) This negative sign for the parameter size could be justified for two reasons. First, the customer choice selected for

the non-shopping centre purchases implies that customers made these purchases at the nearest facility, which had a very small surface area if it were compared to the corresponding size in the shopping centre. Secondly, as previously mentioned at the beginning of this section, the population of the study area presented a high tendency to make purchases outside shopping centres (varying from 43.46% in household goods to 71% in food). This implies that, for the sample used in this study and in line with this behaviour, a larger sized store did not necessarily imply an increase in the attraction that customers perceived, but did indicate a higher value to the proximity factor of the store. Indeed, for the scenarios where the size parameter is positive, the estimated values were low (between 0.089 and 0.337) indicating the greater influence of the transportation cost rather than the size of the facilities themselves. In fact, as Roig-Tierno *et al.* (2013) demonstrated, this result is not contradictory to the opinion of specialists in the field. In this work, a group of experts in retail site selection and marketing analysed different factors that affect the success of a supermarket. They assigned an influence of only 2.44% to factors associated with the size (sales space, number of parking places and number of check-outs), whereas the factors related to the location were observed as having the highest value (75.72%).

The effects of the size and transportation cost on the probabilities of purchase can be compared to the different categories and customer choice options for the non-shopping centre purchases. When the IAB was applied, important differences existed depending on the category. The size concerning the leisure category was the least appreciated (0.0891), whereas the most widely appreciated was clothing (0.3370). The transportation cost effect was quite similar in food, clothing and leisure, but in the household goods category the associated parameter experimented an important decrease: it was around a half of the value of these parameters in the food category. This implies that in the household goods category, customers were willing to make longer trips than for the other categories.

The estimated parameters in the CFB for the household and leisure categories were quite similar. In contrast to these parameter settings, in the food category, customers expressed slightly less aversion to the transportation cost and preferred smaller stores than in the two categories previously mentioned. The behaviour observed in the clothing category was substantially different to the rest of categories, but very similar to that obtained in the IAB.

3.3. Local calibration of the Huff model

While taking into account the non-shopping-centre-purchase scenarios selected in the previous section, the Huff model was locally calibrated by means of GWR. In this case, the process, proposed by Suárez-Vega *et al.* (2015) in the food category, was extended to the other categories. Calculations for GWR were carried out using the R package *GWmodel*¹ (Gollini *et al.*, 2015); likewise, for the GWR model, Gaussian and bi-squared kernels with fixed and adaptive bandwidths were evaluated, with the best bandwidth being selected by minimizing the AICc. In all the scenarios analysed, the adaptive bi-squared kernel was selected.

Local collinearity for the different scenarios was evaluated by means of calculating the local-CNs and VIFs. In most cases there was no evidence of collinearity, except in the food and household categories with the CFB (558 and 744 sample elements with local VIFs higher than 10).

Supplementary Appendix 3 contains the results obtained by GWR for the different categories and non-shopping centre purchase behaviour. All GWR models improved the OLS model, with an average increase in R_a^2 by 0.061 and an average decrease in the AICc by 255.19. This assumption was corroborated by the significance of the FL1, FL2 and FF1 tests. The most significant improvement took place in the estimations of the leisure category when taking CFB into account. In this case, the GWR model explained

¹Available at: <http://cran.r-project.org/web/packages/GWmodel/>.

the variation of the probabilities that were approximately 14% more than the corresponding OLS version. Besides, the FL3 test also suggested that both size and transportation cost coefficients varied over the study area in all the selected scenarios.

The estimated parameters were tested for significance according to the family-wise error rate, calculated using (2) for a $\xi_m = 0.1$. The most common tendency was that the coefficients associated with the transportation time were more significant than those associated with the size. This may suggest that, in the sample analysed, customers give more importance to the transportation costs than to the size of the facility, when deciding on where to make their purchases. In all the scenarios, except in the household category with CFB, the sign of the OLS coefficients coincided with the mean of the significant coefficients of the GWR model. Nevertheless, in the case of the household category with CFB, signs for the size coefficients were positive in contrast to the OLS, in which it was negative. This contradiction may have occurred because, as was previously mentioned, some local collinearity existed in this GWR model.

Given the results obtained by the GWR estimations, a behaviour type for the non-shopping centre purchases was selected so as to be able to predict the parameters associated with the size and the transportation cost in each category. In food and household goods, the IAB was selected because the CFB presented some collinearity problems. For the leisure category, the CFB was selected because it presented better values in R_a^2 and AICc. Finally, in the clothing category, although the different indicators are quite similar, the CFB was selected because it presented a slight improvement with respect to the IAB (in R_a^2 , AICc, and also in the FL1 test.).

Before being able to predict the local parameters for the demand points involved in the Huff model, the prediction accuracy of the OLS and the GWR models was evaluated for each category; to do this, the sample was randomly divided into two subsamples of equal size. One of the subsamples was used to estimate the parameters and the other to validate the predictions. As Table 3 shows, the prediction accuracy of the GWR models was slightly better than that obtained by the OLS (lower values for the RMSPE and MAPE). The uncertainty of the prediction was also better for most of the GWR models, although the difference is very small (MZS close to zero and SDZS close to one). We must take into account that the random selection of the subsample did not make allowances for the spatial distribution of the sample and this issue may negatively affect the GWR results.

3.4. Estimation of the total capture for the new shopping centre

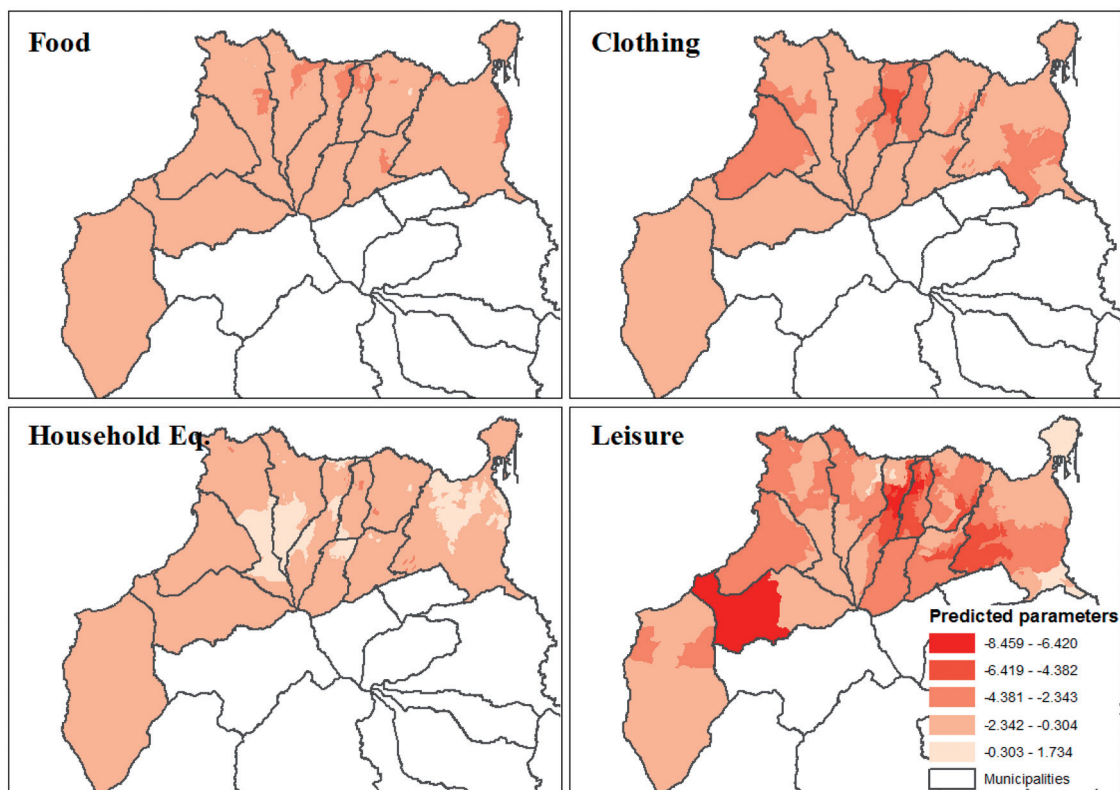
The parameters associated with the demand nodes in each category were predicted for each demand point by means of the proposed GWR models. The predicted parameters for each demand point i and category c were then substituted in (8) to estimate the probabilities p_{ijc} .

Figures 2 and 3 present the distribution of the prediction parameters λ_{ic} and α_{ic} for each category that were then be used for estimating probabilities p_{ijc} . Maps were prepared using ArcGIS software, allocating to each census unit the predicted value of the parameter for its gravity centre. Figure 2 shows the distribution of the predicted parameters λ_{ic} associated with the transportation cost (the darker the area colour, the less reluctant to travel) for each census unit. In all cases, these parameters presented a negative sign. However, for both food and household categories (when the IAB was selected) the range of variation for the parameter is less than the other two categories (with CFB). The parameter distribution is different for each category, which may suggest the need to differentiate the categories in order to estimate the captures obtained by a shopping centre.

Figure 3 shows the distribution of the predicted parameters α_{ic} associated with the store size. In all cases, changes in sign appeared because significant and non-significant predictions are shown. As

TABLE 3 Accuracy prediction indicators

	OLS				GWR			
	RMSPE	MAPE	MZS	SDZS	RMSPE	MAPE	MZS	SDZS
Food	1.982	1.407	0.006	1.009	1.960	1.380	0.004	0.995
Clothing	2.455	2.034	0.003	0.983	2.433	2.013	0.005	0.986
Household	2.667	1.967	0.189	0.982	2.651	1.972	0.167	0.977
Leisure	1.855	2.385	0.008	1.004	1.739	2.302	0.008	1.016

FIG. 2. Predicted parameter distributions for the transportation cost parameter λ_{ic} .

occurred with the transportation-cost parameters, it seems that there is no tendency for this to happen when the different categories are compared.

Following this, the predicted parameters λ_{ic} and α_{ic} were used to calculate the estimated income in each category throughout the feasible region. This region was defined by eliminating the land uses from the study area which were thought to be incompatible with locating a new shopping centre e.g. water bodies, protected areas, airports, and so on, together with those areas with slopes greater than 12%.

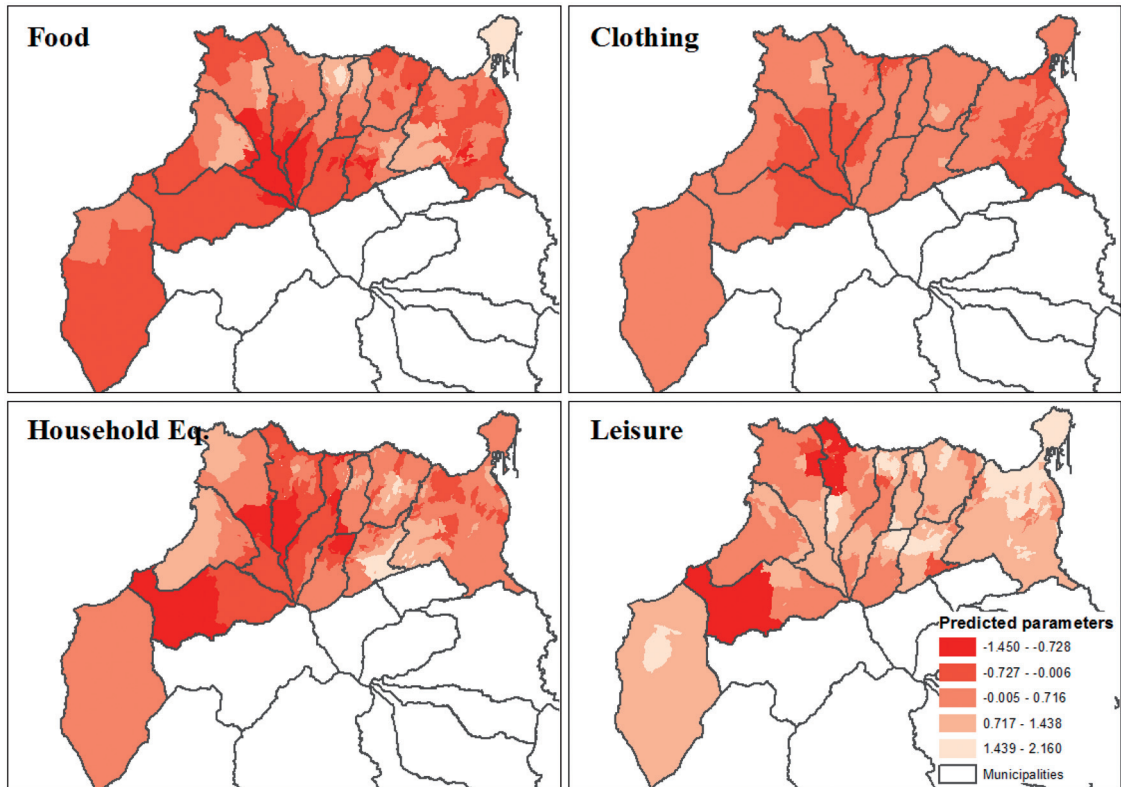


FIG. 3. Predicted parameter distribution for the size parameter α_{ic} .

Afterwards, the said region was divided into plots with a maximum area of 70,000 m² and the centroid of each of these plots was considered as their geographic representative. Therefore, the local estimated income (LEI) for a new shopping centre located at a given plot (with centroid P) with a given set of sizes for the different categories can be obtained by using (12) and the probabilities defined in (8). Moreover, the global estimated income (GEI) for this plot can be calculated by taking the probabilities defined in (5). To apply these formulas, distances d_{ip} were calculated as the transportation time of the fastest route joining point P and demand point i .

The estimated income was obtained for a new shopping centre with a sales surface area similar to that of an existing one, whereby 13,881 m²; 4,206 m²; 11,358 m² and 9,254 m² were allotted to each of the categories—food, household goods, clothing and leisure categories, respectively. The buying power of each demand node in the different categories was calculated according to its population and the average purchase per person in Gran Canaria. The average purchase amounts (in thousands of euros) were 1.514, 0.330, 0.443 and 0.520, respectively (Anuario Económico de la Caixa, 2013). Figure 4 shows the estimated capture and the best location for the new shopping centre considering 0.0, 0.000437, 0.005582 and 0.019364 as lost demand parameters for food, clothing, leisure and household goods, categories, respectively. These lost demand parameters correspond with the 50th percentile of the attraction perceived for each category by the demand nodes from the

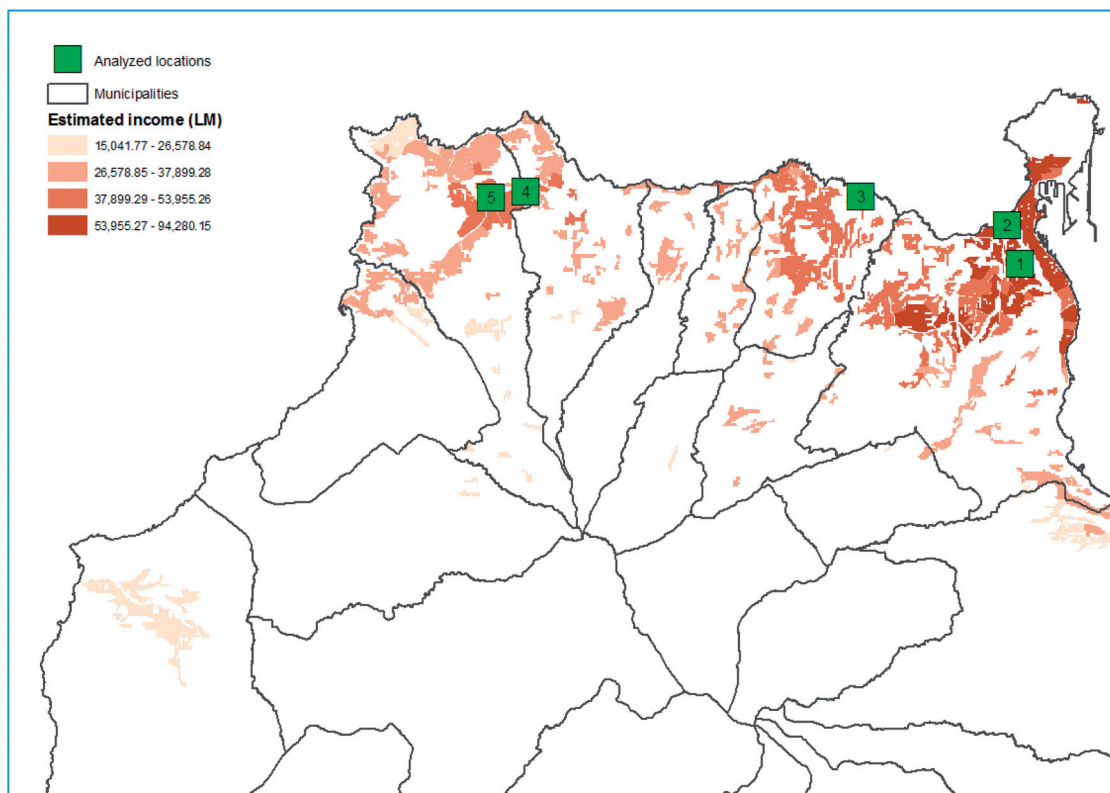


Fig. 4. Estimated income (thousands of €) for the new shopping centre.

existing facilities. In this scenario, the location maximizing the total income is sited in the capital (Loc1 in Fig. 4). Supplementary Appendix 4 contains a deeper explanation of how these parameters were estimated and a sensibility analysis of the expected income with respect to the variation of the lost demand parameters considering five possible locations for the new shopping centre. Differences between incomes obtained by means local and global calibrations are also analysed in this Supplementary Appendix.

4. Concluding remarks and applications

4.1. Conclusions

This article deals with the location of a new shopping centre, whereas previous works related to this problem have considered the demand as a whole, independently of the services provided by the shopping centre, in this article, the demand allocated to the shopping centre was disaggregated in four categories (food, household goods, clothing and leisure). Purchasing in shopping centres has a multipurpose nature because customers may buy from a range of goods and services belonging to different categories in different stores found in the same shopping centre. The issue has been addressed by making an estimation of the purchases made by customers at each shopping centre in each different

category, instead of only making an estimating of the probabilities of going shopping at a shopping centre.

In order to estimate the total income of the shopping centre, the probabilities of customers' purchases in the shopping centres were estimated separately for each category, and finally as an aggregate to give the total sum. These probabilities may depend on different characteristics of the shopping centre, such as its total sales surface or the sizes of the sales areas allocated to each category. For simplicity, in this study we have estimated these probabilities using the Huff customer-choice model, where the attraction perceived by customers towards the shopping centre depends on the store size in each corresponding category on the premises, as well as transport expenses.

As Suárez-Vega *et al.* (2015) proposed, parameters reflecting customers' perception of the store size and transportation costs in the Huff model may differ depending on their locations. In order to take this possibility into account, the Huff model was calibrated both globally (by means of OLS) and locally (using GWR) for each category. In the global model the estimations are some fixed parameters that do not reflect their possible spatial variations due to, for instance, geo-demographic differences. The GWR allows for local estimations of these parameters under the assumption that the closer the sample points, the greater the effect on the estimations.

The study was carried out after conducting a survey where customers were asked what proportion of their purchasing power they usually spent on each category in shopping centres. Since part of their purchases were made in stores outside shopping centres, it was necessary to define the non-shopping centre purchases. In order to do this, two types of customer behaviour were proposed. On the one hand, IAB in which people buy at all stores within a certain area of influence, and on the other, the CFB in which people do their shopping at the closest store. For the case analysed in this article, the IAB best fitted the global estimations in the food and household categories, while the CFB was selected for leisure and clothing.

From the sample used in this study, the non-stationarity of the parameters in the Huff model was proven. Different statistical tests proved that, in this application, GWR performs better than OLS in calibrating the Huff model. The GWR also improved the prediction capacity of the OLS. Therefore, the local calibration method proposed in the article allows the decision makers to locally evaluate how the distance and the size influence customers' preferences.

Due to the fact that some of the stores operating in a shopping centre do not provide essential services, a Huff model with a parameter that absorbs the lost demand has been implemented, as proposed by Drezner & Drezner (2012). This lost-demand parameter was applied to all the categories, with the exception of the food category.

The estimated income for a new shopping centre was calculated for the sites which made up a feasible region considering different lost demand settings and using both the local (LM) and global models (GM). In this application, the best location for the GM was always the same (Loc1), independently of the set of lost demand parameters that were taken into account. Nevertheless, the best location for the LM was Loc1 (in 87.6% of the cases) and Loc2 (situated at a distance of 1620 m) for the rest of the cases. Consequently, the solution to the GM seems to be very robust with respect to the effect of elastic demand in non-essential services. The LM solution also seems to be quite robust, but may fluctuate, for instance, when customer demand is very sensitive to any attraction perceived.

The differences between the estimated income for both the LM and the GM were also evaluated. For this application, the differences were significant, for example, reaching 18.03% and 16.69% for Loc1 and Loc2, respectively. These differences suggest that a wrong choice of model may produce significant deviations in the estimation of the income for a new shopping centre, and therefore, may affect its future viability.

4.2. Possible applications and new lines of research

The Huff model is very popular because it is easily interpreted and the information for its implementation is relatively simple. Moreover, Nakanishi and Cooper's transformations allow its calibration by means of OLS. However, local calibration with GWR does not make it any more complex when collecting data and information. Additionally, local calibration can be achieved using free software. Consequently, the proposed method implies, from our point of view, an improvement on the traditional global procedure frequently used to solve this competitive problem.

Moreover, considering the shopping centre as a set of shops offering services/goods for different categories, instead of as an aggregated unit, makes the problem more realistic. In fact, this option allows the decision maker to differentiate among different purchase habits (preferences and budgets) for each category.

On the other hand, further investigation is needed in order to study the GWR's prediction capacity, especially the way in which the sample is divided in order to obtain both the calibration and validation subsamples. As standard, these are randomly obtained, which undermines the GWR since its performance depends on the spatial distribution of the sample, and as such, random methods do not take this aspect into account.

Ultimately, the layer encompassing the estimated income for the new shopping centre can be embedded in a multi-criteria analysis. This may involve commercial objectives (such as a reduction in cannibalization), efficiency criteria (minimize travel time for clients to stores, improving accessibility) and/or equity aspects (i.e. minimizing the maximum distance to the closest facility). Additional studies could be carried out in order to improve the goodness of fit when the Huff model parameters are estimated. For instance, estimations of the loss demand parameter may contribute to better explaining the purchase probabilities for non-essential services. To do so, an interviewing process may be necessary in order to specifically ask about this matter.

Supplementary Material

Supplementary material is available at <http://www.imaman.oxfordjournals.org/>.

Acknowledgements

The authors thank GRAFCAN (<http://www.grafcan.es/index.php>) for supplying geographical information.

Funding

This work was partially financed by project ECO2014-59067-P from the Spanish Ministry of Economy and Competitiveness and by the Canarias Innova 2020 Program from the University of Las Palmas de Gran Canaria Foundation.

REFERENCES

- ABOOLIAN, R., BERMAN, O. & KRASS, D. (2007a) Competitive facility location model with concave demand. *European J. Oper. Res.*, **181**, 598–619.
- ABOOLIAN, R., BERMAN, O. & KRASS, D. (2007b) Competitive facility location and design problem. *European J. Oper. Res.*, **182**, 40–62
- Anuario Económico de España 2013 (2013). Área de Estudios y Análisis Económico de la Caixa (Ed.).

- ARENTEZE, T. A., OPPEWAL, H. & TIMMERMANS H. J. P. (2005) A multipurpose shopping trip model to assess retail agglomeration effects. *J. Mark. Res.*, **42**, 109–115.
- BERMAN, O. & KRASS, D. (2002) Locating multiple competitive facilities: spatial interaction models with variable expenditures. *Ann. Oper. Res.*, **111**, 197–225.
- BELSLEY, D. A., KUH, E. AND WELSCH, R. E. (1980) *Frontmatter*, in *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Hoboken, NJ, USA: John Wiley & Sons, Inc., doi: 10.1002/0471725153.fmatter.
- BRUNSDON C., FOTHERINGHAM A. S. & CHARLTON M. (1996) Geographically weighted regression: a method for exploring spatial non-stationarity. *Geographical Analysis*, **28**, 281–298.
- BRUNSDON C., FOTHERINGHAM A. S. & CHARLTON M. (1999) Some notes on parametric significance tests for geographically weighted regression. *J. Reg. Sci.*, **33**, 497–524.
- Censo Comercial de Gran Canaria (Commercial Census of Gran Canaria) (2012) Available at <http://www.gobiernodecanarias.org/cicnt/temas/comercio/censocomercial/>.
- CHENG, E. W. L., LI, H. & YU, L. (2007) A GIS approach to shopping center location selection. *Build. Environ.*, **47**, 884–892.
- CHURCH, R. L. (2002) Geographical information systems and location science. *Comput. Oper. Res.*, **29**, 541–562.
- CLARKE, G. (1998) Changing methods of location planning for retail companies. *GeoJournal*, **45**, 289–298.
- CONVERSE, P. D. (1949) New laws of retail gravitation. *J. Mark.*, **14**, 379–384.
- DA SILVA, A. R. & FOTHERINGHAM, A. S. (2016) The multiple testing issue in Geographically Weighted Regression. *Geogr. Anal.*, **48**, 233–247.
- DREZNER, T. & DREZNER, Z. (2012) Modelling lost demand in competitive facility location. *J. Oper. Res. Soc.*, **63**, 201–206.
- FOTHERINGHAM, A. S., CHARLTON, M. & BRUNSDON, C. (1996) The geography of parameter space: an investigation of spatial non-stationarity. *Int. J. Geogr. Inf. Syst.*, **10**, 605–627.
- FOTHERINGHAM, A. S., CHARLTON, M. & BRUNSDON, C. (1997) Two techniques for exploring nonstationarity in geographical data. *Geogr. Syst.*, **4**, 59–82.
- FOTHERINGHAM, A. S., BRUNSDON, C. & CHARLTON, M. (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley.
- GHOSH, A. (1984) Parameter nonstationarity in retail choice models. *J. Bus. Res.*, **12**, 425–436.
- GHOSH, A. (1986) The value of a mall and other insights from a revised central place model. *J. Retailing*, **62**, 79–97.
- GOLLINI, I., LU, B., CHARLTON, M., BRUNSDON, C. & HARRIS, P. (2015) Gwmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *J. Stat. Softw.*, **63**, 1–50.
- GONZÁLEZ-BENITO, O., GREATOREX, M. & MUÑOZ-GALLEGO, P. A. (2000) Assessment of potential segmentation variables. An approach based on a subjective MCI resource allocation model. *J. Retailing Consum. Serv.*, **7**, 171–179.
- HARRIS, P., FOTHERINGHAM, A., CRESPO, R. & CHARLTON, M. (2010) The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Math. Geosci.*, **42**, 657–680.
- HUFF, D. L. (1964) Defining and estimating a trading area. *J. Mark.*, **28**, 34–38.
- LEUNG, Y., MEY, C. L. & ZANHG, W. X. (2000) Statistical test for spacial nonstationarity based on geographically weighted regression model. *Environ. Plann. A*, **32**, 9–32.
- MURRAY, A. T. (2010) Advances in location modeling, GIS linkages and contributions. *J. Geogr. Syst.*, **12**, 335–354.
- NAKANISHI, M. & COOPER, L. G. (1974) Parameter estimation for a multiplicative competitive interaction model-least square approach. *J. Mark. Res.*, **11**, 303–311.
- O'BRIEN, R. M. (2007) A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.*, **41**, 673–690.
- PÁEZ, A., LONG, F. & FABERS, S. (2008) Moving window approaches for hedonic price estimation: An empirical comparison of modelling techniques. *Urban Stud.*, **45**, 1565–1581.
- PÁEZ, A., UCHIDA, T. & MIYAMOTO, K. (2002a) A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for local heterogeneity. *Environ. Plann. A*, **34**, 733–754.

- PÁEZ, A., UCHIDA, T. & MIYAMOTO, K. (2002b) A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests. *Environ. Plann. A*, **34**, 883–904.
- REDONDO, J. L., FERNÁNDEZ, J., ARRONDO, A. G., GARCÍA, I. & ORTIGOSA, P. M. (2012) Fixed or variable demand? Does it matter when locating a facility? *Omega*, **40**, 9–20.
- REILLY, W. J. (1931) *The Law of Retail Gravitation*. New York: Knickerbocker Press.
- ROIG-TIERNO, N., BAVIERA-PUIG, A., BUITRAGO-VERA, J. & MAS-VERDU, F. (2013a) The retail site location decision process using GIS and the analytical hierarchy process. *Appl. Geogr.*, **40**, 191–198.
- SPAULDING, B. D. & CROMLEY, R. G. (2007) Integrating the maximum capture problem into GIS framework. *J. Geogr. Syst.*, **9**, 267–288.
- SUÁREZ-VEGA, R., GUTIÉRREZ-ACUÑA, J. L. & RODRÍGUEZ-DÍAZ, M. (2015) Locating a supermarket using a locally calibrated Huff model. *Int. J. Geogr. Inf. Sci.*, **29**, 217–233.
- SUÁREZ-VEGA, R. & SANTOS-PEÑATE D. R. (2014) The use of GIS tools to support decision-making in the expansion of chain stores. *Int. J. Geogr. Inf. Sci.*, **28**, 553–569.
- SUÁREZ-VEGA, R., SANTOS-PEÑATE, D. R. & DORTA-GONZÁLEZ, P. (2012) Location models and GIS tools for retail site location. *Appl. Geogr.*, **35**, 12–22.
- SUÁREZ-VEGA, R., SANTOS-PEÑATE, D. R., DORTA-GONZÁLEZ, P. & RODRÍGUEZ-DÍAZ, M. (2011) A multi-criteria GIS based procedure to solve a network competitive location problem. *Appl. Geogr.*, **31**, 282–291.
- The International Council of Shopping Centers (2004) *ICSC Shopping Center Definitions*. New York: The International Council of Shopping centers.
- VEGA, R. S., ACUÑA, J. L. G. & DÍAZ, M. R. (2015) Spatial analysis of consumer behavior in a food products market. *Theoret. Empirical Res. Urban Manag.*, **10**, 25–42.
- WHEELER, D. C. (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environ. Plann. A*, **39**, 2464–2481.
- ZHANG, Y., BERMAN, O. & VERTER, V. (2012) The impact of client choice on preventive healthcare facility network design, *OR Spectrum*, **34**, 349–370.