Environment and Planning A, 1989, volume 21, pages 289-298

An empirical study of price correlations: 2. The decrease in price correlation with distance and the concept of correlation length

B Roehner
Laboratoire de Physique Théorique et Hautes Énergies, Université Paris VII, 75251 Paris Cedex 05, France
Received 15 May 1987; in revised form 25 May 1988

Abstract. The dependence of the correlation between wheat price fluctuations on different markets with respect to the distance between those markets is investigated. It is shown that the decrease in the correlation is exponential and is governed by a characteristic distance which is called the correlation length for wheat prices. This is a measure of the level of market integration in a given area. The magnitude of the correlation length is compared for different cereals: wheat, rye, and oats, and the evolution of the correlation length during the 19th century is studied. In particular, it is observed that subsistence crises are characterised by a collective behaviour of the economy which results in a sudden peak in the correlation length. Last, the part played by the spatial correlation between precipitations is investigated. It appears to be rather small; consequently, the price-correlation length is primarily the result of economic factors.

1 Introduction

1.1 *The variation of correlation with geographical distance*

> "Everything is related to everything else, but near things are more related than distant things".

This commonsense statement, which applies to the physical world as well as to the socioeconomic sphere, is sometimes called the first law of geography (Haggett et al, 1977, page 330). It leads us to the central question of this paper: does the correlation between the price of a commodity at different marketplaces decrease with increased distance and, if so, in what way?

To answer this question we have to use a much more extensive data set than the one used for the analysis in the first paper of this series. Indeed, we need here a data set showing prices on several markets separated by distances ranging from a few to several hundred kilometers. As a result, we shall see that the correlation decreases with distance in an exponential way. To support that statement, I shall provide a theoretical argument and a statistical check.

1.2 *Spatial interaction models in geography*

The exponential decrease is precisely the functional form which results from Wilson's (1970) approach of spatial interaction in geography. This approach, which is based upon the principle of maximum entropy, renewed the building of spatial interaction models. My paper is in line with this approach, not only because of the exponential form of the correlation decrease, but also because two other crucial concepts in statistical mechanics are introduced here in a natural way, namely the concept of correlation length, and the notion of collective phenomena known in statistical mechanics as phase transitions.

2 Qualitative discussion of the correlation decrease

2.1 *General form of the correlation decrease*

It is not my purpose in this *empirical* paper to use Wilson's entropy formalism; instead let me make an 'educated guess' that, among all possible forms of interaction with distance (Haggett et al, 1977, page 358), the exponential decrease is the most natural one in the case of correlation.

We can indeed guess the two extreme parts of the correlation curve with distance, namely the left-hand part of the curve, where the correlation should tend to 1 as the distance tends to 0, and the right-hand part of the curve, where the correlation should become very small as the distance becomes very large. Hence the relation between correlation and distance cannot be a linear one, since this would imply a sudden vanishing of the correlation at some *finite* distance.

On the contrary, a form such as

$$c(d) = \exp(-ad)\,, \qquad a > 0\,,$$

where d is distance between markets, perfectly meets the two previous requirements since

$$c(0) = 1\,, \qquad c(\infty) = 0\,.$$

2.2 *The correlation length*

The parameter a is usually referred to as the friction of distance parameter. In this specific context, it is more suggestive to introduce the inverse of a ($L = 1/a$), for the two following reasons. (1) Whereas the parameter a is the inverse of a length, the parameter L is itself a length. It represents the range of the interaction between markets. (2) We shall see that the parameter L has a very simple graphical interpretation (see figure 2 below). However, the order of magnitude of L is not well adapted to the problem under consideration. Indeed, when the distance d increases from 0 to L, the correlation decreases from 1 to

$$c(L) = \exp(-1) = 0.368\,.$$

Such a correlation is very small and is usually not considered as representing a definite interdependence at all. In a specific case which I shall detail below for the region of Bavaria, L will be found equal to 3100 km.

Let us introduce a submultiple of L,

$$l = \frac{L}{100}\,.$$

Parameter l has a very nice interpretation [which we can easily get by developing the function $c(d) = \exp(-d/100l)$ to first order]: when d increases from 0 to 1, the correlation decreases from 1 to 0.99. In other words, a variation of distance of l corresponds to a correlation decrease of 0.01. Although the introduction of l instead of L could appear at this stage as rather technical and artificial, we shall see that l really is a convenient parameter.

To summarise, the correlation reads

$$c(d) = \exp\left(\frac{-d}{100l}\right), \qquad (1)$$

where l is the *correlation length*.

3 Statistical check of the exponential decrease of the correlation with distance

We have to carry out two objectives. First, check that the exponential decrease of equation (1) is indeed compatible with the data. Second, give a way to estimate the correlation length l. These two points received considerable attention in the context of gravity models (Cliff et al, 1974; Griffith and Sheppard, 1975; Johnston, 1973). Johnston, in particular, discussed carefully what precautions should be taken in order to estimate the friction of distance parameter.

3.1 *The procedure and the kind of data we need*

The *principle* of the procedure, as illustrated in figure 1, is simple. We shall consider a number of markets and evaluate for each market pair its correlation $c_{ij}(d_{ij})$; then we plot the numbers $\ln[c_{ij}(d_{ij})]$ against the corresponding distances d_{ij} and we try a *linear fit*.

The *realisation* of the procedure requires care, for the following reasons:

(1) One should not forget that a correlation such as $c_{ij}(d_{ij})$ is a statistical variable; as a consequence, it is known only up to a confidence interval which is represented in figure 1 by an error bar. For the linear fit to have a meaning at all, those error bars should be short. In other words, the computation of the correlation must be based on a sufficiently large sample of (time-series) values. In practice at least fifty values would be reasonable. It is clear that this requirement is difficult to meet when one is using annual prices. But, when one is using monthly prices, a window of a few years would be sufficient. A window of eight years will be used in this paper.

(2) For the linear fit, to give the correlation length with a reasonably narrow confidence interval, the number of the points in figure 1(b) should be large enough, say, at least thirty points. Since the number of points is the number of market pairs, that is, $\frac{1}{2}n(n-1)$, where n is the total number of markets, we need a number of markets of the order of ten.

To summarise, we need monthly prices on about ten different markets; if the period of interest has an amplitude of about fifty years, this represents $50 \times 12 \times 10$ ($= 6000$) prices, as compared with $50 \times 1 \times 5$ ($= 250$) prices in paper 1.

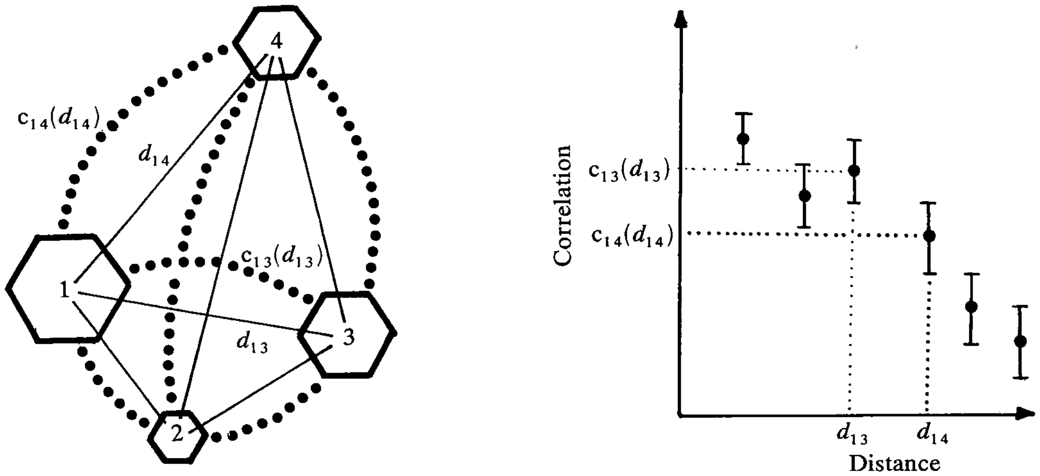


Figure 1. Procedure for analysing the decrease with distance of the correlation.

3.2 The case of Bavaria, wheat, 1825–33

Bavaria was considered (figure 2) because the statistics needed were published in printed and easily accessible form (Seuffert, 1857). I considered the period 1825–33 to draw figure 3, because this was a standard time period without any special problems. Of course, in section 4, I shall give the same results for the whole period 1815–55 for which statistics are available.

Figure 3 shows that a linear fit is indeed reasonable. More precisely, the correlation coefficient, R, giving the goodness of fit is equal to 0.85 and the correlation length l is 31 ± 5 km.

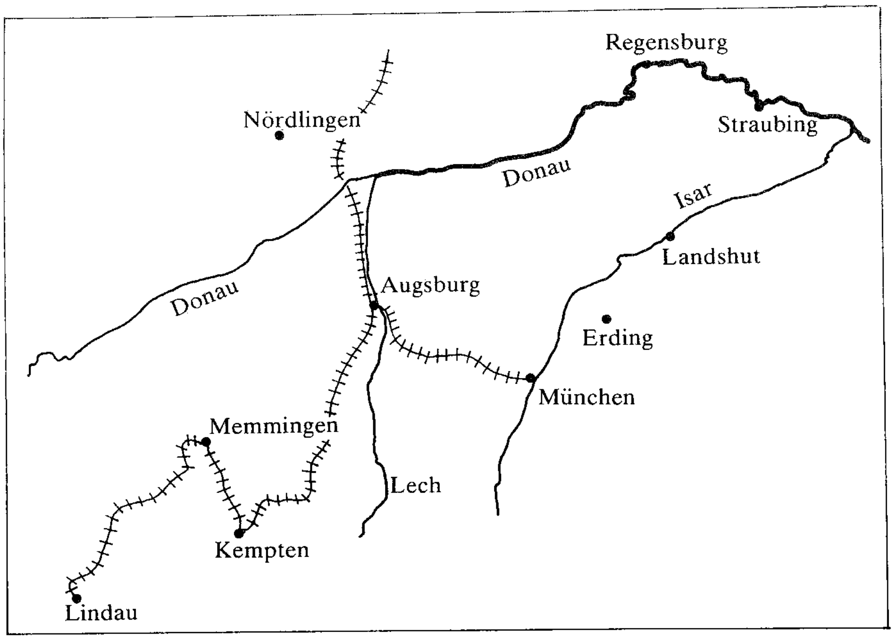
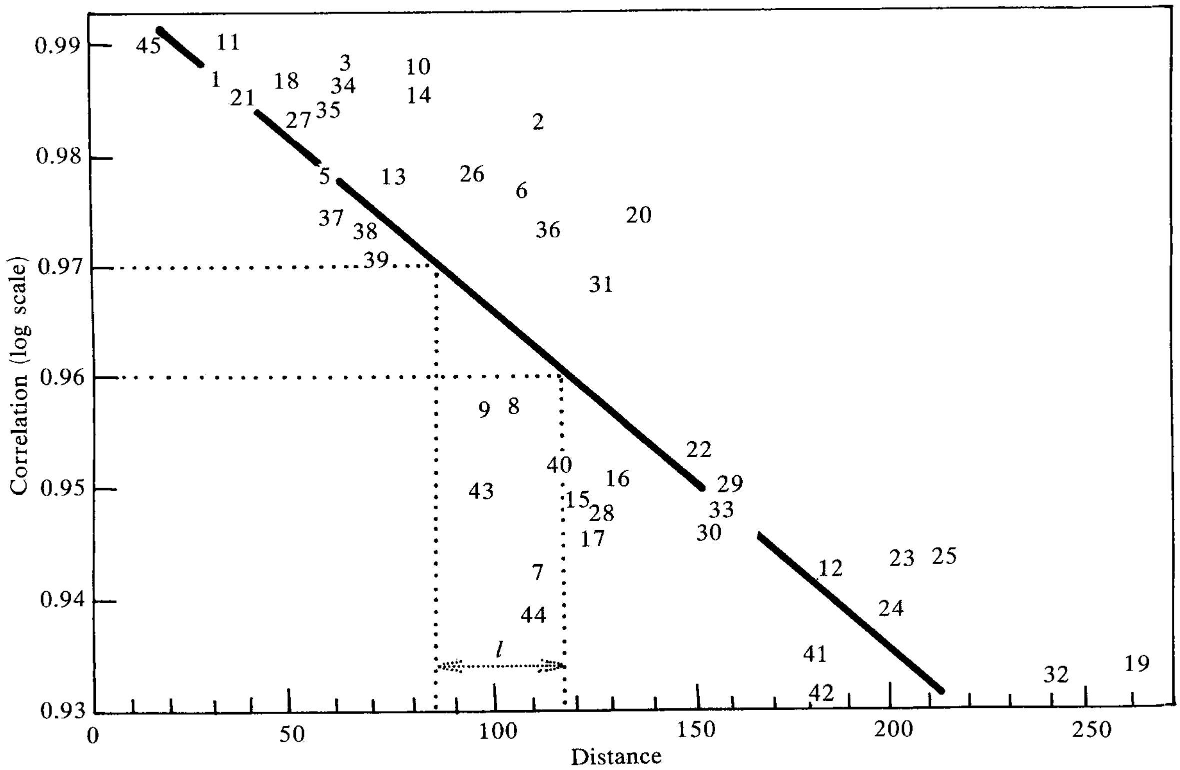


Figure 2. Map of south Bavaria with selected marketplaces.



Key to correlations

München	Erding	1	Straubing	Landshut	18	Lindau	Augsburg	31
	Straubing	2		Lindau	19		Regensburg	32
	Landshut	3		Augsburg	20		Nördlingen	33
	Lindau	4		Regensburg	21		Memmingen	34
	Augsburg	5		Nördlingen	22		Kempten	35
	Regensburg	6		Memmingen	23	Augsburg	Regensburg	36
	Nördlingen	7		Kempten	24		Nördlingen	37
	Memmingen	8	Landshut	Lindau	25		Memmingen	38
	Kempten	9		Augsburg	26		Kempten	39
Erding	Straubing	10		Regensburg	27	Regensburg	Nördlingen	40
	Landshut	11		Nördlingen	28		Memmingen	41
	Lindau	12		Memmingen	29		Kempten	42
	Augsburg	13		Kempten	30	Nördlingen	Memmingen	43
	Regensburg	14					Kemptem	44
	Nördlingen	15				Memmingen	Kempten	45
	Memmingen	16						
	Kempten	17						

Figure 3. Relation between correlation and distance between markets.

Let me summarise this result in the form of the following proposition:
The decrease of the correlation of wheat prices c(d) with distance, d, is given by

$$c(d) = \exp\left(\frac{-d}{100\,l}\right),$$

where l is the correlation length.

Of course, such a conjectural relationship has still to be tested for different countries or products. This will be the purpose of subsequent papers.

4 Evolution of the correlation length

Sliding the observation window through the whole period 1815–55 gives the time evolution of the correlation length. This coincides with the objective of the previous paper, namely the analysis of market integration, but this time, with a much better insight. Let us recall that in paper 1 (section 3.2.1), we used as a measure of market integration, the average value of the correlation over all market pairs. In the light of the present analysis, we can see that this average value could provide only an approximate estimate.

The other side of the coin is that the period 1815–55 for which data are available here is too short to give a full description of the whole market-integration process.

4.1 *Bavaria, wheat, 1815–55*

Figure 4(a) gives the evolution of the wheat-price correlation length in Bavaria and figure 4(b) gives the correlation coefficient, R, which characterises the goodness of the linear fit. Nine markets have been used, München, Erding, Landshut, Straubing, Regensburg, Augsburg, Kempten, Lindau, and Memmingen. Two features emerge from figure 4(a): a prominent peak around 1820, and a steady increase from 1823 to 1850.

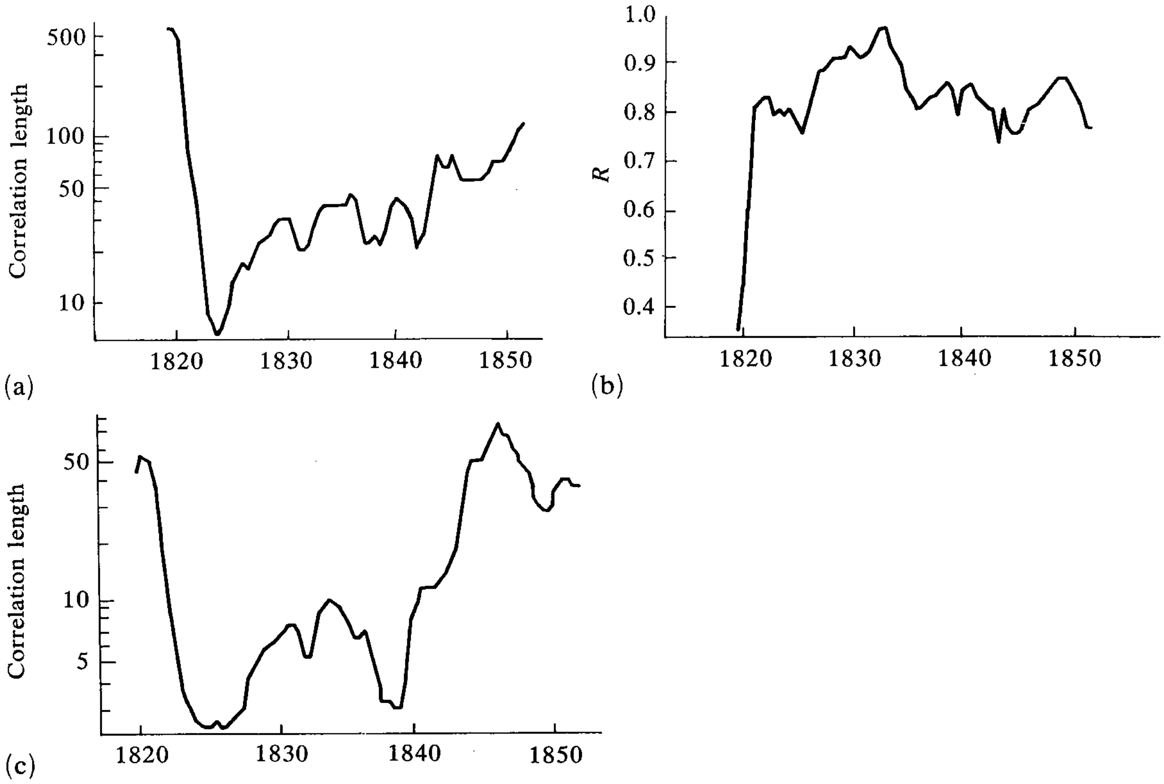


Figure 4. (a) Wheat correlation length for Bavaria (log scale), (b) goodness of the correlation–distance fit, measured by the correlation coefficient R, (c) oats correlation length for Bavaria (log scale).

4.1.1 *The collective behaviour of 1817–21* The peak is caused by the 'year without a summer' (Stommel and Stommel, 1979). Indeed the bad weather during the spring and summer of 1816 produced a very severe shortage of cereals. This was actually the last large-scale famine in Western Europe. In such circumstances wheat prices rose simultaneously everywhere. Instead of being subjected to local, more or less related fluctuations, all markets were suddenly submerged by one wave of price rises. In statistical physics such an effect is called a phase transition and is characterised by a correlation length which is very large. Here too, the correlation length jumps to a level which is about twenty times above its 'normal' level.

Remark To be able to observe such an effect, one has to use a rather narrow correlation window. In figure 4(a), a window of eight years was used. Since the collective behaviour effect lasted only from 1816 to 1822, a larger window would mix high correlations from that interval with much lower ones corresponding to the years after 1822. As a result, the peak would be considerably damped. As an extreme case, if one were to use annual prices, one would need a window of at least twenty years, and the peak would then be completely hidden.

4.1.2 *The market-integration process* The increase after 1823 reveals the market integration occurring in Bavaria at that time. In comparison with the correlation curves of paper 1, figure 4(a) is a kind of magnifying glass in both space and time. In space, because it describes a single region instead of all Germany; in time because the use of monthly data gives the possibility of observing the evolution over time in more detail.

As can be seen, there are at least three ups and downs. However, the explanations proposed in paper 1, namely tariff and railroad modifications, are not applicable here. Indeed, Bavaria was at that time a politically unified kingdom and second, railways appeared only after 1850, that is, at the very end of the period. Thus, the interpretation of the medium-term fluctuations of the correlation length remains open to question.

4.1.3 *The goodness-of-fit curve* Figure 4(b) gives the evolution of the correlation factor R which characterises the goodness of the linear fit. It appears that R is nearly always greater than 0.80, which is quite satisfactory, with the exception of the interval corresponding to the 'collective behaviour peak'. This could be expected, however, because when all the markets are moving in phase, the relation between correlation and distance no longer has any meaning.

4.1.4 *Influence of the window width* As already noted, changing the window width will certainly result in a modification of the correlation length curve. Roughly speaking, the wider the window, the smoother the curve; a wide window will average over many extreme transient situations. I have not displayed the curve of the correlation length for a window of twelve years, for instance, because it is very similar to that of figure 4(a) (eight years). If, however, we compare both curves the following observations can be made. (1) The two principal features previously mentioned, namely the peak around 1820 and the increase from 1825 to 1850, are still present. (2) The order of magnitude of the correlation length is the same except, as expected, for the height of the 1820 peak which is lower. (3) Many of the medium-term fluctuations are common to both curves. (4) The most obvious difference between the curves lies in the relative magnitude of some medium-term fluctuations and in a slight time shift of one curve relative to the other.

4.2 *Bavaria, oats, 1815-55*
Figure 4(c) gives the evolution of the correlation length for oats. The correlation window and the markets are the same as for wheat [figure 4(a)]. Three features emerge: a peak in the period 1817-24 which is not, however, as high as the one for wheat, a trend towards increase, and the rather short correlation length, which varies between 3 km and 70 km as compared with an interval of about 10 km to 150 km for wheat. The first and third points may be explained by the fact that oats were less sensitive to demand than wheat because they were mainly used for horse feeding. We shall meet the same observation again in the next paragraph.

4.3 *Germany, wheat, rye, and oats, 1790-1910*
Above (section 3) I have stressed the drawbacks which could result when one uses annual data instead of montly data, namely much larger confidence intervals for the correlation length and thus much more uncertainty in its evolution. Nevertheless, such a procedure will be used in this paragraph for the following reasons: first, it will specify the hazards of that procedure; second, it will show what can be learned about the correlation length from the *annual* price series, already used in paper 1, which cover the whole 19th century and the whole of Germany.

The two curves of figures 5(a) and5(b) correspond to a correlation window of twenty-one years and give the evolution of the correlation length for wheat and rye; the curve for oats will also be considered but is not reproduced here because it is very similar.

It appears that the correlation length was increasing during the 19th century for the three products: from 10 km to about 1000 km for wheat, from 10 km to about 400 km for rye, from about 5 km to about 200 km for oats. This growth and the respective orders of magnitude of the correlation lengths are in accordance with what common sense would suggest. Indeed, as far as their price is concerned the three cereals rank in the following order (München, 1840)

wheat	157 RM	per 1000 kg,
rye	106 RM	per 1000 kg,
oats	90 RM	per 1000 kg.

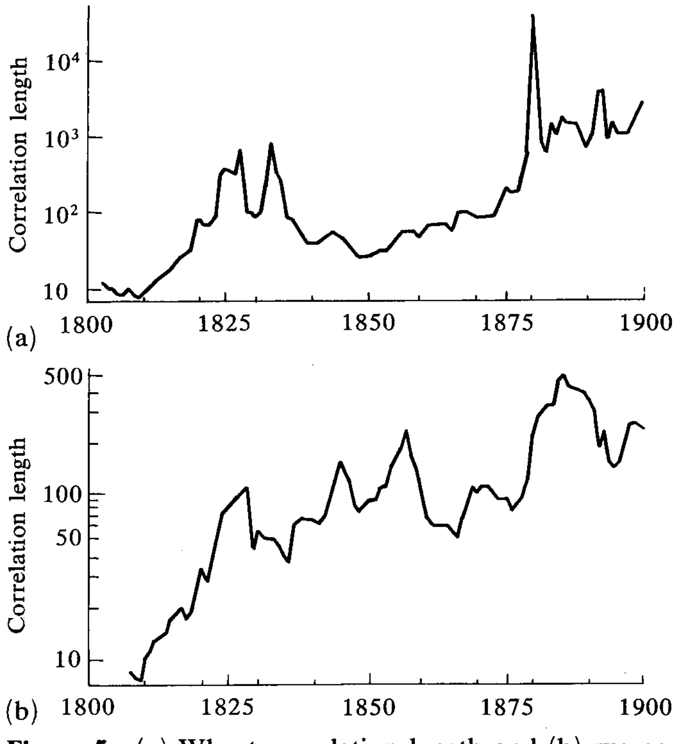


Figure 5. (a) Wheat correlation length and (b) rye correlation length for Germany (log scale).

And, as already pointed out in paper 1, the price level is the first incentive to a higher volume of trade. The social importance of oats is even smaller than its price would suggest since it was at that time mainly used for horse feeding. This can be seen in its correlation length level.

Let us now consider the drawbacks that arise from the inadequacies of the present statistics: annual instead of monthly prices and too few markets, seven marketplaces for rye (that is, twenty-one pairs), six for oats (fifteen pairs), and five for wheat (ten pairs). The consequences are the following. The confidence intervals for the resulting correlation lengths are rather broad. The correlation length curves are much more sensitive to the choice of the correlation window than was the case in section 2.

Thus, only the trend of the curves of figure 5 should be considered significant. The medium-term variations reflect statistical fluctuations. The fact that, even with such poor data, reasonable results could be obtained, at least for the overall evolution, is probably a good test of the 'robustness' of the concept of correlation length.

5 The influence of weather

5.1 *Weather, yields, and prices*

As far as grain *yields* are concerned, weather has undoubtedly had a great influence. This problem has been studied by several authors; for instance by Hooker (1907), Beveridge (1921), Machali (1931), Timoshenko (1944), and more recently by Pfister (1986). A possible correlation between prices and rainfall has been studied (Beveridge, 1922). The respective influences of yield and trade have been carefully analysed by Tits-Dieuaide (1975) for Flanders in the 15th century. Hooker points out that rain is the most important factor at least in Western Europe where there is an oceanic climate, and that the correlation between rainfall summed over the whole cereal year and yield is about -0.65.

The relation of grain *prices* to weather is not as clear, for at least three reasons. First, there are possible variations in the cultivated acreage from year to year as well as variations of demand. These factors probably played a minor role in the 19th century economy. Second, the breakdown of the sales over the whole harvest year (from August to July) may vary from one year to another because of profit maximisation. Third, we must take into account the influence of the grain trade which induces a dependence between prices on distant markets. This is precisely the factor we are interested in.

5.2 *The precipitation-correlation length*

In this paper, I shall concentrate on the influence of *rainfall* on wheat prices for two reasons. First, rain is, as mentioned previously, the most important of the meteorological factors. Second, together with temperature, it is the only meteorological variable for which 19th-century records are available. The correlation between temperature and grain yields is, however, very low (Hooker, 1907).

My reasoning will be the following. The precipitations on one place A are clearly related to the precipitations on a neighbouring place B, and it is reasonable to think that the correlation should decrease when the distance between A and B increases. In other words, it is possible to define and compute a correlation length for precipitations in exactly the same way as we did for prices. Now, suppose just for a moment that the relation between wheat prices and rainfall is almost deterministic. In this case, the rainfall-correlation length and the price-correlation length will be almost equal. More generally, comparing the magnitudes of the correlation lengths will show what part of the wheat-correlation length could *possibly* be attributed to spatial correlations of rainfalls.

To carry out that programme, the first step was to get adequate data. The ideal would have been to find precipitation data for exactly the same places and periods as the price series. This was not possible, however, since precipitation statistics for Germany began around 1850. From Clayton (1944) the following series of monthly precipitations data were selected, all of them for the period 1850-1920: Berlin, Königsberg, Triev, Wien, Breslau, Frankfurt, Gütersloh, Utrecht, Zürich. As can be seen German cities were supplemented with Wien, Utrecht, and Zürich, which are close to Germany, in order to obtain a reasonable number of places for a correct evaluation of the correlation length.

Since its precise form is of little interest, the curve of the correlation length is not displayed. Instead its principal characteristics will be described. Let me first point out that the correlation R is everywhere greater than 0.60 and on the average is of the order of 0.75. Thus, the results for the correlation length are indeed reliable and this is quite natural since we are again working with monthly data and with as many as thirty-six pairs.

Two observations can be made. First, the correlation length is always shorter than 20 km. Actually it is confined in a very narrow interval between 6 km and 20 km, with an average of 10 km. It is thus of the order of the correlation length for grain prices at the very beginning of the 19th century. Second, there is no increase trend and there are no sudden peaks.

As far as the increase of the correlation length for grain prices is concerned, it is thus clear that only economic factors are responsible for it.

6 Conclusion

In this paper, I have shown that the correlation between grain prices on different markets decreases exponentially with distance between markets. This decrease is governed by a characteristic length which was called the correlation length.

The next step would be to deduce such an empirical law from the economic interactions between markets. In other words, we are facing here a field of prices, now what are the field equations?

Acknowledgements. The author would like to express his gratitude to his two anonymous referees and to Dr Pumain. Thanks to their remarks, the present papers have been notably improved.

References

Beveridge W H, 1921, "Weather and harvest cycles" *Economic Journal* **31** 429-452

Beveridge W H, 1922, "Wheat prices and rainfall in western Europe" *Journal of the Royal Statistical Society* **85** 412-459

Clayton H H, 1944 *World Weather Records* (Smithsonian Institution, Washington, DC)

Cliff A D, Martin R L, Ord J K, 1974, "Evaluating the friction of distance parameter in gravity models" *Regional Studies* **8** 281-286

Griffith D, Sheppard E S, 1975, "Map pattern and friction of distance parameters. Reply to comments by R J Johnston and L Curry" *Regional Studies* **9** 285-288

Haggett P, Cliff A D, Frey A, 1977 *Locational Models* (Edward Arnold, London)

Hooker R H, 1907, "Correlation of the weather and crops" *Journal of the Royal Statistical Society* **19** 1-51

Johnston R J, 1973, "On friction of distance parameters and regression coefficients" *Area* **5** 187-191

Machali M, 1931 *Prevision du Rendement du Blé d'après les Éléments Météorologiques* thesis, University of Paris, Paris

Pfister C, 1986, "Climatic fluctuations and their impact upon grain prices, 16th to 20th century", paper presented at the 9th International Congress of Economic History, Bern; copy available from the author, 4 Engehaldenstrasse, CH-3012 Bern; to appear in *Annales Économie, Société, Civilisation*

Seuffert G K L, 1857 *Statistik des Getreide- und Viktualien-Handels im Königreiche Bayern* (Weisz, München)
Stommel H, Stommel E, 1979, "The year without a summer" *Scientific American* **240**(6) 134-140
Timoshenko V P, 1944, "Interregional correlations in wheat yields and outputs" *Wheat Studies of the Food Research Institute, Stanford University* **20**(6) 213-259
Tits-Dieuaide M J, 1975 *La Formation des Prix Céréaliers en Brabant et en Flandre au XV^{e} Siècle* thesis, Centre d'Histoire Économique et Sociale, Free University of Brussels, Brussels
Wilson A G, 1970 *Entropy in Urban and Regional Modelling* (Pion, London)