ON THEORY AND METHODS

© *Investigaciones Regionales. 21 – Pages 11 to 18*
Section On theory and methods

On Some Analytical Statistics for Geographic Patterns: From Non-linearity to Linearity

Jean H. P. Paelinck *

ABSTRACT: In Getis and Paelinck (*L'Espace Géographique, 2004, No 1*) some analytical indices for geographic patterns were proposed; one of them was a Chaitin conditional complexity index, *c*, based on the observed coordinates. This index was reanalyzed, and showed a large variability as a function of those coordinates. A new index of «peakiness», *p*, is proposed, tested, and applied to French data relating to «upper» employment in 37 areas of the Rhône-Alpes region (centered around Lyons).

JEL Classification: C0, R1.

Keywords: Complexity, concentration, patterns.

Algunos estadísticos para los patrones geográficos: de lo no lineal a lo lineal

RESUMEN: Getis y Paelinck (*L'Espace Géographique, 2004, N.º 1*) proponen algunos indicadores analíticos apropiados para analizar los patrones geográficos. Uno de ellos consiste en un índice de complejidad condicionada tipo «Chaitin» (*c*) basado en las coordenadas geográficas observadas. Este índice ha sido estudiado posteriormente mostrando una gran variabilidad en función de dichas coordenadas. En este artículo se propone un nuevo índice de «*peakiness*» (*p*) y se analiza su comportamiento utilizando para ello las cotas superiores de empleo de 37 áreas de la región francesa de Rhône-Alpes (en los alrededores de Lyon).

Clasificación JEL: C0, R1.

Palabras clave: Complejidad, concentración, estructuras.

---

* *George Mason University. School of Public Policy.*
We gratefully acknowledge the cooperation of our colleagues Bernard Coutrot and Alain Sallez in providing the data of section 4.

Received: 2 february 2011 / Accepted: 27 july 2011.

1. Introduction

In Getis and Paelinck (2004) a series of analytical indices for geographic patterns were proposed, i.a. characteristics of spatial statistical distributions, concentration and dispersion indices, geophenograms, cluster indicators.

In this paper one indicator will be taken up again, to wit a conditional complexity index, of which the spatial behavior will be studied; as unexpected characteristics creep up, another index will be developed, which takes into account «peakiness» of geographical space.

Next sections will be devoted to those topics, with conclusions and references following.

2. Conditional complexity analysis

An index of conditional complexity (Chaitin, 1975; Wolfram, 2002, pp. 552 a.f.) can be defined as:

$$c = (t - 1) / (t_{max} - 1) \quad (1)$$

were t is the number of non-zero terms of a polynomial, $p(x,y)$, in coordinates x and y, and t_{max} their maximal number, in fact the number of observations; obviously $0 \leq c \leq 1$.

A special aspect of interest is the presence of a certain number of «peaks», so a case was set up with 25 points randomly scattered (Figure 1 and Table 1); the left hand column lists the abscissae, the right hand one the ordinates.

Figure 1. Coordinates

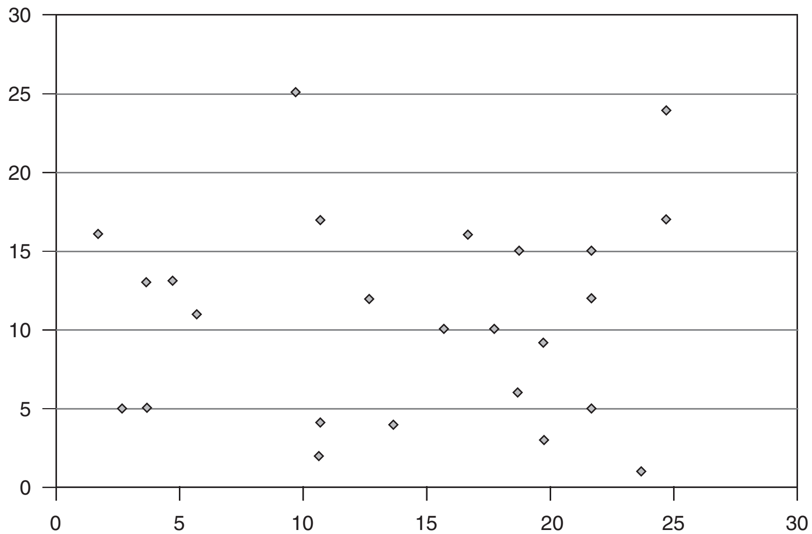

Table 1. Coordinates

x	*y*
3	5
11	2
13	12
22	15
22	5
20	3
25	24
10	25
22	12
2	16
4	13
25	17
20	9
6	11
17	16
19	6
4	5
24	1
16	10
18	10
19	15
5	13
14	4
11	4
11	17

To compute the complexity index, a system of 25 linear equations has to be solved in order to derive the polynomial coefficients; its degree here is 6. If every location gets the same mass - e.g. 1 - only the constant of the polynomial is non-zero, and $c = 0$; but if the whole mass of 25 is located in one point, the results may differ considerably, as Table 2 shows.

Table 2. Some results

Observation	*t*	*c*
1	2	0.0417
3	25	1
7	17	0.7083
8	15	0.6250
18	18	0.7500

The observation numbers correspond to the rows of Table 1. The diversity in results is due, on the one hand to the values of the coordinates, on the other to the non-linearity of the polynomial; the most curious case is that of observation 1, where a sort of discrete Dirac function could be described by only 2 polynomial terms, hence the search for a complementary indicator for «peaky» landscapes.

3. A new spatial pattern indicator

Indeed, observed spatial patterns are often non-smooth, in the sense that a few extremely high values of the variable analyzed are present at some (random) locations, amidst an overwhelming majority of relatively lower values; Figure 2 further down is an excellent illustration of this fact. It has incited to develop an appropriate indicator for such situations; the proposed indicator is the following one:

$$p = \left[\sum_i m_j^k - n(m/n)^k\right] / \left[m^k - n(m/n)^k\right] \quad (2)$$

where the m_i are different masses, m the total mass, n the number of locations, and $k > 1$, the latter characteristic allowing possibly present peaky masses to dominate smaller ones. Obviously, if the whole mass is concentrated in one point, $p = 1$; in case of a homogeneous spread, $p = 0$. For the case of 25 locations with 5 peaks of 5 units each, and $k = 1.1$, $p = 0.4599$, which still shows a fair degree of «peakiness».

Table 3 shows some «peakiness» indices for different locations, three concentrated and the last one dispersed; peakiness is generally high.

Table 3. Peakiness indices for different «peaky» situations

Observations	*p*
3, 15, 19, 20, 24	1
5, 6, 13, 16, 18	1
10, 11, 14, 22, 25	.7083
1, 7, 8, 12, 18	.7500

The logic of the indicator has been exposed above; the real test of its analytical power resides in its application to empirically observed situations, as will be done hereafter.

4. Application to French data

In Coutrot, Paelinck and Sallez (2009) a spatial econometric study was made of the dynamics of locations of «advanced» employment in urban areas of South-Eastern France, the so-called «Rhone-Alpes» region; the method used was that of potentialized partial finite difference equations. Table 4 reproduces the data, Figure 2 the regions and Table 5 hereafter gives the c- and p-values for three years.

Table 4. «Upper» urban employment of *37 French urban areas, employment in 1982, 1990 and 1999; longitudes and latitudes*

French urban areas	*1982*	*1990*	*1999*	*Long*	*Lat*
Lyon	43,364	65,004	75,935	4.84	45.75815
Grenoble	15,891	22,612	28,202	5.72046	45.18487
Saint-Etienne	5,284	7,228	8,084	4.42236	45.42559
Genève(CH)-Annemasse	981	2,640	2,933	6.24884	46.18764
Valence	3,033	4,328	5,008	4.91851	44.92424
Chambéry	2,656	3,520	4,149	5.91044	45.58507
Roanne	1,440	1,664	1,688	4.08665	46.04374
Saint-Chamond	875	996	1,192	4.51223	45.46843
Thonon-les-Bains	690	1,036	1,037	6.49760	46.37112
Romans-sur-Isère	689	816	1,114	5.04985	45.05926
Villefranche-sur-Saône	1,038	1,408	1,602	4.73478	45.98457
Cluses	925	1,392	1,626	6.58466	46.06109
Montélimar	638	840	955	4.74990	44.55462
Vienne	672	932	1,130	4.88886	45.52256
Saint-Juste-Saint-Rambert	443	688	846	4.25215	45.49209
Voiron	395	636	848	5.58996	45.38055
Sallanges	317	528	604	6.60885	45.93178
Aubenas	365	536	445	4.39640	44.60769
Aix-les-bains	522	780	855	5.91287	45.69171
Annonay	328	464	603	4.6468	45.24522
Roussillon	362	504	430	4.82239	45.38275
Albertville	274	472	471	6.41332	45.66381
Bourgon-Jallieu	497	688	735	5.27602	45.60308

Table 4. *(Continue)*

French urban areas	*1982*	*1990*	*1999*	*Long*	*Lat*
Montbrison	242	368	407	4.07652	45.60295
Privas	249	352	374	4.60051	44.7193
Tournon-sur-Rhône	204	332	270	4.81975	45.0534
Tarare	213	244	343	4.42655	45.90881
Livron-sur-Drôme	98	136	130	4.83280	44.79212
Rumilly	146	204	333	5.94732	45.85452
Saint-Marcellin	172	192	197	5.31961	45.15225
Chamonix-Mont-Blanc	91	160	192	6.93263	45.92758
Saint-Jean-de Maurienne	118	144	239	6.35145	45.27329
La-Tour-du-Pin	116	152	223	5.45322	45.57263
Pierrelatte	539	740	595	4.69467	44.36283
Feurs	167	192	176	4.23356	45.73358
Bourg-Saint-Maurice	85	132	134	6.76867	45.66473

Figure 2. The Rhone-Alpes region and its main urban centers

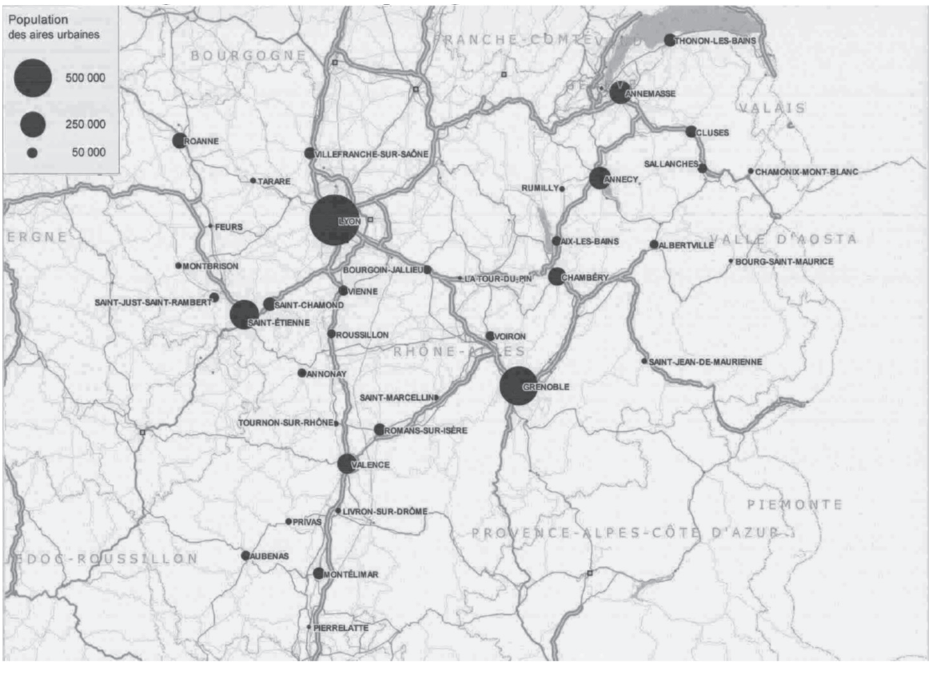

Table 5. *C*- and p-values for the French data

Years	*p*	*c*
1982	0.4411	0.9117
1990	0.4478	0.9117
1999	0.4547	0.8611

One notices a slight increase in the *p*-values, to be compared, for their order of magnitude, with the result shown at the end of section 3; as to the *c*-values, they are systematically high, with a decrease in 1999. These latter values should be compared to the results obtained in section 5 hereafter.

5. An alternative to polynomial complexity

As has been said before, results on complexity are i.a. influenced by the non-linearity of the polynomial; this leads to the idea of investigating a linear variant, and as geographical space is involved, coordinates of the own area and its nearest neighbors in decreasing order could be considered.

That this would not immediately solve the problem can be shown as follows. Translate both coordinates by a same amount, instead of a system matrix **A** one would now have to consider a system:

$$(\mathbf{A} + \lambda \mathbf{J})\mathbf{x} = \mathbf{b}^* \tag{3}$$

where *J* is a full square unit matrix; the solution to the system is now:

$$\mathbf{x} = (\mathbf{I} - \lambda \mathbf{A}^{-1}\mathbf{J}^{-1})\mathbf{A}^{-1}\,\mathbf{b}^* \tag{4}$$

which reduces to the original solution for $\lambda = 0$. So an additional transformation might be to take the values of the deviations to the center of gravity, to generate a quasi-symmetry.

The idea was applied to observations 1, 2, 3, 10, 11 of Table 1 with the following results, total mass having been allocated successively to those observations.

Table 6. Linear complexity analysis

Observation	c_l*: original data*	c_l*: deviations from c.g.*
1	1	1
2	0.6	1
3	1	1
10	1	1
11	0.6	1

One will notice that indeed the deviation method conserves the $0-1$ complexity. This phenomenon was tested on the data of table 1, and the result was the same for complete concentration in one spot, i.e $c = 1$; that result could be easily inferred from the complete inverse matrix of expression (4), which did not contain a single zero.

Applied to the French data the value of c for the successive years 1982, 1990 and 1999 was invariably 0.9444, showing a higher degree of complexity than the one obtained polynomially (Table 4), which confirms the dominant position of the three major cities: Lyons, Grenoble and Saint-Etienne.

6. Conclusions

The exercises presented concern an additional set of indicators, to be used in a geophenogram manner, as was mentioned earlier.

Linearity and non-linearity problems have been recognized at an early stage in spatial econometric analysis (see e.g. Paelinck and Klaassen, 1979, pp. 6-9); there is some parallel with the ex ante – ex post distinction (ex ante behavioral relations in spatial econometrics are more often than not non-linear, or even non-convex – see for instance the generalized Weber location problem – while ex post resulting flows – e.g. transport flows – can be modeled linearly.

In the present exercise the approach is about what can be called «analytical descriptions»; in some cases they should have a strong non-linear character (see the peakiness index of section 3), while in others linear transformations might be in order (as was the case in section 5).

The real issue is to derive a specification that matches the problem at hand; in the present study it has been tried to demonstrate how analytical descriptions can be appropriately defined as a function of the spatial patterns to be analyzed.

7. References

Chaitin, G. J. (1975): «Randomness and Mathematical Proofs», *Scientific American,* vol. 232, n.º 5, pp. 47-52.

Coutrot, B.; Paelinck, J. H. P., and Sallez, A. (2009): «On Potentialized Partial Finite Difference Equations: Analyzing the Complexity of Knowledge-Based Spatial Economic Developments», *Région et Développement,* n.º 29, pp. 237-264.

Getis, A., and Paelinck, J. H. P. (2004): «On Analytical Descriptions of Geographical Patterns», *L'Espace Géographique*, n.º 1, pp. 61-68.

Paelinck, J. H. P., and Klaassen, L. H. (1979): *Spatial Econometrics*, Saxon House, Farnborough, UK.

Wolfram, S. (2002): *A New Kind of Science,* Wolfram Media Inc., Champaign, Ill.