École Polytechnique Fédérale de Lausanne

Laboratory of Geographic Information Systems

Geographic concentration of economic activities: on the validation of a distance-based mathematical index to identify optimal locations

Master Thesis

Olivier Monod

Supervised by: **Dr. Stéphane Joost (LaSIG, EPFL)**
Albert Gaspoz (SDE, Canton du Valais)
Prof François Golay (LaSIG, EPFL)

Spring semester 2011

List of figures

List of tables

Abstract

The present study proposes a validation of a mathematical index Q able to identify optimal geographic places for economic activities, solely based on the location variable. This research work takes its roots in the 1970s with the statistical analysis of spatial patterns, or analysis of point processes, whose main goal is to understand if a resulting spatial distribution of points is due to chance or not. Indeed point objects are commonplace (towns in regions, plants in the landscape, galaxies in space, shops in towns) and the development of specific mathematical tools are useful to understand their own location processes. Spatial point deviations from purely random configurations may be analyzed either by quadrat or by distance methods. An interesting method of the second category – the cumulative function M – was developed recently for evaluating the relative geographic concentration and co-location of industries in a non-homogeneous spatial framework. On this basis, and having quantified retail store interactions, The French physicist Pablo Jensen elaborated the Q-index to automatically detect promising locations. To test the relevance of this quality index, Jensen used location data from 2003 and 2005 for bakeries in the city of Lyon and discovered that between these two years, shops having closed were located on significantly lower quality sites. Here, using bankruptcy data provided by the Registrar of companies of the State of Valais in Switzerland and by the City Council of Glasgow in Scotland, we implemented a method based on univariate logistic regressions to systematically test for the relevance of the Q-index on the many commercial categories available. We show that the Q-index is reliable, although significance tests did not reach stringent levels. Access to trustable bankruptcy data remains a difficult task.

Key index words: Location, GIS, Economic activities, Complex systems, Q-index, Distance-based methods, Geographic concentration, M function, Spatial point processes

1 Introduction

The space in which the *homo oeconomicus* acts is not homogeneous. Political boundaries, land use legislations, both natural and artificial physical structures generate spatial opacity and heterogeneity. In such a space, better referred to as the *economic landscape*, setting up a new business at a location that maximizes all profits requires to take into account a huge number of variables ranging from transportation systems to consumer location through various other factors. Since the father of location theorists Johann Friedrich von Thünen (Isard 1956), published "Der Isolierte Staat" in 1826, location science has not stopped increasing the complexity of its models with the hope that the associated predicting capacity would increase too .

Since then, major advances in location analysis have come about through the development of mathematical models to solve spatial planning problems. In 2006, statistical physicist Pablo Jensen went back to the roots of real estate business and its fundamental mantra: location, location, location. He developed an innovative distance-based index based on the work of Marcon and Puech (Marcon and Puech 2010) to show that the use of location data alone is sufficient to reveal many important facts about the spatial organization of retail trade (Jensen 2006). He also demonstrated that optimal locations can be found based solely on the location of existing activities using simple and computationally inexpensive methods that are theoretically funded (Jensen 2006; Jensen 2009; Jensen and Michel 2009; Marcon and Puech 2010).

Brief review of classic location theories

By the end of the 1950s, in their series of review papers "Spatial Structure of the economy I, II and III", Garrison and Marble (1957) took stock of the situation about the issue to know what determines the spatial arrangement (structure, pattern, or location) of economic activities (Garrison and Marble 1957). They reviewed books by Isard (Isard 1956), Dunn (Dunn 1954), Greenhut (Greenhut 1956), Ponsard (Ponsard 1955), Boustedt (Boustedt 1957) and noticed a large variety of approaches. But the main bases of the theory of the location of economic activities was here, distributed among these different important publications.

Later, several authors reported on literature explicitly addressing the strategic nature of facility location problems that took place from the end of the 1950s (Hamacher and Nickel 1998; Owen and Daskin 1998; Revelle and Eiselt 2005; Revelle, Eiselt et al. 2008). They characterized the broad array of approaches for analysis and modeling in location science. They also emphasized the fact that finding robust facility locations is a difficult task, requiring that decision makers account for uncertain future events.

Recent developments

Lately, economists focused on the spatial dimension of economic activities and attempted to calculate concentration indices, what appeared to be a challenging task. In 2004, Combes proposed a few criteria to build good spatial concentration indices (Combes 2004). Two main approaches trickled down from previous researches: Kernel Density Functions (Silverman 1986) and cumulative functions (distance-based), on which we focus in this work. While the problem of measuring spatial concentration is a quite recent preoccupation in economics, it has a longer

history in statistical analysis of spatial patterns (Ripley 1977) and statistical physics. Discipline in which the works of Marcon and Puech (Marcon and Puech 2010) and Jensen (Jensen 2006; Jensen 2009) take their roots.

Spatial point patterns and concentration indices

The absolute density index such as the K and L functions (Ripley 1976) can be used in order to measure concentration of points. These functions do not adapt well to human-related behaviors due to a major difference concerning the underlying spaces. Space can be considered as homogeneous whereas human-related space (referred to as the *territory* by geographers) is strongly heterogeneous.

A solution proposed by Marcon to overcome the limits of the K and L functions is to measure a *relative* concentration (Marcon and Puech 2010) to be understood as a measure of the deviation of the empirical distribution from a purely random distribution (Jensen and Michel 2009). This is a powerful idea whose main quality is to free density indices from the homogeneous space hypothesis. It also provides a tool not affected by the Modifiable Unit Area Problem affecting other indices like the ones based on Kernel Density Estimators (Openshaw 1984). The only parameter affecting the M- and Q-index models is the search Radius. Discussion about the Radius value is discussed by Jensen (2005).

The principle of the relative measure of concentration is to count the number of neighbors of specific category A in a circle of a given radius and to compare it to the total number of neighbors inside this circle (simple ratio).This provides a local concentration index to compared to a global one afterward: the total number of sampled points belonging to category A divided by the total number of sample points. The measure of relative concentration, called the M-index by Marcon, is the local ratio divided by the global one. The index quantifies the links intensity and its repulsive or attractive nature affecting different categories.

Preliminary results from Marcon (Marcon and Puech 2010) rapidly inspired much of Jensen's work and he developed two by-products of the M-index. First, he applied a community structure analysis algorithm, to the M-matrix in order to reveal underlying communities. The groups formed by the algorithm provide a qualitative insight into the economic structure of the area of interest. Second, he developed an optimal location index, called Q-index hereafter, which is based on the idea that the best location for an activity is the one that maximizes the number of friendly neighbors (positive M value) and minimizes unfriendly neighbors (negative M value).

Jensen showed that both community structure analysis and Q-index performed very well but he did not carry out a systematic validation procedure for many different economic activities (Jensen 2006; Jensen 2009). Stauffer made promising tests on data for the city of Geneva but also on limited data sets (Stauffer 2009). Omlin applied the Q-index to linguistic problematic without a validation perspective (Omlin 2010).

Goal: validation and application

The main objective of this work is to validate the Q-index on large data set describing the economic landscape of the canton of Valais and of the city of Glasgow. Indeed, the validation proposed by Jensen in 2006 was empirical and based on a single retail store category (Jensen 2006; Jensen 2009). Here we compare the Q-index for commercial activities that went bankrupt during a time interval to the Q-index for persistent activities. We used univariate logistic

regressions for several retail and non-retail economic activities, with a focus on primary sector. The working hypothesis is that low Q-indices are significantly correlated with high frequencies of bankruptcies. The second objective is to produce de Q-index map for the 6 most important cities of the Canton of Valais: Sion, Monthey, Martigny, Sierre, Viège and Brig. And finally, in order to complement the analysis of the economic landscape we apply a community structure analysis on the M-index matrix that reveals underlying communities.

2 Data

Data required to calculate the M-index and the Q-index are commercial activities (shops) characterized by geographic coordinates, a qualitative economic category and a unique identifier, collected at time t0. This dataset is referred to as "global". The validation procedure proposed in this study requires 2 additional data sets with similar attributes: the first one, referred to as "bankruptcies", contains activities that disappeared during a time interval starting at t0 and ending at t1; the second one, is referred to as "persistent" and contains the activities that remained active during the same time interval. Hereunder, we present the main characteristics of economic data provided by the Canton of Valais in Switzerland and by the City Council of Glasgow in Scotland. It is worth to mention that classification systems for our 2 study area are not the same. It is therefore very difficult to make comparison between them, what would be very useful.

Federal Registry of Enterprises and Establishments (REE) for the Canton of Valais, Switzerland

The geographic reference is obtained using geocoding at the postal address. This spatial reference is well suited to an urban context but not relevant in rural areas where houses often do not have a street number or even a street name. In the Rhone Valley where most of the retail activities are concentrated, more than 85% of units have a geographic reference. Here below is a general view of the Rhône valley (Source: Google Earth).

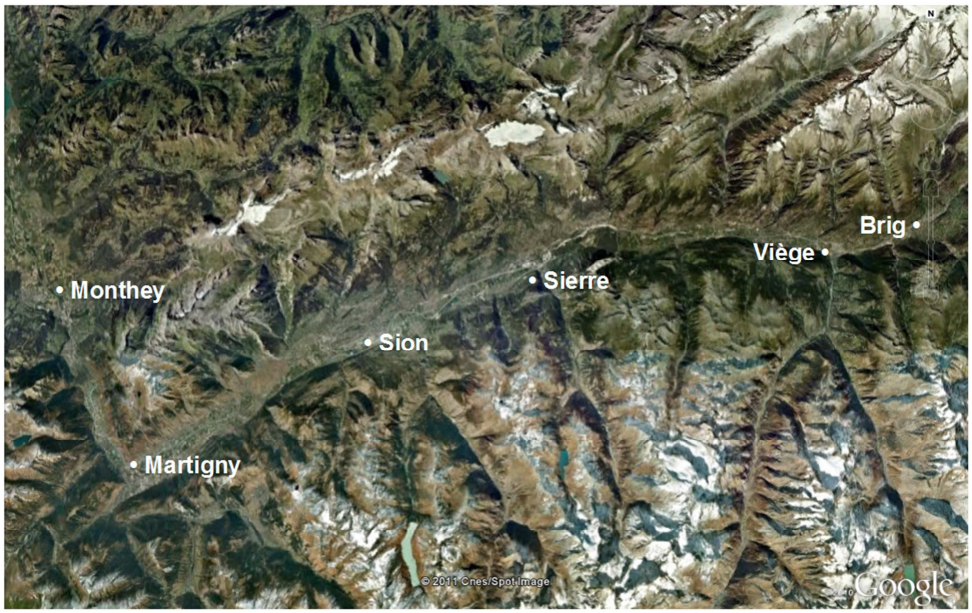

Figure 1: General view of the Rhône valley

This dataset is available for 4 different dates (01.01.2009, 01.01.2010, 01.01.2011, 29.04.2011)

what allows studying 3 different time intervals. This permitted to identify economic activities that went bankrupt. A preliminary analysis was conducted to be sure that the bankruptcy extraction procedure was possible. In the software, the name “bankruptcies” is equivalent to the one of “failures”. The number of points that have geographic reference varies around 22’000. The bankruptcies extraction procedure has some limits. For instance, it is not possible to discriminate between activities that failed and were not replaced (the location remains empty) and activities that were replaced by new activities within the same category or not. This might introduce a bias by adding set activities that did not really failed to real bankruptcies.

The classification system is the General Classification of Economic Activities 2008 (NOGA2008) from Swiss Statistics (SwissStatistics 2008). This classification is very detailed (6 digits) and also proposes a “ready-made” aggregated system from 6 to 2 digits. An even more aggregated classification called NOGA “branches” exists. It classifies economic activities into 20 groups and contains all economic sectors (primary, secondary and tertiary).

Here, we test the Q-index for activities typical for dense urban centers. At the NOGA2 level, the categories fulfilling at best these requirements are the 47th (all retail) and 56th (all restaurants and pubs). Going “down” to NOGA3 the best documented in both global and bankruptcies data sets are: 477 (all others retail), 476 (Cultural and leisure goods retail), 472 (specialized food retail) but we provide the results for all noga-3 47x categories available in the data set. Matching table between NOGA codes and full text names can be found in appendix A.

Commercial Activities from Glasgow City Council

The geographic reference was also obtained using geocoding at the postal address. We have at disposal bi-temporal information (2004 and 2009), with a notable withdraw being that the number of bankruptcies per category is very small. This data set contains only activities from primary sector. The classification system is established by the Glasgow City Council. The identification of bankruptcies is based on a “Vacant Shop” category corresponding to unoccupied shops. This means that a shop occupied in 2004 which became a “Vacant Shop” in 2009 was considered as a bankruptcy. Unfortunately, only 591 bankruptcies could be extracted from this data set, what represents a small number of bankruptcies per category. The reliability of this extraction strategy remains affected by uncertainties.

Classification system

The classification of activities into categories influences the output of the algorithms and the results of the validation process. There is an opposition between the quantitative significance and the aggregation level of the classification. A higher aggregation level implies a bigger sample size per category. The more aggregated the classification system, the less detailed the image of the economic landscape we get, with the risk of mixing activities that experience varying spatial dynamics.

3 Method

M-index

The M-index deals with the problem of quantifying deviations from empirical points distributions from purely random and non-interacting distributions (Jensen 2006; Jensen 2009). Additional information on spatial point processes can be found in previous publications by Marcon and Puech (Marcon and Puech 2010) and Ripley (Ripley 1976; Ripley 1977). Here is the mathematical expression of the M-index used in Jensen and Michel (Jensen and Michel 2009) and implemented in the software developed for the present study.

Notations:

r:	radius
N_t:	There are N_t sites, of which N_A sites are of type A, and N_B sites are of type B
$N_t(S,r)$:	The total number of sites distinct from S that are at a distance lesser than r of site S
$N_A(S,r)$:	Number of A sites in the same region
$N_B(S,r)$:	Number of B sites in the same region
S:	Refer to the “Site” that is the circle of radius r around the point. Note that S is never counted in those quantities, whatever its state.

Remark: The notation Nt (D) (resp. $N_A(D)$ and $N_B(D)$) will denote the total (resp. A and B) number of sites in a subset D of T . Thus for instance $N_A(S, r)$ stands for $N_A(B(S, r) \setminus \{S\})$, where B(S, r) denotes the ball centered at S with radius r (Jensen and Michel 2009).

M-index

The M-index is calculated using two indices. The first is the intra coefficient representing the interaction between activities belonging to the same category. It composes the diagonal of the M-index matrix. The second is the inter-coefficient representing the interaction for activities belonging to different categories. Both intra and inter coefficient contain an aggregated information about the repulsive or attractive nature of the spatial interaction between activities.

Intra coefficient M_{AA}: a measure of spatial interaction between activities belonging to the same category

Jensen introduces the M_{AA} coefficient as follows. Let us assume that we are interested in the distribution of N_A points in the set T, represented by the subset $\{A_i,i =1 ...N_A\} \subset T$. The reference law for this set, called pure random distribution, is that this subset is uniformly chosen at random from the set of all subsets of cardinal N_A of T: this is equivalent to an urn model with N_A draws with no replacement in an urn of cardinal N_t. Intuitively, under this (random) reference law, the local concentration represented by the ratio $N_A(A_i,r)/N_t(A_i,r)$ of stores of type A around a given store of type A should, in average, not depend on the presence of this last store, and should thus be (almost) equal to the global concentration N_A/N_t”. (Jensen 2006; Jensen 2009)

Estimator	Interpretation	
$M_{AA} = \frac{N_t - 1}{N_A(N_A - 1)} \sum_1^{N_A} \frac{N_A(A_i, r)}{N_t(A_i, r)}$	$M_{AA} = 1$	no aggregation tendency
	$M_{AA} > 1$	aggregation tendency
	$M_{AA} < 1$	dispersion tendency

Inter coefficient M_{AB}: a measure of spatial interaction between activities belonging to the different categories

Jensen introduces the M_{AB} coefficient as follows. In order to quantify the dependency between two different types of points, we set the following context: the set T has a fixed subset of N_A stores of type A, and the distribution of the subset $\{B_i, i = 1 ... N_B\}$ of type B stores is assumed to be uniform on the set of subsets of cardinal N_B of T \{A1,...,AN_A}. Just as in the intra case, the presence of a point of type A at those locations, under this reference random hypothesis, should not modify (in average) the density of type B stores: the local B spatial concentration (N_B(Ai,r)) / (N_t(Ai,r) - N_A(Ai,r)) should be close (in average) to the concentration over the whole town, (N_B) / (N_t - N_A) (Jensen 2006; Jensen 2009).

Jensen (2009) proposed a slight modification of his indicator. This modification only has an influence in cases where the number of points of the source category considered is very large compared to N_B(S,r).

Estimate	Interpretation	
$M_{AB} = \frac{N_t}{N_A N_B} \sum_1^{N_A} \frac{N_B(A_i,r)}{N_t(A_i,r)}$ (2006)	$M_{AB} = 1$	no aggregation tendency
	$M_{AB} > 1$	aggregation tendency
$M_{AB} = \frac{N_t - N_A}{N_A N_B} \sum_1^{N_A} \frac{N_B(A_i,r)}{N_t(A_i,r) - N_A(A_i,r)}$ (2009)	$M_{AB} < 1$	dispersion tendency

In this definition, the fraction 0/0 is taken as equal to 1 in the right hand term. Under the pure randomness hypothesis, it is straightforward to check that the average of this coefficient is equal to 1: for all r > 0, we have E[Maa] = 1". (Jensen 2006; Jensen 2009). We also set 0/0 M_{AB} values by 1 in our implementation.

Issues occurring with low density data sets

If points with no neighbors are present in the point subset, the computation of both M_{AA} and M_{AB} coefficients will generate many zero-by-zero divisions. In this case, the M-index value is replaced by 1 and this tends to soften the M-index values. This should not occur in dense urban areas for which the M-index is intended for but is likely to happen when the subset relates to lower density areas. A straightforward way out of the issue consists in filtering the data in order to remove isolated points.

Testing for the hypothesis of pure randomness in point patterns

Using analytical expression for the variance of the inter and intra coefficients, Jensen and Michel (2010) defined a way to test for the pure randomness hypothesis of point patterns. This procedure is useful to evaluate the maximal spatial extent of spatial interaction between activities at a local level.

Hypotheses are the following:

- H_0^B: the locations of B sites are purely random
- H_0^A: the location sites A are purely random

And the testing procedure:

- If $|a_{AB} - 1| > q_\alpha^{AB}$ reject hypothesis H_0^B
- If $|a_{AA} - 1| > q_\alpha^{AA}$ reject hypothesis H_0^A

Where $q_\alpha^{AB} = \sqrt{\sigma^2(a_{AB})/\alpha}$ and $q_\alpha^{AA} = \sqrt{\sigma^2(a_{AA})/\alpha}$ and $\sigma^2(a_{AB})$ is the variance of the M-index (Jensen and Michel 2009) and $a_{AB} = \log(M_{AB})$, and $a_{AA} = \log(M_{AA})$. We calculate the variance empirically in our case.

Quality Index Q

Jensen proposed to calculate a location quality index using the information about spatial interaction among categories contained in the M-index (Jensen 2006; Jensen 2009). The main idea is that the best location is the one that maximizes the number of friends (attraction) and minimizes the number of enemies (repulsion) within the radius of influence.

Different mathematical expressions of this index have been proposed in the literature (Jensen 2006; Jensen 2009; Omlin 2010). Here, we have used a weighted sum of the log of the M_{AB} and M_{AA} coefficients, defined as a_{AB} and a_{AA} values multiplied by the corresponding number of neighbors:

$$Q_i(x,y) = \sum_{1}^{N_{cat}} a_{ij} n_{ij}(x,y)$$

Validation

The main goal of this work is to demonstrate that the probability for a commercial activity to go bankruptcy is significantly higher for a low Q-index value than for a high one. We use logistic regressions to this end. A significantly negative B1 coefficient in the regression would reveal the inadequation of a specific location for a particular commercial activity. We set the binary output as 1 for activities that went bankruptcy (bankruptcies) and 0 for the ones that persisted during the corresponding time interval. We follow here closely the validation procedure proposed by Hosmer and Lemeshow (2000). The logistic regression model uses the logit transformation:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

That is then linearized.

$$g(x) = \ln[\pi \frac{(x)}{1-\pi(x)}] = \beta_0 + \beta_1 x$$

The importance of this transformation is that $g(x)$ has many of the desirable properties of a linear regression model. The logit, $g(x)$, is linear in its parameters, may be continuous, and may range from $-\infty$ to $+\infty$ depending on the range of x (Hosmer and Lemeshow 2000).

Three main output variables from the logistic regression are produced and analyzed:

- p-value: quantifies the probability to reject the null hypothesis
- WALD: quantifies the importance of B1 relative to its variance. Tells if β_1 is statistically significant
- McFadden pseudo-R2: evaluates the goodness of the fit.

Community Structure Analysis

We apply the Louvain method. This algorithm is able to process large data sets very quickly. It has been implemented in various languages and community structure analysis software, including Matlab and C. In the Louvain Method, the results optimize the partition for the modularity criteria, explained as follow.

The modularity of a partition is a scalar value between -1 and 1 that measures the density of links inside communities as compared to links between communities (Girvan and Newman 2002; Newman 2006). In the case of weighted networks (weighted networks are networks that have weights on their links, such as the number of communications between two mobile phone users) it is defined as (Park and Newman 2004):

$$Q = \frac{1}{2m}\sum_{i,j}[A_{ij} - \frac{k_i k_j}{2m}]\delta(c_i, c_j)$$

Where A_{ij} represents the weight of an edge between i and j, $k_i = \sum_j A_{ij}$ is the sum of the weights of the edges attached to vertex i, $c_i$is the community to which vertex iis assigned to the $\delta-$ function $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise and $m = \frac{1}{2}\sum_{ij} A_{ij}$ (Blondel, Guillaume et al. 2008).

Distance-based measure of density

In order to evaluate the local variation of point concentration in a geographical space, we first count the number of neighbors around each point in the data sets within a circle of radius r for each category. Mapping these values emphasizes the homogeneous/heterogeneous character of the study area.

Note concerning the methods implementation

All methods described here were implemented in PointPatternAnalyst in Maltab® programming language, with an integrated graphical user interface. The complete matlab source code and a standalone executable can be obtained on demand.

4 Results

Q-index maps for the city of Sion

Q-index maps were produced for the 5 major cities (Sion, Martigny, Monthey, Sierre, Brig) and integrated to the electronic Atlas of the Canton of Valais. They are available online at following address: http://www2.unil.ch/eatlasvs/geoclip2/.

The maps presented hereunder were produced using QGIS a powerful and user friendly open source GIS software.

Q-index maps for the city of Sion are calculated with a 25 meter resolution grid and a 100m radius value (Jensen, Boisson et al. 2005; Jensen 2006). Blue cells show positive Q values, red cells negative Q values, and pale yellow highlights neutral areas. The map for the global retail category (figure 1) clearly shows that the most interesting locations are located downtown, in the old part of the city. It has to be noted that the spatial distribution of retail high Q values shows only 1 favorable area. There is no secondary zone located elsewhere in town. Figure 5 also shows retail activities but with a radius of 50 meters, and in this case too, the retail area turns out to be continuous with no clear secondary clusters. This map illustrates the typical urban landscape met in the city of Sion affected by varying (low, medium, high) Q-index for commercial activities. High values are found in pedestrian-friendly (20km/h speed limitation) area with a high density of both retail shops and restaurants/pubs. Medium values are located in less friendly area: a quite high traffic road, with difficult pedestrian crossing. The lowest values are located near industries and noisy transportation infrastructures.

Retail and Restaurants (Figure 2) show very similar patterns: these categories are close neighbors. Car shops (Figure 3) get better Q values in the suburbs. This fits well the reality in Sion: most car shops are located south of the Rhone River within the industrial area but a few remain in town. Compared to chemical industry, car shops get higher Q-indices downtown because they are usually located at a shorter distance. Indeed, the large chemical industry located around the city of Monthey (one of the city included in the data set) is located far from the city center because the environmental depredation induced by this activity is massive and nobody wants it in its backyard.

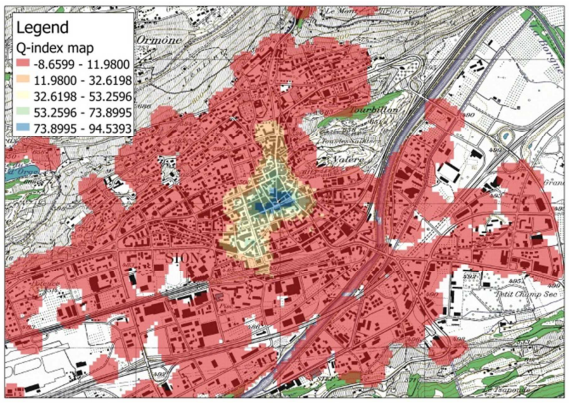

Figure 2: Q-index map for retail R=100m

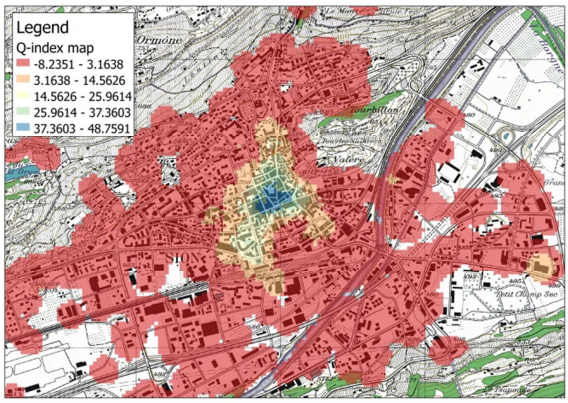

Figure 3: Q-index map for pubs and restaurants R=100m

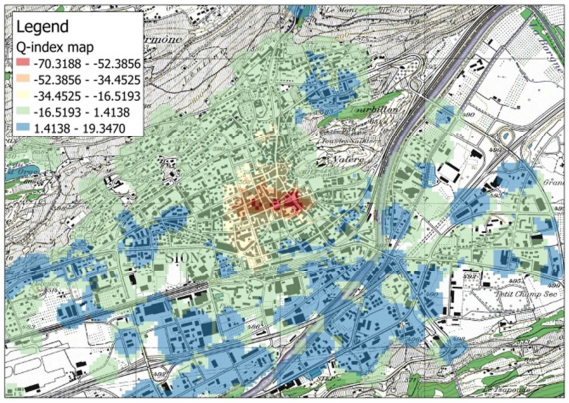

Figure 4: Q-index map for car shops R=100m

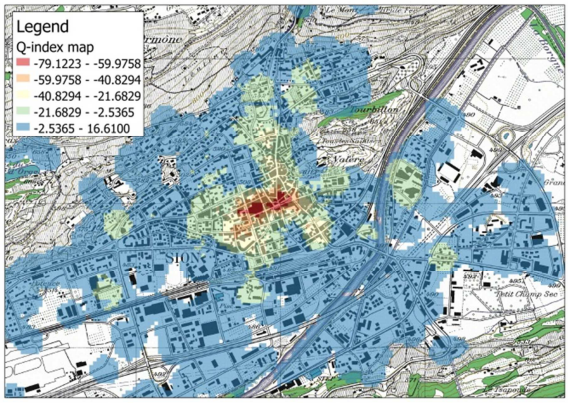

Figure 5: Q-index map for chemical industry R=100m

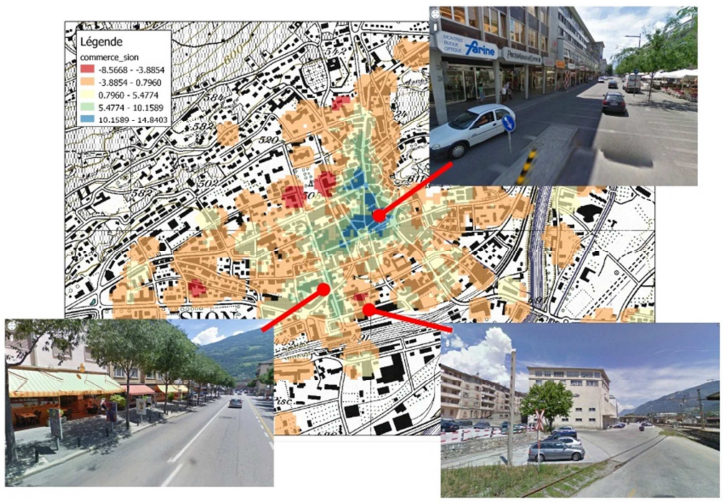

Figure 6: Q-index map for retail in Sion, R=50m

Logistic regression

In the data set from Canton of Valais, there are many categories for which we could run the validation procedure and get very satisfactory results. We decided to publish results for the categories related to dense urban centers and showing the largest number of bankruptcies.

The scatterplots (figures 6, 7, 8) show in most cases a bankruptcy's distribution that is more concentrated on the left (the lower Q values) than the distribution of persistent activities. This provides an indication that bankruptcies are on average affected by lower Q-values than persistent activities.

Results clearly show that the Q-index has a good predicting capability for the "retail" and "restaurant and pubs" categories. All regression curves for the Canton of Valais are decreasing functions of Q, meaning that the probability for a bankruptcy to occur reduces as Q increases. The β_1 estimates (Table 1) for all noga-2 level categories (47, 56) show a significant p-value based on student statistics at the 95% threshold. At noga-3 level, we also observe negative β_1 coefficients for all categories belonging to the noga-2 retail category (47). The noga-3 level categories with highest number of bankruptcies in our dataset are the following (Table 2): 477,476, 475, 473, 472, and 471. For categories 477, 476, 471 the p-value is very significant, below a 99% confidence level. For categories 472, 475, 473, p-values are not significant. Categories 475 and 473 are very poorly documented and this might explain the corresponding high p-values. For category 472 (specialized food shop), β_1 coefficient is close to zero with a high p-value.

Finally, the only category for which the probability to find a failure increases with Q is not statistically significant (large p-value).

Category noga-2 level	**ß1**	**P-value ß1 (student stat.)**	**Accepted at 95% threshold?**	**# Fail.**
NOGA47 (retail)	-1.67	1.43e-6	yes	441
NOGA56 (restaurants & pubs)	-1.28	2.35e-2	yes	274

Table 1: Statistical assessment variables for the 2 best documented retail noga-2 categories

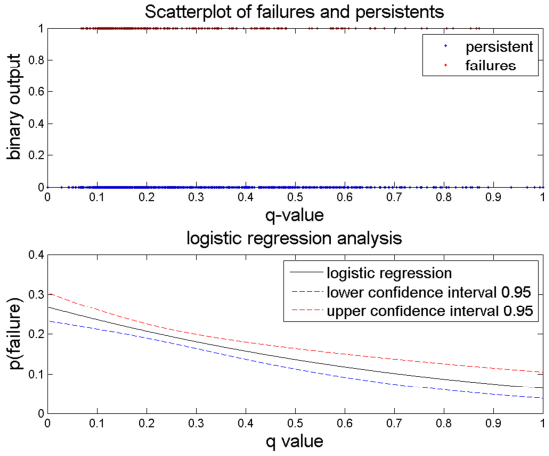

Figure 7: Logistic regression for noga-2 level category 47, radius 100m

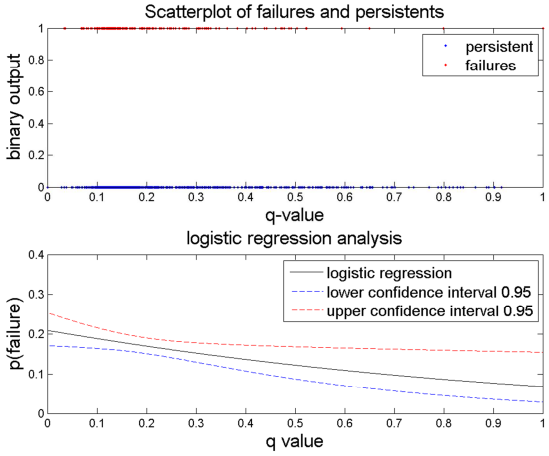

Figure 8: Logistic regression for noga-2 level category 56, radius 100m

Category noga-3 level	**ß1**	**P-value ß1 (student stat.)**	**Accepted at 95% threshold?**	**# Bank.**
NOGA 477 (other ret.)	-1.99	1.03e-4	yes	142
NOGA 472 (specialized food ret.)	-0.64	3.85e-1	no	89
NOGA 476 (Cultur. & leisure ret.)	-2.59	5.74e-3	yes	71
NOGA 471 (not specialized ret.)	-5.27	3.91e-3	yes	52
NOGA 475 (Household equipment)	-1.47	1.96e-1	no	41
NOGA 473 (Fuel retail)	-4.47	7.70e-2	no (90% ok)	14

Table 2: Statistical assessment variables for the 3 best documented retail noga-3 categories

Note: the 6 logistic regression plots are reproduced in appendices B.

Validation of the Q-index for Glasgow data set

For Glasgow, significant results are obtained only for one aggregated retail category approximately corresponding to the noga-2 47 category, composed mainly by activities like clothes shops, computer shops, shoes shop, bakers, butchers, etc…When applying the validation approach to the original classification system, results turn out not to be significant, most probably because of the very low number of bankruptcies in retail activities.

Category	**ß1**	**P-value ß1 (student stat.)**	**Accepted at 95% threshold?**	**# Bank.**
Grouped retail	-1.23	1.03e-3	yes	188

Table 3: Statistical assessment variables for Glasgow global retail category

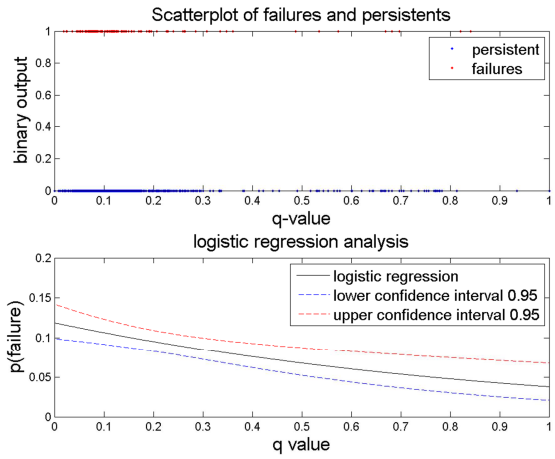

Figure 9: Logistic regression for the aggregated retail category

Validation of the Q-index for Glasgow data set

Category	**β_1**	**P-value ß1 (student stat.)**	**Accepted at 95% threshold?**	**# Bank**
Other grocery retailer	3.79	5.4e-1	no	48
Furniture retailer	-4.06	5.09e-2	no (94% ok)	25
Women, Children clothes retail	-0.80	3.4e-1	no	24
General clothing retail	-1.75	7.93e-2	No (90% ok)	18

Table 4: Statistical assessment variables for Glasgow

Note: the 4 logistic regression plots are reproduced in appendices C.

Community structure analysis

The communities discovered by the algorithm (figure 9) well fit field observations. Group 1 is composed of activities typical from dense urban centers: retail, restaurants, engineers and architects, and human health. These categories require high customer density because profitability depends on a critical number of consumers but not on characteristics of the physical space: a shop or a restaurant can be operated in very standard buildings. They also benefit a lot from the presence of attractive neighbors as described in (Jensen, Boisson et al. 2005). Group 2, is typical of suburban activities: construction, wholesale, accommodation, car shops, etc...

Group 3 is composed of 82% of agricultural activities. This is clearly a group in itself, with a high dispersion tendency and strongly dependent on the underlying physical space (room left, soils, exposition, slope, etc.). The presence of the Sport and Leisure category in this group emphasizes the need for space of the concern activities and their tendency to disperse spatially. Group 4 mostly contains activities that are not dependent on physical space constraints or on the presence of a critical consumer mass within a short distance and look first for large offices at low prices.

The spatial distribution of the points reveals their group appartenance (figure 9) and emphasizes the underlying spatial relations: Group 4 always appears at the suburbs while Group 1 matches the dense urban center. Group 2 is located in-between. Group 3 is spread differently from other for the upper-described because of its need for large surfaces. These results confirm the idea, that location alone is able to reveal much of the interaction across categories.

Valais

group: 1	**%**	**group: 2**	**%**	**group: 3**	**%**	**group: 4**	**%**
Retail	23	Construction	29	Agriculture	82	Banks	19
Restaurants & pubs	15	Accommodation	13	Post service	3	Informatics	19
Human Health	8	Car shops	12	Sport and Leisure	3	Social services	16
Wholesale	7	Associations	92	Insurances	2	Other technical activities	15
Engineers & Arch.	7	Transportation	7	Printing services	1	Other financial services	13
Real estate	5	Building services	6	Work agencies	1	Breeding	7

Table 5: Groups formed by the Louvain method on noga-2 for 2011 global data set

Map of the groups formed by the Louvain method applied to the M-index for Sion

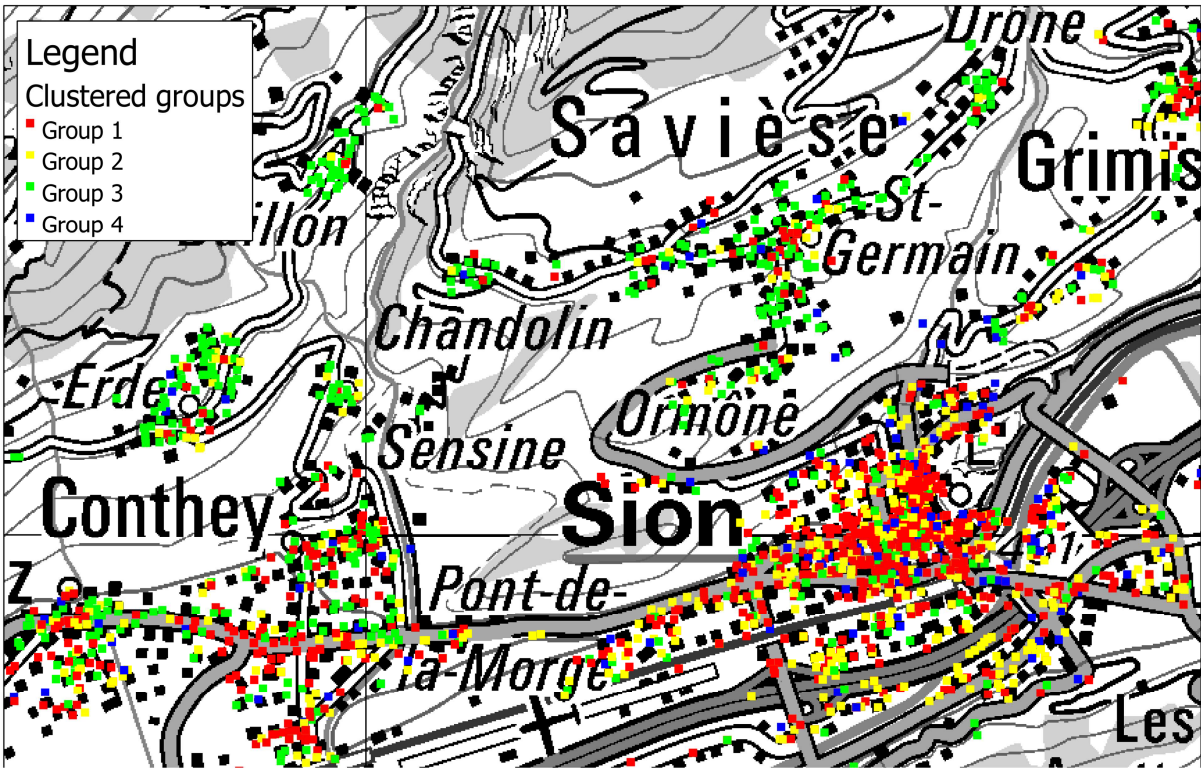

Figure 10: Map of the clustered categories from the Louvain method

Variance analysis

We attempted to reproduce the variance analysis of the intra coefficient (Jensen 2009; Jensen and Michel 2009) for the City of Sion. Our results differ in the following way. First, we only obtain values from a 150m radius as this is the smallest distance observed between 2 bakeries in Sion.

Remind that hypotheses are the following (Jensen and Michel 2009):

- H_0^A: the location sites A are purely random

And the testing procedure:

- If $|a_{AA} - 1| > q_\alpha^{AA}$ reject hypothesis H_0^A

Figure 10 shows that no bakeries are located in Sion at a distance smaller than 170m. After 300m, the curve bends to 1. This means that the spatial pattern approaches a random one. The variance calculated here is empirical and do not correspond to the one defined by Jensen (2009). Here the variance reduces as the radius size increases.

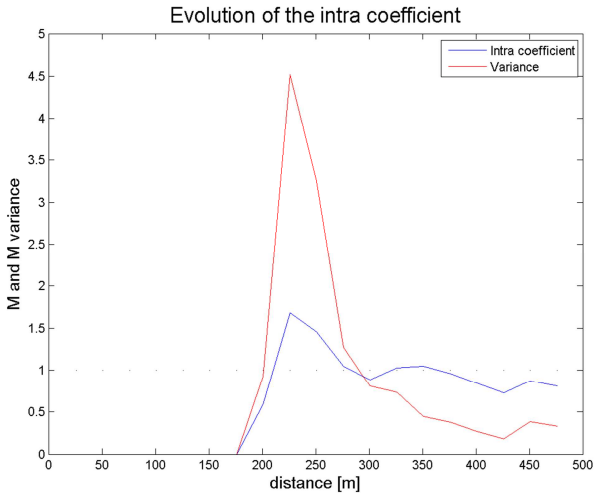

Figure 11: M-index variance analysis for bakeries in Sion

Illustrating the heterogeneity of space

Figure 11 to 14 plot the number of neighbors for each point in the original data set for 4 distinct categories in the city of Sion: retail, pubs and restaurants, architect and engineers, human health. The size of the points is proportional to the number of neighbors. These maps highlight the space's heterogeneity or in other words the non-stationarity of the density of points in these different categories.

The spatial distribution of the density of engineers and architects appears to be homogenous over the territory. On the contrary, retail, pubs and restaurants and human health clearly activities show different local regimes.

Using distance-based density estimator is a very fast way to get an initial picture of the economic landscape and is a good entry point to the M and Q-index study. Here we clearly see that variations of local concentration are important.

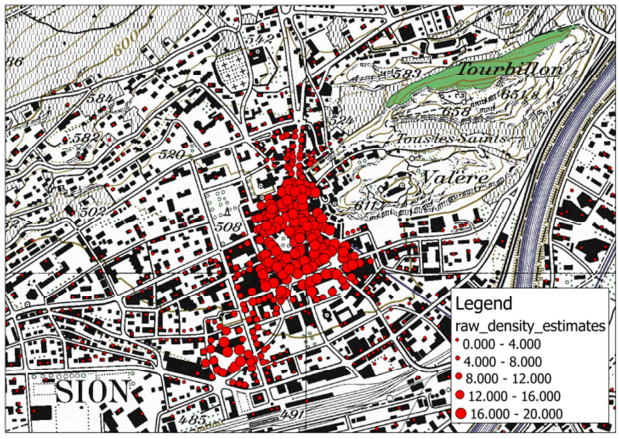

Figure 12: Raw density for retail

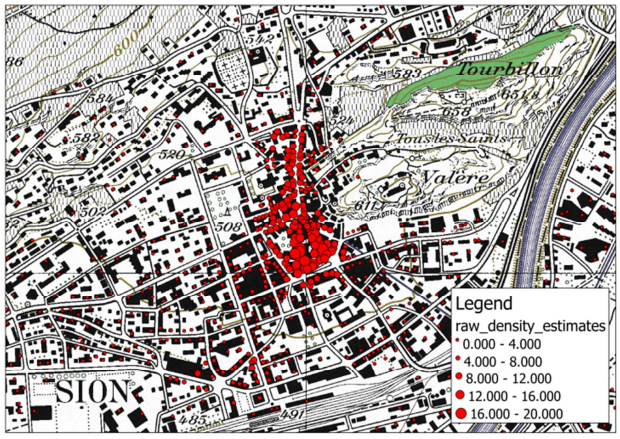

Figure 13: Raw density for pubs and restaurants

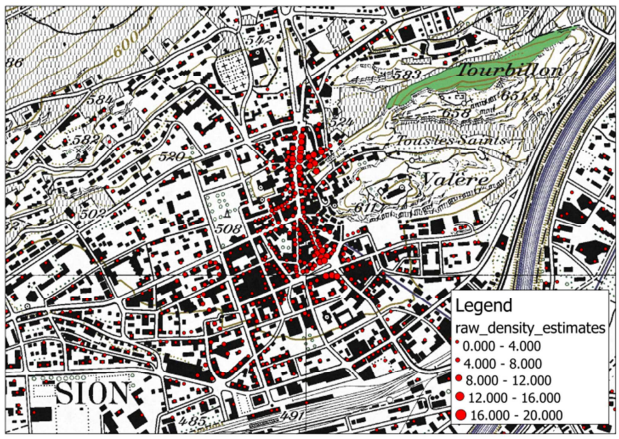

Figure 14: Raw density for engineers and architects

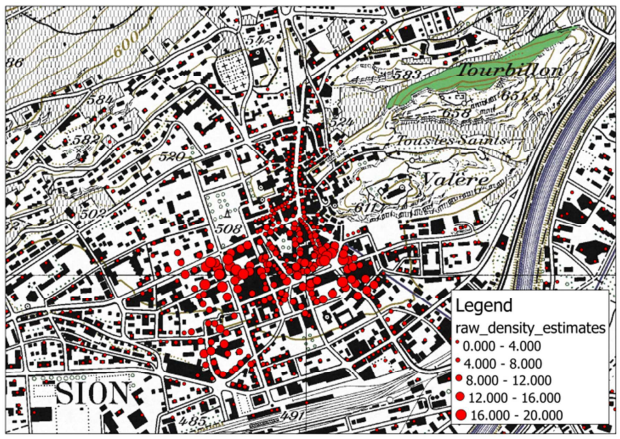

Figure 15: Raw density for human health

5 Discussion

While the Canton of Valais provided data usually very difficult to obtain, the geographic characteristics of the study area significantly differs from big and dense cities like Lyon, Geneva or Barcelona analyzed by Jensen (2006) and Stauffer (2009). In the Canton of Valais, we have a longitudinal territory dotted with multiple relatively small centers (Sion: 3000 activities). These centers are the 6 major towns: Sion, Monthey, Martigny, Sierre, Viège and Brig.

The main problem caused by this spatial configuration is the existence of many regions with very low densities of points. To counter this drawback, we implemented a density filter able to remove points in the global data set showing less than N neighbors. The results for the Canton of Valais were obtained with a value of N = 1, meaning that only points showing more than 2 neighbors were taken into consideration. Plotting the point pattern before and after density filter revealed that even such a modest filtering produced a noticeable effect. This took the data set closer to a standard larger city's configuration.

The validation results obtained are satisfactory, but are not as good as expected (the signal obtained is not as clear as expected). The main issue encountered with the present validation approach is about the retrieval of bankruptcy cases. For a reason that remains unclear, state agencies keeping track of bankruptcies are not coordinated with agencies whose task is to collect data about economic activities. They often don't share common identifiers what makes data handling very time-consuming. A long data formatting process was required to produce a usable data set for the Canton of Valais. As for the Glasgow data set, a few doubts subsist regarding the robustness of bankruptcy information deducted from the 2 yearly states available. In particular, whether the "vacant shop" category really represents a surrogate to bankruptcies remains discutable.

Since 2006, Jensen's research on the Q-index has been implemented in software used by the *Chambre du Commerce et de l'Industrie (CCI*) of the city of Lyon. The application is named LoCo and reads location data of stores to provide in a few seconds the top quality regions. LoCo helps retailers to find good locations. Unfortunately, this software source code is not open. We therefore developed an alternative free solution named "PointPatternAnalysis (PPA)" within the Matlab Computing framework, what is likely to simplify further developments. *PPA* offers much functionality: density filtering, validation procedure, community structure analysis, distance-based density estimator, M-index variance analysis and interaction with GIS software thanks to ESRI shapefile output format. Among these functionalities, the most important is the validation procedure: assuming bankruptcy data do exist, the user not only gets a Q-index map but also estimates the associated predicting capability. Knowing the quality of the output is as important as the output itself. Moreover, thanks to the Matlab Compiler ability to deploy .NET components and c-shared libraries, integration with GIS is conceivable in the short term. The code structure is detailed in appendices D.

As geographic information specialists, we favored the development of features usually encountered in the GIS world. On the one hand, we developed the use of geo-visualization tools, very helpful to apprehend complex spatial data sets. On the other hand, we programmed a by-product able to plot for each point of the global data set the number of neighbors for a given category. This tool makes it possible to show the spatial behavior of the interaction between 2 activities (or of the intra-interaction case), and to highlight possible different **local** regimes.

Indeed, the M-index **globally** quantifies the interaction between 2 commercial activities, but is not able to differentiate local regimes. The solution proposed here permits to identify the existence of local regimes, but optimally further research is required to process both global and local M-indices.

This add-on well complements the M-index variance analysis proposed by Jensen and Michel (Jensen and Michel 2009). Indeed, one could state that the specific point where the variance curve crosses the M-index curve (see figure 10) is the distance beyond which we move from a local to a global scale for 1 (intra) or two commercial categories. For the city of Sion, this occurs with a radius of 300m what could correspond approximately to the extent of the densest part (the very heart) of the city with the highest density of retail shops, restaurants and pubs.

6 Conclusion

The Q-index approach proposed by Jensen (2006) and based on the M cumulative function (Marcon and Puech 2010) rests on a single but powerful idea: location alone represents aggregated information about many characteristics of the environment of a given economic activity. This model requires a minimum of attribute data (ID, longitude, latitude, commercial category).

In this study, using bankruptcy data provided by the Registrar of companies of the State of Valais in Switzerland and by the City Council of Glasgow in Scotland, we showed that the Q-index is reliable. Although significance tests did not reach stringent levels, most of regression models computed for the different commercial categories under study reveal higher probabilities of bankruptcies where Q-index values are low. The model works especially well for the NOGA2 « retail » and « restaurants and pubs » categories with all indicators showing very significant values.

There is much room left for further investigation about M index, Q index and community structure analysis. It would consist in carrying out more tests on the Q-index with very large data sets (hundreds of thousands of points) on cities of international importance.

Acknowledgements

I want here to thanks all the enthusiast and helpful people who in a way or another helped me during this master project:

- Dr. Stéphane Joost for the wise and lucid advises and his constant presence
- Albert Gaspoz from the Canton of Valais administration for the warm welcome in *Business Valais* and his help in understanding of the geo-economic context of the Canton of Valais
- All the people of LaSIG who answered nicely to my questions and helped me bringing the Q-index maps to GeoClip plateform.
- And of course all friends and family members who have been supporting in any manner my studies at EPFL.

7 References

Blondel, V. D., J. L. Guillaume, et al. (2008), *Fast unfolding of communities in large networks*, Journal of Statistical Mechanics-Theory and Experiment(10).

Boustedt , O., Ranz (1957), *Regionale Struktur- und Wirtschaftsforschung***,** Aufgaben und Methoden Breinen: Walter Dorn.

Combes, P.-P. (2004), *The spatial distribution of economic activities in the European Union*, Handbook of Urban and Regional Economics, Elsevier. **4**.

Dunn, E. S. (1954), *The Location of Agricultural Production***,** Gainesville, University of Florida Press.

Garrison, W. L. and D. F. Marble (1957), *The spatial structure of agricultural activities*, Annals of the Association of American Geographers 47(2): 137-144.

Girvan, M. and M. E. J. Newman (2002), *Community structure in social and biological networks*, Proceedings of the National Academy of Sciences of the United States of America 99(12): 7821-7826.

Greenhut, M. L. (1956), *Plant Location in Theory and Practice: The Economics of Space*, Chapel Hill, The University of North Carolina Press.

Hamacher, H. W. and S. Nickel (1998), *Classification of location models*, Location Science(6): 13.

Hosmer, D. W. and S. Lemeshow (2000), *Applied Logistic Regression,* New York, John Wiley & sons INC.

Isard, W. (1956), *Location and Space Economy*, New York, John Wiley and Sons.

Jensen, P. (2006), *Network-based predictions of retail store commercial categories and optimal locations*, Physical Review E 74(3): 4.

Jensen, P. (2009), *Analyzing the Localization of Retail Stores with Complex Systems Tools*, Advances in Intelligent Data Analysis Viii, Proceedings 5772: 10-20.

Jensen, P., J. Boisson, et al. (2005), *Aggregation of retail stores*, Physica a-Statistical Mechanics and Its Applications 351(2-4): 551-570.

Jensen, P. and J. Michel (2009), *Measuring spatial dispersion: exact results on the variance of random spatial distributions*, The Annals of Regional Science: 30.

Marcon, E. and F. Puech (2010), *Measures of the geographic concentration of industries: improving distance-based methods*, Journal of Economic Geography 10(5): 745-762.

Newman, M. E. J. (2006), *Modularity and community structure in networks*, Proceedings of the National Academy of Sciences of the United States of America 103(23): 8577-8582.

Omlin, S. (2010), *Analyzing the localization of language features with Complex Systems Tools and predicting language vitality*, International Conference "Cognitive Modelling in Linguistics" cml-2010 Dubrovnik, Croatia 11.

Openshaw, S. (1984). *The modifiable areal unit problem*, GeoBook.

Owen, S. H. and M. S. Daskin (1998), *Strategic Facility Location: A Review*, European Journal of Operational Research(111): 24.

Park, J. and M. E. J. Newman (2004), *Statistical mechanics of networks*, Physical Review E 70(6).

Ponsard, C. (1955), *Economie et Espace: Essai d'induction du facteur spatial dans l'analyse économique*, Paris.

Revelle, C. S. and H. A. Eiselt (2005), *Location analysis: A synthesis and survey*, European

Journal of Operational Research - EJOR 165(1): 20.
Revelle, C. S., H. A. Eiselt, et al. (2008), *A bibliography for some fundamental problem categories in discrete location science*, European Journal of Operational Research - EJOR 184(3): 31.
Ripley, B. D. (1976), *The second-order analysis of stationary point processes*, Journal of Applied Probability 13: 13.
Ripley, B. D. (1977), *Modelling spatial patterns*, Journal of the Royal Statistical Society B 39: 40.
Silverman, B. W. (1986), *Density estimation for statistics and data analysis*, Monographs on Statistics and Applied Probability: 22.
Stauffer, J. (2009). *A network-based approach to predict commercial activities optimal location and infer urban centrali ty indices*, Master Thesis: 82.
SwissStatistics (2008). NOGA 2008, *Nomenclature générale des activités économiques*, Statistics. Neuchâtel: 36.

8 Appendices

A) Matching table between Noga codes and category text description

Noga Code	Text description
47	Retail
43	Construction
56	Restaurants and Pubs
477	Other retail in specialized shop
476	Cultural and leisure goods retail in specialized shop
475	Household goods retail in specialized shop
474	Computer and communication goods retail in specialized shop
473	Fuel retail in specialized shop
472	Food retail in specialized shop
471	Not specialized retail

A) Logistic regression plots for noga 3 level for the Canton of Valais

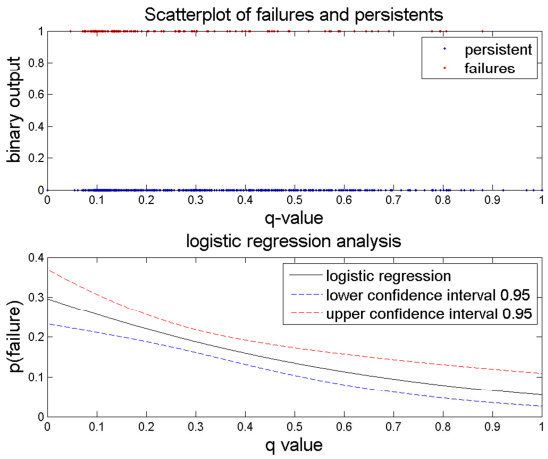

Figure 16: Logistic regr. for noga-3 cat 477, r=100m

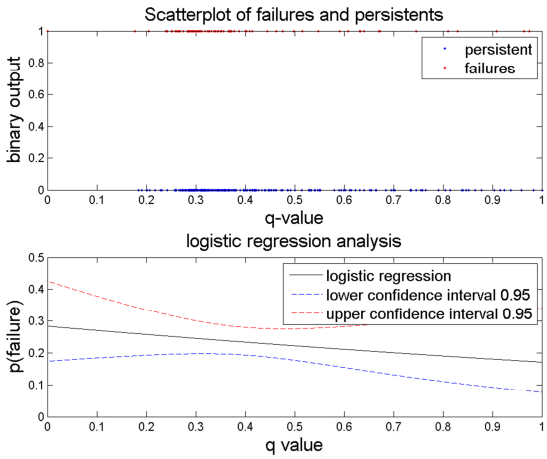

Figure 17: Logistic regr. for noga-3 cat 472, r=100m

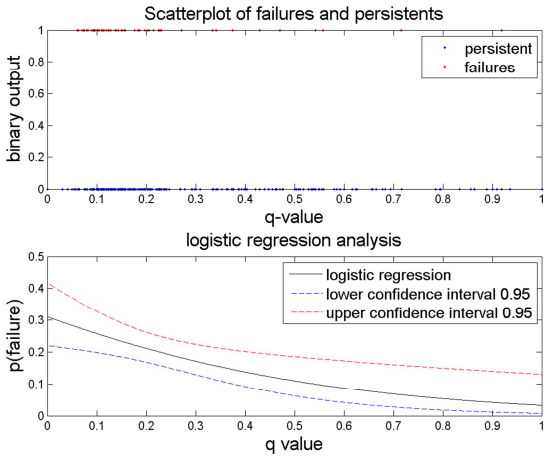

Figure 18: Logistic regr. for noga-3 cat 476, r=100m

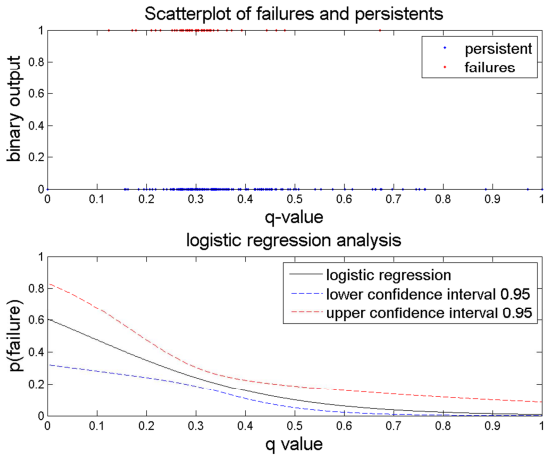

Figure 19: Logistic regr. for noga-3 cat 471, r=100m

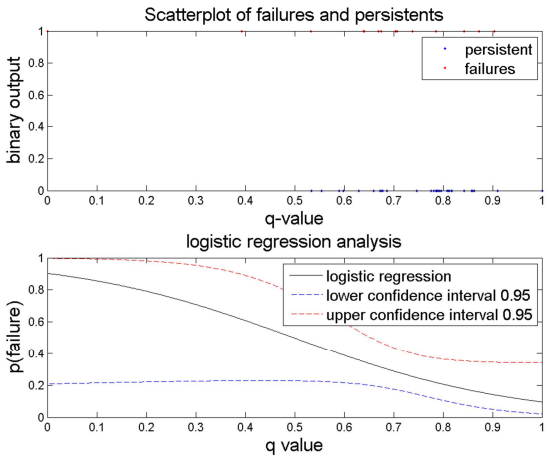

Figure 20: Logistic regr. for noga-3 cat 475, r=100m

B) Logistic regression plots for Glasgow

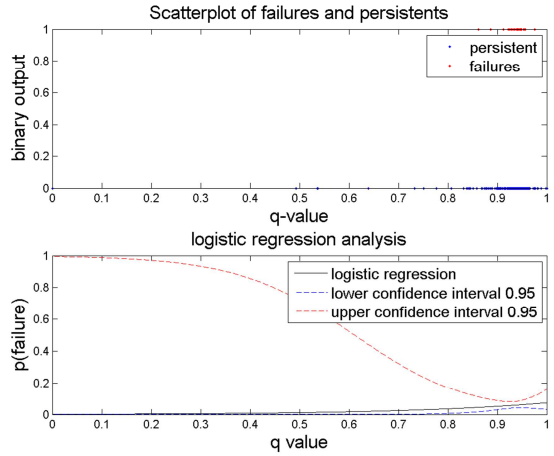

Figure 21: Logistic regression for the "other grocery retailer"

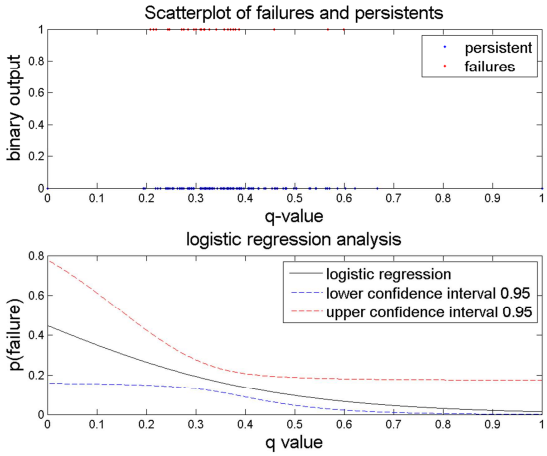

Figure 22: Logistic regression for the "Furniture retail" category

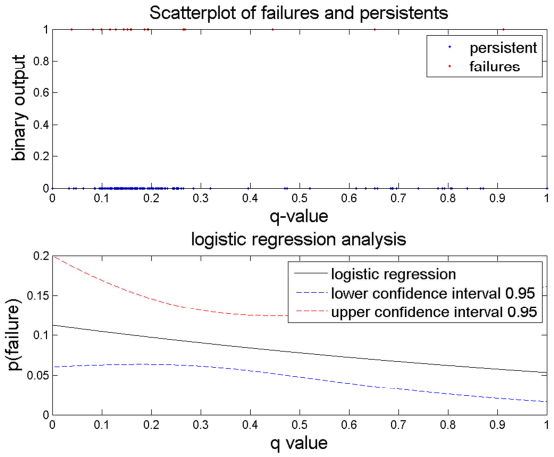

Figure 23: Logistic regression for the "Women and Child clothes retail" category

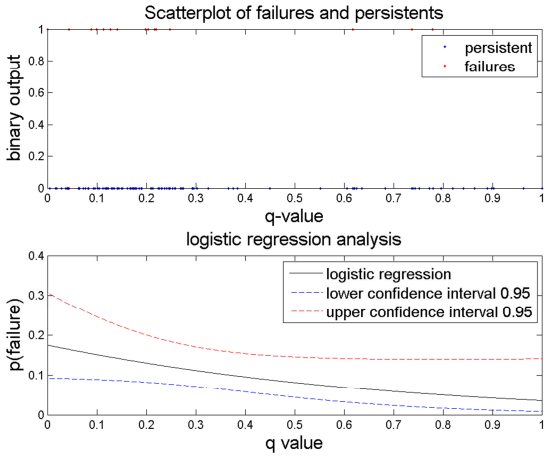

Figure 24: Logistic regression for the "General clothes retail" category

C) Technical note on software implementation

Language and coding approach

The project is coded in Matlab ® Language. This scripting language offers a good balance between an easy scripting language and computing performances. Furthermore, it is widely spread among the academic community, thus increasing the probability of further developments. Matlab also offers the possibility to deploy application using the Matlab Compiler Runtime that create multi-platform standalone executable that can be used without license for non-commercial purposes. A alternative could have been to make the developments directly inside a Geographic Information System. During pre-study phase, a script was developed for Postgres/PostGIS but it did not outperform Matlab in term of computational speed and the lack of existing statistical tool for Generalized Linear Models made this alternative less attractive. Furthermore, Matlab offers a simple but powerful tool for building GUI. Nevertheless, interfacing our project with existing GIS software is easily possible as Matlab Compiler can generate for instance c-shared libraries or .NET assemblies.

General consideration for further developments

The coding approach is of the procedural type. While it is now possible to make Object Oriented Programming (OOP) in Matlab, it usually tends to slower the computations but offers many other advantages. The project was kept as tidy as possible with rigorous functions hierarchy and minimizing memory use. Still, very large data sets (>100000 points) may lead to long computation time. A good way out of this problem could be to code the most demanding function in the c-mex language that is a low-level and very powerful language. But then, the application would requires compiling c-mex function for each computer on which it is used and is Operating System dependent.

The points layer a stored as arrays. This is not very flexible and storing is as a structure class would be a useful improvement as accessing to specific field would be less subject to mistakes. For instance, accessing to “Y” coordinates is now done using “array(:,4)”, using a structure, it would be “array.Y” , with less confusion risks and simplify the inputs/outputs, especially when using ESRI shapefile formats.

Key Features

M-index

Based on a set of geolocalized points having an identifier (ID) and an economic activity category as identifier, the application is able calculate the M-index as defined in (Jensen and Michel 2009). The only parameter is the radius.

Q-index

Using the M-index, the Q-index can be calculated for the Global points themselves or a user defined point data set in order to create a Q-index map, exported top GIS-compatible format

ESRI shapefiles. Also a complete bi- and mono-temporal validation procedure can be done in order to evaluate the predicting capability of the Q-map produced.

Community structures analysis

This feature unfolds possibly existing communities (groups of categories that obeys to similar spatial dynamics) in the M-index matrix using the Louvain Method (Blondel, Guillaume et al. 2008).A matching table showing the group affiliation of each category in the Global data set and its associated map is produced.

Density Estimator

This little tool is a distance-based density estimator that counts the number of neighbors per category around each point of the input data set and produce 2 outputs: the raw and the relative density index. The relative is raw density divided by the total number of neighbors (all categories aggregated).

Variance Analyst

This tool compare the M-index value to its variance for different radius size, allowing to check for pure randomness hypothesis (Jensen, 2009).

Required licenses

A standard Matlab license suffices to run all part of the application except the export to ESRI shapefiles what require to have the Matlab Mapping Toolbox installed. When using the deployed version, no license is required.

A small additional tool was built for developers who do not have the Mapping Toolbox license but need to export the outputs to shapefiles. This utility is csv2shp.exe application enables the user to select a .csv file containing X Y coordinates and additional numeric fields and to export them as point or polygon (squares) shapefiles. This tool is still in development phase at the time of writing and to not support string field names.

Functions procedural hierarchy: MAIN_fcn.m dependent functions

The diagram below explicit the calls between functions excluding the ones related to GUI. Calls to internal Matlab functions are not represented. The GUI basically stores user-defined parameter in variables and structures and passes them to MAIN_fcn.m function. Thus, MAIN_fcn.m function can easily turned back into an GUI-independent code by editing manually all required parameter values.

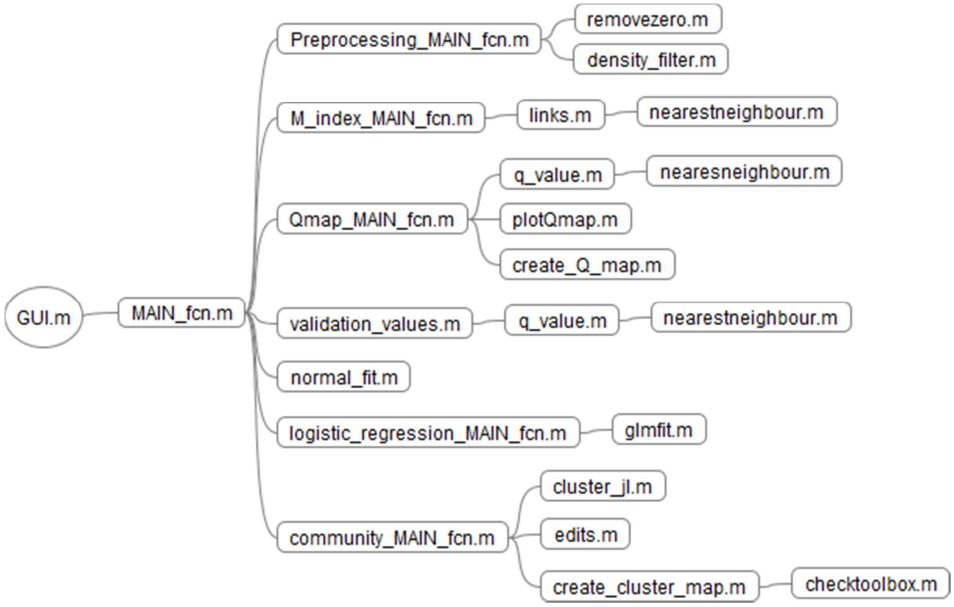

Figure 25: Procedural hierarchy for MAIN_fcn.m

Functions procedural hierarchy for Main GUI dependent functions

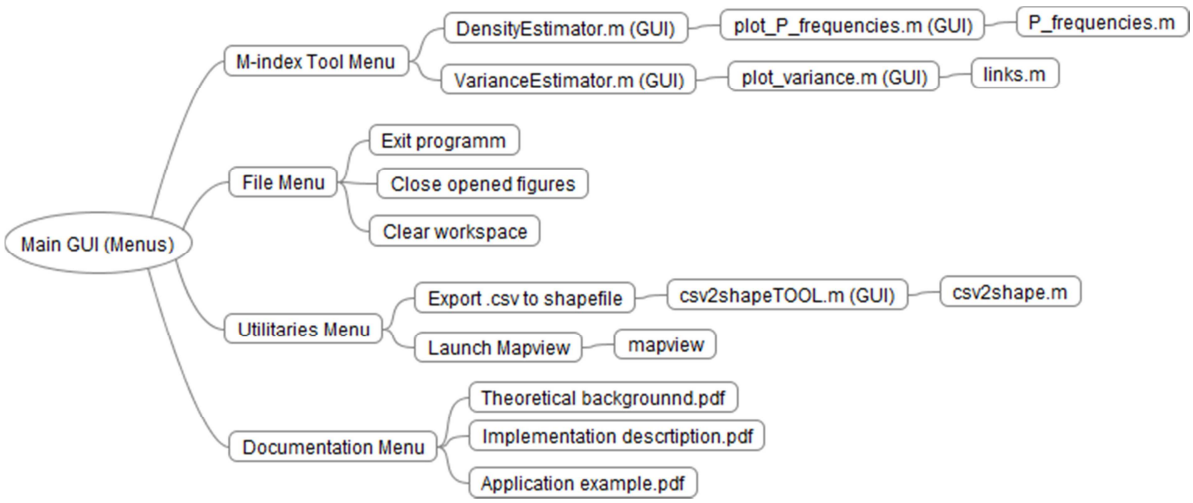

Figure 26: Procedural hierarchy for Main GUI dependent functions

Functions description
The detailed function description can be found inside the m-files which are richly commented. Here is a little abstract for the most important functions.

Name	**Description**
GUI	Base function for the GUI. Calls MAIN_fcn.
MAIN_fcn.m	Main function. Input parameters can be hard-coded there.
Preprocessing_MAIN_fcn.m	Main functions for pre-processing. Here could be added more function
remozero.m	simple function removing the points that have 0 0 cordinates. If 0 coordinates exist in you data set, you will have to modify the code of set these coordinates to 0+something
density_filter.m	density filtering of the main_layer data set. It counts the number of neighbours around a point and remove it if it has less than n neighbours.
M_index_MAIN_fcn.m	main function for M index calculations
links.m	calculates the Mindex after the Jensen (2009) Formula.
nearestneighbour.m	search for number of neighbour within radius R around the point.
Qmap_MAIN_fcn.m	calls q_value for the global data set and calls plotQmap to plot Qmap inside Maltab and create_Q_map.m to generate esri shapefiles output
validation_values.m	calculates the q values for the persistents and the failures using q_value. plots the corresponding histograms
q_value.m	calculate the Q value accorded to Monod (2011) after Omlin (2010) adpapted form Jensen (2009) ! ;-) calls nearestneighbour.m function
normal_fit.m	fit a normal distribution to the histograms of both persistents and failures and compare their mode's position
logistic_regression_MAIN_fcn	proceed logistic regression using the output of validation_values.m and edits results
community_MAIN_fnc.m	Main function for community structure analysis
cluster_jl.m	Louvain Method algorithm in Maltab language.
mode_filter.m	mode filtering of the output from the Louvain Method. Simple code that set the point value to the most frequent of the other points arount it including itself
edits.m	Edits for the clustering results. Graphical representations requires some reclass operation

Application folders and functions location

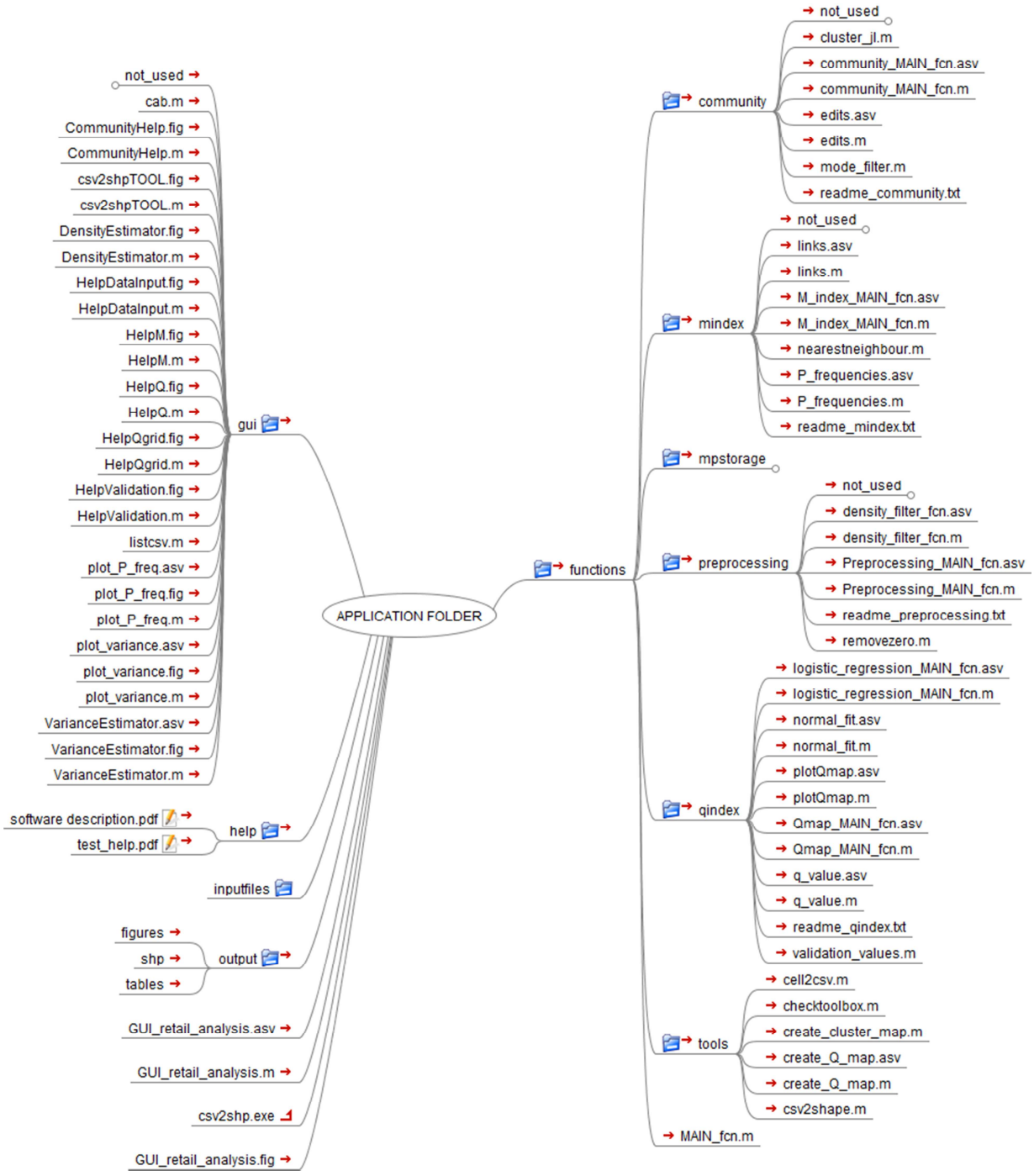

Figure 27: Application folder structure

Graphical User Interface (GUI)

Running numerous simulations by modifying parameters inside code can be a cumbersome task, especially for people not too familiar to coding. Thus a Graphical User Interface was developed with detailed help menus and checks for inconsistent user choices. At startup, options panel that are not required are hidden to the user in order to enhance user-friendliness. Also, inconsistent inputs are treated as errors. The input responsible for the error is then turns to red color until modified by the user.

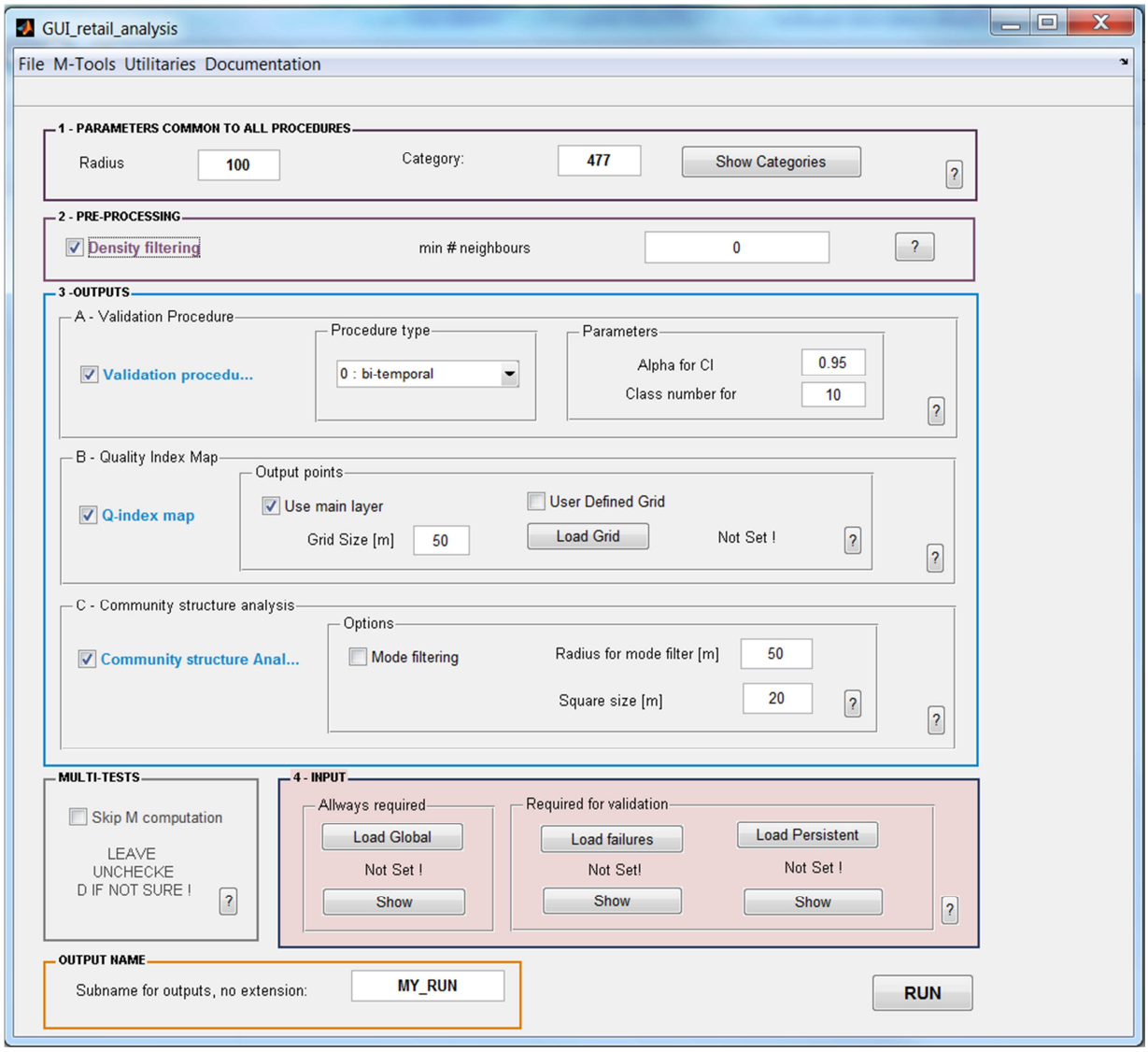

Figure 28: GUI with all options visible

Inputs and options related to GUI panels

NOTE: inputs and option related to GUI menus are pretty simple and therefore not detailed here.

Geographic Data

Variable	Name/Object in GUI	Description	Type	Format
main_layer	*Load Global* push button, panel 4	Global point layer. Array that must have the following fields in following order: Category, ID, X, Y	Category (numeric), ID (numeric), X(numeric) Y(numeric).	All ASCII formats; Matlab .mat binary, MS Excel .xls, .xlsx
disparition	*Load Failures* push button, panel 4	Failure point layer with the same structure as the global point layer	Idem	idem
remaining	*Load Persistent* push button, panel 4	Persistents point layer with the same structure as the global point layer	idem	idem
usergrid	*Load Grid*, panel 3-B	Points for which the user wants to calculate the Q-index	X(numeric) Y(numeric)	idem

Global parameters: common to all calculations: structure G

Variable	Name/Object in GUI	Description	value
G.analysis	*Skip M Computation* check box, multi-test panel	Re-compute M-index and re-apply density filter if chose by user	0: calculate everything from the beginning (recommended!) 1: use existing M-index and points from mpstorage folder
G.radius	*Radius [m]* edit text, panel 1	The search radius in meter	M-index, Q-index, Density Filter, Mode Filter
G.validation_category	*Category* edit text, panel 1	Category of interest	Any category available (numeric)
G.outputname	*Subname* panel output	Subname for the simulation	String

Options for preprocessing: structure P

Name	Name/Object in GUI	Description	values
P.density_filter	*Density filtering* check box, panel pre-processing	Apply density filter	0: no filtering 1: filtering

P.threshold	*Min # neighbors* edit text, panel pre.proc.	Number of neighbors under which the point is removed	0-n

Options for Q-index: class Q (Q-index map) and V (Validation Procedure)

Name	**Name/Object in GUI**	**Description**	**values**
Q.qmap	*Q-index map* check box, panel 3-B	Create Q-index map	0: no Q map calculated 1: Q map calculated
V.Validation	*Validation procedure* check box panel 3-A	Run validation procedure	0: no validation 1: validation (requires failures and persistents sets)
V.validation_option	*Bi-temporal* popup menu, panel 3-A	Type of validation procedure	0: multi-temporal 1: mono-temporal
V.confidence level	*Alpha for CI*, edit text, panel 3-A	Confidence level for logistic regression	[0-1]
V.nclasshisto	*Class number for histogram*, edit text, panel 3-A	Number of class for histograms	[1 – n]
Q.qmapres	*Gridsize [m]*, edit text, panel 3-B	Square polygon size for the interpolated Q-index map	Positive [m]

Options for community structure analysis: structure C

Name	**Name/Object in GUI**	**Description**	**values**
C.clustering	*Community structure anaylsis*, check box, panel 3-C	Run network clustering algorithm	0: no network clustering 1: network clustering
C.resample	*Mode Filtering*, check box, panel 3-C	Apply mode filter	0: don’t apply 1: apply
C.Resample_radius	*Radius for mode filter,* edit text, panel 3-C	Search radius for mode filtering	Positive [m]
C.clutermapres	*Square Size [m]*, edit text, panel 3C	Size of the squares for the shapefile export	Positive [m]

Input and options not accessible through GUI

Name	**Description**	**values**
C.consider_self_weight	Parmeter for the Louvain Method (cluster_jl.m)	0: don’t consider self-weights 1: consider self-weights (default)

Outputs

Outputs are generated in formats widely spread for maximal flexibility. 3 main types of outputs are generated.

- Images: .png, .fig.
- Tables: .csv
- Geographic data: esri shapefiles .shp

For the Q-index map, bilinear interpolation that is done on q values calculated for the global points layer if the user do not provide its own grid. The interpolation might produce huge output files depending of the square size chose by user and the size of the study area. To avoid crashes that may occur when writing shapefiles, the maximum number of points for shapefile export is set to 100'000. This is hard-coded in create_Q_map.m function for now. In any case, .csv file is produced.

Figure 29: Output description

Functions retrieved from Mathworks File Exchange (FEX) or other servers

nearestneighbour.m:
www.mathworks.com/matlabcentral/fileexchange/12574-nearestneighbour-m

cab.m:
Mathworks FEX,
http://www.mathworks.com/matlabcentral/fileexchange/?term=tag%3A%22cab%22

cluster_jl.m:
http://www.inma.ucl.ac.be/%7Eblondel/research/louvain.html not currently maintained

cell2csv:
mathworks FEX
http://www.mathworks.cn/matlabcentral/fileexchange/4400-cell-array-to-csv-file-cell2csv-m