

Spatial Dimension of Choice Models and Zoning Processes within a Transportation System

Il ruolo della dimensione spaziale nei modelli decisionali e nei processi di zoning in ambito trasportistico

Luigi D'Ambra

Department of Mathematics and Statistic
University of Naples *Federico II*
dambra@unina.it

Guido Rodia

Department of Mathematics and Statistic
University of Naples *Federico II*
rodia@unina.it

Francesca Pagliara

Department of Transportation Engineering
University of Naples *Federico II*
fpagliar@unina.it

Riassunto: Nel presente lavoro sarà considerato il ruolo della componente spaziale all'interno dei modelli di utilità aleatoria per la definizione della domanda di trasporto, si confronteranno una serie di metodologie che tengono in debita considerazione tale componente. L'aspetto spaziale interverrà nella preliminare definizione dell'insieme di scelta del decisore razionale e nella successiva modellizzazione dei comportamenti dello stesso. La componente territoriale verrà esaminata: sia come componente specifica, come fattore individuale che influenza il comportamento di decisori razionali nelle scelte che essi si trovano a dover effettuare in un complesso sistema di trasporti; sia a livello generale, come vincolo spaziale che agisce sulle zone di interesse degli spostamenti.

Keywords: spatial association, choice set, local factorial analysis, cluster analysis, zoning, mixed spatial logit, Kernel logit.

1. Introduction

A transportation system can be seen like the set of different social, economic and infrastructural components: these components determine from one side the demand of movement among different points of the territory; from the other side they determine the supply of transportation services for the satisfaction of this demand. The transportation system is therefore constituted by two integrated subsystems: the system of the demand and the supply ones (Cascetta E., 2001).

These subsystems are tightly correlated, because all movement components are influenced by the characteristics of the infrastructures and by offered transportation services; as the performances of the supply system are influenced by the number of consumers using it in a determined period of reference and in a considered area: that is the transportation demand.

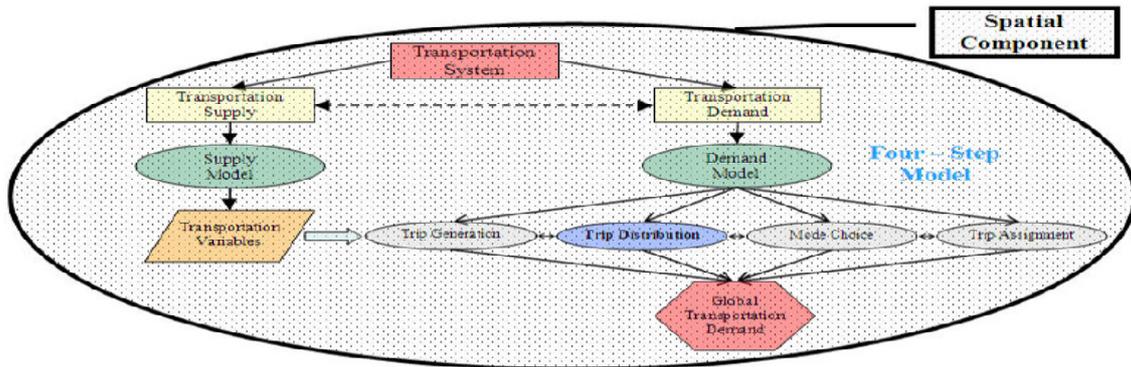
The objective of this paper is to underline the contribution of the spatial component inside the discrete choice models for the definition of transportation demand. In order to do this, we will be introduced and compared a series of techniques that consider the spatial effects.

The problem of the spatial aspect within the discrete choice models finds no much attention in literature (Fleming M., 2004), the few existing works are, above all, related to binary choice models. Following, we will focus the attention on choice models gradually more complexes and, above all, more and more "flexible", so that they are able to well describe, the choices of a rational subject. Particularly, we will underline the ability of these models to foresee the demand of mobility, that develops a central role for the observation and the modelling of a transportation system.

The global transportation demand is a complex and multidimensional quantity: for this reason we proceed to a decomposition of the global demand function in the product of interconnected sub models, each one related to a dimension of choice. The well known sequence of sub models, report in the **figure 1**, is known as four-step model. Among the sub models individualized we will consider the distribution model, that provides the percentage of subjects they bring them in a definite destination zone.

It is clear how the spatial component strongly characterizes these models and that we cannot put aside without consider it in the evaluation of a complex transportation system and in the choices of a rational decision-maker.

Figure 1: *The spatial component and the four-step model for estimation of the transportation demand.*



The regional component will be at first considered at "level of zones", for the determination of the choice set of the decision-maker, preliminarily to the specification of the distribution model. An integrated approach will be proposed among multidimensional data analysis and global and local indicators of spatial association, to individualize and aggregate homogeneous classes of traffic zones.

Then, we will analyze some techniques, in which the spatial dimension is directly present in the model specification, underlining the constrains of the discrete choice models and looking out upon the possibility to overcome them using articulated and complex logit functional forms. In this context we will consider the spatial component, that will be both present to individual and regional level.

2. The choice set specification

Within the definition of the global transportation demand, decomposed in its constitutive components in the four-step model, we will dwell upon the destination model (location choice model). It provides the percentage $p^i[d/osh]$ of individuals of a category (i),

moving themselves from an origin (o), for a motive (s), in a certain timeframe (h), towards a destination (d).

The first aspect to be established for the definition of the model is the choice set. The interest for the choice set is motivated by the errors that can result if it's not correctly specified. Moreover we have to consider the problem that, within the destination models, the available alternatives is rather wide, implying an elevated computational cost, as well as a difficulty in the model calibration, that can result less explanatory. To solve these problems the literature has proposed a series of solutions, all based on the reduction of this choice set, supposed equal for all the rational decision-maker.

McFadden D. (1978) already proposed the possibility to select a casual subset of alternatives, within the universe of the possible choices, under the hypothesis that the error terms of the perceived utility were independent and identically distributed (i.i.d.). Another interesting approach was been provided from Ben-Akiva M. (1985): he determined an importance sampling, constituted by all the alternatives chosen and a casual set of alternatives not selected.

All these methods contemplate a reduction of choice set based on the removal of some zones, it is clear that these approaches involve the automatic elimination of some information, that can eventually distort the model.

Another set of techniques bases itself on the zoning, or rather on the individualization of macro areas, obtained by the aggregation of elementary destination zones. The option of the zoning aggregation, compared to their elimination, it is also justified by the important property of the logit destination choice models. In effect, under some hypotheses on the explanatory variables inserted in the model, it's possible to verify the property of the invariability towards the regional level of aggregation; this property allows to get results not influenced by the adopted level of spatial aggregation. In the zoning processes it is important to consider the source of potential error, resulting from the constitution of artificial areal unit. The *Modifiable Areal Unit Problem* (MAUP) can conduct to scale and/or to aggregation problems; in the aggregation process, these drawbacks must be overcome with the imposition of appropriate weight matrixes that keep in consideration the different structural characteristics of the zones.

A large literature exists regarding to the zoning techniques too: some authors propose the definition of opportune proximity indexes (spatial and not spatial indexes) for measure the similarity degree of the zones, and progress to the following aggregation (Getis A., Ord J. K., 1992); other authors use multivariate analysis models to identify homogeneous macro areas inside and heterogeneous macro areas among them. In a such context, it results interesting the technique that foresees the use of a spatial zoning algorithm based on the factorial analysis, and a subsequent cluster analysis.

A relative proposal consists in the integration between the two approaches just described, using jointly techniques of multidimensional analysis and indexes of spatial association together. Particularly, a local factorial analysis, proposed by Lebart L. (1969), can be hypothesized using contiguity matrixes on spatial dates, with territorial constrains composed from Boolean matrixes; otherwise, the following developments, like the conditioned correspondences analysis (Escoufier B., 1990), that allows to reduce the influence of the local variations; even the factorial techniques, in which the regional proximity matrixes are based on the global and local indexes of spatial association (Anselin L., 1995).

We could propose a further development: it consists in the use of a matrix $\mathbf{G} = \mathbf{F} - \mathbf{P}$ as the matrix of the scalar products inside the principal component analysis relative to a

reference subspace (PCAR) (D'Ambra L., Lauro C. N., 1982). Particularly, we consider the proximity operator \mathbf{G} , that is a semi-defined positive matrix, obtained from the difference among the diagonal matrix \mathbf{F} , of elements $f_{ij} = \sum_j w_{ij} d_j^*$, sum of the weights related to the adjoining zones to every zone i , and the matrix $\mathbf{P} = \mathbf{W}\mathbf{D}^*$, with w_{ij} element of the Boolean contiguity matrix \mathbf{W} and d_j^* element of the diagonal matrix of the "spatial" weights \mathbf{D}^* . It's possible, in this way, to enlarge the PCAR approach : replacing the matrix of the scalar products with the operator of proximity \mathbf{G} and, considering the diagonal matrix $\bar{\mathbf{D}}$ of generic element equal to the reciprocal of the number of observations $1/n$, we can obtain $\mathbf{X}(\mathbf{X}'\bar{\mathbf{D}}\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{D}}(\mathbf{F} - \mathbf{P})\bar{\mathbf{D}}\mathbf{X}(\mathbf{X}'\bar{\mathbf{D}}\mathbf{X})^{-1}\mathbf{X}'$. We can reach the same result following the approach considered by Méot *et al.* (1991): using the principal component analysis (PCA) on the triplet $(\mathbf{X}, \mathbf{Q}, \bar{\mathbf{D}})$ (dates, metrics, weighs), the authors search the metric \mathbf{Q} so that results: $\min = \|\mathbf{G} - \mathbf{X}\mathbf{Q}\mathbf{X}'\|^2$. The matrix \mathbf{Q} is equal to $(\mathbf{X}'\bar{\mathbf{D}}\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{D}}\mathbf{G}\bar{\mathbf{D}}\mathbf{X}(\mathbf{X}'\bar{\mathbf{D}}\mathbf{X})^{-1}$, for which we will get the expression $\mathbf{X}(\mathbf{X}'\bar{\mathbf{D}}\mathbf{X})^{-1}\mathbf{X}'\bar{\mathbf{D}}\mathbf{G}\bar{\mathbf{D}}\mathbf{X}(\mathbf{X}'\bar{\mathbf{D}}\mathbf{X})^{-1}\mathbf{X}'$, that coincides with the preceding formulation.

Once that we have determined the latent components through the "spatial" factorial analysis, can be executed a subsequent automatic classification. Such multivariate approach, conjoint to the use of spatial proximity indexes, allows the formation of homogeneous cluster, constituted by elementary traffic zones adjacent between them. In this manner, the insertion of the spatial component inside the behavioral models, it happens already in the preliminary phase of the choice set specification.

A different approach, without passing for the factorial analysis, it is given by the use of spatial cluster methods. These techniques are based on the use of spatial autocorrelation indicators, as the Moran's I and Geary's C indexes, to individuate homogeneous classes obtained with the comparison between the structure of autocorrelation of the zones.

They exist, besides, a family of iterative techniques, that foresees the estimation of different models, everyone with a different set of alternatives. One of these techniques expects the estimation of the systematic utility function parameters, calibrating a model on a limited number of casually selected destinations zones; subsequently the technique proceeds to the aleatory extraction of other zones and it calibrate the model again, getting other coefficients. Such procedure is repeated N times, reaching the estimation of N sets of parameters (one for each model): the "final" set of parameters will be obtained as average of the N initial sets. However, this approach has an inconvenient: extracting every choice set it doesn't give a different weight to different zones that have different features; it creates therefore, in this way, a model whit evident distortions.

An alternative approach consists of estimating so many binary logit models, how many the possible zones j of destination are. The choice set for each model will be: "inside or out the zone j ." In such case, different sets of parameters will be obtained and these parameters will have a distribution with a certain average and varianza. Then it's possible to choose the average of the different sets or a more appropriate position index, as "final" set of parameters.

To our notice, a technique to integrate the precedent approach, coherently with the interpretation of the binomial logit coefficients, it consists, in to doesn't use simply the average of the esteemed coefficients, but to calculate the average of such coefficients also considering the marginal effects, $p_i(1 - p_i)\beta_k$.

3. Destination models and spatial components

Destination models belong to the wider family of random utility models. In the latter the rational decision-maker (user or group) makes a choice among the possible alternatives on the basis of the information he has and, taking into account the constraints, try to maximise his perceived utility.

The perceived utility U_{ij} , relative to user i and choice j , can be expressed by the sum of the systematic utility V_{ij} and by a random residual ε_{ij} . The systematic utility V_{ij} represents the expected value of the perceived utility among all the users with the same choice set I^i of decision-maker i ; while the random residual ε_{ij} defines the deviation from the average value of the perceived utility of decision-maker i . Therefore $U_{ij} = V_{ij} + \varepsilon_{ij}$, with $\forall j \in I^i$.

In the distribution model, the choice of a destination imposes spatial and temporal constraints, which eliminate the choice of all the other possible destinations from the choice set. The best functional form to be used to specify the probability of choosing a given alternative is the Multinomial Logit Model (MNL). In this context, the MNL model will be used for more complex functional forms which are able to incorporate the spatial component. It is important to define some assumptions which are at the basis of random utility models and which are particularly important for destination models.

Independence from Irrelevant Alternatives (IIA). This assumption imposes that the error terms of the utility function of each choice alternative are i.i.d. according to a Gumbel random variable. This means that the probability relative to the choice of two traffic zones is independent from the introduction or elimination of other zones in the choice set. However, this hypothesis is not always valid in distribution models.

Users homogeneity. In behavioural models it is assumed that rational decision-makers react homogeneously to the attributes of the choice alternatives, independently from their socio-economic characteristics. This assumption is often violated and in order to avoid this problem homogenous users' groups are formed and each model is then calibrated on each group.

Spatial implications. Transport demand is influenced by three spatial problems: dependence, heterogeneity, heteroskedasticity (Bhat C. R., Zhao H., 2001). The spatial dependence describes the presence of not-observed factors, like territorial, which influence travel behaviour. The spatial heterogeneity highlights that the relationship between the dependent variable and those independent can vary on the basis of the different zones, therefore it may not exist a global relationship, but different local relationships. Another source of spatial distortion, the spatial heteroskedasticity, reflects the possibility that the not observed factors variability can influence the choice process in a different way according to the reference zone.

It is evident that the spatial component creates some serious limits, which should be considered in the model specification. In the following sections it will be introduced in the model specification the territorial aspects: both at an individual level, trying to specify the proximity effects among the users (Mohammadian A. *et al.*, 2003); and at a zone level, by considering the contiguity constraints among the destinations (Miyamoto K. *et al.*, 2004).

3.1 Multinomial Spatial Logit

The introduction of the spatial component is motivated not only by logic or conceptual aspects but it has also important implications on the estimates properties. In particular,

Goetzke F. (2003) highlights that the elimination of the territorial component from the discrete choice model leads to parameters estimates which are not correct; vice versa, the introduction of an autoregressive component in the systematic utility part, determines correct estimates.

The first proposal at an individual level, which defines a spatial multinomial logit (MNSL), considers the introduction of spatial constraints directly in the initial data matrix, i.e. as an explicative variable of the decision-maker choice, at the same level of the other independent variables. In particular, it can be assumed that the systematic component, V_{ij} , is made up of two parts: a linear function in the parameters made up of the observed attributes on the decision-maker i and relative to alternative j ; a second term which captures, on the other hand, the spatial dependency among individuals. Therefore, the perceived utility is:

$$U_{ij} = V_{ij} + \varepsilon_{ij} = \left(\sum_j \beta_j X_{ij} + \sum_{k=1}^K \rho_{ikj} y_{kj} \right) + \varepsilon_{ij}$$

where β_j are the coefficients to be estimated, which characterise the set of explicative variables, X_{ij} , observed for each user i and alternative j ; the parameters ρ_{ikj} are the elements of the matrix coefficients which express the proximity between two individuals. In particular, the influence that the choice of decision-maker k has on individual i ; y_{kj} is a dummy variable equal to 1 if the decision-maker k chooses alternative j , 0 otherwise. In spatial statistics ρ_{ikj} is usually computed as exponential of the negative distance (d_{ik}), which separates two decision-makers. The latter is a function of the parameters λ and γ to be estimated with the maximum likelihood method. The total influence of all the decision-makers K on the preference of individual i , who chooses alternative j , is equal to:

$$Z_{ij} = \sum_{k=1}^K \rho_{ikj} y_{kj} = \sum_{k=1}^K \lambda \exp\left(-\frac{d_{ik}}{\gamma}\right) y_{kj}$$

The probability that a decision-maker i chooses alternative j , among all the possible alternatives belonging to the choice set I^i , depends on the fact that the perceived utility U_{ij} , relative to alternative j , is the highest among all the perceived utilities: $P_{ij} = P[U_{ij} > U_{ih}] = P[(V_{ij} + \varepsilon_{ij}) > (V_{ih} + \varepsilon_{ih})] = P[V_{ij} - V_{ih} > \varepsilon_{ih} - \varepsilon_{ij}] ; [\forall h \neq j] \in I^i$.

If it is assumed that the errors ε are i.i.d. according to a Gumbel variable, therefore it is possible to derive the MNL model. In fact the Gumbel variable has the important property of stability with respect to maximization; moreover if ε_1 e ε_2 are i.i.d. according to a Gumbel variable, therefore the difference $\varepsilon = \varepsilon_1 - \varepsilon_2$ has a logistic distribution. It is possible to derive the probability that individual i chooses alternative j :

$$P_{ij} = \frac{\exp(\theta V_{ij})}{\sum_{j \in I^i} \exp(\theta V_{ij})} \quad \forall j \in I^i \quad (1)$$

and the systematic utility function V_{ij} of the MNL model is of the spatial type:

$$V_{ij} = \sum_j \beta_j X_{ij} + Z_{ij} = \sum_j \beta_j X_{ij} + \sum_{k=1}^K \rho_{ikj} y_{kj}$$

In particular, the parameter estimate ρ , through γ e λ , is carried out by maximising the log-likelihood function, which for a sample of size N , is:

$$\ln(L(\beta)) = \sum_{i=1}^N \sum_{j \in I^i} \ln P_{ij}^{y_{ij}} = \sum_{i=1}^N \sum_{j \in I^i} y_{ij} \left[V_{ij} - \ln \left(\sum_{j \in I^i} \exp(V_{ij}) \right) \right] \quad (2)$$

where y_{ij} is a dummy variable, relative to decision-maker i , equal to 1 if the decision-maker i has chosen alternative j , 0 otherwise. The method used is the Newton-Raphson one which maximises equation (2) with respect to β , while γ and λ can be obtained by computing the first and second derivatives. It is interesting to note that the parameters' estimate γ and λ can be obtained also through an iterative procedure by assigning a series of values to the parameter γ , at the same time the λ value can be estimated like a simple MNL parameter and by stopping the procedure when the best fitting of the model is obtained.

A further development of the multinomial spatial logit model has been introduced by Nelson, G. *et al.* (2004). In this approach a conditional multinomial spatial logit is considered (MNSCL), where apart from the decision-maker i characteristics, other variables relative to the rational individual i and choice alternative j are considered. Using this model it is possible to estimate, separately from the observed characteristics on each zone, which type of rational decision-makers each destination zone is more willing to attract.

In the MNSCL model, the linear component of the systematic utility of the MNSL is made up of two parts (always linear), therefore it follows:

$$x_{ij}\beta_j = h_{ij}\eta + g_i\delta_j$$

where h_{ij} represents the characteristics of individual i relative to alternative j ; while g_i are the individual characteristics of the rational decision-maker i independent from choice j . The perceived utility U_{ij} has the form:

$$U_{ij} = V_{ij} + \varepsilon_{ij} = \left(\sum_j h_{ij}\eta + \sum_j g_i\delta_j + \sum_{k=1}^K \rho_{ikj} \gamma_{kj} \right) + \varepsilon_{ij}$$

with ρ_{ikj} equal to the proximity between two individuals i and k . The probability that decision-maker i chooses alternative j is given by equation (1).

If the models presented consider the spatial effects, which influence the choice of a destination zone; they are not always able to solve the IIA property. In the literature a set of methods have been proposed which try to avoid such limit. Examples are Nested logit models; the ordered generalized extreme value models; the paired combinatorial logit models and the logit models with random coefficients.

None of these models, which can be included in the wider class of generalized extreme value models, is flexible enough to be able to overcome the IIA assumption on random residuals. In particular, in a recent paper (Browstone D, Train K., 1999), it has been demonstrated that the nested logit model is not sufficiently flexible to forecast the user's behaviour in the transportation context. The only models which are able to overcome such limits and to approximate the behaviours of a decision-maker, who has to make a discrete choice, are the multinomial probit and the mixed logit models.

Probit models present quite complex functional forms, based on multiple integrals with open forms. In the following section, mixed models will be considered and, by overcoming the IIA property, they offer a good support for forecasting the choice behaviour. In these models the territorial constraint will be introduced, taking into account the spatial aspects. Finally they will be considered in the wider family of the Kernel logit models.

3.2 Mixed Logit and constraints at zone level.

In the following a series of models will be presented based on the mixed multinomial logit (MMNL) approach which considers the spatial aspects in the choice context. In these models the territorial component does not take into account the proximity among individuals, but the spatial contiguity among the destination zones. This approach allows to use a system of weights, within the models themselves, which integrate the territorial constraints in the choice process.

In particular, the spatial matrix of weights \mathbf{W} , where the generic element can be computed as reciprocal of the power of the distance between two zones j and h , i.e. $w_{jh} = 1/d_{jh}^\alpha$, is introduced in the model within the autoregressive component.

These approaches start from a MNL model with an autoregressive deterministic component and with spatial weights, where the perceived utility, expressed through a matrix, is $\mathbf{U} = \mathbf{V} + \boldsymbol{\varepsilon} = \boldsymbol{\rho}_1 \mathbf{W}_1 \mathbf{V} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

The spatial component is introduced in the stochastic component of the MMNL model with autoregressive errors, the perceived utility is given by:

$$\mathbf{U} = \mathbf{V} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\xi} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\eta} + \boldsymbol{\rho}_2 \mathbf{W}_2 \boldsymbol{\xi} + \boldsymbol{\mu} \quad (3)$$

Another approach considers an integrated mixed logit model with spatial effects. In the model proposed by Miyamoto K. *et al.* (2004), the spatial component is introduced in two steps: by considering the spatial interaction among the data, i.e. in the systematic utility; by examining the spatial autocorrelation in the error term. This model has the following form:

$$\mathbf{U} = \mathbf{V} + \boldsymbol{\varepsilon} \begin{cases} \mathbf{V} = \boldsymbol{\rho}_1 \mathbf{W}_1 \mathbf{V} + \mathbf{X}\boldsymbol{\beta} & \text{spatial and autoregressive} \\ & \text{deterministic component} \\ \boldsymbol{\varepsilon} = \boldsymbol{\eta} + \boldsymbol{\xi} = \boldsymbol{\eta} + \boldsymbol{\rho}_2 \mathbf{W}_2 \boldsymbol{\xi} + \boldsymbol{\mu} & \text{spatial and autoregressive} \\ & \text{error term} \end{cases}$$

where \mathbf{U} is the perceived utility; \mathbf{V} is the systematic utility component; $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ are the parameters for the spatial effects, relative to the systematic component and the error term; \mathbf{W}_1 and \mathbf{W}_2 are the spatial weight matrices; $\boldsymbol{\beta}$ is the vector of parameters of the explicative variables \mathbf{X} ; $\boldsymbol{\varepsilon}$ is the error component of the perceived utility and it is made up of $\boldsymbol{\eta}$ with null mean, i.d. like Gumbel variables, independent from the observed data and from the computed parameters; of another part $\boldsymbol{\xi}$, with null mean as well, with a distribution like $f(\boldsymbol{\xi}|\boldsymbol{\Omega})$, with $\boldsymbol{\Omega}$ set of parameters of the distribution, it depends from the structure of the observed data, from the estimated coefficients, and from the choice alternatives. This latter component $\boldsymbol{\xi}$, in the presented model, is autoregressive and it is influenced by spatial factors: $\boldsymbol{\xi} = \boldsymbol{\rho}_2 \mathbf{W}_2 \boldsymbol{\xi} + \boldsymbol{\mu}$, with $\boldsymbol{\mu}$ vector of multi-normal random variables with null mean and variance-covariance matrix $\boldsymbol{\Sigma}$. If it is assumed that

$\Sigma = \sigma^2 \mathbf{I}$, then it follows $\boldsymbol{\mu} = \sigma \boldsymbol{\tau}$, and σ is a scalar which expresses the standard deviation and $\boldsymbol{\tau}$ is the vector with elements distributed as a standard normal variable.

The parameter estimates of the model can be obtained by using the maximum likelihood method, which leads to the formulation of multidimensional integrals which are not in a closed form. Recent researches have developed simulation techniques able to approximate such integrals and lead to the definition of a log-likelihood function to be maximised (Bolduc D., 1992).

Another family of models, which explains the choice behaviour, is that of the Kernel logit models. Such models are based on a functional form of the logit type and consider extensions to the residual part which is made up of two components: one is probit, the other is distributed with a Gumbel variable. They integrate both the logit aspects and those of the probit. The probit component, which has a multivariate form, capture the interdependence among the alternatives. The matrix form of the perceived utility (Ben-Akiva M. *et al.*, 2001) is equal to $\mathbf{U} = \boldsymbol{\beta}\mathbf{X} + \mathbf{F}\boldsymbol{\xi} + \boldsymbol{\eta}$, with $\boldsymbol{\xi} = \mathbf{T}\boldsymbol{\zeta}$ where: \mathbf{F} is the matrix of factor loadings; $\boldsymbol{\xi}$ is the error with multivariate distribution, which is decomposed into a triangular matrix \mathbf{T} , which takes into account the heteroskedasticity, and a set of factors $\boldsymbol{\zeta}$ i.d. with a Gauss random variable; while $\boldsymbol{\eta}$ is again a vector of i.i.d. Gumbel variables.

A special form of the Kernel logit model can be obtained when the error term is of the autoregressive type, this formulation, particularly indicated in the models where the choice set is wide, leads to a MMNL model with autoregressive errors and with $f(\boldsymbol{\xi}|\boldsymbol{\Omega})$ of the probit type of equation (3). In fact, if a generalised autoregressive process is considered (GAR) of the first order of the errors $\boldsymbol{\zeta}$, with $\boldsymbol{\xi} = \boldsymbol{\rho}_2 \mathbf{W}_2 \boldsymbol{\xi} + \mathbf{T}\boldsymbol{\zeta}$, it follows that $\mathbf{U} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\eta} + \boldsymbol{\rho}_2 \mathbf{W}_2 \boldsymbol{\xi} + \mathbf{T}\boldsymbol{\zeta}$.

With the aim of verifying the assumptions which are at the basis of the destination models, through the use of the integrated mixed spatial logit model, an application is proposed in the following based on shopping trips in the metropolitan area of Naples. The database has been provided by the Transportation Department of Campania district. For the sake of simplicity it has been used just one explicative variable relative to the number of commercial activities in a given zone.

Concerning the weight matrix \mathbf{W} , it has been used a Boolean matrix ($\mathbf{W} = \mathbf{W}_1 = \mathbf{W}_2$), where \mathbf{W}_1 and \mathbf{W}_2 can be different in the two components of the model, i.e. deterministic and stochastic and must be evaluated, by considering distance functions such as: $w_{jh} = 1/d_{jh}^2$, or $w_{ij} = \exp(-d_{ij}^{1.5})$, which maximise the explicative capacity of the model.

In the application a comparison has been made between a classical MNL, a MNL with spatial deterministic and autoregressive component, a MMNL with spatial and autoregressive error component and, finally with the integrated spatial mixed logit model. For each model for several values of ρ_1 and ρ_2 , it has been estimated the coefficient β and the procedure has been stopped when the model has better reproduced reality.

In **table 1** for each model the estimated parameters are reported and the likelihood ratio is reported as well. It is clear the better performance of the fitting capacity of the integrated spatial mixed logit model. In particular, the use of a spatial mixed logit model allows the introduction of the parameters ρ_1 and ρ_2 , which give a measure of the degree of the impact of the territorial aspects on the systematic component of the model. Positive and statistically significant parameters characterise the proximity among the zones which influence the modelled choice process, relative to the observed data and/or not

observed. Vice versa in the case of negative coefficients. Another application with a larger number of variables will be presented at the SIS meeting.

Table 1: Parameters estimates

	MNL	MNL with V autoregressive and spatial	MMNL with ε autoregressive and spatial	MMNL integrated
β	4,01 (21,69)	4,13 (21,45)	5,02 (18,68)	5,73 (18,07)
ρ_1	-	0,084	-	0,18
ρ_2	-	-	0,64	0,78
L ratio	0,252	0,255	0,253	0,296

References

- Aluja T., Lebart L., (1984) Local and Partial Principal Component Analysis and Correspondence, COMPSTAT, Physica Verlag, Vienna, 113-118.
- Anselin L. (1995) Local Indicator of Spatial Association – LISA, *Geographical Analysis*, Ohio State University Press, Vol 27, 2, 93-115.
- Ben-Akiva M., Lerman S. (1985) *Discrete choice analysis, theory and application to travel demand*, The MIT Press, Cambridge, USA.
- Ben-Akiva M., Bolduc D., Walker J. (2001) Specification, Identification and Estimation of the Logit Kernel Model, working paper, Massachusetts Institute of Technology.
- Bhat C. R., Zhao H. (2001) The spatial analysis of activity stop generation, presentation at the *Transportation Research Board*, Washington, DC.
- Bolduc D. (1992) Generalized Autoregressive Errors in the Multinomial Probit Model, *Transportation Research B*, 26-2, 155-170.
- Brownstone D., Train K. (1999) Forecasting New Product Penetration with Flexible Substitution Patterns, *Journal of Econometrics*, Vol. 89-1, 109-129.
- Cascetta E. (2001) *Transportation Systems Engineering: Theory and Method*, Kluwer Academic Publishers, Dordrecht.
- D'Ambra L., Lauro N. C. (1982) Alcune estensioni dell'analisi in componenti principali in rapporto a sottospazi di riferimento, *Alcuni lavori di analisi statistica multivariata*, Leoni R. (a cura di), SIS Firenze, 41-67.
- Escofier B, Benali H., Bachar K., (1990) Comment introduire la contiguïté en analyse des correspondances? Application en segmentation d'image, *Statistique et Analyse des Données*, Vol 15, 3, 61-92.
- Fleming M. (2004) Techniques for estimating Spatially Dependent Discrete Choice Models, in: L. Anselin, R. Florax and S. Rey (Eds.), *Advances in Spatial Econometrics*, Springer, 145-167.
- Getis A., Ord J. K. (1992) The Analysis of Spatial Association by Use of Distance Statistics, *Geographical Analysis*, 24, 189-206.
- Goetzke F. (2003) Are Travel Demand Forecasting Models Biased because of Uncorrected Spatial Autocorrelation?, presented at the *North American Meeting of the Regional Science Association International*, Philadelphia, PA.
- McFadden D. (1978) Modelling the choice of residential location, in: A. Karlqvist, L. Lindqvist, F. Snickars and J.W. Weibull (Eds.), *Spatial Interaction Theory and Planning Models*, North-Holland, Amsterdam.
- Méot A., Chessel D., Sabatier R. (1991) Analyse des données spatio-temporelles et opérateurs de voisinage, *Biométrie et environnement*, J.D. Lebreton & cols (Eds), 29, Masson, Paris.
- Miyamoto K., Vichiensan V., Shimomura N., Páez A. 2004 Discrete Choice Model with Structuralized Spatial Effects for Location Analysis, *Transportation Research Record: Journal of the TRB*, National Research Council, Washington, 1898, 183-190.
- Mohammadian A., Kanaroglou P. S. (2003) Applications of Spatial Multinomial Logit Model to Transportation Planning, conference paper *10th International Conference on Travel Behaviour Research*.
- Nelson G., De Pinto A., Harris V., Stone S. (2004) Land use and road improvements: a spatial perspective, *International Regional Science Review*, 27-3, 297-325.