J. Phys. A: Math. Gen. **31** (1998) L361–L365. Printed in the UK PII: S0305-4470(98)89047-1

LETTER TO THE EDITOR

Evidence for global mixing in real influenza epidemics

Eric Bonabeau†§, Laurent Toubiana‡‖ and Antoine Flahault‡

† Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
‡ INSERM-U444, B3E-ISARS, CHU Saint-Antoine, 27 rue Chaligny, 75012 Paris, France

Received 6 November 1997, in final form 6 February 1998

Abstract. The spatiotemporal behaviour of the spread of influenza in France has been studied, and algebraic spatial correlations (with exponent $\alpha = -0.34$) spanning the whole territory have been found to be present as soon as the number of reported cases begins to increase, about 15–25 weeks before the peak of the epidemic. This result is surprising, as one would expect long-range correlations, if any, only in the vicinity of the maximum incidence, whereas our observations suggest that there exists an underlying non-trivial spatial structure at the very beginning of the observed epidemic. The observed long-range correlations are in fact present in the spatial distribution of the population. Correlations in the number of cases normalized by local population density are characterized by $\alpha \approx 0$. This suggests that the spread of the epidemic is statistically uniform in space over a complex substrate that already contains the observed long-range correlations.

1. Introduction

Epidemic models have been of interest to physicists for many years. Their emphasis has been on critical properties and universality classes [1]. Models developed by epidemiologists have been aimed at predicting epidemics, undertaking immunization programmes, or evaluating the efficiency of vaccination for them [2, 3]. Understanding how infectious diseases spread in space and time is essential. In both contexts, there exist many models of epidemic spread, but only a few incorporate real data, even fewer include heterogeneities (although there are exceptions [3], that incorporate, for example, the age structure of the population, or the networks of interactions, etc), and very few deal with the spatial component of epidemics [4]. This can be understood as it is hard to find long and reliable epidemic time series, and virtually impossible to find good spatiotemporal data with reasonably fine-grained sampling scales, of the order of one week in time, and of a few kilometres in space. The existence of a large network of general practitioners (GP) in France (about 500) evenly distributed over the French territory since 1984 allowed us to obtain high-quality data (there exist detailed reports for a large number of detected cases) on a weekly basis on a relatively small spatial scale (the data can be considered reliable on an average scale of 20 km) [5]. In the last few years, the size of the data set has increased considerably to reach 2000 reports within an epidemic period of about 15 weeks.

A question of particular interest is whether or not space is relevant to the spread of epidemics. If space is irrelevant, mean-field models, that rely on global mixing, are appropriate; otherwise, explicit modelling of space and of how epidemics diffuse

§ E-mail address: bonabeau@satafe.edu
‖ E-mail address: toubiana@b3e.jussieu.fr

0305-4470/98/190361+05$19.50

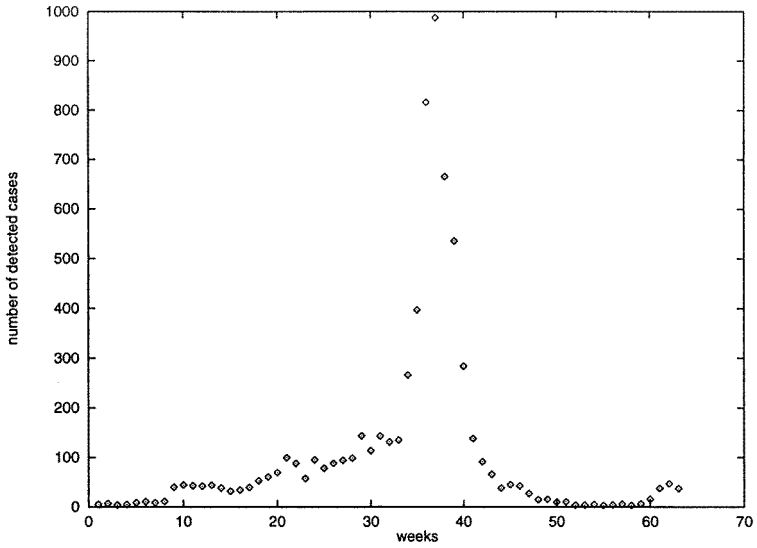

Figure 1. Time series of the number of cases of influenza reported by the Sentinelles network of GPs during the winter 1994–95 epidemic. The duration of the time series is 63 weeks (from the 30th week of 1994 to the 40th week of 1995), and the peak of the epidemic occurs at week 37.

is necessary. There has been evidence in the past that space is indeed relevant, and that epidemics spread with spatial waves from an initial epidemic centre [4]. However, transportation systems have changed in the last 30 years, promoting exchanges both among and within countries, such systems may have been amplified to the point where the timescale of global mixing is so short that it dominates over local dynamics. We focus here on influenza, the spread of which we analyse from a 'physics' perspective. Results are given for the epidemic of winter 1994–5, but essentially the same conclusions can be drawn for all other epidemic periods since 1984.

In order to study the role of space, we examined the spatial correlation structure of the epidemic, by measuring spatial correlations in the number of reported cases. We find long-range correlations, but, surprisingly, these algebraically decaying correlations are present as soon as the number of cases begins to increase, that is, about 15–25 weeks before the epidemic reaches its peak, which means that such correlations can be observed over a period of 30–50 weeks per year, as they also persist for 15–25 weeks after the peak. Figure 1 shows the time series of the total number of cases reported by the GP network over the whole territory, and figure 2 shows the density–density autocorrelation function at scale s (that is, when spatial data are binned in square patches of linear size s km, hereafter called s-patches) $C_s(h,t) = \langle\langle N_s(r,t)N_s(r+y,t)\rangle_r\rangle_{|y|=h}$, where t is the time (in weeks) starting from $t = 1$ (36 weeks before the peak of the epidemic), $N_s(r,t)$ is the number of cases reported during week t in the s-patch centred on location r, $\langle\ldots\rangle_r$ denotes averaging over r and $\langle\ldots\rangle_{|y|=h}$ averaging over $|y| = h$. We find that $C_s(h,t) \propto h^{\alpha}$ with $\alpha = -0.34 \pm 0.12$, from week 9 to week 49, up to a cut-off value $h_c \approx 1000$ km, which is of the order of the French metropolitan territory's maximum linear size. Before week 9 and after week 49, correlations are either short-ranged or exhibit no structure. Although the origin of the cut-off size h_c cannot be determined with certainty because this would require data from neighbouring countries, there is a good chance that it results from the

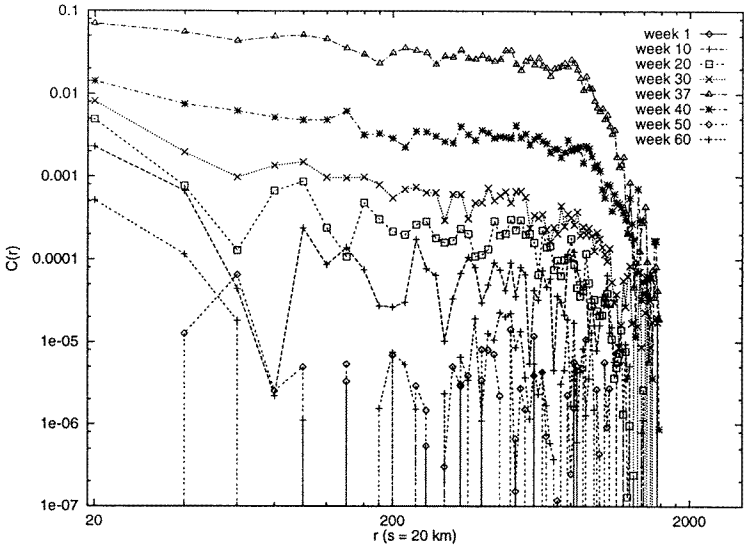

Figure 2. Spatial autocorrelation function $C(r)$ (r in km) of the number of reported cases for $s = 20$ km at weeks 1, 10, 20, 30, 37, 40, 50, 60.

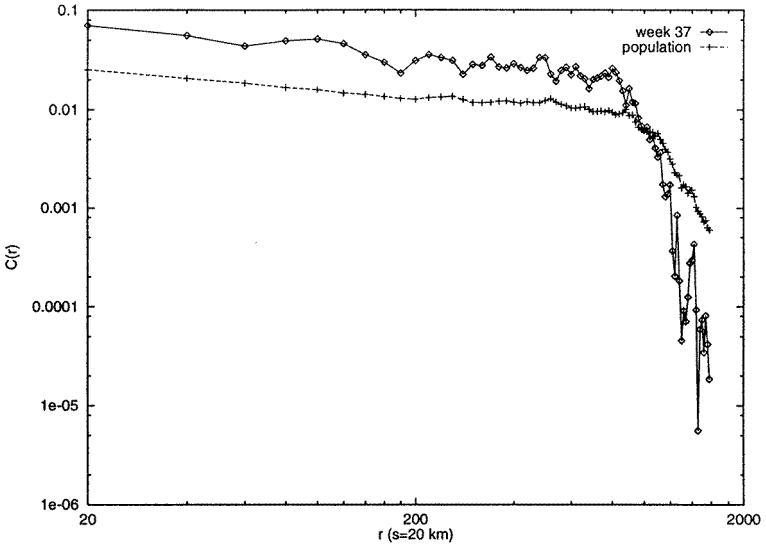

Figure 3. Spatial autocorrelation functions of the number of reported cases at week 37, of the distribution of GPs, and of the distribution of population for $s = 20$ km.

spatial limitations of the data set, and that, if data were available for territories beyond the French borders, correlations would extend to these territories. A fractal distribution of dimension D_0 should satisfy $C_s(h,t) \propto h^{D_0-d}$ in a d-dimensional space [6], which indicates that the distribution of cases is either fractal or multifractal, the fractal dimension being $D_0 = 1.66 \pm 0.12$. What is the origin of these long-range correlations, which are usually associated with critical points? One possible answer is that the epidemic process is in a sense continuously critical. Obviously the structure of the spatial population distribution

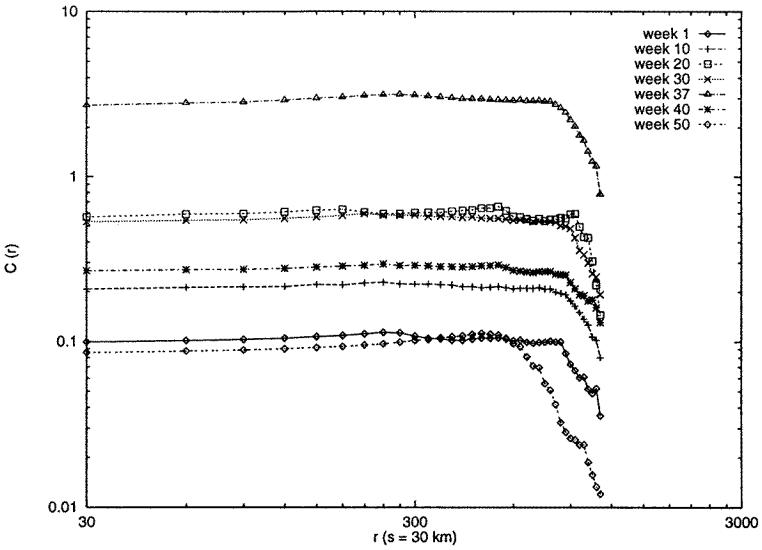

Figure 4. Spatial autocorrelation function of the number of reported cases divided by the size of the local population for $s = 30$ km at weeks 1, 10, 20, 30, 37, 40, 50.

influences the correlation properties of the spatial distribution of cases of influenza. It is therefore necessary to examine the role played by this underlying substrate.

Data obtained from the French Institut Géographique National (IGN) allowed us to study the properties of the spatial distribution of the population in France (from the last 1990 census survey). Figure 3 shows the correlations found in this data, compared with the correlations obtained at the peak of the epidemic. Figure 3 is consistent with the hypothesis that both curves exhibit approximately the same correlations ($\alpha = -0.31 \pm 0.03$ for the population distribution, consistent with what is found for the UK and the USA [7]). It suggests therefore that all of the correlations present in the case distribution can be due to the structure of the population distribution. In order to further test our assumption about the origin of the correlations, we have measured the spatial correlations of $n_s(r, t) = N_s(r, t)/P_s(r)$ where $P_s(r)$ is the total population in the s-patch centred around r. Figure 4 shows that the correlations in $n_s(r, t)$ can be reasonably well characterized by $\alpha \approx 0 (\alpha = 0.04 \pm 0.03)$. This result suggests that the distribution of cases is randomly distributed over the territory, with the number of cases in each patch being approximately proportional to the population (or weight) of the patch). It is tempting to conclude that the data reflect a random distribution of cases according to demographic weight, and that the epidemic propagates mainly through a global, and not only local, mixing process, certainly because of global transportation systems [8]. A mean-field regime is reached before local dynamical heterogeneities have time to dominate. Space therefore appears to be weakly relevant and models assuming perfect spatial mixing are appropriate.

We thank the sentinel general practitioners who collected the data. The Sentinelles network is part of the French Communicable Diseases Network (FCDN), developed at INSERM U 444 in collaboration with the Réseau National de Santé Publique (Public Health Network) and the Direction Générale de la Santé (Ministry of Health). EB is supported by the Interval Research fellowship at the Santa Fe Institute.

References

[1] Grassberger P 1983 *Math. Biosci.* **63** 157
Grassberger P 1985 *J. Phys. A: Math. Gen.* **18** L215
Cardy J L and Grassberger P 1985 *J. Phys. A: Math. Gen.* **18** L267
Boccara N, Cheong K and Oram M 1994 *J. Phys. A: Math. Gen.* **27** 1585
Johansen A 1994 *Physica* **78D** 186
Rhodes C J and Anderson R M 1996 *Phys. Lett.* A **210** 183
Rhodes C J and Anderson R M 1996 *Nature* **381** 600

[2] Kermack W O and McKendrick A G 1927 *Proc. R. Soc.* A **115** 700
May R M and Anderson R M 1984 *Math. Biosci.* **72** 83
May R M and Anderson R M 1991 *Infectious Diseases of Humans. Dynamics and Control* (Oxford: Oxford University Press)

[3] Hethcote H W 1978 *Theor. Pop. Biol.* **14** 338
Hethcote H W and Van Ark J W 1986 *Math. Biosci.* **84** 84
Hethcote H W and Levin S A 1989 *Applied Mathematical Ecology* ed S A Levin, T G Hallam and L J Gross pp 193–211 (Berlin: Springer)
Mollison D 1977 *J. R. Stat. Soc.* B **39** 283
Grenfell B T *et al* 1995 *Stat. Meth. Med. Res.* **4** 160
Post W M, DeAngelis D L and Travis C C 1983 *Math. Biosci.* **63** 289

[4] Cliff A D and Haggett P 1988 *Atlas of Disease Distributions: Analytic Approaches to Epidemiological Data* (Oxford: Blackwell)
Cliff A D 1995 *Epidemic Models—Their Structure and Relation to Data* ed D Mollison (Cambridge: Cambridge University Press)
Baroyan O V *et al* 1971 *Adv. App. Probab.* **3** 224
Murray G D and Cliff A D 1975 *Inst. Br. Geog.* **2** 158
Noble J V 1974 *Nature* **250** 726
Ryachev L A and Longini I M 1985 *Math. Biosci.* **75** 1

[5] Valleron A-J *et al* 1986 *Am. J. Public Health* **76** 1289

[6] Sahimi M 1993 *Rev. Mod. Phys.* **65** 1393

[7] Appleby S 1996 *Geograph. Anal.* **28** 147

[8] Flahaut A *et al* 1988 *Stat. Med.* **7** 1147
Flahaut A, Deguen S and Valleron A-J 1994 *Eur. J. Epidemiol.* **10** 471