CrossMark

Brief communication

Snow's case revisited: new tool in geographic profiling of epidemiology

Alessio Papini [a,*], *Ugo Santosuosso* [b]

[a] *University of Florence, Department of Biology, Firenze, Italy*
[b] *University of Florence, Department of Clinical and Experimental Medicine, Firenze, Italy*

ARTICLE INFO

ABSTRACT

Geographic Profiling technique is used to find the origin of a series of crimes. The method was recently extended to other fields. One of the best renowned data in epidemiology is that by John Snow during an outburst of cholera in London. We wrote Python scripts to perform the analyses to apply the Geographic Profiling for individuating the starting origin of an infection by using the old Snow's data set. We modified the method by applying a weight to each point of the map where cases of cholera were reported. The weight was proportional to the number of cases in a given location.

This modification of the Geographic Profiling method allowed to individuate in the map an area of maximum probability of the infection source, which was a few meters wide and including the historically known source of cholera, that is the "classical" water pump at Broad Street.

The method appears to be a useful complement in order to individuate the source of epidemics when available data about the cases of the infections can be summarized on a map.

© 2016 Sociedade Brasileira de Infectologia. Published by Elsevier Editora Ltda. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Geographic Profiling (GP) is an analytic tool widely used in criminology in order to identify on a map an area of highest probability assumed to contain the origin of linked events, typically crimes executed by a serial offender.[1] The method was extended from criminology to other fields where it was possible to identify a series of linked events which might have originated from a starting point in the space (represented on a two dimensional map). Fields of application other than criminology have been: invasion by alien species,[2–5] bumblebees foraging and nest location,[6,7] and infectious diseases targeting.[8,9]

GP uses the coordinates on the mapped events, creating a probability surface, the so-called geoprofile.[1] The geoprofile does not indicate the exact origin of the events, but rather prioritize a series of geographical points, based on the data.[1]

* Corresponding author.
  E-mail address: alpapini@unifi.it (A. Papini).

The geoprofile will provide on the map a decreasing probability density of finding the source of the events drawn on the map.[1]

The model does not search simply the geographical center of the events, but instead it considers a distance-decay function, such that the probability of an event will be lower by increasing the distance from the center of origin; and a buffer zone, within which the probability of an event tends to zero.[1] The distance-decay function is related to maximizing parsimony in movement, in economical and energy terms. Surprisingly, these functions revealed to be found not only for humans (criminals), but also even for invasive (not human) species[2,3] and infectious diseases.[8–10]

The need for analytical tools to recognize the source of the spreading of "something" (generally a threat) has always been an important task.[11] One of the best known cases is, in epidemiology, that of cholera outbreak in London, 1854, studied by John Snow[12] and widely cited as a seminal work in spatial epidemiology[13] [13 and references therein]. Dr. Snow tagged the cholera cases and the water pumps on the map of London and searched for the area with the highest number of cases, so discovering that the origin of the outbreak (the so-called focus of infection) was a contaminated water pump in Broad Street. The tagged cholera cases drawn by Snow on the map of London can be converted in a data set of coordinates, that was already used by Le Comber et al.[8] to test the GP method for targeting infectious diseases. Le Comber et al.[8] were able to mark a restricted area in the map of London containing the famous water pump of Broad Street (see Fig. 1C and D in their article). These authors used as input data the individual addresses where case of deaths due to cholera had occurred, that is 321 addresses, while the total number of cases amounted to 575, since more than one case might have occurred at the same address. Le Comber et al.[8] used this approach "to avoid the possible problem of spatial temporal non-independence due to secondary infections at a given address". Our approach included, instead, all cases assigning a weight to each point (addresses) proportional to the number of cases. We overlooked possible secondary human-to-human contagions, since cholera should not easily transmit from person-to-person, while its transmission is known to be more food- or water-born.[14] For this reason, we interpreted more than one case in the same address as independent events and hence summable.

Therefore, here we propose a new method of applying GP in which a different weight is assigned to each point of the map proportionally to the number of cases occurred in each point.

Methods

The data about the positions of cases on the map were acquired with Neuronmorpho (http://www.southampton.ac.uk/~dales/morpho/), a plugin of ImageJ (National Institute of Health; http://rsb.info.nih.gov/ij/), that can read a map position with a mouse click, building a csv file containing the coordinates point by point. Weights were added manually. Our method calculates the GP by weighting each point of the map in direct proportionality with the number of cases occurred in a given point of the map. That is, some points of the map are

more important than others. The data were analyzed with a Python script (Geoprof3.0.2.py).

Crucial for the GP analysis is the assignment of the values B, corresponding to the radius of the buffer zone.[2] In our analysis we used $B = 30$, corresponding to a buffer zone of 30 pixels (about 15 m on our map), that is quite small, with respect to other GP analyses in other fields, such as those on malaria cases in Cairo.[9] We evaluated more B values, calculating the impact on the analysis. The GP technique is described in detail in Papini et al.[3] The variable B (the buffer zone) is of course dependent on the map magnification and on the map resolution, since B is expressed in pixels, while the actual meaning of the buffer zone can be understood only if expressed in meters or km.

The Python scripts were written by the authors and can be retrieved from the site www.unifi.it/caryologia/PapiniPrograms.html. The scripts were executed with Python 2.7.3 (http://www.python.org/), running in Ubuntu 12.04 LTS operating system, kernel 2.6.32. The Python (>=2.6 version) programs need NumPy (http://www.numpy.org/), SciPy (http://www.scipy.org/), Matplotlib (http://matplotlib.org/), Scikit-learn (http://scikit-learn.org), and Python Image Library – PIL – (http://www.pythonware.com/products/pil/) libraries installed. A note about the software is provided as Supplementary material (SoftwareUsesupplementary.pdf).
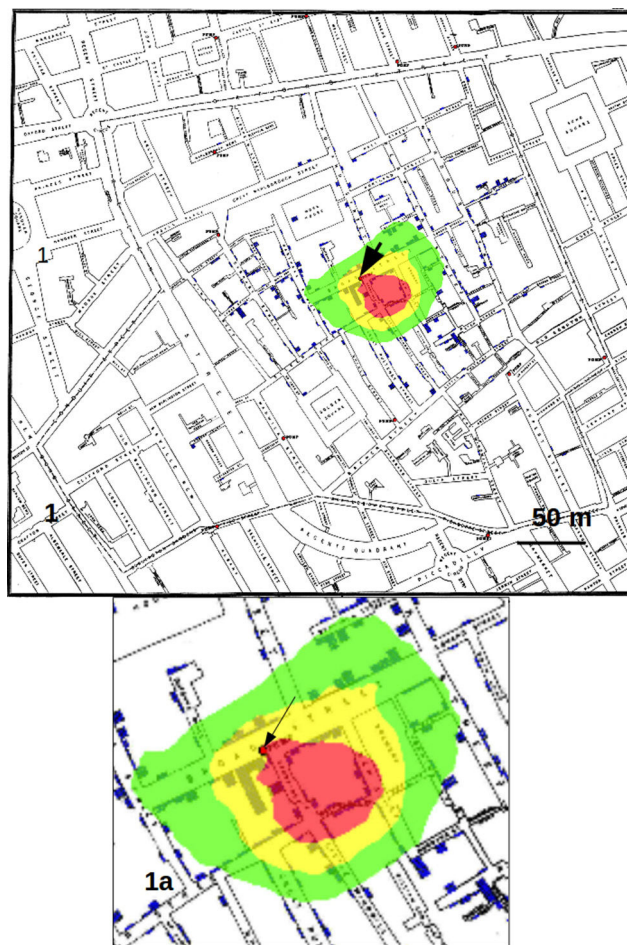

Fig. 1 – **Results obtained by considering only the addresses on the map as data sets. No weight is assigned to each address on the basis of the number of recorded cases.**

Results and discussion

Fig. 1 shows the results obtained by considering only the addresses on the map as data sets, corresponding to the analysis by Le Comber et al.,[9] that is, no weight was assigned to an address on the basis of the number of recorded cases. In Fig. 2 we show the GP analysis with weights assigned to each point of the map on the basis of the number of cases. The result is quite striking, since the red area, representing the area of the map with the points with 95% of highest probability comprised the pump of Broad Street. This area was about 30 m in diameter. With respect to the method that does not consider the number of cases as weights (shown in Fig. 1), the total area of highest probability of the presence of the source was hence much smaller.

Counting the pixels with highest probability of finding the source of the crimes, we found that the red pixels (those with highest probability) decreased substantially passing from considering only the addresses to using the whole data set with weights, that is from 36533 to 10068 (visible from the reduction in dimension of the red area from Fig. 1 to Fig. 2).
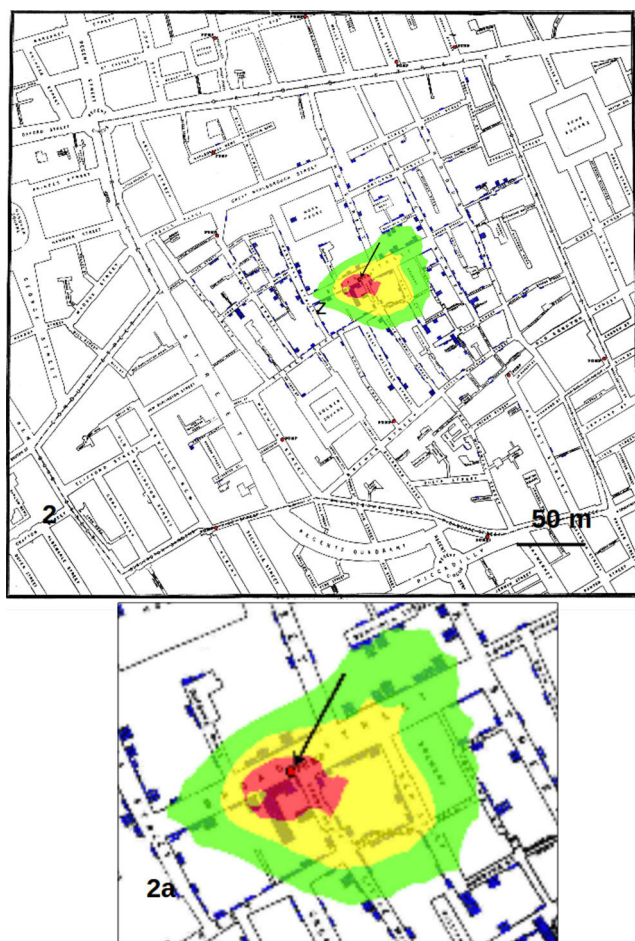
Fig. 2 – GP analysis with weights assigned to each point of the map on the basis of the number of cholera cases. The red area (that with highest probability to find the infection source) is only about 30 m in diameter and it comprises the famous pump of Broad Street.

Calculating each case as a single point, also if located in the same position on the map (that is at the same address), produced an area of red pixels only slightly higher with respect to the use of weights (data not shown).

Calculating the distance on the map, the GP analysis with weights produced an area of maximum probability of finding the source of about 30 m in diameter, which contains the well known source of cholera cases in London, that is the famous pump of Broad Street recognized by Snow.[12] This result shows that the use of weights proportional to the number of cases in each address largely increase the precision of the analysis, that is, it reduces the area of maximum probability where to look for the source with respect to other GP techniques as those employed by Le Comber et al.[9] and Verity et al.[11]

Conclusion

The weighted geoprofiling can be a useful method to identify a center of origin of an outbreak of a disease, in cases when more cases of infection can be found in the same point of the map (normally corresponding to a residence), largely reducing the priority points and hence showing the highest precision in delimiting the source search area.

The use of weights for more cases of infections at the same address, can be a good choice only in cases where secondary person-to-person infections can be considered not probable (as it is likely the case of cholera), otherwise, as stated by Le Comber et al.[9] it is necessary to use as input data each address (point on the map) as points with the same weight = 1.

Funding

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bjid.2016.09.010.

REFERENCES

1. Rossmo DK. Geographic profiling. Boca Raton, FL: CRC Press; 2000.
2. Stevenson MD, Rossmo DK, Knell RJ, Le Comber SC. Geographic profiling as a novel spatial tool for targeting the control of invasive species. Ecography. 2012;35:1–12.
3. Papini A, Mosti S, Santosuosso U. Tracking the origin of the invading *Caulerpa* (Caulerpales, Chlorophyta) with geographic profiling, a criminological technique for a killer alga. Biol Invasions. 2013;15:1613–21.

4. Cini A, Anfora G, Escudero-Colomar LA, et al. Tracking the invasion of the alien fruit pest *Drosophila suzukii* in Europe. J Pest Sci. 2014;87:559–66.

5. Santosuosso U, Papini A. Methods for Geographic Profiling of biological invasions with multiple origin sites. Int J Environ Sci Technol. 2016;13:2037–44.

6. Raine NE, Rossmo DK, Le Comber SC. Geographic profiling applied to testing models of bumble-bee foraging. J R Soc Interface. 2009;6:307–19.

7. Suzuki-Ohno Y, Inoue MN, Ohno K. Applying geographic profiling used in the field of criminology for predicting the nest locations of bumble bees. J Theor Biol. 2010;265:211–7.

8. Le Comber SC, Rossmo DK, Hassan AN, Fuller DO, Beier JC. Geographic profiling as a novel spatial tool for targeting infectious disease control. Int J Health Geogr. 2011;10:35.

9. Smith CM, Downs SH, Mitchell A, Hayward AC, Fry H, Le Comber SC. Spatial targeting for bovine tuberculosis control: can the locations of infected cattle be used to find infected badgers? PLOS ONE. 2015;10:e0142710.

10. Le Comber SC, Stevenson MD. From Jack the Ripper to epidemiology and ecology. Trends Ecol Evol. 2012;27:307–8.

11. Verity R, Stevenson MD, Rossmo KD, Nichols RA, Le Comber SC. Spatial targeting of infectious disease control: identifying multiple, unknown sources. Methods Ecol Evol. 2014;5:647–55.

12. Snow J. Snow on cholera. A reprint of two papers by John Snoe, MD, together with a biographical memoir by BW Richardson, MD, and an introduction by Wade Hampton Frost. New York: The Commonwealth Fund; 1936.

13. Shiode N, Shiode S, Rod-Thatcher E, Rana S, Vinten-Johansen P. The mortality rates and the space-time patterns of John Snow's cholera epidemic map. Int J Health Geogr. 2015;14:21.

14. Sack DA, Sack RB, Nair GB, Siddique AK. Cholera. Lancet. 2004;363:223–33.