Cholera in the 19 th century: Constructing epidemiological risk with complexity methodologies

Éric Daudé, Emmanuel Eliot, Emmanuel Bonnet

Cholera in the 19th century: Constructing epidemiological risk with complexity methodologies

Éric Daudé, Emmanuel Eliot, Emmanuel Bonnet

***Abstract*— Risk epidemic and complexity are linked by space and interactions. First, space matters in risk situations because of its ability to hold concurrently and simultaneously favorable conditions for a future emergence or re-emergence of epidemics. Second, space matters as a mediator of interactions, social as environmental, and at different levels. Risk is dynamic and its spatio-temporal dimension increases difficulties to catch it. Empirical data lack off precision to follow epidemiological outbreak. Complex system theory and connected methodologies can help us to enlighten this empirical failure.**

First, we present some knowledge about risk, health and complexity. Second we present social and spatial data based on the first epidemic of cholera in the city of Rouen, in 1832. Third we propose two models to explore the diffusion of this epidemic.

Index Terms— Risk, epidemic, dynamical system, cellular automata, modeling, simulation.

I. Introduction

THE spatial analysis of risk may be defined as the investigation of probability of being affected by a hazard in space and in time. This type of analysis requires a deep focus on the multilayered and complex combinations of indicators that are located in space. The analysis is ever difficult for certain risks but it reaches a peak when it concerns epidemiological ones. Because mobility and the ways people move in space and time is a major factor in the dynamic, especially in the case of an epidemic, the investigation of epidemiological risk faces three major problems:

- First, the need of understanding the ways people move and interact with space. Mobility is socially constructed and its patterns vary in history and according to social and cultural characteristics.
- Second, the question of the emergence of risk and its location. Because the intensity of the epidemic depends upon the level of interactions between people, we need to consider the accessibility of spaces in regards with others.
- Third, the evaluation of the temporality of the epidemic.

We propose to explore these problems on the basis of the analysis of the second cholera pandemic that affected most parts of France at the beginning of the 19th century. We examine it on the basis of an ecological approach of the epidemic but also by adding complexity theories analysis. The study will take advantage of these frameworks in order to avoid and overcome data bias.

In the analysis, space and risk are strongly related. Risk appears in space and can be created by space: not only considered as a support but also as an 'incubator' of risk situations. Risk is thus space related: presence and density of the *vibrio cholera* are dependent upon both aquatic reservoir and upon the more or less high concentration of population in the environment. In addition, risk occurs at different scale (world, nations and cities) and involves many actors: disease control, doctors and inhabitants. Risks are dynamics: going from the emergence of the virus to the pandemic may reflect this.

Self-organization theory [1] is adapted to explain emergence of risk for which local disruptions may product global and unpredicted events [2]. The self-organization theory identifies processes which allow describing behaviour at a global level, persisting in time and space, from numerous interacting entities located at one or several lower levels. Most of these interactions are local one and such systems are characterized by an absence of planning: no global control which would pilot such structure, such behaviour, or such form. Activity of

The authors would like to thank the IRSHS and the region Haute-Normandie for the funding of the program from which this work stems from.

Éric Daudé is with Rouen University (email: Eric.Daude@univ-rouen.fr), Emmanuel Eliot is with Le Havre University (email: Emmanuel.eliot@univ-lehavre.fr) and Emmanuel Bonnet is with Caen University (email: Emmanuel.bonnet@univ-caen.fr). Authors are with UMR-CNRS-IDEES.

such system, dynamic and open to their environment, is in evolution. Evolution between attractors can be cyclic. Such systems are characterized by phases of intense activities: evolution of societies is tagged by wave of huge pandemics. Otherwise, system can evolve towards a stationary state, converged at an attraction point and absorbing progressively its activity. Activity of a system, in an epidemic perspective, can lead it through different states through the time. This switch from a state to another is situated close to a bifurcation point that may lead towards chaos. In an earlier work, we explored the different phases of activity of the logistic function often linked with diffusion processes [2]. When the system evolves from a bifurcation threshold, the transition from one state to another qualitatively similar refers us to the concept of resilience. The stability of self-organized systems refers to the possibility of change which explains that all living systems go through distinct phases during their activities. These phases are theorised by the criticality [1], which shows that all self-organized systems evolve towards a critical state and that a small and local disruption is enough to produce huge alterations. This event is characterized by a system which goes into a phase of mutual and global interaction during which level of connections and interdependences is maximal: this is the case of pandemics.

If they are useful in a heuristic context, such concepts are however difficult to use when one wants to apply them or to spot them in an empirical way. For example, how evaluating the intensity of relationships between elements at the same level and between elements at different levels? These uncertainties lead us to propose simple models of diffusion based at the same time on empirical evidence and theoretical knowledge.

II. THE GLOBAL AND NATIONAL CONTEXTS OF THE EPIDEMIC AT THE BEGINNING OF THE 19TH CENTURY

After decades, routes of the epidemic have been recomposed on the basis of archives, reports and medical information. The second epidemic seems to have started in the British colony of Bengal in 1826. In 1837 the west coast of Mexico, the Anglo Egyptian protectorate and the French colonies of the North West of Africa seemed to have reported the last cases in the known world at the beginning of the 19th century (figure 1).

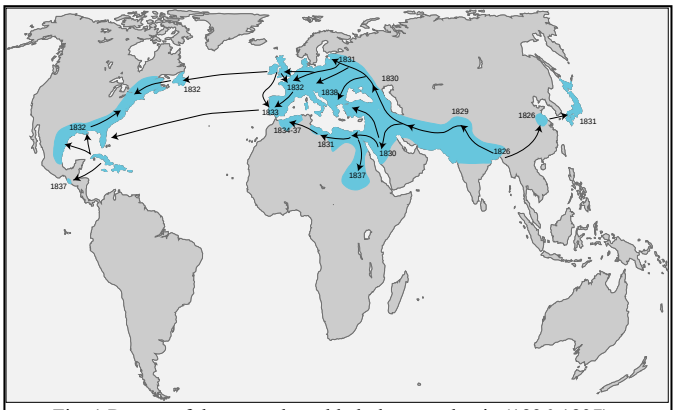


Fig. 1 Routes of the second world cholera pandemic (1826-1837).

Although the etiologic of *vibrio cholerae* was unknown yet, details about the transmission disease were reported by doctors of the British Raj since the beginning of the 19th century. As many other unknown diseases, cholera produced social reactions. In the French context of the 1830's, the epidemic broke out in a period of political troubles, which contributed to reinforce both the political conflicts and the social representations. Officially, the epidemic broke out in Paris in April 1832 and spread until the month of November. However, high rates of mortality due to diarrhoea were already reported in the northern parts of France by the end of 1831 [4].

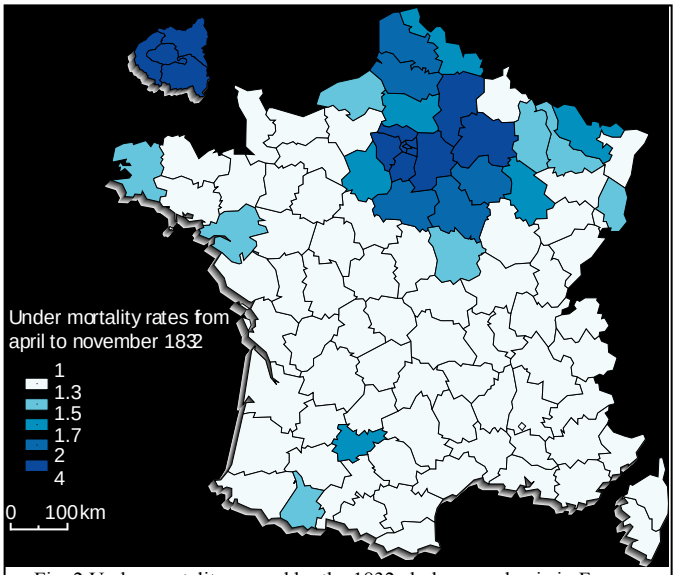


Fig. 2 Under mortality caused by the 1832 cholera pandemic in France.

The 'département' of the 'Seine-Inférieure' was not the most severely affected by the epidemic

according to the official sources, but located between Paris and the English coastline - two major epicentres of the epidemic - this region is interesting for analysing the course of the epidemic at several levels (figure 2).

III. The data

A. The Cholera data

All the cholera data were collected in the archive services of the region between 2006 and 2007. The collection covers the period 1832-1893, i.e. from the second to the fifth pandemic. The sources of information are mixed: medical reports, medical topographies, municipal and administrative reports. The present paper focuses on the second pandemic (« the cholera morbus epidemic ») and in the main regional centre of the department: Rouen. The epidemic reached this city in April 1832 and left it in October 1832. In Rouen, the analysis of the epidemic is based on two complementary materials: a medical topography done by the chief doctor of the hospital, Dr. Hellis (figure 3) and the report of the municipal officers. Obviously, the collected data suffer from bias.

Fig. 3 Dr. Hellis' map of cholera's cases.

First, in 1832, the etiology of the disease was unknown yet. The second pandemic was the first that reached France in the beginning of the 19th century and the causes and the ways of treatments were totally unknown, although information circulated all over the colonial empires. The main debate was based on the explanation of the causes of the disease: either contagious or not. This debate fed all the policies and fuelled all the theoretical conflicts until the end of the 19th century. The discovery of the cholera organism by F. Pacini in 1854 and after all by R. Koch in 1883 however improved the knowledge about the disease.

Second, the data were produced by a health care system that was centralized at the national and departmental levels. It produces statistics and topographies that gave an overview of the epidemic but it does not have the possibility to evaluate the under-reporting. Moreover, sources of report were very mixed: doctors, sanitary and municipal officers. In addition, the topographies (map) done on cholera aimed at proving the interpretation of the disease. In other words, all the cartographic methods were used in order to show that the epidemic comes from outside and was imported by seamen. However, comparisons between local data based on hospital and municipal reports with the Hellis'map seem to converge. As a conclusion, the map and the associated data are in fact the only available source able to trace the first cholera epidemic in Rouen.

B. From data to visualization and interpretation

We used different methodologies to integrate historical data from the archives services. The most important difficulties using archives data is the lack of statistical and geographical information. The first step was to consider if this collect was representative about the disease. The second step aimed at validating the data by checking their localization and their translation between manuscript and database. The third step aimed at calibrating the data with geographical information. Thus, we needed to reconstruct the geography of the 19th century by recreating the different administrative levels, by modifying the city's names and by identifying some city groupings.

The interpretation of the epidemic necessitates a better understanding of the topography and of the social geography of Rouen. Different surveyor's maps were available about the city. This step of modelling allowed us to construct environmental factors which were necessary to understand the context of the epidemic. Therefore, the hydrology, the topography, the open spaces, the fountains, and some public places (market) were integrated in a Geographical Information System (GIS). Based on

the available parishes, charity expenditures, rental values of housing, population density were integrated and compiled into indicators in the GIS. These information were based on academic and archives information that were compiled by historians [5].

For the geocoding, we have used the current street names for the first treatment, and for the unavailable data, we have compared current and past street names. After the first and the second treatment, 90% of the archive data were located.

The analysis of each cholera case is difficult because the information on the map were insufficient. So, by using some spatial analysis treatment we have produced a density map of the cholera cases in the city (figure 4).

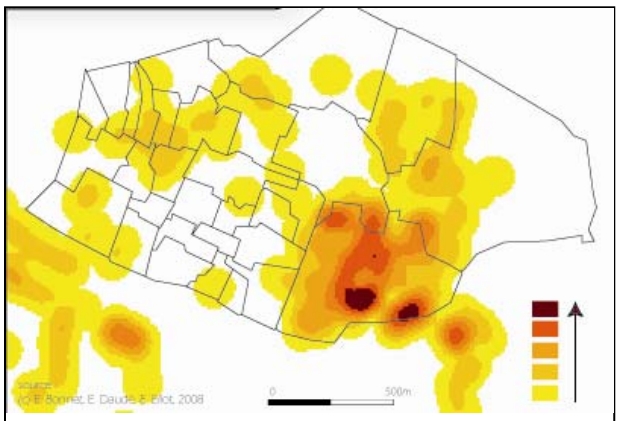


Fig. 4 Density map of cholera's cases.

There were many interpolation methods that provided this type of representation. We chose a Smooth Surface Reconstruction because this method uses natural neighbor interpolation, works in any dimension and allows dealing with non uniform samples. All these treatments allow identifying the most affected places of the epidemic.

A cluster of cases is reported in the South-eastern parts of the city. By using animations based on the weekly available data, we also identify the diffusion of cholera in the western and north western parts of Rouen. An analysis of the mapping shows a relation between the cholera cases and socio-economic indicators, the aquatic environment and the density of population. Based on this interpretation of the epidemic, we aim at describing and explaining the dynamic of this epidemic at an infra-urban level. In the next section of this paper, we describe two models that capture these social and spatial aspects.

IV. Mathematical and Computational Exploration of the Cholera Diffusion

First, we define a model based on ordinary differential equations. This macro model aims at understanding the general mechanisms of the spread of cholera. A second model is then presented, based on cellular automata. It takes into account both spatial and social heterogeneity as well as local interactions.

A. A basic model of contagion

The basic classical SIR model of Kermack-McKendrick [6] has to be modified to take into account the indirect process of contagion [7], i.e. the ingestion of contamined water. In this case propagation is contingent on the existence of a mediator which is the *vibrio cholera*. In the model, population is divided in four groups: Susceptible, Infected, Removal and Death. During an epidemic, a fraction of population is contaminated by the virus, mostly by ingestion of infected water. Once infected, people becomes actors of the propagation because they produce and reject *vibrio cholera* in the environment. After few days, evolution of the infection may lead to death or recovering, depending mostly on the health state of the individual and of the care conditions. We capture all this elements in the following model.

$$S(t+dt) = s(t) - r.S(t).dt \quad \text{avec } r = \beta.f(C) \quad (1)$$

$$I(t+dt) = I(t) + (r.S(t) - \gamma.I(t)).dt \quad (2)$$

$$R(t+dt) = R(t) + (\gamma.\alpha.I(t)).dt \quad (3)$$

$$D(t+dt) = D(t) + (\gamma(1-\alpha).I(t)).dt \quad (4)$$

$$C(t+dt) = C(t) + (e.I(t) + \lambda.C(t)).dt \quad (5)$$

In equation (1), parameter r represents the proportion of susceptible which is infected by the *vibrio*. Parameter β is the probability to interact with an infected aquatic environment and $f(C)$ is the probability to be contaminated by the vibrio, which is a function of the quantity of *vibrio* cholera (C) in water.

In equation (2), γ represents the proportion of infected case which get out from infection. Part of them will remove from infection ((α) in equation (3)) and other will die ($(1-\alpha)$ in equation (4)).

In equation (5), λ represents the loss rate of *vibrio cholera* in the aquatic environment and *e* the growth

rate of C due to the excretions of each infected case. The flowchart of this model is presented in figure 5.

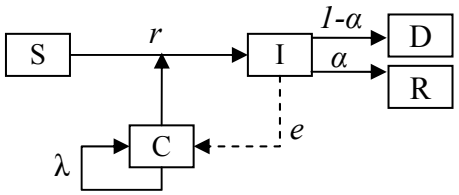


Fig. 5: Model diagram. Susceptible (S) can become infected (r) as they have a probability to be exposed to vibrio cholera (C). Infected people can die (D) or recovered (R) at rates (1-α) and α. While infected, people produce *vibrio* at a rate of *e*. Population of *vibrio cholera* declines at a rate λ.

We aim now at modifying this macro model in order to take into account the spatial dimension and the heterogeneity of the population. Implicit hypothesis of this model is a perfect mixing of the population with environment, but aquatic environment is not present everywhere in the city. The second implicit hypothesis is the homogeneity of the distribution and of the type of inhabitant, which is not the case.

B. A cellular automata model of contagion

A conceptual framework has been developed to capture the structures and dynamics which occur in the propagation processes [1, 8]. This framework is applied in the context of a cellular automaton. Structure of this model is defined by three elements:

- *Elementary entities*: a cell represents a square of one hundred meters. A cell can be an environmental cell {river, green space, public building...} or a 'social' one. In this last case, it has variable states which represent social attributes, such as number of inhabitants and the level of income. The domain is then the cellular grid (47x29) which shapes the city of Rouen.

- *Propagation channel*: it is the local spatial interaction structure of the cellular automata, i.e. the Moore neighborhood. Each cell can then interact with its 8 neighboring cells.

- *Virus*: It is the driver of the diffusion. This particle is generated by the fraction of infected people and is transported both by the environment and the infected cases. Each cell has a variable state which stocks an amount of the vibrio.

The diffusion dynamic is related to this structure and is composed of three processes:

- *Emission*: it represents the propagation of the vibrio, from the people to the environment. In the following simulation, each aquatic cell computes the stock of infected people (*I*) in its surrounding and receives a fraction *e* of *vibrio* related to this stock.

- *Circulation*: this process defines the 'life' of the virus in the environment. The circulation of the stock of *vibrio cholera* is mainly aquatic dependent. Three processes define this life: a water cell receives a fraction of virus from upper cells (1) and gives a fraction of virus to lower cells (2) - based on the elevation ground - and the stock declines at a λ rhythm (3);

- *Infection*: the rule of infection is relevant to this process. People interact with their environment and can be in contact with infected water, and be contaminated. In the model, each inhabited cell compute the volume of *vibrio* present in its vicinity - order 4 - and the stock of susceptible population has a risk *r* to be infected by this amount of virus.

Finally, once infected, the number of people which die or recover depend upon two parameters, *d* and *(1-d)*. These parameters, which are proportions, are the same for all cells. This latter process is linked with the circulation process in the sense that it represents an indirect rule life for the *vibrio cholera*.

Figure 6 presents the initial configuration of the model and the localization of the first cholera case in the south-eastern parts of the city.

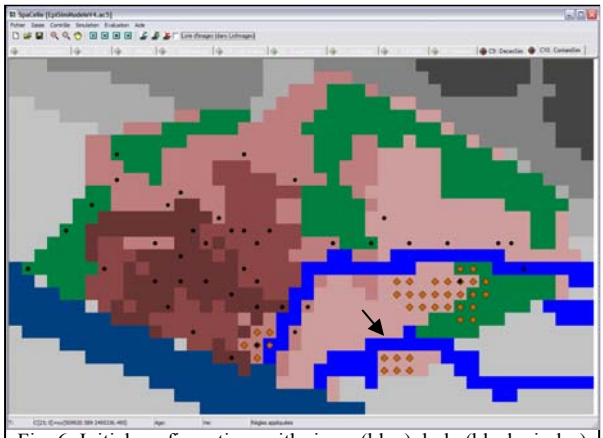


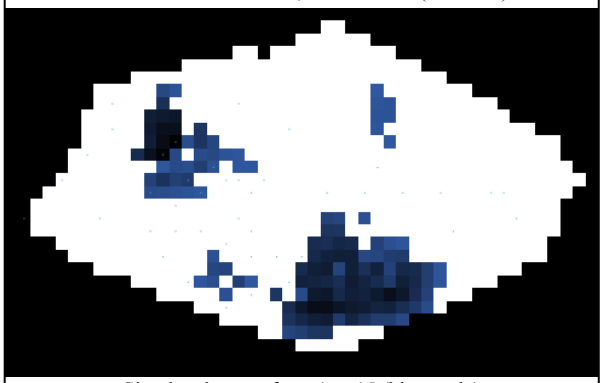
Fig. 6: Initial configuration: with rivers (blue), hole (black circles) green spaces (green), elevations (gray), income distribution (brown) and the first cholera case (noted by an arrow).

In the next section we present some simulations and first results.

V. SIMULATION OF THE CHOLERA DIFFUSION

In this cellular automata model, the two mains factors which are responsible of the propagation are the presence of aquatic environment (river and wells) and the density and the level of income of the population, which is measured by the charity expenditures. Firstly, the proximity to a wet environment increases the probability to get the vibrio. There are two rivers in Rouen, the *Robec* and the *Aubette*, and a high number of cases are reported along and between these two rivers. The main reason of this spatial correlation is the location of mills, spinning and paper mills where many workers were concentrated. Secondly, and correlated to the first, both the high density of population in these areas and the low level of income of these populations increase the probability to be contaminated. Health risk based on the analysis of social and spatial determinants is thus very high in this model, as probably in this past reality.

Observed cases in 1832, from 1 to 15 (red scale)



Simulated cases, from 1 to 15 (blue scale)

Fig. 7: Observations in 1832 and example of simulation.

These first results presented in figure 7 hold attention because there are good qualitative and quantitative correlations between the simulations and the observations.

Thus, the model has to be tested in different ways: What is the sensibility of the results to the parameters? Are the results significantly different if the first case is located in other sites in the city? Have the parameters significantly closed values to the observed ones?

REFERENCES

[1] P. Bak, *Quand la nature s'organise*, Flammarion, Paris, 1999.

[2] E. Daudé, D. Provitolo, E. Dubos-Paillard, E. Gaillard, E. Eliot, P. Langlois, E. Propeck-Zimmermann and T. Saint-Gérand, "Spatial risks and complex systems: methodological perspectives" in From System Complexity to Emergent Properties, Springer series understanding complex systems, in press.

[3] P. Langlois, E. Daudé, "Concepts et modélisations de la diffusion géographique", Cybergéo: revue Européenne de géographie, http://193.55.107.45/articles/364res.htm, n°364, 23 p.

[4] P. Bourdelais P., "La marche du choléra en France, 1832 et 1854", *Annales E.S.C.*, n° 1, pp. 125-142, 1978.

[5] J.-P. Bardet, *Rouen aux XVII^e et XVIII^e siècles. Les mutations d'un espace social*, SEDES, Paris, 197 p., 1983.

[6] W. Kermack and A. McKendrick, "A contribution to the Mathematical Theory of Epidemics," Proceedings of the Royal Society of London*, 115*, pp. 700–721, 1927.

[7] V. Capasso, S.L. Paveri-Fontana, "A mathematical model for the 1973 cholera epidemic in the european mediterranean region," Revue Epidémie et Santé Publique, n°27, pp. 121-132, 1979.

[8] E. Daudé, P. Langlois, "les formes de la diffusion", actes du colloque GéoPoint'04, pp. 171-175, Avignon, 2006.