FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF MATHEMATICS,
STATISTICS AND ACTUARIAL SCIENCES

French Regional Wheat Prices : 1756-1872 Correlation Length

E. Borghers

RESEARCH PAPER 2006-008
February 2006

University of Antwerp, Prinsstraat 13, B-2000 ANTWERP, Belgium
Research Administration - Room B.213
phone: (32) 3 220 40 32
e-mail: joeri.nys@ua.ac.be

D/2006/1169/008

French Regional Wheat Prices : 1756-1872
Correlation Length

S'ils n'ont pas de pain, qu'ils mangent des brioches
Marie-Antoinette

Prof. dr. E. Borghers

University of Antwerp
Faculty of Applied Economics
Department of Mathematics, Statistics and Actuarial Sciences

Key Words : Multivariate Statistics, Correlation Length, Wheat Prices, France

Acknowledgement

The helpful comments and suggestions made by dr. G. Brys are greatly appreciated.

Corresponding Address

University of Antwerp
Faculty of Applied Economics
Room C 442
Prinsstraat 13
B 2000 Antwerp - Belgium

eddy.borghers@ua.ac.be
T +32 (0) 3 220 41 31 (Office)
T +32 (0) 3 220 41 54 (Secretary)
F +32 (0) 3 220 48 57 (Secretary)

Contents

Section 1 : Introduction

In a series of previous papers different aspects of the regional price behavior of wheat on the French market were investigated. The derived results, presented in **Borghers [1, 2, 3, 4 and 5]**, are based on aggregated as well as disaggregated data and are covering the last decades of the *Ancien Régime*, i.e. the period 1756-1790, and the period immediately following the French revolution, i.e. the period 1806-1872.

One of the main and recurrent results from these analyses is the consistent, regular but changing regional patterns in the price behavior of wheat. Furthermore these findings seem to be independent from the specific statistical techniques and measures used in the analysis. It is evident that the nature of these regularities are a great help in better understanding the internal structure of the French wheat market throughout the periods under investigation.

The main purpose of this paper is to investigate whether or not these regional differences could be 'explained' by the geographical distance between these regions. In other words the relationship between the wheat price fluctuations in different markets will be analyzed with respect to the distance between those markets. The basic hypothesis is that the wheat price relationship, measured by the correlation coefficient, will decrease as a function of the growing distance. The dependency of the correlation on distance is known as correlation length. In this respect the correlation length can be seen as an important measure of the level of market integration.

This paper is following a simple structure. In Section 2 some preliminary information will be given about the geographical locations used in the analysis. Once these locations are defined the geographical distances can be derived. An analysis of these distances will be the subject of Section 3. In Section 4 and Section 5 different aspects about the correlation length will be discussed in detail. In the final Section 6 the summarized results will be confronted with some results found in the literature.

Section 2 : Geographical Locations

The phenomenon of a decreased influence due to the growing distance is generally known. It is sometimes called 'the first law of geography' and is based on the idea that everything is related to everything else, but near things are more related than more distant things (**Haggett et al. [7]**).

Prior to investigate the influence of the geographical distance on the local price behavior two problems ought to be solved. The first and main problem is to get the necessary data to describe the geographical distance between the locations under consideration. The second and related problem is that each of the *généralités* and the *départements* should have a precise point of reference in order to enable the recording of these distances.

To start with the second problem it turns out that 14 out of the 33 *généralités* have been given the name of a region rather than the name of the principal town or *chef-lieu de généralités*. As a consequence these 14 *généralités* were renamed with the new name referring to a specific location within the *généralité*.

However this necessarily and inevitable pragmatic simplification cannot be used for the data for Paris and vicinity. The available data are labeled as *'Généralité de Paris'* and *'Paris Ville'*. Whereas the geographical location for *Paris Ville* is not a real problem it is still unclear how and where to allocate the prices for the *généralité de Paris*. Since it is hard to find an objective criterion to allocate the price data for the *généralité de Paris* to a specific geographical location, expressed in longitude and latitude, it was decided to exclude the data for this *généralité* from further analysis and to proceed with only the prices for *Paris Ville.*

The renaming of the 14 *généralités* can be found in Table 1. The list of the 33 *généralités* and further detailed information about this data set is given in Appendix 1 and 2.

Table 1 : Geographical Location - Généralités - Villes - Period : 1756-1790
Longitude and Latitude

			Coordinates (2)	
Généralité (1)		**Ville**	**Lat.**	**Long.**
1.	Alençon	Alençon	48.417	0.083
2.	Alsace	Strasbourg	48.583	7.750
3.	Amiens	Amiens	49.900	2.300
4.	Auch	Auch	43.500	0.600
5.	Bayonne	Bayonne	43.480	-1.480
6.	Bordeaux	Bordeaux	44.833	-0.567
7.	Bourges	Bourges	47.083	2.383
8.	Bourgogne	Dijon	47.333	5.033
9.	Bretagne	Rennes	48.100	-1.667
10.	Caen	Caen	49.183	-0.367
11.	Champagne	Châlons	48.967	4.367
12.	Flandres	Lille	50.650	3.083
13.	Franche-Comté	Besançon	47.233	6.200
14.	Grenoble	Grenoble	45.183	5.717
15.	Hainaut	Valenciennes	50.350	3.530
16.	Languedoc	Montpellier	43.600	3.883
17.	La Rochelle	La Rochelle	46.167	-1.167
18.	Limoges	Limoges	45.833	1.250
19.	Lorraine	Nancy	48.700	6.200
20.	Lyon et Dombes	Lyon	45.767	4.833
21.	Metz	Metz	49.117	6.183
22.	Montauban	Montauban	44.017	1.333
23.	Moulins	Moulins	46.567	3.333
24.	Orléans	Orléans	47.900	1.900
25.	Paris			
26.	Poitiers	Poitiers	46.583	0.333
27.	Provence	Aix-en-Provence	43.530	5.430
28.	Riom-Auvergne	Riom	45.900	3.120
29.	Rouen	Rouen	49.433	1.083
30.	Roussillon	Perpignan	42.700	2.900
31.	Soissons	Soissons	49.617	3.550
32.	Tours	Tours	47.383	0.700
33.	Paris Ville	Paris Ville	48.867	2.333

(1) For the 14 entries with gray background the généralités are allocated at the chef-lieu de généralités.
(2) Longitude expressed as deviation from Greenwich Prime Meridian in decimal degrees.

No comparable problems were encountered with the allocation of the available price data for the period 1806-1872. The data for each of the *départements* were allocated to the main town of the *département*, i.e. the capital city of the *département* bearing the title of *préfecture*. The only problem with the available data for the period 1806-1872 is that from the original available 90 *départements* only 85 series could be retained for further analysis. More information about the price data for this period can be found in Appendix 1 and 3. The 85 remaining *départements* that will be used are tabulated in Table 2.

Table 2 : Geographical Location - Départements - Villes - Period : 1806-1872
Longitude and Latitude

			Coordinates (°)	
Département		**Ville**	**Lat.**	**Long.**
01.	Ain	Bourg-en-Bresse	46.20	5.22
02.	Aisne	Laon	49.57	3.62
03.	Allier	Moulins	46.57	3.33
04.	Alpes de Hautes-Prov.	Digne	44.08	6.23
05.	Hautes-Alpes	Gap	44.55	6.08
07.	Ardèche	Privas	44.73	4.60
08.	Ardennes	Charleville	49.77	4.73
09.	Ariège	Foix	42.95	1.58
10.	Aube	Troyes	48.30	4.08
11.	Aude	Carcassonne	43.22	2.35
12.	Aveyron	Rodez	44.35	2.57
13.	Bouches-du-Rhône	Marseille	43.30	5.37
14	Calvados	Caen	49.18	-0.37
15.	Cantal	Aurillac	44.93	2.43
16.	Charente	Angoulême	45.67	0.17
17.	Charente-Maritime	La Rochelle	46.17	-1.17
18.	Cher	Bourges	47.08	2.38
19.	Corrèze	Tulle	45.27	1.77
21.	Côte-d'Or	Dijon	47.33	5.03
22.	Côtes d'Armor	Saint-Brieuc	48.52	-2.75
23.	Creuse	Guéret	46.17	1.87
24.	Dordogne	Périgueux	45.20	0.73
25.	Doubs	Besançon	47.23	6.20
26.	Drôme	Valence	44.93	4.90
27.	Eure	Évreux	49.05	1.18
28.	Eure-et-Loir	Chartres	48.45	1.50
29.	Finistère	Quimper	48.00	-4.10
30.	Gard	Nîmes	43.83	4.35
31.	Haute-Garonne	Toulouse	43.62	1.45
32.	Gers	Auch	43.50	0.60
33.	Gironde	Bordeaux	44.83	-0.57
34.	Hérault	Montpellier	43.60	3.88
35.	Ille-et-Vilaine	Rennes	48.10	-1.67
36.	Indre	Châteauroux	46.82	1.68
37.	Indre-et-Loire	Tours	47.38	0.70
38	Isère	Grenoble	45.18	5.72
39.	Jura	Lons-le-Saunier	46.68	5.55
40.	Landes	Mont-de-Marsan	43.90	-0.50
41.	Loir-et-Cher	Blois	47.60	1.33
42.	Loire	Saint-Étienne	45.43	4.38
43.	Haute-Loire	Le Puy-en-Velay	45.05	3.88
44.	Loire-Atlantique	Nantes	47.23	-1.58
45.	Loiret	Orléans	47.90	1.90
46.	Lot	Cahors	44.47	0.43
47.	Lot-et-Garonne	Agen	44.20	0.63
48.	Lozère	Mende	44.53	3.50
49.	Maine-et-Loire	Angers	47.48	-0.53
50.	Manche	Saint-Lô	49.12	-1.08
51.	Marne	Châlons	48.97	4.37
52.	Haute-Marne	Chaumont	48.12	5.13
53.	Mayenne	Laval	48.07	-0.75
54.	Meurthe-et-Moselle	Nancy	48.70	6.20
55.	Meuse	Bar-le-Duc	48.77	5.17
56.	Morbihan	Vannes	47.67	-2.73
57.	Moselle	Metz	49.12	6.18
58.	Nièvre	Nevers	47.00	3.15
59.	Nord	Lille	50.65	3.08
60.	Oise	Beauvais	49.43	2.08
61.	Orne	Alençon	48.42	0.08
62.	Pas-de-Calais	Arras	50.28	2.77
63.	Puy-de-Dôme	Clermont-Ferrand	45.78	3.08
64.	Pyrénées-Atlantiques	Pau	43.30	-0.37
65.	Hautes-Pyrénées	Tarbes	43.23	0.08
66.	Pyrénées-Orientales	Perpignan	42.70	2.90
67.	Bas-Rhin	Strasbourg	48.58	7.75
68.	Haut-Rhin	Colmar	48.08	7.35
69.	Rhône	Lyon	45.77	4.83
70.	Haute-Saône	Vesoul	47.63	6.15
71.	Saône-et-Loire	Mâcon	46.30	4.83
72.	Sarthe	Le Mans	48.00	0.20
75.	Paris	Paris	48.87	2.33
76.	Seine-Maritime	Rouen	49.43	1.08
77.	Seine-et-Marne	Melun	48.53	2.67
78.	Yvelines	Versailles	48.80	2.13
79.	Deux-Sèvres	Niort	46.32	-0.45
80.	Somme	Amiens	49.90	2.30
81.	Tarn	Albi	43.93	2.13
82.	Tarn-et-Garonne	Montauban	44.02	1.33
83.	Var	Toulon	43.12	5.92
84.	Vaucluse	Avignon	43.93	4.80
85.	Vendée	La Roche-sur-Yon	46.63	-1.50
86.	Vienne	Poitiers	46.58	0.33
87.	Haute-Vienne	Limoges	45.83	1.25
88.	Vosges	Épinal	48.17	6.47
89.	Yonne	Auxerre	47.80	3.58

(°) Longitude expressed as deviation from Greenwich Prime Meridian in decimal degrees.

Section 3 : Geographical Distances

From Locations to Distances

The next and final step in collecting the necessarily data to enable a further spatial analysis of the influence of distance on the local wheat prices is defining the distance between the locations where the data are assumed to have been collected. Defining these distances is however not evident at all. It is totally impossible to reconstruct these distances as they were experienced in the periods under investigation. It follows that whatever the solution might be the final result will always be an approximation.

One of the options could have been to use the distances of the default type of routes given by **Michelin [16]**. These recommended routes are a compromise between time and distance. A longer journey, i.e. a longer distance, is accepted only if it provides a sufficient time saving. An alternative and even more pragmatic solution is the use of distances as a bird's-eye view. Each of the distances can be calculated given the latitude and longitude of the two locations involved. The calculations can be implemented by using the information given in references **[14]**,**[15]**and **[17]**. Being aware of the crucial role that will be played by the chosen geographical distances the potential capabilities of those distances will be investigated first.

Once all the distances are available they can be arranged in a distance matrix. Each element d_{ij} of a distance matrix represents the distance in km between location i and j. The dimension of these matrices are 32×32 for the *généralités* (1756-1790) and 85×85 for the *départements* (1806-1872) resulting in respectively 1024 and 7225 matrix elements. For at least two reasons a distance matrix can be compared with a correlation matrix. First of all a distance matrix is symmetric, i.e. the distance between region i and j is exactly the same as the distance between region j and region i. A second reason is based on the results obtained for the main diagonal. Whereas the diagonal of the price correlation matrix consists of one's, i.e. the price correlation for a given region with itself is one, the main diagonal of a distance matrix will consist of zero's, i.e. the distance between a given region and itself is zero. It follows that, just as for the price correlation matrix, not all the elements of the distance matrix need to be investigated.

Preliminary Analysis of Distances Data

In the next paragraphs particular attention will be paid to the distances that will be used for the spatial analysis of the regional wheat prices. A first aspect is the comparison of the general descriptive statistics obtained for distances for the *généralités* and the *départements*. These general statistics are tabulated in Table 3.

Table 3 : Distances - Généralités and Départements - General Statistics

General Statistics	Généralités	Départements
Number of Observations	496	3570
Mean	405.7568	392.6574
Confidence Interval -95%	389.0755	386.5993
Confidence Interval +95%	422.4381	398.7155
Minimum	45.9150	16.3985
Maximum	909.2321	948.4740
Range	863.3171	932.0756
Variance	35753.6100	34084.0800
Standard Deviation	189.0862	184.6187
Standard Error	8.4902	3.0899
Skewness Value	0.3459	0.2226
Standard Error	0.1097	0.0410
Kurtosis Value	-0.6003	-0.6853
Standard Error	0.2189	0.0819

Figure 1 : Distances versus Longitude and Latitude - Généralités
Central and Outermost Locations - Quadratic Surface - Contour Plot

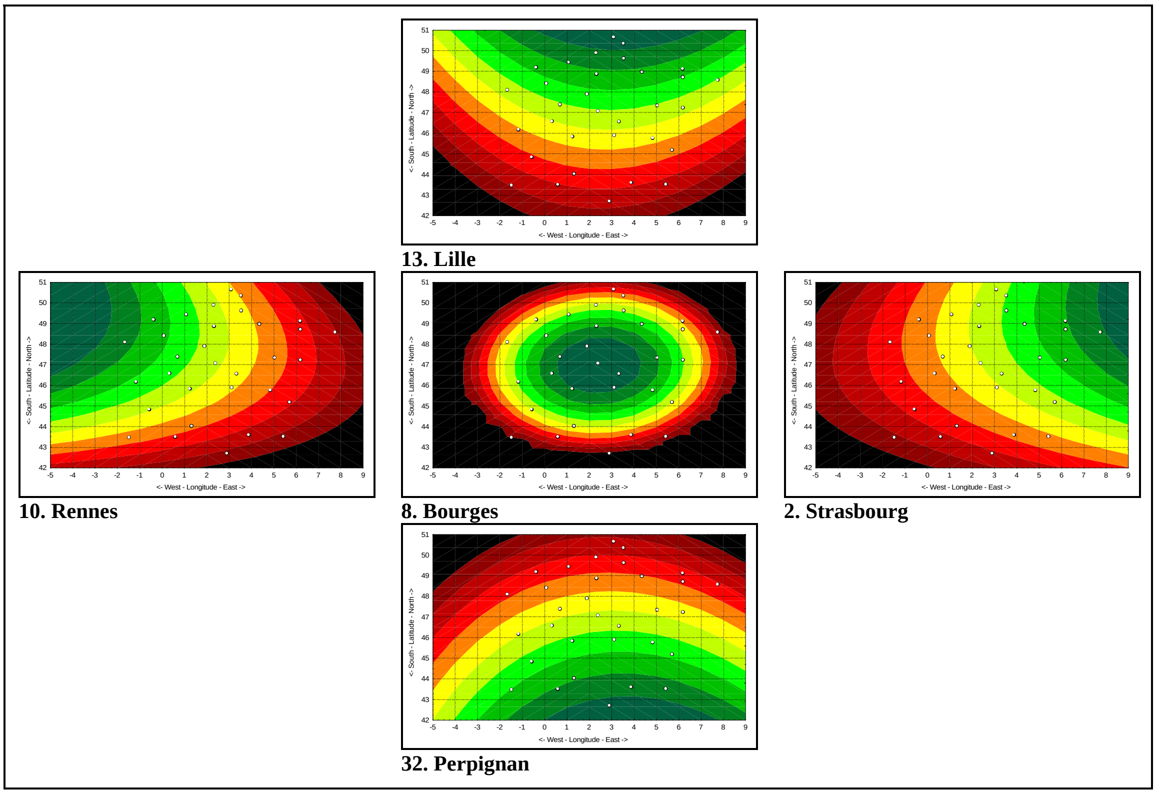
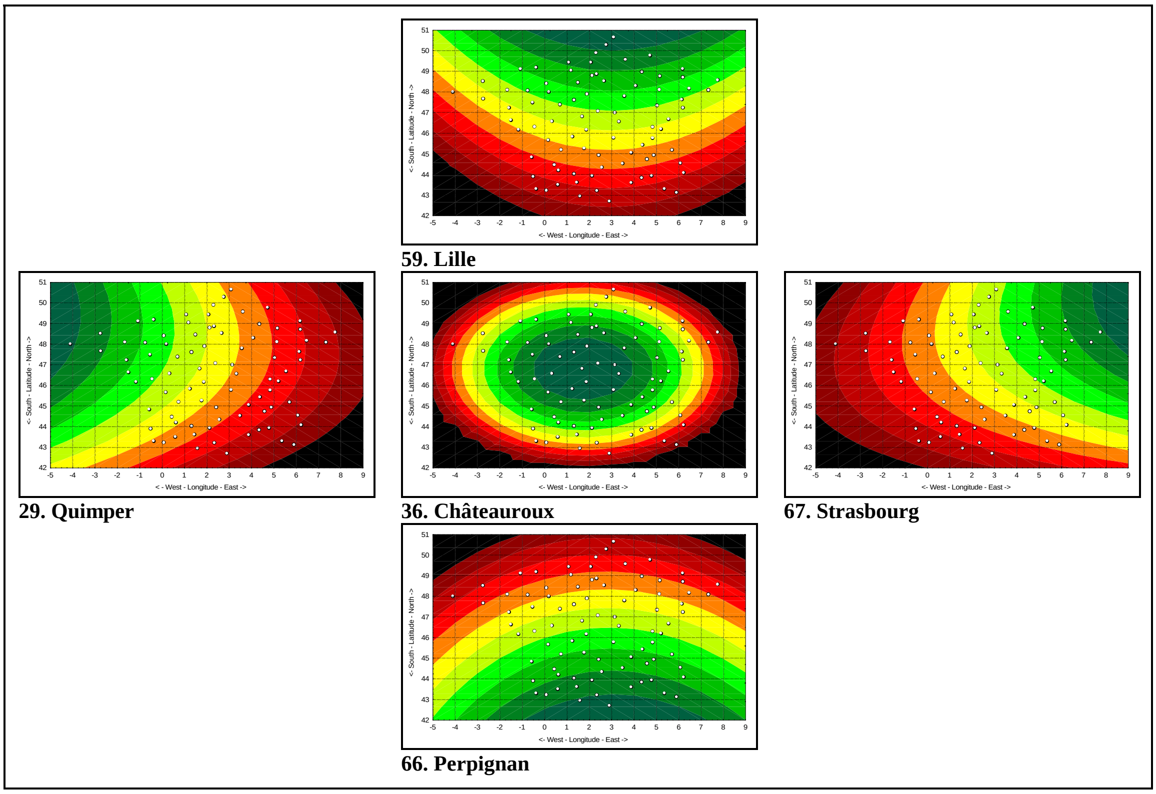


Figure 2 : Distances versus Longitude and Latitude - Départements
Central and Outermost Locations - Quadratic Surface - Contour Plot



From the comparison of these general descriptive statistics the following conclusions can be drawn :

- the mean distance between locations is almost the same, i.e. 405.8 km for the *généralités* against 392.7 km for the *départements*
- the much smaller minimum distance between *départements* is the result of an increased density of the points of measurement, due to an increase of the number of locations from 32 *généralités* to 85 *départements*
- the larger value obtained for the maximum distance between the *départements* can be explained by an increased and extended coverage of the north-west of the French territory by the data set for the period 1806-1872

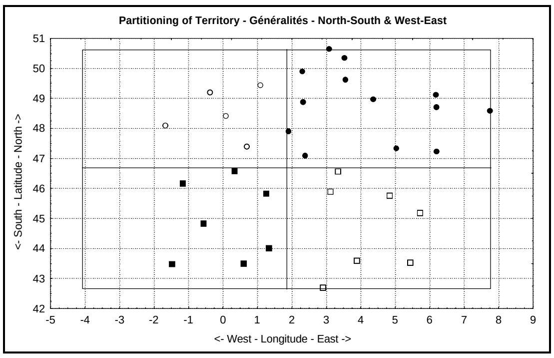
Apart from a few differences the data about the distances for the *généralités* and *départements* are showing at least two important common characteristics. A first common characteristic is the potential general explanatory capabilities of the distances data. This characteristic can be visualized by Figure 1 and Figure 2. These figures consist of the contour plots for the distances of five locations. Four of these locations are the outermost locations of each of the data sets. The fifth location is that location that is the closest to the center of the covered territory. Detailed information about these locations is tabulated by Table 4 and represented by Figure 3. The contour plots are the horizontal projections of the quadratic surfaces fitted through the 3D scatterplots. See **Borghers [5]** for an other example of using a quadratic surface to smooth a 3D scatterplot.

Table 4 : Geographical Locations - Généralités and Départements
Central and Outermost Locations - Longitude and Latitude

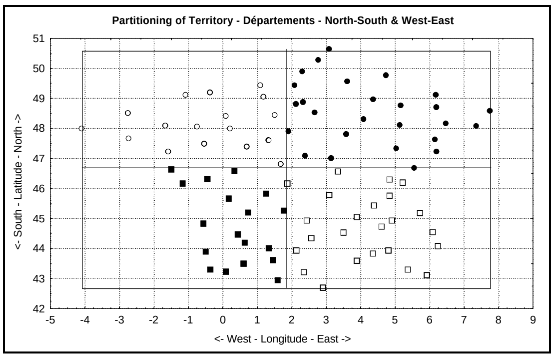
Généralités				**Départements**			
Location		**Longitude**	**Latitude**	**Location**		**Longitude**	**Latitude**
Outermost				**Outermost**			
N	Lille	3.08	50.65	**N**	Lille	3.08	50.65
E	Strasbourg	7.75	48.58	**E**	Strasbourg	7.75	48.58
S	Perpignan	2.90	42.70	**S**	Perpignan	2.90	42.70
W	Rennes	-1.67	48.10	**W**	Quimper	-4.10	48.00
Center				**Center**			
	Bourges	2.38	47.08		Châteauroux	1.68	46.82

Figure 3 : Geographical Locations - Généralités and Départements
Partitioning of Territory - Central and Outermost Locations

Généralités - Period : 1756-1790



Départements - Period : 1806-1872

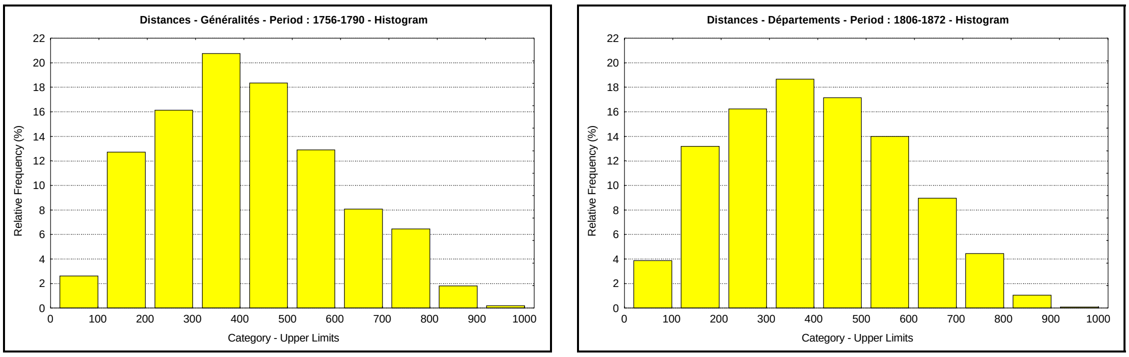


From the comparison of Figure 1 and Figure 2 one can conclude that at least for the five chosen locations the distances among the *généralités* and *départements* can be represented almost equally well by a quadratic surface of the 3D scatterplot of the distances. Generalizing this idea leads to the conclusion that the explanatory capabilities of the 496 distances among the 32 *généralités* can be compared with the 3570 distances among the 85 *départements*.

A second common characteristic is the skewness of the distributions of the distances data. From Table 3 it can be seen that both the data about the distances among the 32 *généralités* and those about the distances among the 85 *départements* are positively skewed, i.e. skewed to the right. This can be visualized by the graphical representations of Figure 4. From these graphs it can be seen that not only both distributions are skewed to the right but that also the general shapes of these distributions are highly comparable.

Figure 4 : Distances - Généralités and Départements - Histogram

Généralités - Period : 1756-1790 **Départements - Period : 1806-1872**

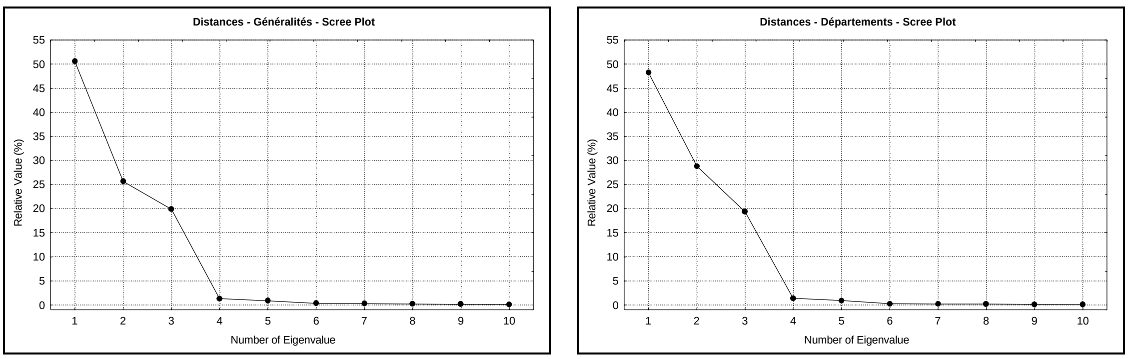


Principal Component Analysis of Distances Data

The potential general explanatory capabilities of the distance data can be analyzed in more detail by using an appropriate multivariate technique. The first technique that will be used is the principal component analysis. As a first step in this analysis a somewhat different approach will be used by calculating the eigenvalues of the correlation matrix of the distances instead of the eigenvalues of the distance matrix itself. The graphical representation of the relative contribution of the ten largest of these eigenvalues are given by Figure 5. From the comparison of the general pattern shown by the scree plots it can be concluded that the majority of the information of both the distance matrices can be represented by only three dimensions. The resemblance of the results obtained for the distances data for the *généralités* and the *départements* becomes even more apparent when the numerical results are compared. The numerical results for the five largest eigenvalues are given by Table 5. Based on the available information it was decided to proceed the analysis of the distances data by using three components.

Figure 5 : Distances - Généralités and Départements - Correlation Matrix Eigenvalues - Scree Plot

Généralités - Period : 1756-1790 **Départements - Period : 1806-1872**



The next step in the analysis is to extract for each of the three eigenvalues the corresponding principal component loadings. These loadings can be seen as the correlation coefficients between the variables and the theoretical constructed orthogonal (uncorrelated) explanatory components. In order to facilitate the interpretation of these component loadings the three axes were rotated. Among the various rotational strategies the varimax rotation was used. This rotation method is aimed at maximizing for each of the components the variances of the squared raw component loadings across the variables. The goal of the rotation is to obtain a clear pattern of loadings, i.e. to achieve a more interpretable simple structure by changing the actual coordinates of the points without changing the relative locations of the points to each other. This simple structure will be characterized by components with high loadings for some of the variables and low loadings for the others. The thus derived rotated solution is presented by Table 7 and Table 8. The absolute and relative contribution of the rotated components in explaining the variance of the distances is given by Table 6.

Table 5 : Distances - Généralités and Départements - Principal Components Eigenvalues - Original Solution - Absolute and Relative Contribution

	Généralités				Départements			
	Non-Cumulated		Cumulated		Non-Cumulated		Cumulated	
Eigenvalue	Value	%	Value	%	Value	%	Value	%
1	16.18	50.58	16.18	50.58	41.02	48.26	41.02	48.26
2	8.21	25.64	24.39	76.22	24.49	28.82	65.51	77.07
3	6.36	19.89	30.75	96.11	16.45	19.35	81.96	96.43
4	0.43	1.33	31.18	97.44	1.17	1.38	83.14	97.81
5	0.27	0.86	31.45	98.30	0.79	0.93	83.93	98.74

Table 6 : Distances - Généralités and Départements - Principal Components Eigenvalues - Varimax Solution - Absolute and Relative Contribution

	Généralités				Départements			
	Non-Cumulated		Cumulated		Non-Cumulated		Cumulated	
Eigenvalue	Value	%	Value	%	Value	%	Value	%
1	14.65	45.78	14.65	45.78	31.25	36.76	31.25	36.76
2	9.44	29.50	24.09	75.28	23.67	27.85	54.92	64.61
3	6.67	20.84	30.75	96.11	27.04	31.81	81.96	96.43

From the results obtained for the rotated component loadings several interesting conclusions can be derived. The first and main conclusion is the interpretation of each of the components. As can be seen from the results the first component is highly positively correlated with a first subset of regions, highly negatively correlated with a second subset and almost not correlated at all with a third subset. This first component is a typical example of a bipolar component. From a closer inspection of the regions involved it can be seen that the positive pole of the first component is highly positively correlated with regions in the north and north-east and the opposite negative pole with regions in the south and south-west. Also the second and the third principal component represent a simple structure. The second component can typically be associated with regions located in the eastern part of the territory while the third component is highly and positively correlated with regions in the western part of France.

The here given interpretation of the components illustrates that the principal component analysis can be considered as a data reduction technique as well as a classification analysis. From the results of Table 5 and Table 6 it follows that almost 97% of the information content of the distances among the 32 *généralités* and among the 85 *départements* can be reduced to only three orthogonal components. In addition to the drastic reduction of the dimension of the original distances data the principal component analysis can also be used to classify the 32 *généralités* and the 85 *départements* into a much smaller number of larger geographical regions. These larger regions are shown on Figure 6.

The classification of the *généralités* and *départements* accordingly to the obtained component loadings is not always a clear-cut operation. This can best be illustrated by using the results for the *généralité* # 9 (Dijon). As can be seen from Table 7 the component loading for both the first and the second component is 0.69. This result might look as an ambiguous situation. However from a closer look at the geographical location of this particular *généralité* it turns out that it is situated exactly at the border of either the region in the north-east, i.e. associated with the positive pole of the first component, or the region in the south-east, i.e. described by the second component. In Table 7 and Table 8 the comparable cases are marked by printing the component loadings on a light gray background. It can be verified that all these cases are situated at the border of one of the classification regions. As an illustration of the effect of an alternative classification of the ambiguous results an alternative representation of the results for the *généralités* is given as the second column of Figure 6. For convenience this column is labeled as 'Border Correction'.

Table 7 : Distances - Généralités - Rotated Principal Components
Component Loadings and Explained Variances

#	Component Loading			Explained Variance			
	1	3	2	1	3	2	Tot.
2	0.91	-0.14	0.31	0.83	0.02	0.10	0.94
3	0.78	0.57	-0.21	0.60	0.32	0.04	0.97
9	0.69	0.07	0.69	0.48	0.00	0.48	0.97
12	0.93	0.32	0.12	0.86	0.10	0.01	0.97
13	0.85	0.43	-0.23	0.72	0.19	0.05	0.96
14	0.71	-0.15	0.66	0.50	0.02	0.44	0.96
16	0.88	0.40	-0.18	0.77	0.16	0.03	0.97
21	0.94	0.01	0.27	0.89	0.00	0.07	0.96
23	0.96	0.05	0.18	0.93	0.00	0.03	0.96
33	0.88	0.43	-0.08	0.78	0.18	0.01	0.97
35	0.71	0.69	-0.04	0.50	0.47	0.00	0.97
4	-0.95	-0.08	0.24	0.90	0.01	0.06	0.96
5	-0.97	0.09	0.04	0.94	0.01	0.00	0.95
7	-0.92	0.30	0.12	0.84	0.09	0.02	0.95
17	-0.67	-0.41	0.58	0.45	0.17	0.33	0.95
19	-0.67	0.52	0.48	0.45	0.27	0.23	0.96
24	-0.91	-0.07	0.35	0.83	0.00	0.12	0.96
32	-0.80	-0.37	0.41	0.65	0.14	0.17	0.96

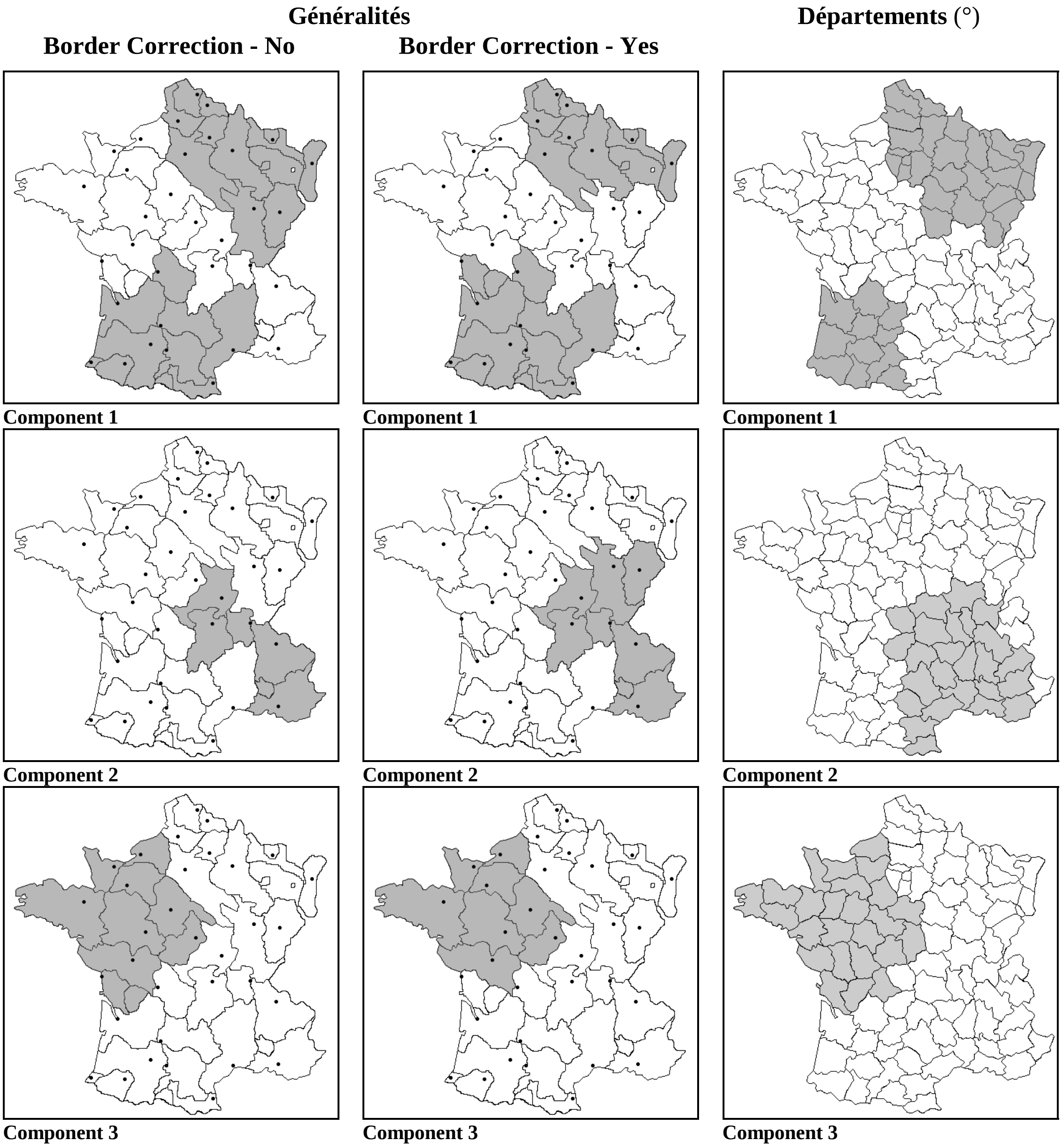
#	Component Loading			Explained Variance			
	1	3	2	1	3	2	Tot.
1	0.23	0.94	-0.18	0.05	0.88	0.03	0.97
8	0.14	0.77	0.60	0.02	0.60	0.36	0.97
10	-0.07	0.93	-0.28	0.01	0.86	0.08	0.95
11	0.35	0.86	-0.30	0.12	0.74	0.09	0.96
18	-0.67	0.70	-0.01	0.45	0.50	0.00	0.95
26	0.40	0.88	0.19	0.16	0.77	0.04	0.97
28	-0.47	0.84	0.20	0.22	0.70	0.04	0.97
31	0.59	0.75	-0.24	0.35	0.56	0.06	0.97
34	-0.06	0.98	0.12	0.00	0.96	0.02	0.98
15	-0.08	-0.42	0.87	0.01	0.18	0.76	0.95
22	-0.01	-0.21	0.96	0.00	0.04	0.92	0.96
25	0.08	0.38	0.91	0.01	0.15	0.82	0.97
29	-0.47	-0.53	0.65	0.22	0.28	0.42	0.93
30	-0.32	0.21	0.91	0.10	0.04	0.83	0.97

Table 8 : Distances - Départements - Rotated Principal Components
Component Loadings and Explained Variances

#	Component Loading			Explained Variance			
	1	3	2	1	3	2	Tot.
2	0.87	0.32	-0.33	0.75	0.10	0.11	0.96
8	0.93	0.17	-0.27	0.86	0.03	0.07	0.96
10	0.95	0.27	-0.00	0.90	0.07	0.00	0.97
21	0.90	-0.01	0.41	0.81	0.00	0.17	0.97
25	0.86	-0.22	0.41	0.74	0.05	0.17	0.96
39	0.73	-0.20	0.63	0.54	0.04	0.39	0.97
51	0.95	0.21	-0.16	0.90	0.05	0.03	0.97
52	0.98	0.04	0.12	0.95	0.00	0.01	0.97
54	0.98	-0.08	0.00	0.95	0.01	0.00	0.96
55	0.98	0.07	-0.06	0.96	0.01	0.00	0.97
57	0.98	-0.04	-0.08	0.95	0.00	0.01	0.96
58	0.65	0.51	0.53	0.43	0.26	0.29	0.98
59	0.79	0.35	-0.46	0.62	0.12	0.22	0.95
60	0.70	0.54	-0.42	0.49	0.29	0.18	0.96
62	0.77	0.39	-0.46	0.59	0.15	0.21	0.95
67	0.95	-0.22	0.08	0.89	0.05	0.01	0.95
68	0.93	-0.24	0.17	0.86	0.06	0.03	0.95
70	0.92	-0.17	0.29	0.85	0.03	0.08	0.96
75	0.73	0.57	-0.31	0.54	0.33	0.10	0.96
77	0.78	0.56	-0.20	0.61	0.31	0.04	0.97
78	0.70	0.61	-0.31	0.49	0.37	0.10	0.96
80	0.73	0.47	-0.46	0.53	0.22	0.21	0.96
88	0.96	-0.15	0.14	0.91	0.02	0.02	0.96
89	0.89	0.40	0.15	0.79	0.16	0.02	0.98
9	-0.73	-0.12	0.64	0.53	0.01	0.42	0.96
24	-0.64	0.48	0.58	0.41	0.23	0.33	0.97
31	-0.73	-0.03	0.66	0.53	0.00	0.43	0.96
32	-0.80	0.06	0.56	0.63	0.00	0.32	0.95
33	-0.77	0.47	0.37	0.60	0.22	0.13	0.96
40	-0.84	0.24	0.43	0.70	0.06	0.19	0.95
46	-0.76	0.28	0.55	0.58	0.08	0.30	0.96
47	-0.77	0.18	0.58	0.59	0.03	0.33	0.96
64	-0.84	0.13	0.46	0.71	0.02	0.22	0.94
65	-0.83	0.07	0.51	0.68	0.01	0.26	0.95
82	-0.72	0.06	0.67	0.52	0.00	0.44	0.96

#	Component Loading			Explained Variance			
	1	3	2	1	3	2	Tot.
14	0.32	0.76	-0.53	0.10	0.58	0.28	0.97
16	-0.58	0.69	0.38	0.34	0.48	0.15	0.97
17	-0.52	0.83	0.01	0.27	0.69	0.00	0.96
18	0.50	0.76	0.38	0.25	0.58	0.14	0.98
22	-0.07	0.81	-0.54	0.01	0.66	0.29	0.95
27	0.55	0.69	-0.43	0.31	0.47	0.19	0.96
28	0.56	0.75	-0.30	0.31	0.57	0.09	0.97
29	-0.24	0.79	-0.50	0.06	0.63	0.25	0.94
35	-0.05	0.86	-0.46	0.00	0.75	0.21	0.96
36	0.18	0.90	0.37	0.03	0.81	0.14	0.98
37	0.15	0.98	-0.08	0.02	0.95	0.01	0.98
41	0.37	0.91	-0.06	0.14	0.84	0.00	0.98
44	-0.25	0.90	-0.30	0.06	0.81	0.09	0.96
45	0.58	0.79	-0.09	0.34	0.63	0.01	0.98
49	-0.06	0.95	-0.27	0.00	0.90	0.07	0.97
50	0.21	0.79	-0.55	0.04	0.62	0.30	0.97
53	0.06	0.89	-0.41	0.00	0.80	0.17	0.97
56	-0.24	0.84	-0.44	0.06	0.70	0.19	0.95
61	0.27	0.85	-0.41	0.07	0.73	0.17	0.97
72	0.21	0.91	-0.32	0.04	0.83	0.10	0.98
76	0.56	0.65	-0.48	0.31	0.42	0.23	0.96
79	-0.43	0.88	0.08	0.19	0.77	0.01	0.97
85	-0.41	0.88	-0.15	0.17	0.77	0.02	0.96
86	-0.24	0.95	0.15	0.06	0.89	0.02	0.98
87	-0.39	0.66	0.62	0.15	0.44	0.38	0.97
1	0.54	-0.22	0.79	0.29	0.05	0.62	0.96
3	0.50	0.37	0.77	0.25	0.13	0.59	0.98
4	-0.03	-0.52	0.82	0.00	0.27	0.68	0.95
5	0.06	-0.49	0.84	0.00	0.24	0.71	0.95
7	-0.08	-0.33	0.92	0.01	0.11	0.85	0.96
11	-0.65	-0.18	0.72	0.42	0.03	0.52	0.97
12	-0.52	-0.06	0.84	0.27	0.00	0.70	0.98
13	-0.26	-0.48	0.81	0.07	0.23	0.65	0.95
15	-0.44	0.10	0.88	0.19	0.01	0.78	0.98
19	-0.48	0.34	0.79	0.23	0.12	0.63	0.98
23	-0.10	0.69	0.70	0.01	0.48	0.49	0.98
26	0.02	-0.35	0.92	0.00	0.12	0.84	0.96
30	-0.32	-0.38	0.85	0.10	0.14	0.72	0.96
34	-0.42	-0.34	0.83	0.17	0.11	0.68	0.97
38	0.20	-0.41	0.86	0.04	0.17	0.75	0.96
42	0.12	-0.20	0.95	0.01	0.04	0.91	0.97
43	-0.11	-0.17	0.96	0.01	0.03	0.93	0.97
48	-0.32	-0.19	0.91	0.11	0.04	0.84	0.98
63	0.02	0.20	0.97	0.00	0.04	0.94	0.98
66	-0.61	-0.27	0.72	0.37	0.07	0.52	0.97
69	0.32	-0.22	0.90	0.11	0.05	0.81	0.96
71	0.56	-0.13	0.80	0.31	0.02	0.64	0.97
81	-0.63	-0.07	0.75	0.40	0.00	0.57	0.97
83	-0.22	-0.52	0.79	0.05	0.27	0.63	0.95
84	-0.24	-0.42	0.85	0.06	0.17	0.73	0.96

Figure 6 : Distances - Généralités and Départements - Rotated Principal Components Component Loadings



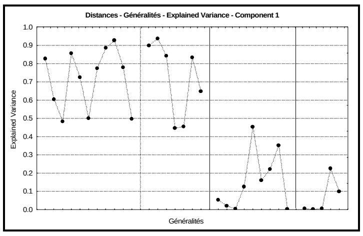
(°) The *départements* excluded from the analysis, i.e. Alpes-Maritimes, Savoie and Haute-Savoie, are left blank. See Appendix 3.

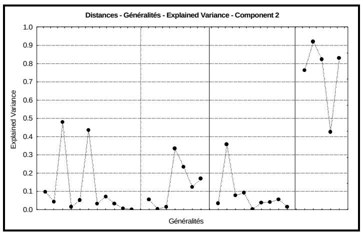
The final conclusion that can be drawn from the results shown by Figure 6 is the striking resemblance between the results obtained for the distance data for the *généralités* and the *départements*. Accounting for both the mentioned problem of the classification of the border regions and the rather important difference in the partition of the French territory the resemblance of the results obtained for analysis of the distances between the *généralités* and the *départements* cannot be denied. The discriminating behavior of the three components is identical for both data sets.

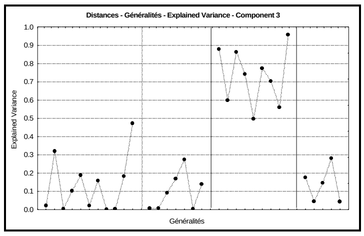
This last argument can further be clarified and illustrated by Figure 7. Each graph of Figure 7 is representing the proportion of the variance explained by the corresponding component. These proportions are the squared component loadings given in Table 7 and Table 8. It can be seen that for both the distances of the *généralités* and the *départements* each component accounts for at least 50% of the variance of a specific and disjunct subset of the variables.

Figure 7 : Distances - Généralités and Départements - Rotated Principal Components Explained Variances

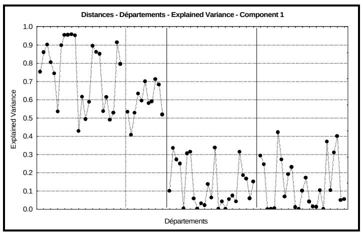
Component 1	Component 2	Component 3

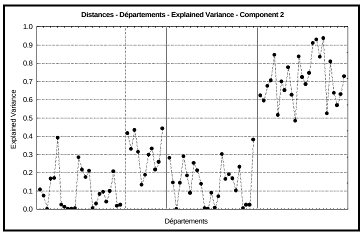


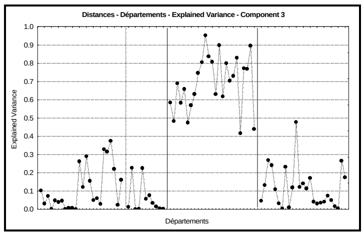




Généralités **Généralités** **Généralités**







Départements **Départements** **Départements**

Cluster Analysis of Distances Data

As a second multivariate technique cluster analysis can be used to analyze in more detail the distances data. It must be stressed that clustering analysis has nothing to do with statistical significance testing in the traditional sense. This multivariate technique is nothing else than a collection of different algorithms aiming at putting objects into clusters. In this sense the use of this technique is highly appropriate in the explanatory phase of research when not any a priori hypothesis is available. In the line of the previous applications in **Borghers [3]** and **[4]** two specific clustering methods will be used, i.e. the k-means clustering and the joining or tree clustering method.

One may think of the k-means clustering method as an analysis of variance 'in reverse'. The algorithm starts with k random clusters. Then the points are iteratively adjusted by moving them between the clusters in order to minimize the variability within each cluster and to maximize the variability between the k clusters. The standard output for the final solution of the k-means clustering consists of the members of each of the k clusters and the distance of each of the members from the respective cluster center. To facilitate the comparison of the results with those obtained for the principal components analysis it was decided to apply the k-means clustering method with four clusters. The results for the distance data of the *généralités* and the *départements* are tabulated in respectively Table 9 and Table 10.

Table 9 : Distances - Généralités - k-Means Clustering Members of Clusters and Distances from Cluster Center

Généralité #	Ville	Cluster 1	2	3	4
1	Alençon	103.3			
3	Amiens	54.1			
10	Rennes	155.0			
11	Caen	79.5			
13	Lille	110.6			
16	Valenciennes	104.9			
31	Rouen	35.9			
33	Soissons	87.9			
35	Paris Ville	77.1			
2	Strasbourg		101.0		
9	Dijon		98.6		
12	Châlons		90.3		
14	Besançon		77.3		
21	Nancy		35.8		
23	Metz		59.7		
4	Auch			75.4	
5	Bayonne			147.1	
7	Bordeaux			129.8	
15	Grenoble			172.2	
17	Montpellier			73.6	
24	Montauban			69.2	
29	Aix			119.8	
32	Perpignan			88.7	
8	Bourges				62.4
18	La Rochelle				143.4
19	Limoges				71.6
22	Lyon				141.1
25	Moulins				74.1
26	Orléans				109.7
28	Poitiers				72.1
30	Riom				79.0
34	Tours				82.7

Table 10 : Distances - Départements - k-Means Clustering
Members of Clusters and Distances from Cluster Center

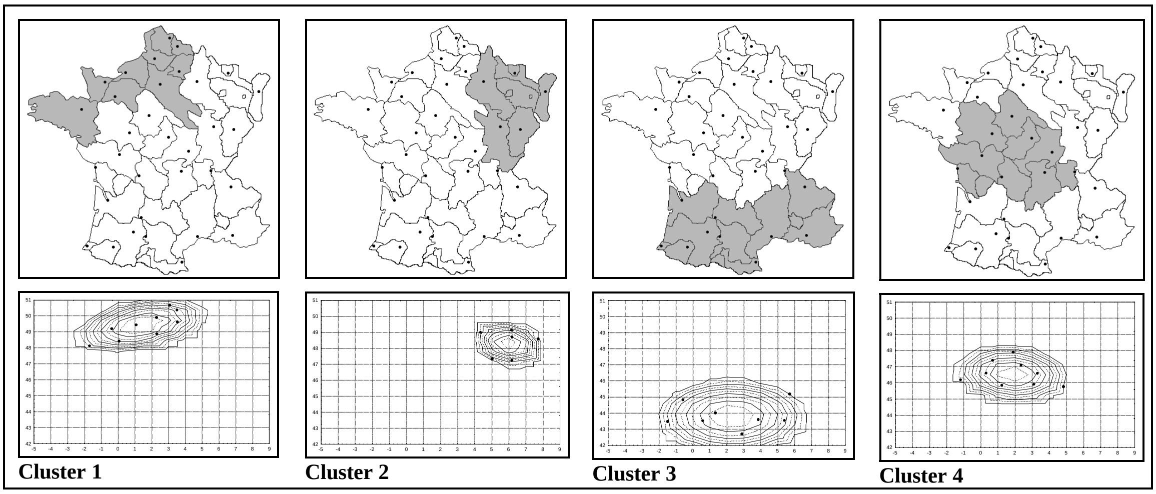
Département #	Ville	Cluster 1	Cluster 2	Cluster 3	Cluster 4
2	Laon	64.9			
8	Charleville	71.6			
10	Troyes	68.1			
21	Dijon	115.4			
25	Besançon	120.1			
51	Châlons	34.0			
52	Chaumont	65.2			
54	Nancy	72.2			
55	Bar-le-Duc	37.5			
57	Metz	79.6			
59	Lille	141.1			
60	Beauvais	107.9			
62	Arras	120.1			
67	Strasbourg	142.8			
68	Colmar	122.1			
70	Vesoul	99.3			
75	Paris	104.4			
77	Melun	101.9			
78	Versailles	112.6			
80	Amiens	111.3			
88	Épinal	87.3			
89	Auxerre	110.1			
1	Bourg-en-Bresse		91.9		
3	Moulins		127.9		
4	Digne		134.2		
5	Gap		102.8		
7	Privas		46.2		
13	Marseille		158.4		
15	Aurillac		91.7		
19	Tulle		118.8		
23	Guéret		143.2		
26	Valence		44.8		
30	Nîmes		100.0		
34	Montpellier		114.5		
38	Grenoble		67.9		
39	Lons-le-Saunier		125.5		
42	Saint-Étienne		45.0		
43	Le Puy-en-Valay		42.8		
48	Mende		63.1		
58	Nevers		157.6		
63	Clermont-Ferrand		89.5		
69	Lyon		62.7		
71	Mâcon		95.9		
83	Toulon		188.8		
84	Avignon		98.9		
87	Limoges		149.3		
9	Foix			79.6	
11	Carcassonne			77.7	
12	Rodez			91.0	
16	Angoulême			139.4	
24	Périgueux			107.9	
31	Toulouse			36.3	
32	Auch			42.0	
33	Bordeaux			86.8	
40	Mont-de-Marsan			64.3	
46	Cahors			51.8	
47	Agen			34.2	
64	Pau			87.9	
65	Tarbes			76.2	
66	Perpignan			121.6	
81	Albi			59.9	
82	Montauban			34.5	
14	Caen				94.4
17	La Rochelle				113.5
18	Bourges				147.7
22	Saint-Brieuc				144.1
27	Évreux				104.1
28	Chartres				97.4
29	Quimper				208.7
35	Rennes				74.6
36	Châteauroux				134.6
37	Tours				77.5
41	Blois				95.1
44	Nantes				69.6
45	Orléans				111.5
49	Angers				40.6
50	Saint-Lô				103.3
53	Laval				38.9
56	Vannes				123.2
61	Alençon				52.2
72	Le Mans				46.2
76	Rouen				119.3
79	Niort				107.4
85	La Roche-sur-Yon				91.4
86	Poitiers				105.7

The interpretation of the tabulated results for the 4-means clustering can be substantially improved by using the graphical representations of Figure 8. As a first attempt the members of each of the four clusters are represented on a map of the French territory. Since for each of the four clusters also the distance of the members to the center of the cluster is available a second graphical representation can be constructed. These graphs are 3D contour plots. Each of these contour lines represent an equal distance to the cluster center and are the result of projecting the quadratic smoothed 3D scatterplot of the distances tabulated in Table 9 and Table 10. The main advantage of the contour plots is that they give a rather precise idea of the exact geographical location of the cluster centers.

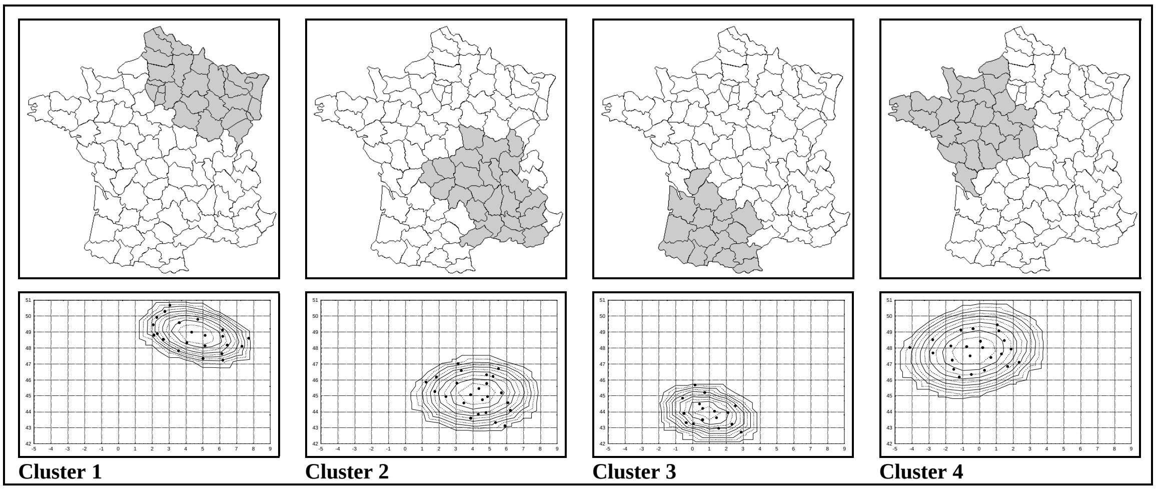
The main characteristics resulting from the comparison of the maps and the contour plots of Figure 8 can easily be summarized. The distances among the 32 *généralités* are represented by four clusters. These clusters can be labeled as a north-west cluster (cluster 1), a north-east cluster (cluster 2), a central cluster (cluster 4) and a southern cluster (cluster 3). For two out of the four clusters representing the distances among the 85 *départements* the same labeling can be used, i.e. cluster 4 can be labeled as a north-west cluster and cluster 1 as a north-east cluster. A somewhat different naming must be used for the two remaining clusters. Whereas the southern part of the French territory is represented by one single cluster as far as the *généralités* is concerned the whole southern region is represented by two clusters for the distances among the *départements*, i.e. a south-west cluster (cluster 3) and a south-east cluster (cluster 2). A detailed visual inspection of the graphical representations of Figure 8 reveals that the *départements* covering the territory described by the central cluster for the *généralités* are mainly partitioned among the north-west and the south-east cluster.

Figure 8 : Distances - Généralités and Départements - Cluster Analysis
k-Means Clustering

Généralités



Départements



The second clustering method that will be used is the joining clustering also known as the tree clustering method. The main purpose of this method is to join together the objects (distances) into successively larger clusters by using an appropriate measure of similarity. In other words by proceeding in this way more and more objects are linked together resulting in larger and larger clusters of increasingly dissimilar elements. The algorithm that will be used is the method proposed by **Ward [13]**. This method is using an analysis of variance approach to evaluate the distance between clusters. Furthermore this method tends to create clusters of small size.

The main and typical output of this specific type of clustering is the hierarchical tree. This tree can be represented graphically by the vertical icicle plot or dendrogram. A dendrogram is a bar chart like device that graphically shows how the groups or clusters coalesce through the analysis. The vertical axis of this graph denotes the linkage distance. In other words for each node, i.e. were a new cluster is formed, one can read off the criterion distance at which the respective elements are linked together into a new single branch. The lower on the vertical scale the more homogeneous the groups. When the data consist of a clear structure, characterized by objects that are similar to each other, this structure will be reflected in the hierarchical tree as distinct branches.

Figure 9 : Distances - Généralités - Tree Clustering - Dendrogram
Ward’s Method - Euclidean Distances

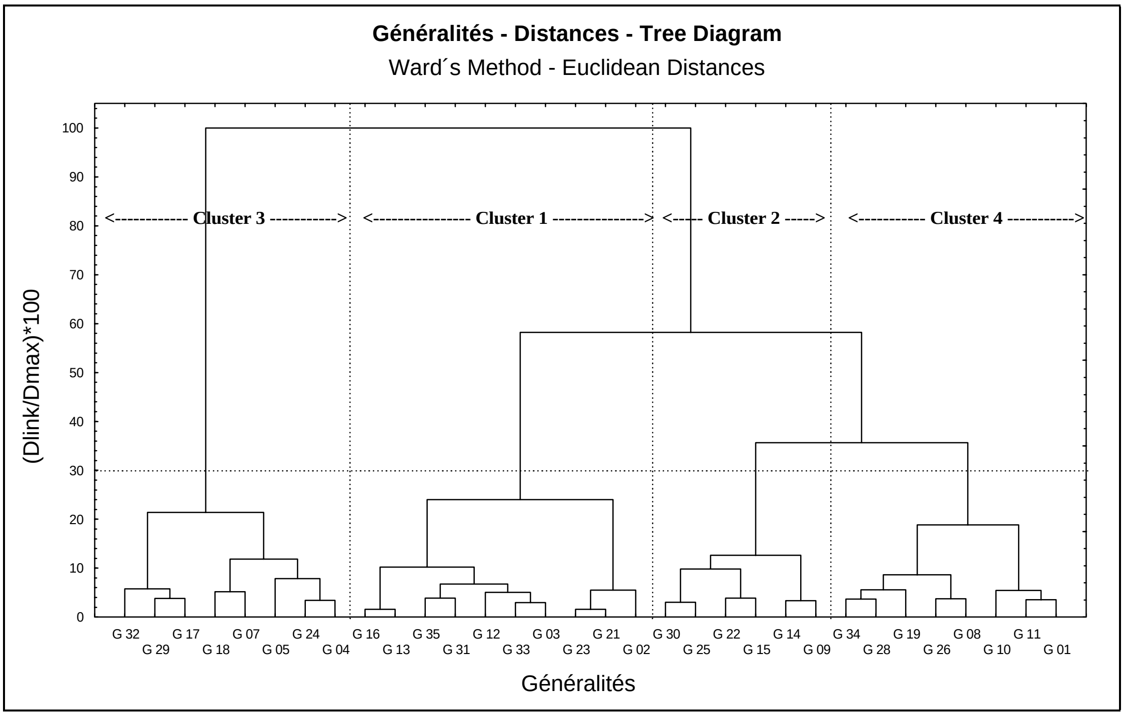
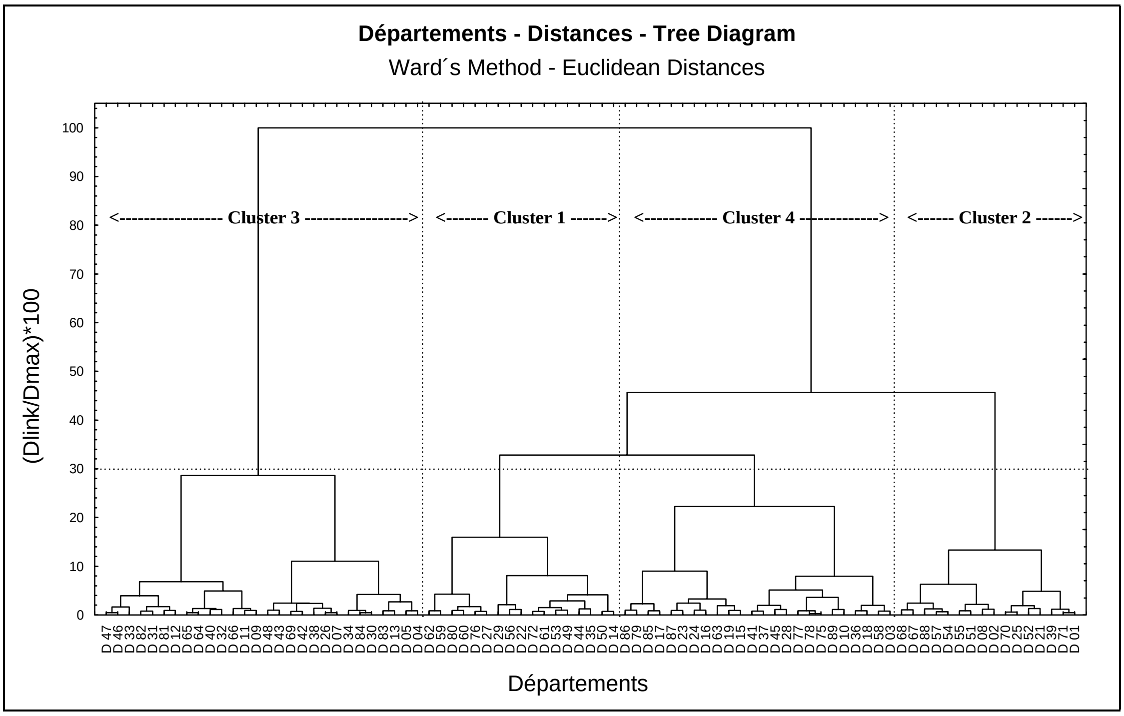


Figure 10 : Distances - Départements - Tree Clustering - Dendrogram
Ward’s Method - Euclidean Distances



The dendrograms for both the distance data sets of the *généralités* and the *départments* are given as respectively Figure 9 and Figure 10. A comparison of these tree diagrams leads to some interesting conclusions. A first and main characteristic is that both diagrams are sharing a comparable simple hierarchical structure. This structure consists of two dominant branches, i.e. a smaller branch represented at the left of the diagrams and a larger and more complex structured branch represented at the right.

Table 11 : Distances - Généralités - Tree Clustering
Ward's Method - Euclidean Distances

Généralité #	Ville	Cluster 1	2	3	4
2	Strasbourg	■			
3	Amiens	■			
12	Châlons	■			
13	Lille	■			
16	Valenciennes	■			
21	Nancy	■			
23	Metz	■			
31	Rouen	■			
33	Soissons	■			
35	Paris Ville	■			
9	Dijon		■ 0		
14	Besançon		■ 0		
15	Grenoble		■ 0		
22	Lyon		■ 0		
25	Moulins		■ 0		
30	Riom		■ 0		

Généralité #	Ville	Cluster 1	2	3	4
4	Auch			■	
5	Bayonne			■	
7	Bordeaux			■	
17	Montpellier			■	
18	La Rochelle			■	
24	Montauban			■	
29	Aix			■	
32	Perpignan			■	
1	Alençon				■ 0
8	Bourges				■ 0
10	Rennes				■ 0
11	Caen				■ 0
19	Limoges				■ 0
26	Orléans				■ 0
28	Poitiers				■ 0
34	Tours				■ 0

Table 12 : Distances - Départements - Tree Clustering
Ward's Method - Euclidean Distances

Département #	Ville	Cluster 1	2	3	4
14	Caen	■			
22	Saint-Brieuc	■			
27	Évreux	■			
29	Quimper	■			
35	Rennes	■			
44	Nantes	■			
49	Angers	■			
50	Saint-Lô	■			
53	Laval	■			
56	Vannes	■			
59	Lille	■			
60	Beauvais	■			
61	Alençon	■			
62	Arras	■			
72	Le Mans	■			
76	Rouen	■			
80	Amiens	■			
1	Bourg-en-Bresse		■		
2	Laon		■		
8	Charleville		■		
21	Dijon		■		
25	Besançon		■		
39	Lons-le-Saunier		■		
51	Châlons		■		
52	Chaumont		■		
54	Nancy		■		
55	Bar-le-Duc		■		
57	Metz		■		
67	Strasbourg		■		
68	Colmar		■		
70	Vesoul		■		
71	Mâcon		■		
88	Épinal		■		

Département #	Ville	Cluster 1	2	3	4
4	Digne			■	
5	Gap			■	
7	Privas			■	
13	Marseille			■	
26	Valence			■	
30	Nîmes			■	
34	Montpellier			■	
38	Grenoble			■	
42	Saint-Étienne			■	
43	Le Puy-en-Velay			■	
48	Mende			■	
69	Lyon			■	
83	Toulon			■	
84	Avignon			■	
9	Foix			0 ■	
11	Carcassonne			0 ■	
12	Rodez			0 ■	
31	Toulouse			0 ■	
32	Auch			0 ■	
33	Bordeaux			0 ■	
40	Mont-de-Marsan			0 ■	
46	Cahors			0 ■	
47	Agen			0 ■	
64	Pau			0 ■	
65	Tarbes			0 ■	
66	Perpignan			0 ■	
81	Albi			0 ■	
82	Montauban			0 ■	
3	Moulins				■
10	Troyes				■
15	Aurillac				■
16	Angoulême				■
17	La Rochelle				■
18	Bourges				■
19	Tulle				■
23	Guéret				■
24	Périgueux				■
28	Chartres				■
36	Châteauroux				■
37	Tours				■
41	Blois				■
45	Orléans				■
58	Nevers				■
63	Clermont-Ferrand				■
75	Paris				■
77	Melun				■
78	Versailles				■
79	Niort				■
85	La Roche-sur-Yon				■
86	Poitiers				■
87	Limoges				■
89	Auxerre				■

A second characteristic becomes evident by concentrating on the hierarchical structure of these more complex branches. The interpretation of these branches can be simplified by splitting up the branches in accordance with the hierarchical structure. This partitioning leads to the identification of three clusters. These three clusters, completed with the regions belonging to the left simpler branch, result in four clusters for each of the two distance data sets. Precise information about the partitioning into four clusters is superimposed on the tree diagrams given in Figure 9 and Figure 10. The partitioning of the hierarchical structures into clusters can also be represented by the Table 11 and Table 12. The main purpose for reporting these tables is to facilitate the comparison of the classification scheme derived from the principal component and cluster analysis. In order to improve the comparison the following symbols were used to construct the Table 11 and Table 12 :

- to indicate the members of the clusters the symbol '■' is used
- the combined symbol '■ O' indicates that the corresponding regions belong to the same branch in the hierarchy
- the symbol 'O ■' is used to indicate that the corresponding regions can also be seen as a belonging to a separate cluster

The interpretation of the identified clusters can be facilitated even further by representing the members of each cluster on a separate contour map of the French territory. The resulting four maps for the distances data of the *généralités* and the *départements* are given as respectively the second and fourth column of Figure 11. The maps of the second and fourth column can be supplemented with the maps obtained for the k-means clustering method given by Figure 8. The resulting graphical representations not only enable the comparison of the results obtained for each of the periods 1756-1790 (*généralités*) and 1806-1872 (*départements*) but also the comparison between the results obtained for each of the two clustering techniques, i.e. the k-means clustering method and the tree clustering approach represented by **Ward**'s algorithm. An additional advantage of the combined results given by Figure 11 is that they can easily be compared with the results obtained for the principal component analysis that are represented by Figure 6.

Conclusions for the Analysis of Distances Data

A first conclusion of the analysis is that the results are accentuating a double diagonal structure of the distances data, i.e. a first diagonal running from north-west to south-east and a second diagonal running from north-east to south-west. This is not surprising given the fact that the diagonal structured clusters are wholly in line with the longest distances between *généralités* and *départements*.

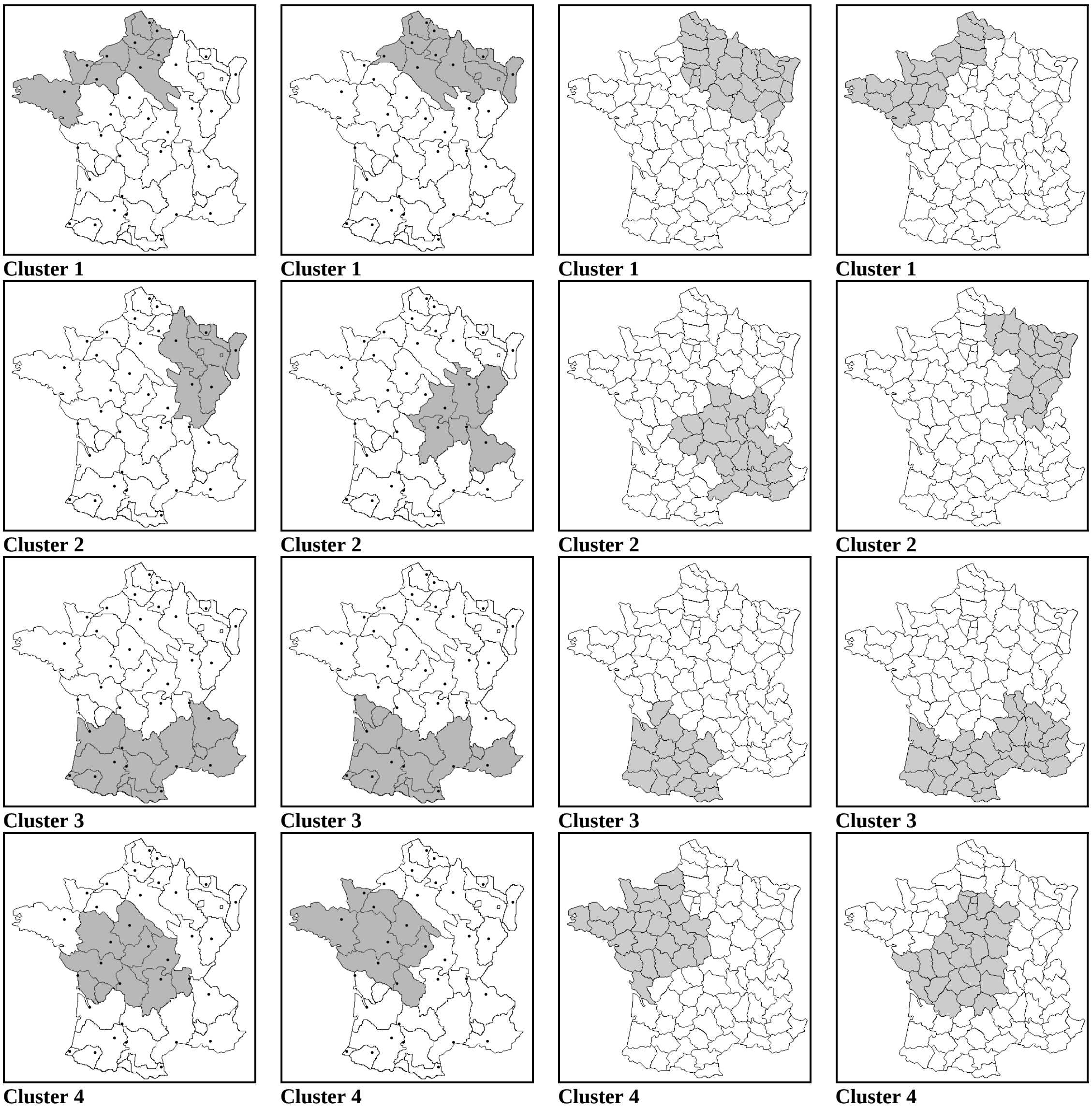
A second conclusion is that for both the *généralités* and the *départements* the results from the principal component analysis are confirmed by the results obtained for both of the clustering methods.

A third conclusion is about the differences between the results obtained for the two clustering methods. As can be seen from Figure 11 the results given by the k-means clustering for the *généralités* can best be compared with the results for the *départements* obtained by Ward's method while the results for the *généralités* obtained by Ward's method are consistent with the results for the *départements* obtained by the k-means clustering.

A last and summarizing conclusion that follows from the analysis is that even with only 32 points of measurement the underlying structure of the distances among the *généralités* can be compared with the structure of the distances among the 85 points of measurement for the *départements*. Although this generalizing conclusion is not really based on proper statistical testing it seems to be tenable and highly reliable.

Figure 11 : Distances - Généralités and Départements - Cluster Analysis
k-Means Clustering and Ward's Tree Clustering

Généralités		**Départements** (°)	
k-Means Clustering	**Ward's Method**	**k-Means Clustering**	**Ward's Method**



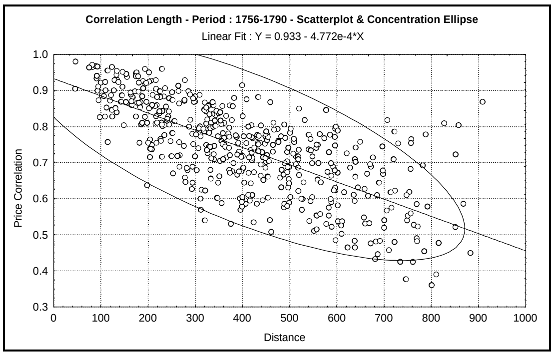
(°) The *départements* excluded from the analysis, i.e. Alpes-Maritimes, Savoie and Haute-Savoie, are left blank. See Appendix 3.

Section 4 : Distances versus Price Correlation 1

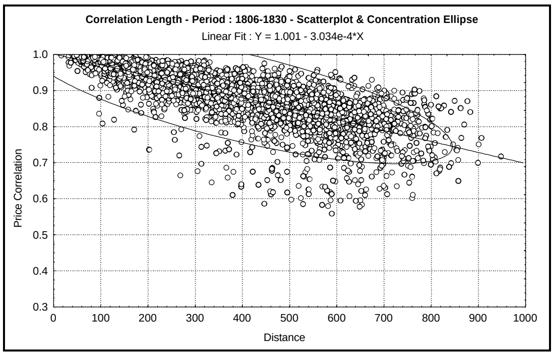
In the following paragraphs the relationship between the price correlations of wheat and the distance between the regional points of measurement will be analyzed in more detail. In other words the dependence of the correlation between the price fluctuations of wheat in different markets will be investigated with respect to the distance between these markets. The main and underlying assumption used is that the degree to which prices in different markets or regions fluctuate in unison reflects the degree to which those markets are associated or even integrated. It is evident that market integration need not to be the only cause of an increased synchronization of local price movements. Major exogenous events felt throughout all the regions may result in similar short-time price movements being experienced in regional markets that are even not integrated. Extreme weather conditions affecting whole regions and resulting in abundant or bad harvests might be a good and realistic example.

Global Results - Whole period

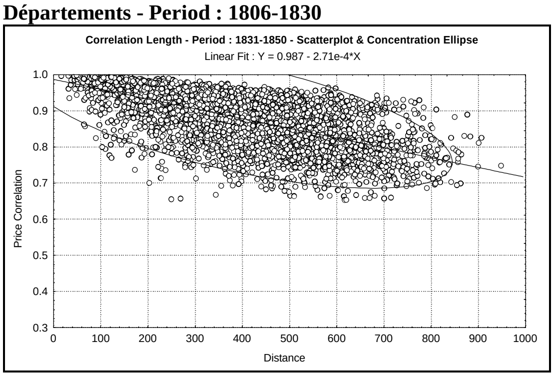
**Figure 12 : Price Correlation versus Distance - Global Results - Subperiods
Linear Fit - Scatterplot - Concentration Ellipse and Regression Line**



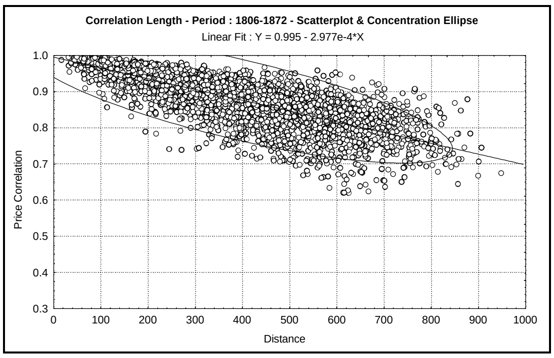
Généralités - Period : 1756-1790



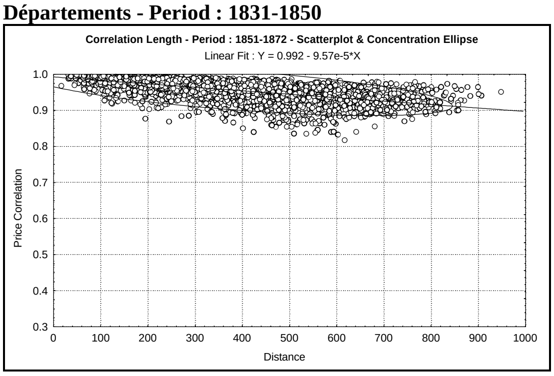
Départements - Period : 1806-1830



Départements - Period : 1831-1850



Départements - Period : 1806-1872



Départements - Period : 1851-1872

The first results are called ‘Global Results’ and are related to the entire sample of *généralités* and *départements*. A graphical representation of the relationship between correlations and distances is given in the left column of Figure 12. The given graphs consist of a scatterplot, a concentration ellipse and the linear regression line estimated by the method of ordinary least squares. Numerical results about the estimation can be found in Table 13. The ellipse is based on the assumption that the two variables are generated by a bivariate normal distribution. It shows the prediction interval for a single new observation given the parameter estimates defining the bivariate distribution. From these results it can be seen that the estimation of the regression lines is based on the elements above or below the diagonal of the symmetric correlation and distance matrices. This means that the zero distances and the unity correlations are excluded from the estimation.

Table 13 : Price Correlation versus Distance - Global Results - Subperiods
Linear Fit - Numerical Results (°)

		Linear Least Squares Fit				
Period	**Distance km**	**Slope**	**St. Error**	**t-Value**	**Nobs**	$\bar{R}^2$
1756 - 1790	**0 - 1000**	-0.000477	0.000021	22.7336	496	0.5103
	0 - 250	-0.000628	0.000106	5.9167	121	0.2208
	250 - 500	-0.000438	0.000081	5.3882	229	0.1095
	500 - 750	-0.000568	0.000132	4.2916	120	0.1277
	750 - 1000	*0.000664*	*0.000621*	*1.0686*	*26*	*0.0056*
1806 - 1830	**0 - 1000**	-0.000303	0.000005	65.9307	3570	0.5491
	0 - 250	-0.000307	0.000017	18.1296	918	0.2633
	250 - 500	-0.000341	0.000017	19.8972	1609	0.1972
	500 - 750	-0.000241	0.000032	7.4321	934	0.0549
	750 - 1000	*-0.000147*	*0.000160*	*0.9217*	*109*	*0.0000*
1831 - 1850	**0 - 1000**	-0.000271	0.000006	48.3534	3570	0.3957
	0 - 250	-0.000339	0.000028	11.9304	918	0.1335
	250 - 500	-0.000280	0.000023	12.4237	1609	0.0871
	500 - 750	-0.000266	0.000034	7.8094	934	0.0604
	750 - 1000	*-0.000166*	*0.000139*	*1.1953*	*109*	*0.0040*
1851 - 1872	**0 - 1000**	-0.000096	0.000002	47.3890	3570	0.3861
	0 - 250	-0.000113	0.000009	12.0273	918	0.1354
	250 - 500	-0.000121	0.000008	15.2833	1609	0.1264
	500 - 750	-0.000058	0.000013	4.5802	934	0.0210
	750 - 1000	*0.000072*	*0.000056*	*1.2902*	*109*	*0.0061*
1806 - 1872	**0 - 1000**	-0.000298	0.000004	71.2202	3570	0.5869
	0 - 250	-0.000322	0.000016	19.7011	918	0.2969
	250 - 500	-0.000351	0.000016	21.9656	1609	0.2304
	500 - 750	-0.000240	0.000029	8.3304	934	0.0683
	750 - 1000	-0.000338	0.000125	2.7022	109	0.0551

(°) Slope coefficients not significant different from zero are printed in italics.

Table 14 : Price Correlation versus Distance - Global Results - Subperiods
Distance in km for Correlation Reduction of 0.01 - Rounded

	Period				
Distance km	**1756-1790**	**1806-1830**	**1831-1850**	**1851-1872**	**1806-1872**
0 - 1000	21	33	37	104	34
0 - 250	16	33	29	88	31
250 - 500	23	29	36	83	28
500 - 750	18	41	38	172	42
750 - 1000	*-15*	*68*	*60*	*-139*	30

The global results obtained for the *généralités* (1756-1790) and the *départements* (1806-1872) are sharing at least one common characteristic. From the left column of Figure 12 it can be seen that the larger the distance between the locations the larger will be the variance around the fitted linear regression line. Apart from this common characteristic the results are also showing a striking difference between the two periods. This difference can best be illustrated by comparing the slope parameters of the estimated linear regression lines. From Table 13 it can be seen that for the period 1756-1790 an increase of the distance between the locations with one km will result in a reduction of the price correlation with 0.000477 while for the period 1806-1872 this reduction is only 0.000298.

In order to improve the accessibility of the estimated results for the slope parameter one can use the information presented by Table 14. This table gives the distance in km in order to reduce the price correlation between two locations with 0.01. From these results it follows that for the *généralités* this reduction is reached after 21 km while for the *départements* this reduction will only be reached after 34 km. Under the assumption that the influence of distance on the correlation of wheat prices can be used as a measure of market integration one can conclude that the wheat market in the period 1806-1872 is characterized by a stronger integration than in the period 1756-1790.

Global Results - Subperiods

One drawback of the previous global results is the high degree of aggregation of the available sample information. In a first attempt to disaggregate the data it was decided to split up the period 1806-1872 into the three subperiods of comparable size, i.e. 1806-1830, 1831-1850 and 1851-1872. The global results for each of these three subperiods are graphically represented as the right column of Figure 12 and tabulated in Table 13 and Table 14.

These results reveal at least two interesting conclusions. A first conclusion is the resemblance of both the graphical and the numerical results obtained for the subperiods 1806-1830 and 1831-1850 with those obtained for the whole period 1806-1872. Whereas for the whole period a reduction of the price correlation of 0.01 is obtained after 34 km this distance is 33 km and 37 km for respectively the first and second subperiod. A second conclusion is the totally different behavior of the relationship between price correlations and distances obtained for the subperiod 1851-1872. For this period the distance to reduce the correlation with 0.01 is 107 km, i.e. almost three times the distance obtained for the first and second subperiod.

From all these results it follows that :

- the results obtained for the entire period 1806-1872 are definitely not representative for the subperiod 1851-1872
- a real break-through of the integration of the wheat market took place in the period 1851-1872

Global Results - Whole Period & Subperiods - Scatterplot

In the previous paragraphs the results for the correlation-distance relationship were presented by using a traditional scatterplot. An inconvenient drawback of a scatterplot is the exact or partial overlap of the scattered points of the graph. When multiple data points with (almost) the same coordinates are plotted on a scatter graph it is (almost) impossible to know how many data points are present unless some additional method is used to encode that information into the data graphic. A first solution for the exact overlap of graph locations is to use the jittering technique, i.e. adding a small amount of random noise to the data before graphing. In other words the data points on the graph are randomly shifted a slight amount so no symbol obscures another. A second graphical method that can relieve the exact and even partial overlap consists in using the so called sunflower technique. A sunflower symbol is a dot with short radiating lines called petals. The number of these radiating lines corresponds to the number of data points represented by the symbol (See **Cleveland [6, p. 161 and p. 163]**.

If however the scatterplot consists of a large number of points none of these methods will help to reveal the density of points. Since the number of points for the scatterplots of Figure 12 is 496 for the *généralités* and 3570 for the *départements* an alternative graphical representation would be a valuable help. An alternative method to avoid the problem of overlap is to consider the data as discrete information. An advantage of this rather drastic changing approach is that the frequency characteristics of the bivariate data will be (over) emphasized. The practical consequence of this approach is that the distance and correlation variable are both categorized in arbitrary defined intervals resulting in the bivariate frequency tables represented in Table 15. In order to get a more detailed tabulated representation of these bivariate distributions it was decided to use a fixed interval length of 200 km for the distance variable. In other words the distances were categorized in five intervals of 200 km each. In order to facilitate the interpretation a varying interval length was used to categorize the price correlations.

To visualize the cross-tabulation of the values of the two variables given by Table 15 a three-dimensional histogram can be used. Each of the bivariate histograms, given in the left column of Figure 13, can be considered as a conjunction of two simple univariate histograms. The two univariate histograms are combined in such a way that the frequencies of the co-occurrences of values of the two variables analyzed can easily be examined. One of the main major reasons why frequency distributions are of particular interest is that from the shape of the distributions one can learn about the nature of the examined variables. However the overall shape and the global descriptive characteristics of bivariate distributions can be much easier explored by using a graphical representation than by consulting tabulated numerical results.

Table 15 : Price Correlation versus Distance - Bivariate Frequency Distribution

Généralités - Period : 1756-1790

Correlation Interval	Distance Interval km 0-200	200-400	400-600	600-800	800-1000	0-1000
0.9 - 1.0	35	10				45
0.8 - 0.9	37	73	11	1	3	125
0.7 - 0.8	3	66	71	14	1	155
0.6 - 0.7	1	26	46	21		94
0.5 - 0.6		8	26	20	2	56
0.4 - 0.5			1	15	2	18
0.3 - 0.4				1	2	3
0.0 - 1.0	76	183	155	72	10	496

Départements - Period : 1806-1830

Correlation Interval	Distance Interval km 0-200	200-400	400-600	600-800	800-1000	0-1000
0.9 - 1.0	602	845	193	15		1655
0.8 - 0.9	21	404	766	267	14	1472
0.7 - 0.8	1	17	143	165	19	345
0.6 - 0.7		10	30	37	9	86
0.5 - 0.6			7	5		12
0.0 - 1.0	624	1276	1139	489	42	3570

Départements - Period : 1831-1850

Correlation Interval	Distance Interval km 0-200	200-400	400-600	600-800	800-1000	0-1000
0.95 - 1.00	440	359	40			839
0.90 - 0.95	108	432	340	63	1	944
0.85 - 0.90	48	220	229	53	3	553
0.80 - 0.85	17	163	232	116	9	537
0.75 - 0.80	10	72	199	147	17	445
0.70 - 0.75	1	24	79	88	8	200
0.65 - 0.70		6	20	22	4	52
0.00 - 1.00	624	1276	1139	489	42	3570

Départements - Period : 1806-1872

Correlation Interval	Distance Interval km 0-200	200-400	400-600	600-800	800-1000	0-1000
0.9 - 1.0	596	801	178	12		1587
0.8 - 0.9	27	435	726	264	8	1460
0.7 - 0.8	1	40	222	178	26	467
0.6 - 0.7			13	35	8	56
0.0 - 1.0	624	1276	1139	489	42	3570

Départements - Period : 1851-1872

Correlation Interval	Distance Interval km 0-200	200-400	400-600	600-800	800-1000	0-1000
0.98 - 1.00	418	259	22	1		70
0.96 - 0.98	139	601	309	46	6	1101
0.94 - 0.96	50	271	362	108	10	801
0.92 - 0.94	12	95	246	180	13	546
0.90 - 0.92	4	32	105	105	8	254
0.88 - 0.90		14	49	42	5	110
0.86 - 0.88	1	4	30	3		38
0.84 - 0.86			11	2		13
0.82 - 0.84			5	1		6
0.80 - 0.82				1		1
0.00 - 1.00	624	1276	1139	489	42	3570

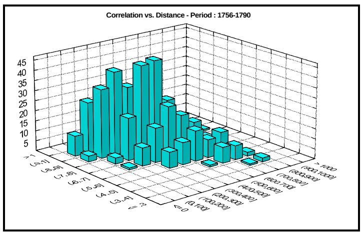
An important limitation of a three-dimensional graph is the fact that it is sometimes difficult or even impossible to see or interpret all of the data graphic when viewed from one single angle. A simple technique to overcome this problem is to rotate the entire graph. However even without such a rotation the interpretation of the general shape of the bivariate distribution of the correlation and distance variables can still be further improved. By using the appropriate smoothing and interpolation technique the bivariate histogram can be transformed to a 3D surface plot. A surface graph is a three-dimensional wireframe graph in which the areas between the wires (lines) are opaque. The construction of the surface plots, given in the second column of Figure 13, is characterized by the combined use of both wireframes and different colors or shades of gray. The lines of the wireframes have the form of a fishnet and are connecting equal X and Y values. The different shades of gray represent the contour of equal Z values, i.e. the frequencies represented on the vertical axis.

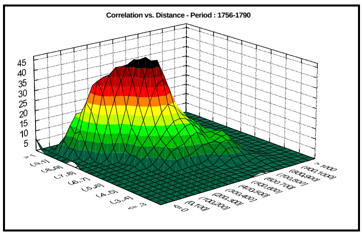
As can be seen from the plots of Figure 13 the back transformation from the discrete to the continuous metric resulted in a much better interpretation of the general shape of the observed discrete bivariate distribution, resulting from the simultaneous behavior of the regional price correlations on one side and the distances between these regions on the other. However the price that had to be paid is the loss of a small amount of information, smoothed away by the transition from the discrete to the continuous environment.

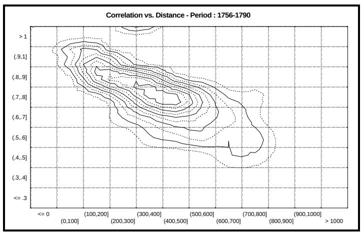
An alternative graphical representation that can be useful for the interpretation of the simultaneous behavior of the price and distance variables is the two-dimensional contour plot. This graph, represented in the third column of Figure 13, is the result of the projection on the horizontal X-Y plane of the gray shaded contour lines of the 3D surface plot of the second column. In this respect a 2D contour plot can be seen as a two-dimensional version of the 3D surface plot. By supplementing these 2D contour plots with the fitted linear regression line one gets the graphical representations given by Figure 14. These graphs are combining the advantages of a regular scatterplot with the characteristics of the density of the bivariate relationship between price correlations and distances.

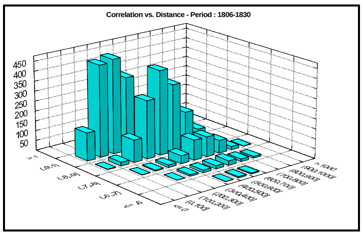
Figure 13 : Price Correlation versus Distance - Bivariate Distribution 1
Bivariate Histogram - Surface Plot - Contour Plot

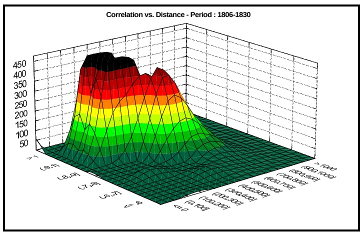
Bivariate Histogram	Surface Plot	Contour Plot

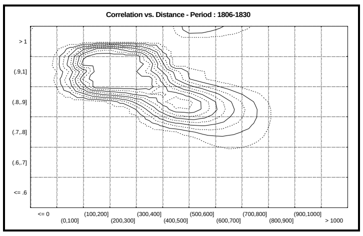

Généralités - Period : 1756-1790

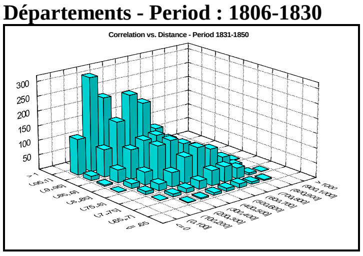

Généralités - Period : 1756-1790

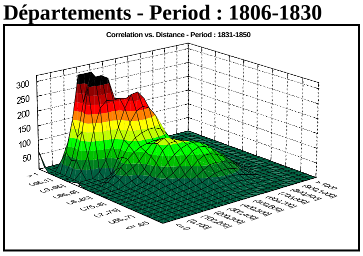

Généralités - Period : 1756-1790

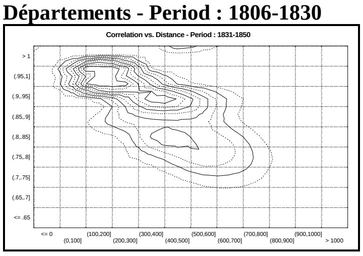

Départements - Period : 1806-1830

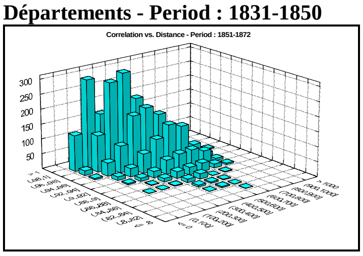

Départements - Period : 1806-1830

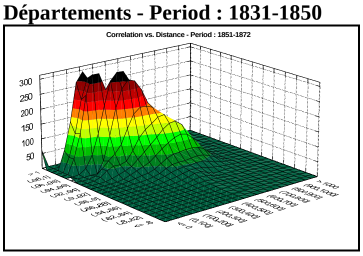

Départements - Period : 1806-1830

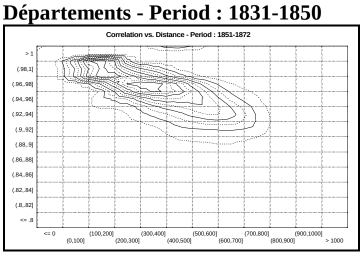

Départements - Period : 1831-1850

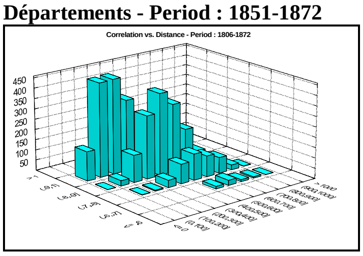

Départements - Period : 1831-1850

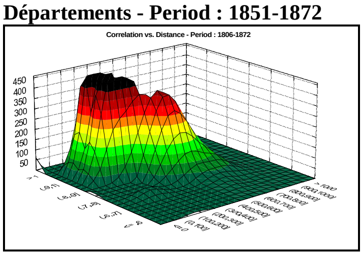

Départements - Period : 1831-1850

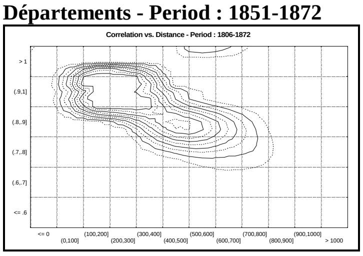

Départements - Period : 1851-1872


Départements - Period : 1851-1872


Départements - Period : 1851-1872


Départements - Period : 1806-1872

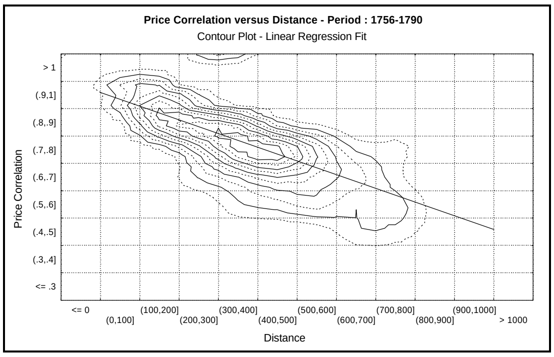

Départements - Period : 1806-1872


Départements - Period : 1806-1872

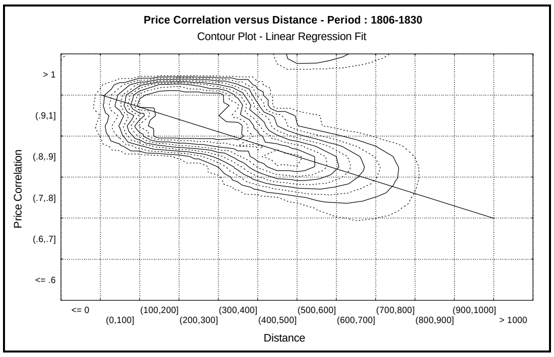
Global Results - Distance Subintervals

A second attempt to reduce the aggregation is to split up the total distance of 1000 km into four intervals of 250 km each. The main purpose of using these four subintervals is to verify if the estimated reduction of the price correlation is representative for the whole distance range. In other words with this attempt one can verify whether or not one and the same linear relation could be used for the whole range of distances. Combining the disaggregation of the total distance range into subintervals with the disaggregation into subperiods leads to the detailed numerical results reported in Table 13 and Table 14. Graphical representations for these cases can be found in Figure 15. The first row of Figure 15 consists of the same graphs as those represented by Figure 12. They can be used to be compared with the graphs for each of the subintervals. Apart from the estimated regression line and the concentration ellipse for the whole sample the latter consist of the scatterplot with accompanying ellipse and the linear fit for the subinterval.

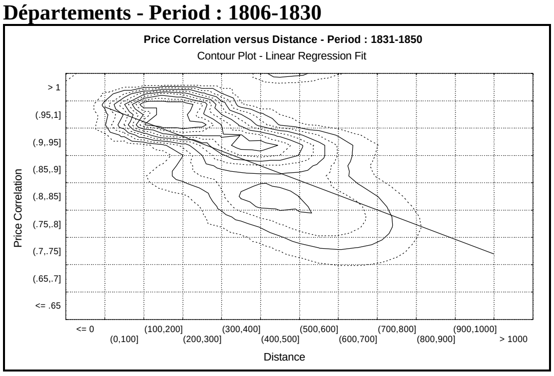
Figure 14 : Price Correlation versus Distance - Bivariate Distribution 2
Contour Plot - Linear Regression Fit



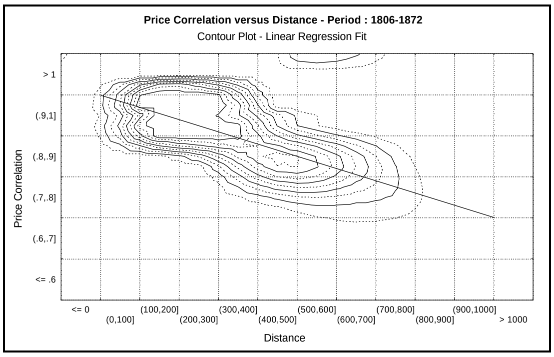
Généralités - Period : 1756-1790



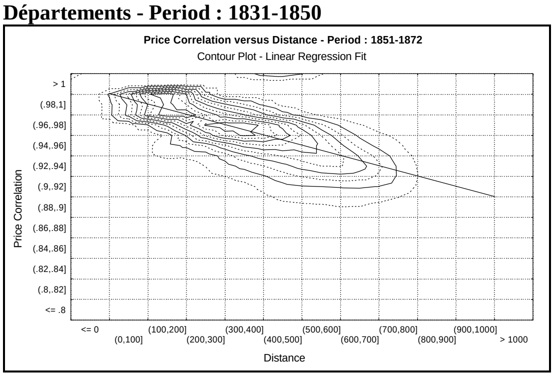
Départements - Period : 1806-1830



Départements - Period : 1831-1850



Départements - Period : 1806-1872



Départements - Period : 1851-1872

The most relevant results obtained for the subintervals can be summarized as follows. In the first place the results obtained for distances between 750 and 1000 km need to be mentioned. The only significant linear relationship for this distance interval was obtained for the whole sample for the period 1806-1872. For all other cases the slope parameter for this interval was not significant different from zero. No wonder that in two cases, i.e. for the period 1756-1790 and 1851-1872, the estimation resulted even in a positive slope. Since for the period 1851-1872 all the estimates are based on 109 observations it is hard to see how these (disappointing) results could be attributed to the smaller number of observations. Therefore one can conclude from these results that the linear relationship between price correlations on one side and distances larger than 750 km on the other is at least highly questionable.

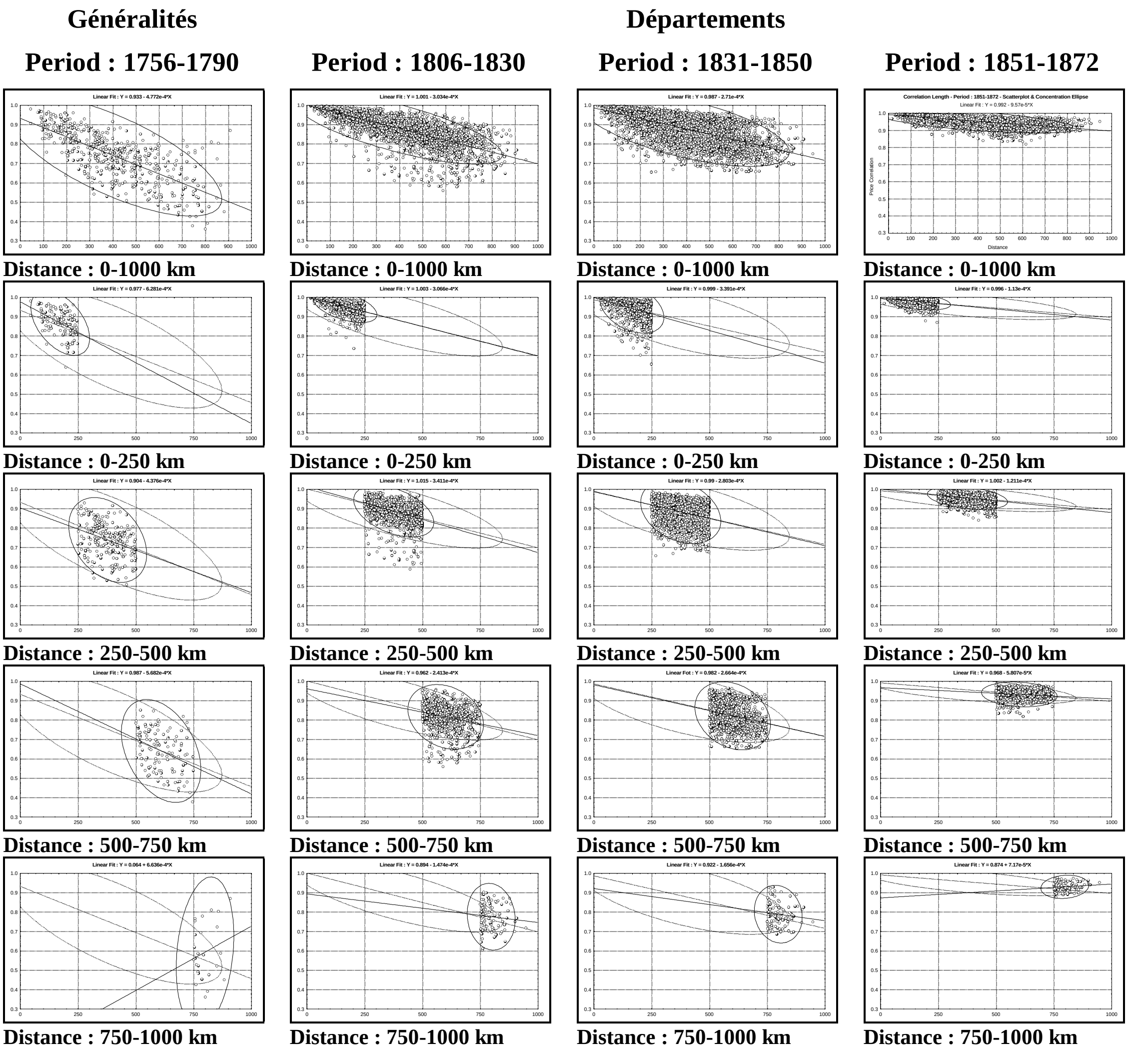
The second conclusion relates to the results for the first three intervals. For all of these results the specific concentration ellipse for the interval lies almost entirely within the constructed ellipse describing the entire scatterplot. The second characteristic of these results is that for each of the periods and subperiods the estimated slope parameters are rather close to each other. The only exception is the slope parameter obtained for the subperiod 1851-1872 and more in particular for distances between 500 and 750 km. In this case a reduction of the price correlation with 0.01 is only obtained after 172 km.

Local Results - Subperiods

In the previous paragraphs the results were based on the entire sample. The derived linear relationship between price correlations and distances refers to all the *généralités* and all the *départements*. In the next paragraphs this linear relationship will be applied to each of the locations. In other words instead of using all the elements below or above the diagonal of the correlation and distance matrices a specific column of the correlation matrix will be related to the corresponding column of the distance matrix. Therefore the thus derived results will be called 'Local Results'.

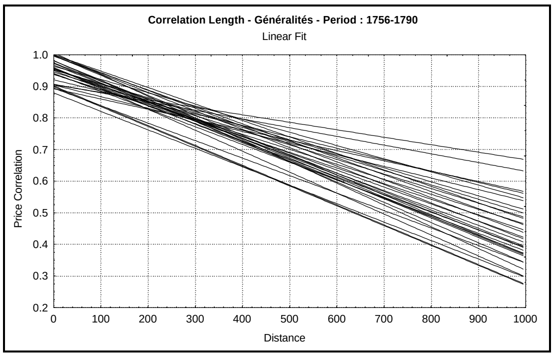
The main consequence of this approach is that for each period and subperiod there will be as many relationships as there are locations. In order to avoid too much clutter only the estimated regression lines will be graphically represented. The resulting graphs are given in Figure 16. In an effort to reduce the abundance of local results it was decided not to report the estimated slopes but to give only the main summarizing descriptive statistics. The numerical results are tabulated in Table 16.

Figure 15 : Price Correlation versus Distance - Global Results - Subperiods
Linear Fit - Scatterplot - Concentration Ellipse and Regression Line

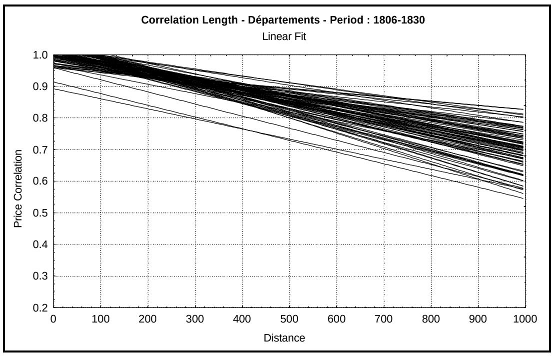


The interpretation of the summarizing local results leads to an important conclusion about the behavior of the correlation decrease for the period 1831-1850. One of the conclusions about the global results was that the influence of the distance on the price correlation for this period was very close to the results obtained for the subperiod 1806-1830. From the results for the local samples however it can be seen that this conclusion stands as far as the mean of the slope parameter is concerned but definitely not if also the variability is taken into account. From the graphical representations of Figure 16 and the numerical results of Table 16 it follows that the variability of the local slope parameters for the period 1831-1850 is larger than for the period 1806-1830, i.e. 0.000120 against 0.000077, and even larger than for the period 1756-1790, i.e. 0.000120 against 0.000113. Based on these results one could conclude that the period 1831-1850 can be seen as a transition period in a growing integrating wheat market.

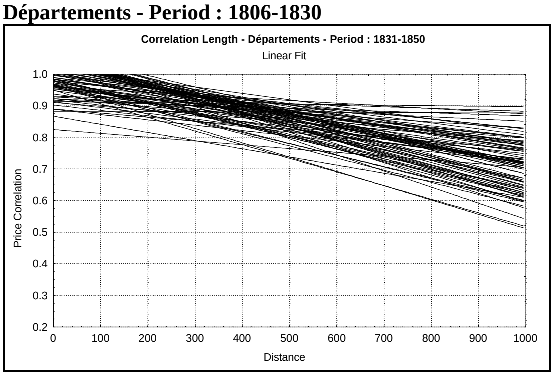
Figure 16 : Price Correlation versus Distance - Local Results - Subperiods
Linear Fit - Regression Line



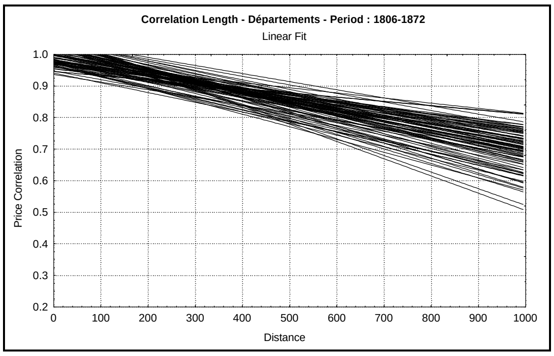
Généralités - Period : 1756-1790



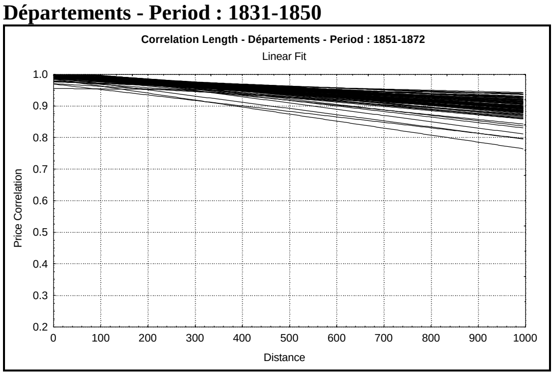
Départements - Period : 1806-1830



Départements - Period : 1831-1850



Départements - Period : 1806-1872



Départements - Period : 1851-1872

A last remark must be made about the comparability of the results based on the aggregated samples and the local results which were derived from the disaggregared samples. Whereas the unity correlations and the zero distances were left out from the aggregated samples these values were included in the local samples.

Linearity of Correlation Decrease

In the previous sections a linear relationship was used to describe the association between the price correlations and the distances between the locations. It must be admitted that a linear relationship is not the only possible functional form to describe the correlation decrease. Based on theoretical grounds the hypothesis of a linear relationship might even be rejected since this would imply that at some finite distance the correlation would become zero. A theoretical and more realistic hypothesis would be to describe the relationship between decreasing correlations and distance by an exponential function. The advantage of this hypothesis is that at least the interpretation of the two extreme parts of the theoretical decreasing curve becomes much more realistic, i.e. the left-hand part of the curve will be characterized by the correlation tending to one as the distance tends to zero while at the right-hand part of the curve the correlation will tend to zero as the distance becomes very large.

Table 16 : Price Correlation versus Distance - Global and Local Results - Subperiods Linear Fit - Slope Parameter

		Period				
Procedure	**Results**	**1756-1790**	**1806-1830**	**1831-1850**	**1851-1872**	**1806-1872**
Whole Sample	**Global**					
	Years	35	25	20	22	67
	Equations	1	1	1	1	1
	Observations	496	3570	3570	3570	3570
	Estimate	-0.000477	-0.000303	-0.000271	-0.000096	-0.000298
	Standard Error	0.000021	0.000005	0.000006	0.000002	0.000004
Local Samples	**Local**					
	Years	35	25	20	22	67
	Equations	32	85	85	85	85
	Observations	32	85	85	85	85
	Minimum	-0.000688	-0.000483	-0.000509	-0.000222	-0.000546
	Maximum	-0.000237	-0.000132	-0.000006	-0.000015	-0.000135
	Mean	-0.000522	-0.000305	-0.000278	-0.000102	-0.000305
	Median	-0.000557	-0.000306	-0.000280	-0.000102	-0.000297
	Standard Deviation	0.000113	0.000077	0.000120	0.000036	0.0000082

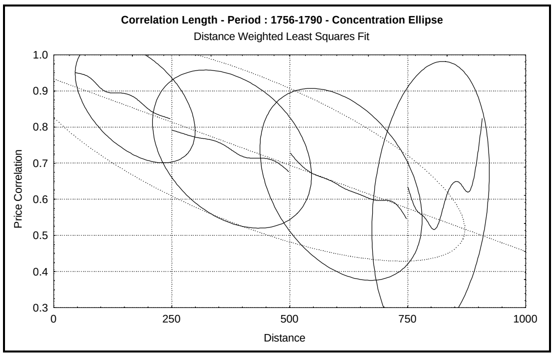
The decision about the most appropriate functional form of the relation between correlation and distance can be based on the outcome of the statistical hypothesis testing of the linear and the exponential alternative. The criterion might be some goodness-of-fit statistic obtained for each of the two functional forms. The results of these testing procedures are not shown for the simple reason that without any exception all these results are supporting the linear relation.

Instead of reporting all the numerical results it was decided to illustrate the appropriateness of the linear relationship by using the graphical representations that can be found in Figure 17 and Figure 18. For each of these graphs the whole sample is represented by the linear fit and the concentration ellipse, both represented by a dotted line. Proceeding as in the previous sections the whole distance range was divided into four subsections of 250 km each. For each of these subsections the sub-samples are represented by their local concentration ellipse and local distance weighted least squares fit. These local results are represented by using a solid line pattern. It must be remembered that the distance weighted least squares fit can be seen as an adaptive and non-parametric fit that can be compared by some weighted moving average procedure. This method is highly appropriate when the variances of the residuals are not constant over the range of the values of the independent variable. The main difference between Figure 17 and Figure 18 is that for the latter a separate graph is used for each of four sub-samples. These separated graphs largely facilitate the comparison of the results obtained for each of the four subperiods.

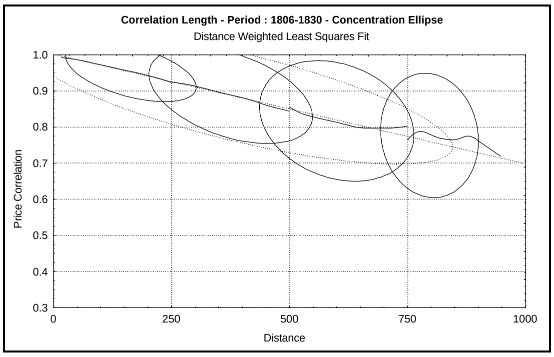
The interpretation of the graphical results represented by Figure 17 and Figure 18 leads to the following conclusions :

- with the exception for the sub-samples representing distances in the interval 750-1000 km the local distance weighted least squares fits are very close to the global linear fit
- the local distance weighted least squares fits for distances in the range from 750 to 1000 km are not always supporting the linear functional form obtained for the other three sub-samples
- in none of the graphs evidence could be found to confirm an exponential relationship

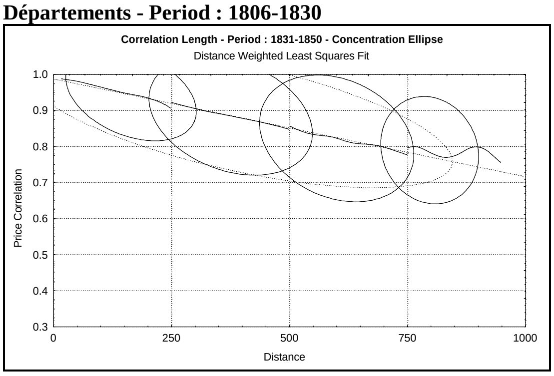
Figure 17 : Price Correlation versus Distance - Généralités and Départements
Distance Weighted Least Squares Fit - 1



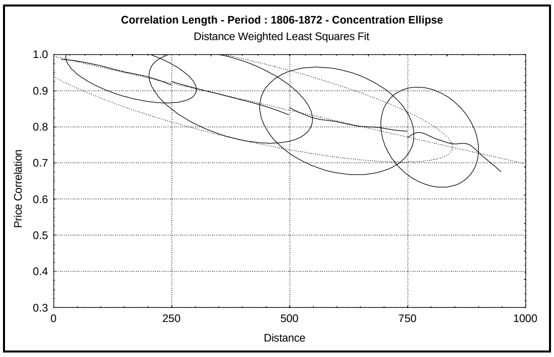
Généralités - Period : 1756-1790



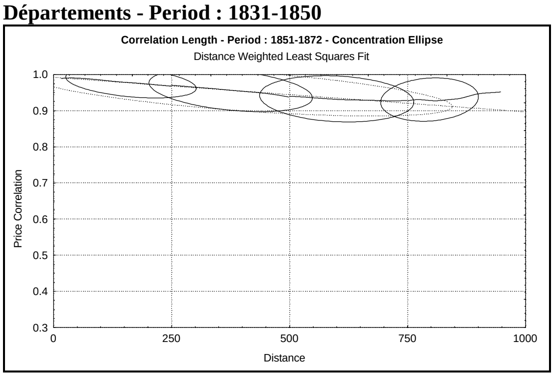
Départements - Period : 1806-1830



Départements - Period : 1831-1850

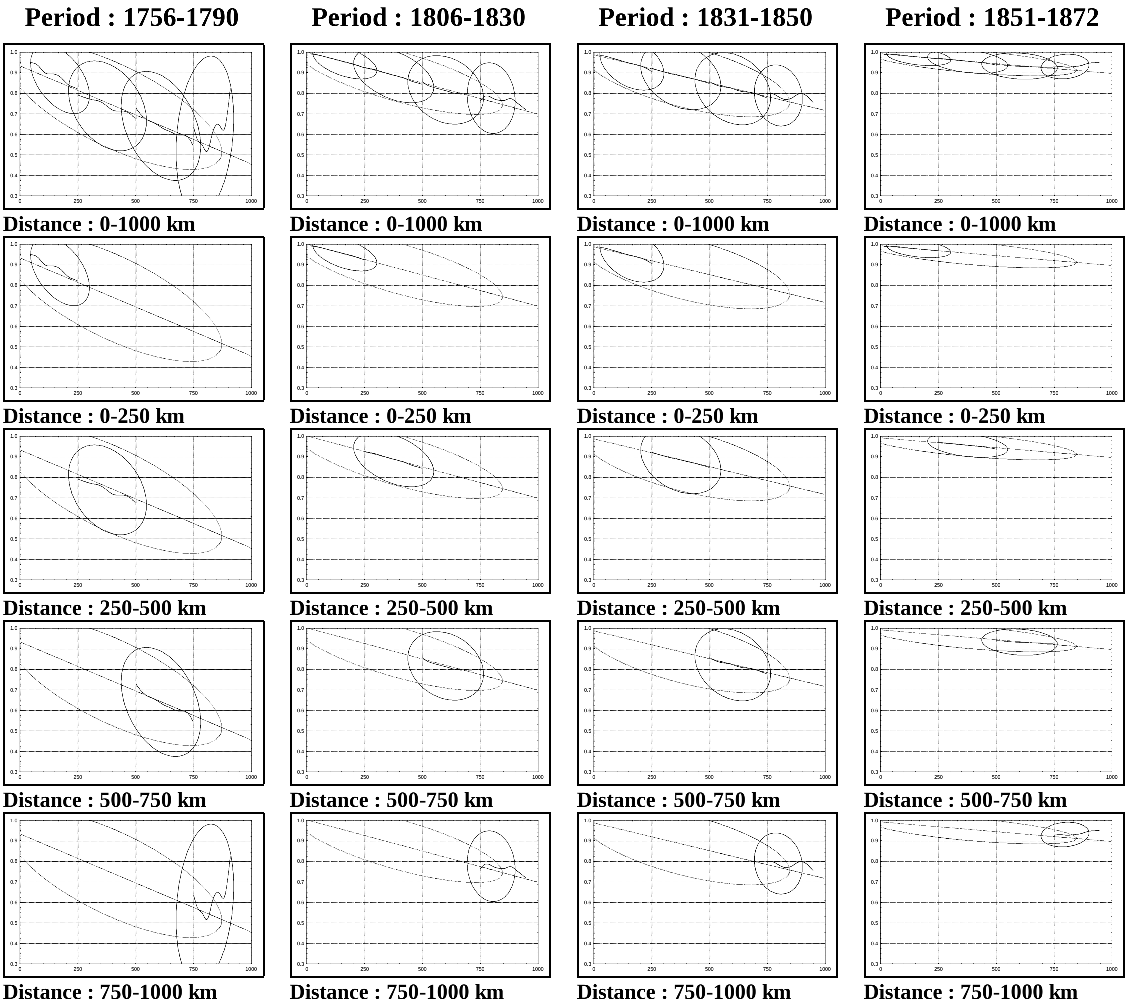


Départements - Period : 1806-1872



Départements - Period : 1851-1872

Figure 18 : Price Correlation versus Distance - Généralités and Départements
Distance Weighted Least Squares Fit - 2

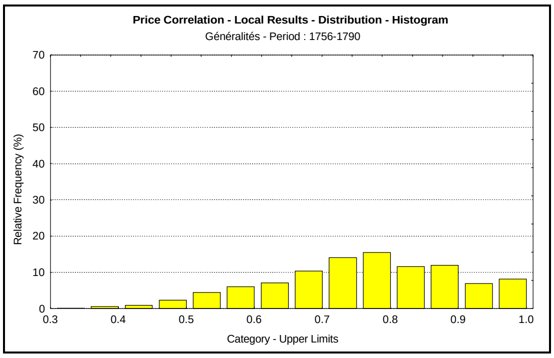


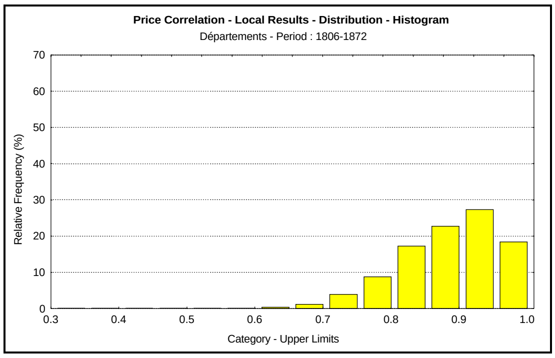
Section 5 : Distances versus Price Correlation 2

In the previous section the linear relationship between the correlation of the prices of wheat and the distances between the locations was investigated. The disaggregation of the available data resulted even in rather specific conclusions. However none of these attempts resulted in conclusions that could be related to the regional behavior and diversity of this relationship. In this section special attention will be paid to the spatial characteristics of this behavior.

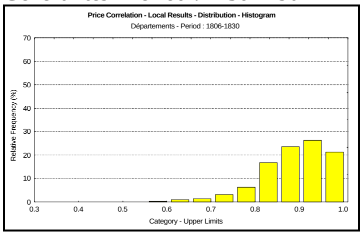
Among the derived results in previous sections one of the main conclusions was the changing pattern of the price correlations when the period 1756-1790 is compared with the period 1806-1872. These results can be summarized by the graphical representations of Figure 19. From these graphs it can be seen that the distribution of the price correlations, while remaining left-skewed, is shifted to the right. In other words the price correlations for the period 1806-1872 are more concentrated around higher values compared by those obtained for the period 1756-1790. The graphs on the second row of Figure 19 are illustrating that this shift is mainly occurring during the subperiod 1851-1872.

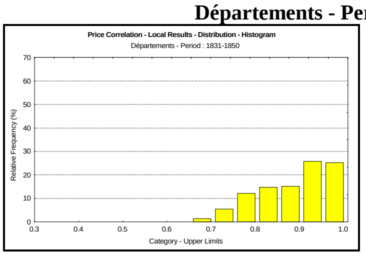
Figure 19 : Price Correlation - Subperiods - Histogram

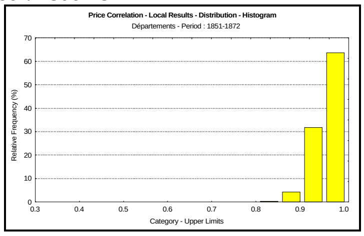




Généralités - Period : 1756-1790 **Départements - Period : 1806-1872**







Départements - Period : 1806-1830 **Départements - Period : 1831-1850** **Départements - Period : 1851-1872**

The general shift to higher values not necessarily means that no regional differences might be found among the correlations. This can be illustrated by the examples given by Figure 20 and Figure 21. These graphical representations consist of contour plots of the spatial distributions of the price correlations for the central and outermost locations for each of the periods 1756-1790 and 1806-1872. It must be reminded that a contour plot can be seen as a two dimensional representation of a smoothed three dimensional scatterplot. From the comparison of the graphs for the illustrative five selected locations one can conclude that the spatial distribution of the correlations not only differ from each other but also between the two periods.

Based on previous reasoning one can conclude that, given the comparable and constant pattern of distances, the relationship between the changing correlations among the prices of wheat and the distances between the locations must also have been changing. This conclusion seems to be justified and can be illustrated by the graphs presented in Figure 22 and Figure 23. For both the periods 1756-1790 and 1806-1872 these graphs represent the linear relation between price correlations and distances for the same locations as those used in Figure 20 and Figure 21.

From the simple visual inspection of these graphs one comes to the conclusion that at least for the five given locations the linear relation between correlations and distances might depend on both the period and the geographical location. The problem remains as to whether or not also differences about this changing relationship could be found that are related to sub-regions of the French territory.

Figure 20 : Price Correlation versus Longitude and Latitude - Period : 1756-1790
Central and Outermost Locations - Quadratic Surface - Contour Plot

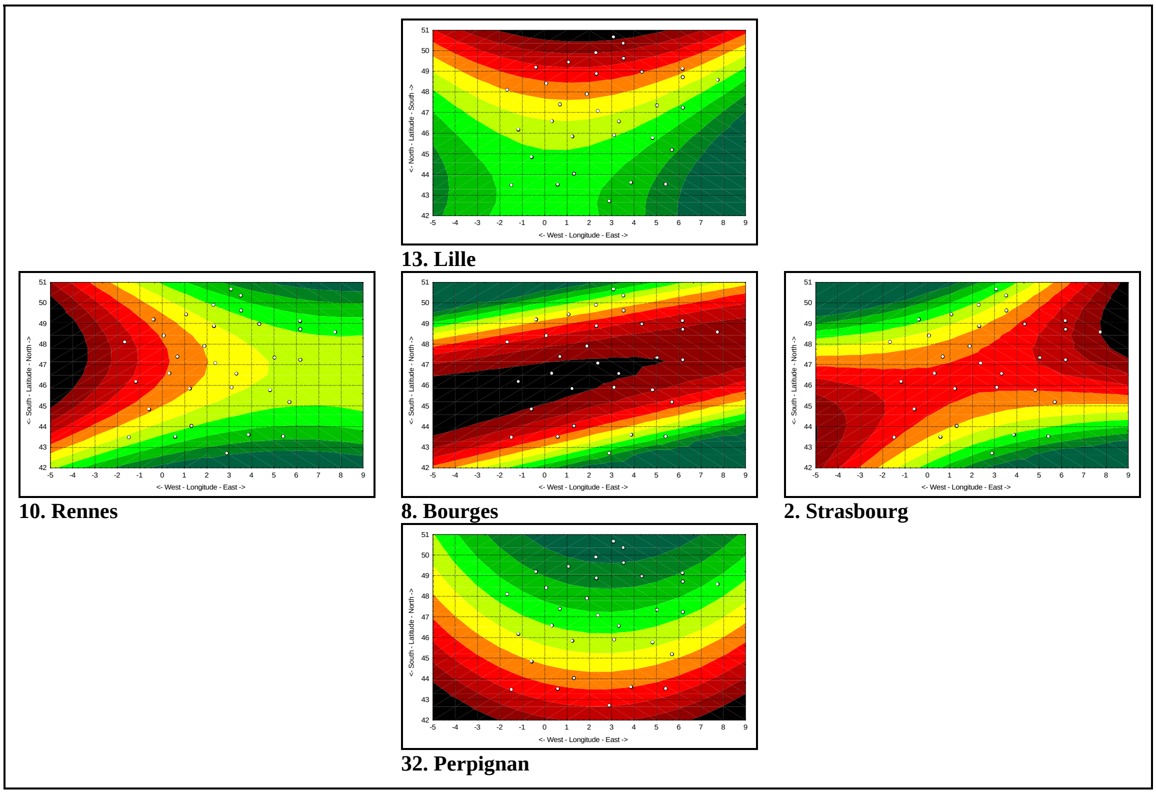


Figure 21 : Price Correlation versus Longitude and Latitude - Period : 1806-1872
Central and Outermost Locations - Quadratic Surface - Contour Plot

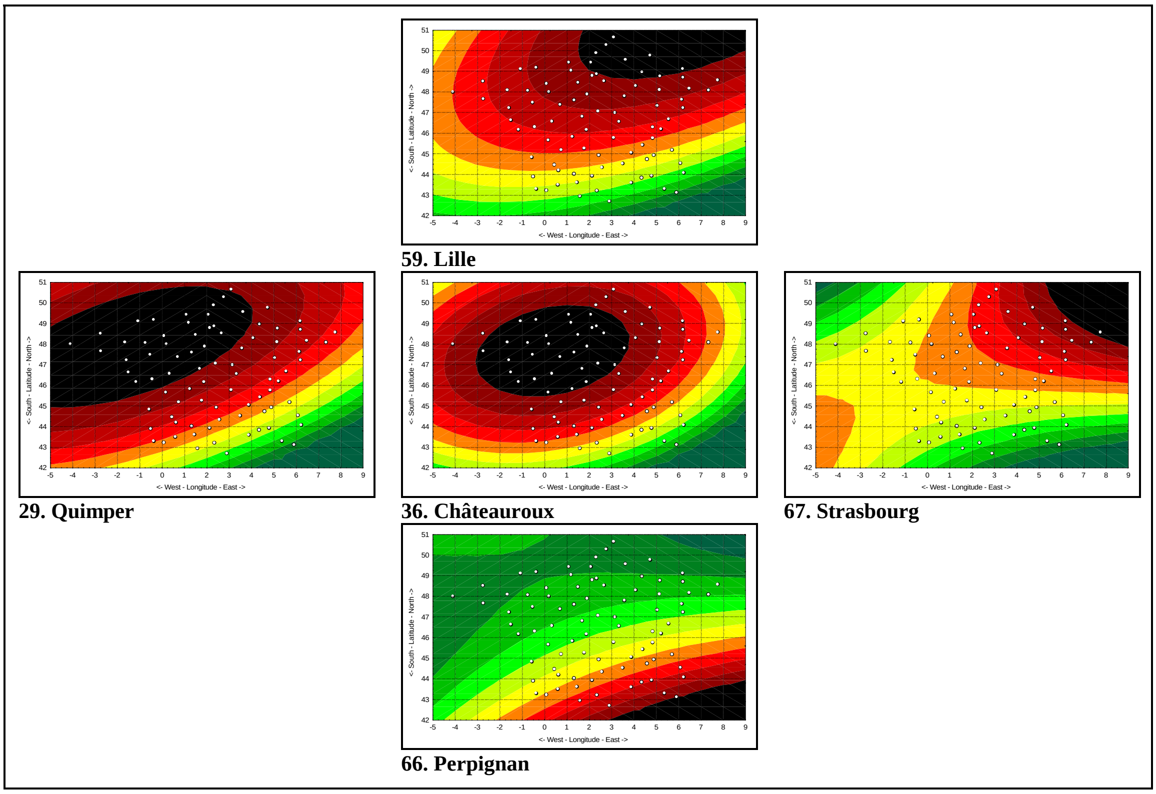
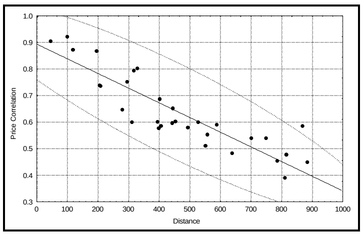
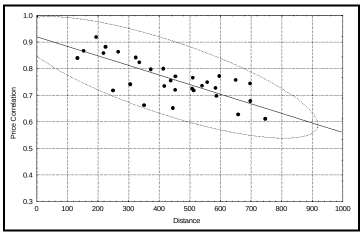


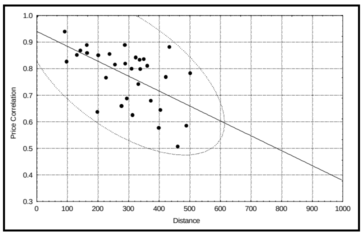
Figure 22 : Price Correlation versus Distance - Local Results - Period : 1756-1790
Central and Outermost Locations - Scatterplot - Ellipse - Linear Fit



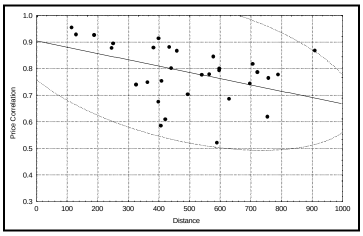
13. Lille



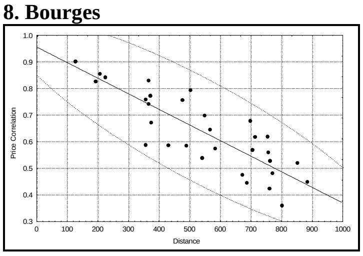
10. Rennes



8. Bourges

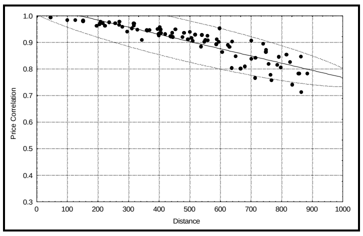


2. Strasbourg

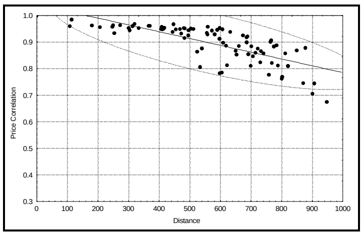


32. Perpignan

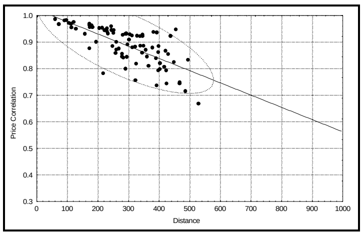
Figure 23 : Price Correlation versus Distance - Local Results - Period : 1806-1872
Central and Outermost Locations - Scatterplot - Ellipse - Linear Fit



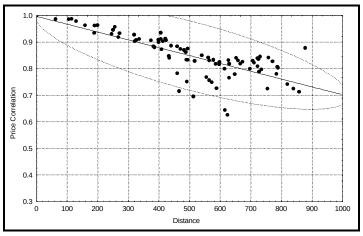
59. Lille



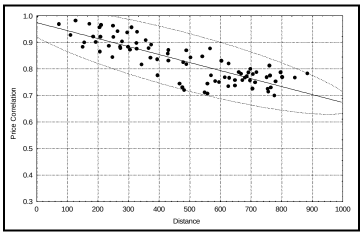
29. Quimper



36. Châteauroux



67. Strasbourg



66. Perpignan

Among the several results about the linear relation between correlations and distances at least two distinct features are of primary importance. The first of these two features is the estimated slope parameter, i.e. the quantified response of an increased distance on the price correlations. The second feature is the R^2-value, measuring the strength of the hypothesized linearity of the relationship. In the following paragraphs attention will be paid to each of these two features.

Correlation Length - Magnitude of Relationship

The use of the local results about the correlation-distance relationship leads to a specific slope parameter for each of the locations, i.e. 32 slope parameters for the *généralités* (1756-1790) and 85 slope parameters for the *départements* (1806-1872). The distribution of these estimated slopes can be found on the first and third row of Figure 24. The slope parameters can be graphically represented by a 3D scatterplot with the axes labeled as longitude, latitude and slope parameter value. In order to further simplify the interpretation this three dimensional scatter can be smoothed by fitting a quadratic surface through the scattered points. The resulting three dimensional graphs are represented as the second and fourth row of Figure 24.

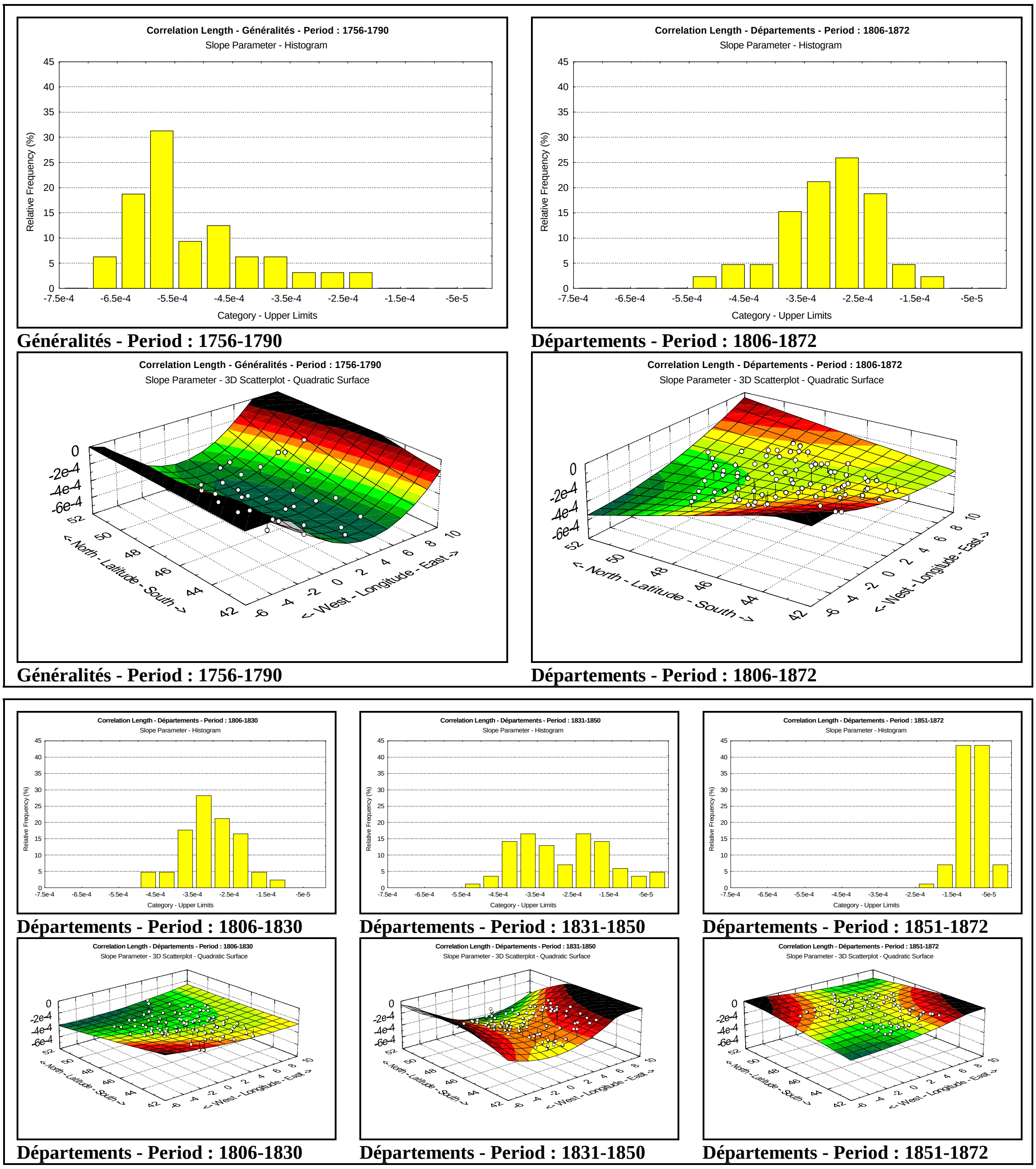
The results presented by Figure 24 can be used to compare the behavior of the slopes obtained before with those obtained after the French revolution. The first comparison that can be made is about the ordinary distribution of the slope parameters. At least two distinct aspects of these distributions can be distinguished. The first aspect is the general form of the distributions. From the graphs it can be seen that whereas the distribution of the 32 slopes for the period 1756-1790 is right skewed the distribution of the 85 slopes for the period 1806-1872 is skewed to the left. The second aspect is the location of each of the distributions. The distribution for the period 1806-1872 is clearly shifted to the right. In other words the response of the distances on the price correlations, measured in absolute value, is considerably smaller after the French revolution than before the revolution.

In order to get more precise results for the 19th century the period 1806-1872 was split up into three subperiods. From these results it can be seen that the distribution for the period 1806-1872 is hardly representative for the whole period. It can only be successfully compared with the distribution for the subperiod 1806-1830 but definitely not with the distributions for the subperiods 1831-1850 and 1851-1872. Both these latter periods are characterized by a totally different shape of the distribution. Whereas the distribution of the slope parameters for the subperiod 1831-1850 is showing a considerable variance almost 90% of the slope parameters for the subperiod 1851-1872 are concentrated within a relative small interval. Besides the concentration it can be seen that this small interval is very close to zero indicating a minor influence of distances on the corresponding price correlations.

By using the three dimensional scatterplots represented at the second and fourth row of Figure 24 some interesting conclusions can be drawn about the spatial distribution of the slope parameters. For the period 1756-1790 the general shape of the quadratic surface fitted through the 3D scatterplot can be seen as a valley. This valley is situated about in the middle of the territory and runs from north to south. This means that for the *généralités* west and east of this valley the response of the distances on the price correlations is smaller than for the *généralités* located in the valley. For the period 1806-1872 the situation is slightly different. Although also for this period the quadratic surface looks like a valley this valley is not as deep and runs from north-west to south-east instead of running from north to south.
Just as for the ordinary distribution of the slope parameters also the spatial distribution for the whole period 1806-1872 is not representative for each of the three subperiods. This can be seen from the graphs on the fourth row of Figure 24. Whereas for the subperiods 1806-1830 and 1851-1872 one can hardly speak of the existence of a valley the quadratic surface for the subperiod 1831-1850 can successfully be compared with the period 1756-1790. In terms of market integration this could mean that for the period 1831-1850 the integration process in the *départements* in the western and eastern part of France was by far ahead of the integration of the *départements* located in the valley, the latter situated in the middle of the territory and running from north to south.

A totally different situation was obtained for the subperiods 1806-1830 and 1851-1872. From the graphs of the quadratic surfaces for these periods it can be seen that these surfaces are almost entirely reduced to flat and horizontal planes. This could be a clear indication for the absence of any significant regional difference with respect to the magnitude of the influence of distances on the price correlations.

Figure 24 : Price Correlation versus Distance - Local Results - Subperiods
Slope Parameter - Histogram - 3D Scatterplot - Quadratic Surface

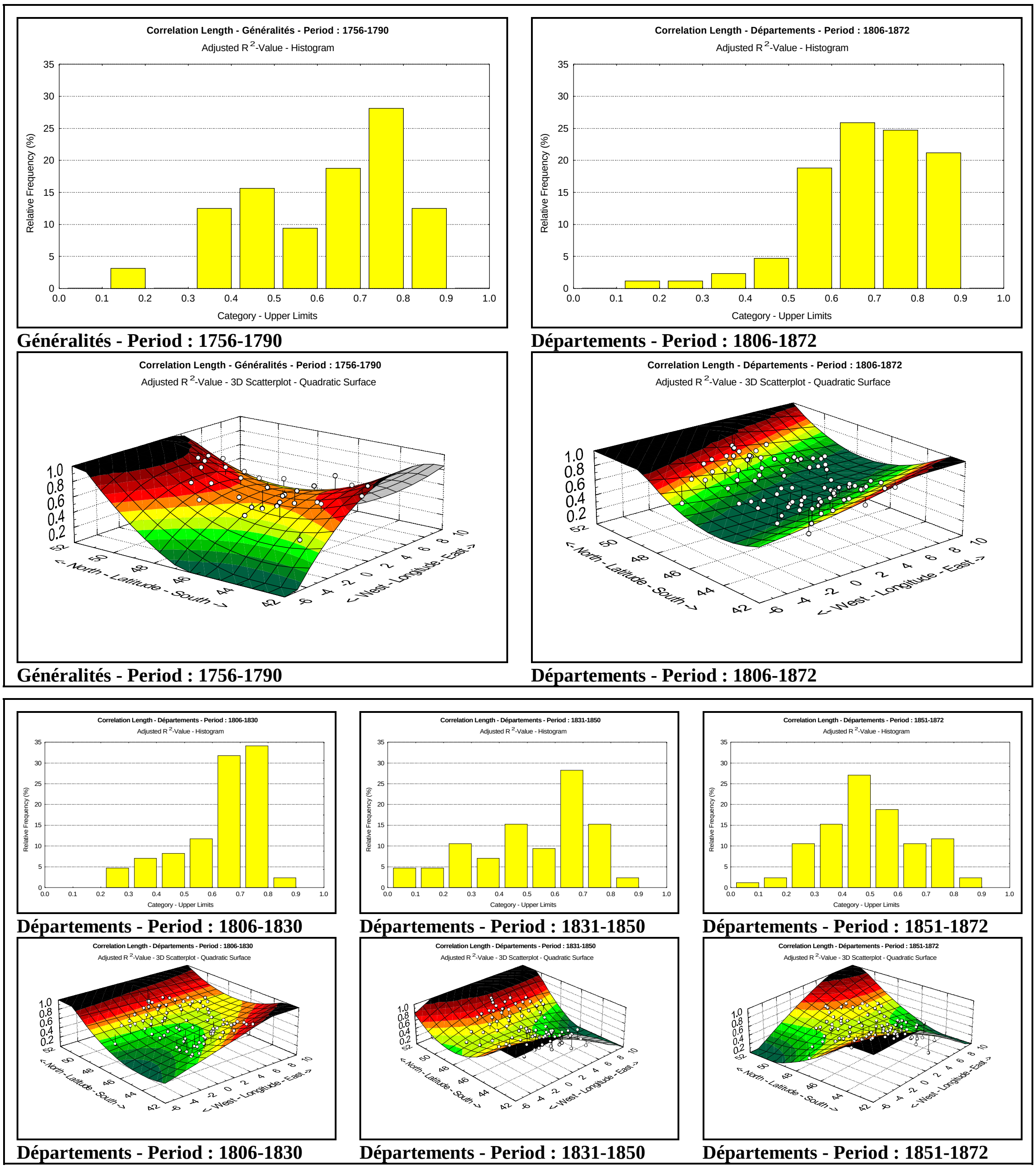


Correlation Length - Strength of Relationship

The approach followed in previous paragraphs to analyze the geographical distribution of the magnitude of the influence of distances on price correlations can readily be applied to analyze in more detail the spatial distribution of the strength of the distance-correlation relationship, the latter being measured by the (adjusted) R^2-values. The results about the spatial distribution of the strength of the relationship are presented in Figure 25. Construction and composition of these graphs are identical to those of Figure 24.

A convenient way to summarize the main results about the spatial distribution of the strength of the distance-correlation relation is to compare the 3D scatterplots and accompanying quadratic surfaces for each of the subperiods. Starting with the graph for the period 1756-1790 it can be seen that the strongest relationships can be found in almost the whole northern region of France and about in the middle of the southern part of the territory. The main difference with the subperiod 1806-1830 is that for the latter period the higher (adjusted) R^2-values are shifted from the south to the south-east. Another systematic change took place during the 1831-1850 subperiod. While the R^2-values remain high in the north the higher values in the south are shifted from the south-east to the south-west. The last subperiod 1851-1872 was characterized by the changing behavior of the R^2-values that took place in the north. Unlike in the previous subperiod the higher values no longer cover the whole northern part but are concentrated in the north-east.

Figure 25 : Price Correlation versus Distance - Local Results - Subperiods
Adjusted R^2-Value - Histogram - 3D Scatterplot - Quadratic Surface

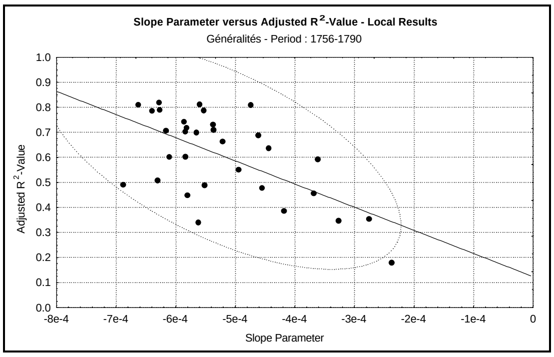


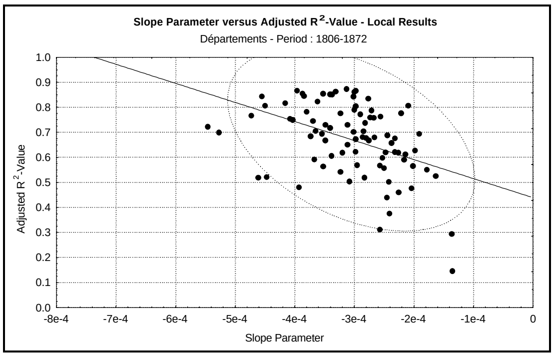
An alternative way to summarize the changing behavior of the strength of the relationship between distances and price correlations is to use the notion of saddle points illustrated in **Borghers [5]**. The quadratic surface for the period 1756-1790 demonstrates a situation characterized by a saddle point which is situated in the center of the French territory. The axes describing the direction of the higher and the lower values of the surface are running from respectively north-south and west-east. The transition from this first period to the last subperiod 1851-1872 can be described as the gradual rotation of these axes to a NE-SW and a NW-SE direction. The intermediate subperiods 1806-1830 and 1831-1850 can then be seen as the intermediate stages of the rotation procedure.

Correlation Length - Magnitude versus Strength of Relationship

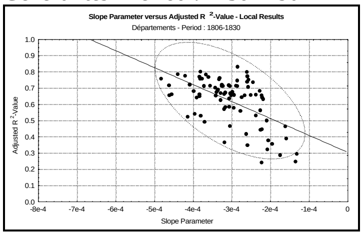
A peculiar side result from the analysis of the magnitude and the strength of the distance-correlation relation is the negative correlation between these two aspects. A simple visual inspection and comparison of the graphical representations of the quadratic surfaces of Figure 24 and Figure 25 leads to the conclusion that slope parameters close to zero can be associated with high (adjusted) R^2-values and that smaller R^2-values correspond with larger (in absolute value) slope parameters. This conclusion can be visualized by Figure 26.

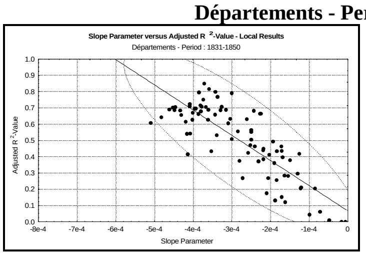
Figure 26 : Price Correlation versus Distance - Slope Parameter versus Adjusted R^2
Local Results - Subperiods - Scatterplot - Concentration Ellipse - Linear Fit

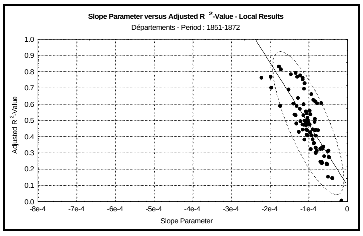




Généralités - Period : 1756-1790 **Départements - Period : 1806-1872**







Départements - Period : 1806-1830 **Départements - Period : 1831-1850** **Départements - Period : 1851-1872**

Section 6 : Final Conclusions and Remarks

Approach

The main subject of this paper was focused on the time varying and regional differences of the integration of the French wheat market for the periods 1756-1790 and 1806-1872. The methodology used was the analysis of the relationship between the correlations among the regional wheat prices and the geographical distances between these regions. Furthermore it was the intention to restrict the analysis to the more traditional statistical descriptive methods and to present at least the derived final results of the analysis in a way that would also be accessible by those who are not familiar with the underlying statistical techniques. This explains the abundant use of graphical representations. This intention is not necessarily in contradiction with the use of the generally less accessible multivariate techniques such as principal component and cluster analysis. The latter techniques do have at least the advantage that the final outcome from these analyses can be presented by relatively easy to understand graphs and even by appropriate maps. At this stage of the analysis it is still unclear how this approach and intention could have been pursued by using techniques such as cointegration or the even more recently developed spatiotemporal methodologies.

Final Results

After the introductory and preliminary information about defining the geographical locations and the resulting distances much attention was devoted on the analysis of these geographical distances data. The reason for the analysis of the distances is twofold. First there is the crucial role played by these distances in defining the relationship between price correlations and these geographical distances. Second there is the fact that the comparison between the two periods is based on different data sets. The main difference between these data sets is the number of locations for which price were recorded, i.e. 32 locations (*généralités*) for the period 1756-1790 against 85 locations (*départements*) for the period 1806-1872. The final conclusion from this analysis was that even with only 32 points of measurement the underlying structure of distances among the *généralités* can be compared with the structure of the distances among the 85 points of measurement for the *départements*.

About the proper analysis of the correlation-distance relationship at least three main results emerging from the analysis must be mentioned. A first conclusion is about the behavior of the relationship. No evidence could be found in favor of an exponential relationship. In both the period before and after the French revolution the relationship between price correlation and distance seems to be characterized by a linear relationship. The only exception is the behavior of the relationship for distances above the 750 km. For this segment no sensible relationship could be detected. A second conclusion is the striking difference between the degree of integration of the wheat market before and after the French revolution. The third and last general conclusion is that a real break-through of the integration of the wheat market took place during the third quarter of the 19th century.

Comparative Results from the Literature

Among the applications of using the correlation length to analyze the integration of markets the reported results about the applications for the integration of the French grain (wheat) markets ought to be mentioned. More in particular reference must be made to the research results published by **Keller and Shiue [7]**, **Roehner [8][9][10]** and **Roehner and Shiue [11]**. There are however fundamental problems in comparing the results derived in this paper with those published in these references. A first difficulty is the different data sets used. A second difficulty is that also the periods under investigation are not entirely comparable. Despite these problems at least two firm conclusions can be drawn from the comparison. The main conclusion is that, without even testing the appropriateness of the form of the relationship, the published results are based on an exponential relationship while in this paper a linear relationship was obtained. A second conclusion from the comparison is that the results obtained by this paper are consistent with the published results in the sense that the market integration of the wheat market in the first half of the 19th century is substantially higher than in the second half of the 18th century.

Appendix 1 : Data

1756<-	Set 1	->1790

1806<-	Set 5	->1872

Data Set 4 : Les Prix du Blé en France dans 32 Généralités du Royaume et la Ville de Paris

Source	**:** See **Labrousse ［18, pp. 106-113］**
Period	**:** 1756-1790
Regions	**:** Généralités du Royaume et Paris Ville - See Appendix 2
Series	**:** 33
Frequency	**:** Yearly
Observations	**:** 35 per series
Price / Unit	**:** Livres et centièmes de livre / Setier de Paris pesant 240 livres

Data Set 5 : Prix Moyens Annuels de l'Hectolitre de Froment par Département Intérieur

Source	**:** See **Labrousse et al. ［19, pp. 45-219］**
Period	**:** 1806-1872 (IX-XIV and 1806-1872) Data for the period IX-XIV are excluded
Regions	**:** Départements Intérieurs - See Appendix 3
Series	**:** 85 - See Appendix 3
Frequency	**:** Yearly
Observations	**:** 67 per series
Price / Unit	**:** Livres et centièmes de livre / Hectolitre

Appendix 2 : Généralités du Royaume et Paris Ville

#	Généralité	#	Généralité
1.	Alençon	**17.**	La Rochelle
2.	Alsace	**18.**	Limoges
3.	Amiens	**19.**	Lorraine
4.	Auch	**20.**	Lyon et Dombes
5.	Bayonne	**21.**	Metz
6.	Bordeaux	**22.**	Montauban
7.	Bourges	**23.**	Moulins
8.	Bourgogne	**24.**	Orléans
9.	Bretagne	**25.**	Paris
10.	Caen	**26.**	Poitiers
11.	Champagne	**27.**	Provence
12.	Flandres	**28.**	Riom-Auvergne
13.	Franche-Comté	**29.**	Rouen
14.	Grenoble	**30.**	Roussillon
15.	Hainaut	**31.**	Soissons
16.	Languedoc	**32.**	Tours
		33.	Paris Ville

Appendix 3 : Départements Intérieurs

#	Département	#	Département	#	Département
1.	Ain	**31.**	Haute-Garonne	**61**	Orne
2.	Aisne	**32.**	Gers	**62.**	Pas-de-Calais
3.	Allier	**33.**	Gironde	**63.**	Puy-de-Dôme
4.	Basses-Alpes	**34.**	Hérault	**64.**	Basses-Pyrénées
5.	Hautes-Alpes	**35.**	Ille-et-Vilaine	**65.**	Hautes-Pyrénées
6.	Alpes-Maritimes (°)	**36.**	Indre	**66.**	Pyrénées-Orientales
7.	Ardèche	**37.**	Indre-et-Loire	**67.**	Bas-Rhin (°)
8.	Ardennes	**38.**	Isère	**68.**	Haut-Rhin (°)
9.	Ariège	**39.**	Jura	**69.**	Rhône
10.	Aube	**40.**	Landes	**70.**	Haute-Saône
11.	Aude	**41.**	Loire-et-Cher	**71.**	Saône-et-Loire
12.	Aveyron	**42.**	Loire	**72.**	Sarthe
13.	Bouches-du-Rhône	**43.**	Haute-Loire	**73.**	Savoie (°)
14.	Calvados	**44.**	Loire-Inférieure	**74.**	Haute-Savoie (°)
15.	Cantal	**45.**	Loiret	**75.**	Seine (°)
16.	Charente	**46.**	Lot	**76.**	Seine-Inférieure
17.	Charente-Inférieure	**47.**	Lot-et-Garonne	**77.**	Seine-et-Marne
18.	Cher	**48.**	Lozère	**78.**	Seine-et-Oise
19.	Corrèze	**49.**	Maine-et-Loire	**79.**	Deux-Sèvres
2A.	Corse-du-Sud (°)	**50.**	Manche	**80.**	Somme
2B.	Haute-Corse (°)	**51.**	Marne	**81.**	Tarn
21.	Côte-d'Or	**52.**	Haute-Marne	**82.**	Tarn-et-Garonne
22.	Côtes-du-Nord	**53.**	Mayenne	**83.**	Var
23.	Creuse	**54.**	Meurthe	**84.**	Vaucluse
24.	Dordogne	**55.**	Meuse	**85.**	Vendée
25.	Doubs	**56.**	Morbihan	**86.**	Vienne
26.	Drôme	**57.**	Moselle (°)	**87.**	Haute-Vienne
27.	Eure	**58.**	Nièvre	**88.**	Vosges
28.	Eure-et-Loire	**59.**	Nord	**89.**	Yonne
29.	Finistère	**60.**	Oise		
30.	Gard				

(°) **Remarks**

#	Département	
6.	Alpes-Maritimes	Excluded - Missing data for the period 1814-1860
2A	Corse-du-Sud	Excluded - Département Extérieur
2B	Haute-Corse	Excluded - Département Extérieur
57.	Moselle	Included - Missing data for the years 1871 and 1872
67.	Bas-Rhin	Included - Missing data for the years 1871 and 1872
68.	Haut-Rhin	Included - Missing data for the years 1871 and 1872
73.	Savoie	Excluded - Data available as from 1861
74.	Haute-Savoie	Excluded - Data available as from 1861
75.	Seine	Included - Missing data for the years 1871 and 1872

References

General

[1] **Borghers, E.**
Cyclical Regional Wheat Prices - France : 1806-1872
Research Paper 2006-001
Department of Mathematics, Statistics and Actuarial Sciences
Faculty of Applied Economics, University of Antwerp, Antwerp, February 2006.

[2] **Borghers, E.**
Time Path of Price Inequality of Regional Wheat Prices - France : 1756-1790 versus 1806-1872
Research Paper 2006-002
Department of Mathematics, Statistics and Actuarial Sciences
Faculty of Applied Economics, University of Antwerp, Antwerp, February 2006.

[3] **Borghers, E.**
French Regional Wheat Prices : 1756-1870 - An Illustrated Statistical Analysis
Research Paper 2006-003
Department of Mathematics, Statistics and Actuarial Sciences
Faculty of Applied Economics, University of Antwerp, Antwerp, February 2006.

[4] **Borghers, E.**
French Regional Wheat Prices : 1756-1870 - Généralités versus Départements
Research Paper 2006-004
Department of Mathematics, Statistics and Actuarial Sciences
Faculty of Applied Economics, University of Antwerp, Antwerp, February 2006.

[5] **Borghers, E.**
French Regional Wheat Prices : 1756-1870 - Graphical Spatial Representation
Généralités versus Départements
Research Paper 2006-007
Department of Mathematics, Statistics and Actuarial Sciences
Faculty of Applied Economics, University of Antwerp, Antwerp, February 2006.

[6] **Cleveland, W.S.**
The Elements of Graphing Data
Wadsworth, Monterey, 1985.

[7] **Haggett, P., Cliff, A.D. & Frey, A.**
Locational Models
Edward Arnold, London, 1977.

[8] Keller, W. & Shiue, C.
Markets in China and Europe on the Eve of the Industrial Revolution
Centre for Economic Policy Research, Discussion Paper no. 4420, June 2004.

[9] Roehner, B.M.
An Empirical Study of Price Correlations
1. How should spatial interactions between interdependent markets be measured ?
Environment and Planning, Series A, vol. 21, 1989, pp. 161-173.

[10] Roehner, B.M.
An Empirical Study of Price Correlations
2. The decrease of price correlation with distance and the concept of correlation length
Environment and Planning, Series A, vol. 21, 1989, pp. 289-298.

[11] Roehner, B.M.
Les méchanismes d'interdépendence spatiale entre marchés du blé au XIXe siècle
Histoire Economie et Société, vol. 13, no. 2, 1994, pp. 343-394.

[12] Roehner, B.M. & Shiue, C.H
Comparing the Correlation Length of Grain Markets in China And France
International Journal of Modern Physics, Series C, vol. 11, no. 7, 2000, pp. 1383-1410.

［13］Ward, J.H.
Hierarchical Grouping to Optimize an Objective Function
Journal of the American Statistical Association, vol. 58, March, 1963, pp. 236-244.

［14］Calculating Distance and Bearing between Two Latitude/Longitude Points
http://www.movable-type.co.uk/scripts/LatLong.html
03/01/2005 - 08:45 am

［15］Using Longitude and Latitude to Determine Distance
Math Forum - Ask Dr. Math - The Math Forum @ Drexel
Drexel University - Philadelphia
http://mathforum.org/library/drmath/view/51711.html
03/01/2005 - 08:43 am

[16] ViaMichelin
http://www.viamichelin.com/viamichelin/fra/tpl/hme/MaHomePage.htm
25/02/03 - 12:08 pm

［17］What is the Best Way to Calculate the Distance between 2 Points ?
GIS FAQ Question Q5.1 - Chamberlain, B.
http://census.gov/cgi-bin/geo/gisfaq?Q5.1
03/01/2005 - 08:54 am

Price Data

［18］Labrousse, C.-E.
Esquisse du Mouvement des Prix et des Revenus en France au XVIII^e Siècle
Collection Scientifique d'Economie Politique III
Tome 1 : Les Prix
Tome 2 : Les Prix (Fin), Les Revenus
Dalloz, Paris, 1933, pp. 106-113.

［19］Labrousse, C.-E., Romano, R. & Dreyfus, F.-G.
Le Prix du Froment en France
Au temps de la Monaie Stable, 1726-1913
Ecole Pratique des Hautes Etudes - VI^e Section
Centre de Recherches Historiques
Monnaie - Prix - Conjoncture IX
SEVPEN, Paris, 1970, pp. 9-11, pp. 23-24 & pp. 27-35.

Software

［20］Statistica
Kernel Release 5.5, Edition '99
Statsoft, Tulsa, 1999.