Dealing with the Imperfection of Historical Hazards Data in GIS

Limits and Solutions to Make Chronological Maps for Decision Makers

Cécile Saint-Marc[1,2], Paule-Annick Davoine[1], Marlène Villanova-Oliver[1]

[1] LIG, STeamer team, Saint-Martin-d'Hères, France
{cecile.saint-marc, paule-annick.davoine}@imag.fr
[2] SNCF Infra, PSIGT-LVE, La Plaine Saint-Denis, France

Abstract. Natural hazard studies are often based on the spatial analysis of remarkable past events. However, historical information contains imperfections in its spatial and temporal locations and in the description of events. In this paper, we want to show the benefits of a system approach, that take the context of use of the visualization into account during the early phase of storing up data in the GIS, in order to make maps with imperfect data. We base on a concrete example: making a map of the chronology of phenomena during a past flood event, for decision makers who are lay users in cartography.

Keywords: historical data; natural hazards; imperfection; uncertainty; chronological map.

1 Introduction

The study of natural risks is often based on the understanding of remarkable past events, with the aim to prevent upcoming hazards. In that context, geo-historical data is used. But this kind of data has many imperfections: it comes from testimonies or archives and so it is heterogeneous, often poorly structured and imperfect in the description, location and dating of events [14].

The interest of Geographic Information Systems (GIS) has been demonstrated in the field of risk study [11]. However, this tool has two main limits in the context of historical data. Firstly, the temporal dimension was only recently implemented in GIS tools and is still poorly incorporated. Secondly, even if rules to visualize uncertainty exist ([8], [22], [12], [1], [17]) and may be applied in classical GIS, imperfection is not treated differently than other attributes of objects. The use of imperfect data in a GIS is a twofold issue: 1) during the modelling and the storage of the data, 2) to include imperfection in visualization.

Our work is based on a case study: the chronology of the flood of October 1940 on the river Le Tech, in the South-West of France. We endeavour to visualize the impacts of major historical floods on the railway system. It requires storing up in the

adfa, p. 1, 2011.

GIS data about the impacts of floods, their causes and responses to the events (decisions that were taken, reconstruction work). The information is intended to be visualized by decision makers who are not used to use complex maps. This perspective of use had an important impact in our approach. Indeed, [16] proposed to adapt the visualization to the users' level of experience with graphic displays and to the type of tasks the displays would have to support.

In this paper, we present the dimension of imperfection in geo-historical data, and propose some solutions to deal with the limits of GIS. Then we explore a chronological visualization of events and some ideas to show imperfection in the map.

2 The Dimensions of Imperfection in Geo-Historical Data

The information concerning risks is multidimensional. It incorporates spatial (location and spatial development), temporal (date, duration, return period), informational (description) and also contextual dimensions (the context of the event) [7]. Imperfection may be found in each one. For example, in the historical data about the flood of 1940 on Le Tech river, we found: "At 4:30pm, le Tech reached almost the underside of the bridge" (vague description), "during the flood from October 17th to 20th of 1940, most of the railway have been destroyed" (imprecise dating and location of impacts). So it is difficult to reconstruct the phenomena and their chronology.

Besides, every phenomenon has a limited reliability because of the nature of sources of information (testimonies, sometimes collected a long time after the event, thus inducing subjectivity). Moreover, the information is incomplete because some clues of old events have disappeared.

While constructing our information system, we chose to store imperfection in natural language (qualitative format) in order to distinguish the different kinds of imperfection. Every field of the GIS layer relative to a dimension of the information was lined with a second field concerning its imperfection.

Several classifications exist to characterise the kind of imperfection affecting data ([18], [10], [4], [9], [13]). We chose to use the classification of [20] that concern a large set of imperfections and that was built specifically to characterise spatial and temporal imperfection of natural risks data. It is presented in Figure 1.

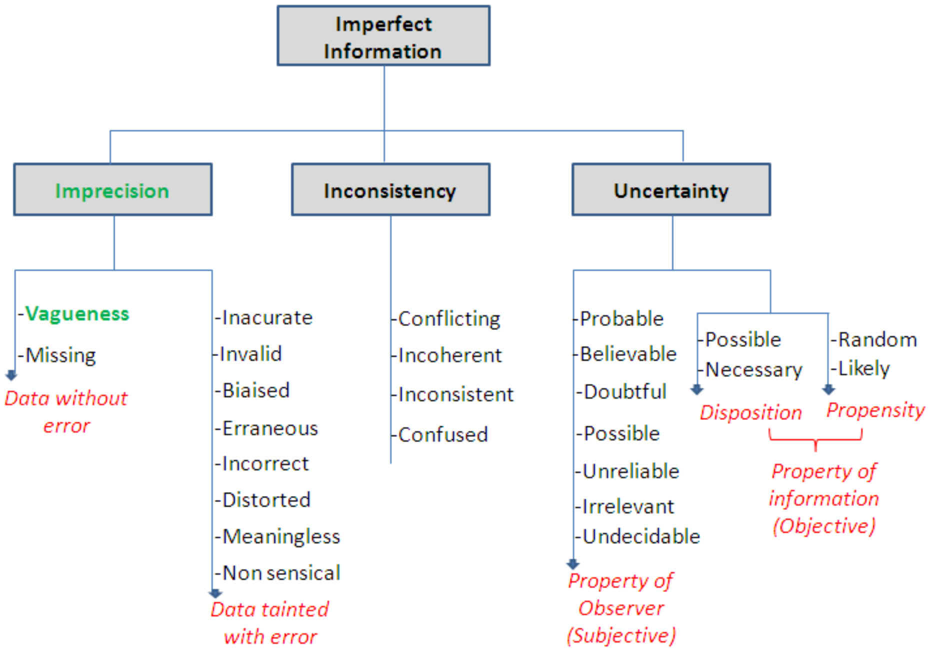

Fig. 1. Taxonomy of imperfect information ([19], reproduced by permission)

This taxonomy distinguishes three types of imperfection: imprecision (the true value is located in a defined subset of values), inconsistency (conflict or incoherence in the values) and uncertainty (partial knowledge about the true value of information). In our dataset, we can see that 50% of the dates are imprecise or uncertain, that lead to difficulties when we want to study the temporality of phenomena (Table 1).

Type	**Space**	**Time**	**Attribute**
Imprecision (3 subtypes)	5	9	8
Inconsistency (2 subtypes)	1	0	1
Uncertainty (6 subtypes)	4	11	7
Total	**10** (24%)	**20** (49%)	**16** (39%)

Table 1. Typology of imperfections in our historical dataset

3 Limits of Classical GIS to Manage Historical Information about Natural Risks

3.1 Semantic versus Geometric Heterogeneity

To store our data, we used 'classical GIS': ArcGis©, QGIS©. In these GIS "one theme is represented by one layer of geographic information, which is a group of elementary objects having the same geometry (point, line or polygon)" [15].

In our GIS, we had to deal with objects that were impacted by floods, with diverse geometries (e.g. station, rail track, city centre, mountainside). We had two options: either to create three layers for the three kinds of geometries or to create a single layer. We chose this second option and stored the impacts as points in the GIS, such as markers that indicate "that object was damaged". This solution had the advantage of keeping the thematic data-logic of the GIS. Second, we thought that punctual locations lightened the map load and so lead to a clearer map for users than polygonal geometries, even if it had not been tested at this time. To change precise linear or surface geometries into points introduce more imprecision in the data. However, we assume that storing punctual locations as polygons would have introduced the same level of imprecision.

For precise geometries, we placed the points at their barycentre or at the nearest point inside the surface or the line. For fuzzy locations, the point was located randomly in the possible area of location, or at the most probable location according to the topographic context when it was possible.

3.2 Date of phenomena and visualization

A set of tests with policy makers, led by [5], show that this audience finds it useful to visualize uncertainty of data, but wants to see information with no ambiguity ("is the future global water balance a problem or not?"). These two statements are paradoxical. We think that this need could be answered by suppressing as much as possible the imperfection in the appearance of 'raw' data, while it is notified in an other way (alternative map, pop-up window). We excluded the idea to assign a period of time (maximum and minimum possible dates) to damages on the rail infrastructure that were a priori one-time. As a consequence, we assign an only dating to each phenomenon of the flood, even if its real date is fuzzy. We also excluded the fuzzy set theory [24] because it was shown that lay users tend to ignore statistical probabilities when making decisions [23].

When the date of a phenomenon was missing, we excluded to assign no date in the GIS because, in an animated-map context, this phenomenon would never appear in the visualization or always stay on it. We used the hydroclimatic context of the flood to deduce probable dates for events. The imperfection of the date was recorded in the attribute table. For example, [21] dealt with a "mudslide in Le Firal" during the flood, but did not write its date. The flood peak of Le Tech River was on October 17th and 18th of 1940, and the rain intensity was maximal on October 17th. So we assigned the probable dating of October 17th to this phenomenon.

3.3 Temporal precision is different between periods

Historical events are described hour by hour in monograph about the event, such as in [21]. However, since the beginning of the decrease of the watercourse level, the temporal granularity decreases to a daily unit. And when the situation become normal again, the date is even rarely quoted. In fact, the temporal precision of data depends on the historical period described by the testimonies. Nevertheless, in our context, it is important to consider the phenomena that occurred after the flood, such as the approach used to restore the traffic.

On the one hand, for an event occurred during the flood period, we considered that “in the afternoon of October 18th” is an imprecise dating, because it is difficult to sort in a chronology with “on October 18th at 4:30pm”. On the other hand, we considered that a dating with a precision of the day, during the post-flood period, was precise because there was rarely more than one event per day in that period. Then, if two events happened the same day in that period, we considered that they happened at the same time.

4 First map of a Historical Chronology

Our data structuring work results in a first map of the chronology of the flood's impacts, which is presented on Figure 2. Several phenomena overlay spatially across time on the map. As a consequence, we chose to figure the dating with the variable 'size', in order to sort overlaid objects by size, as it was proposed by [3] in the use case of surfaces. That permits to see together the older phenomena (the bigger points) behind the more recent ones. Even if the variable 'size' is useful to handle overlaying, it is not adapted to visualize qualitative data [2]. Indeed, it seemed to us that it was difficult to link visually the absolute size of a point from the map to the legend. That is why we also added the variable 'colour': two different hues distinguish the rise and the decrease of the watercourse level, while the intensity of colour increases near the peak of the flood.

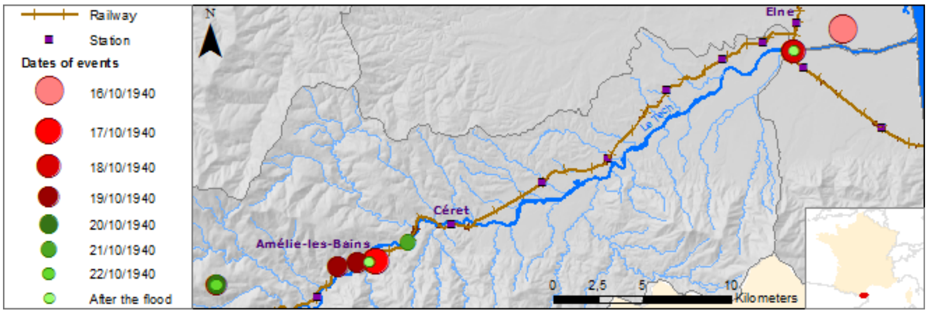

Fig. 2. Map of the chronology of impacts of the flood of October 1940 on the railway system in Le Tech watershed

However, this first map do not show the imperfection of the data. To represent together the value of data and its imperfection is still an open research issue [12]. Some authors proposed solutions to map several kinds of imperfections, together with a quantitative attribute of the data [1]. But most of the current studies consider only one type of imperfection on each map, or even different representations per subtype [13]. How could we visualize imperfection clearly for decision makers, if different kinds of imperfection exist together in the dataset? And what if several dimensions of the dataset are tainted jointly with imperfection?

Some courses of actions exist to answer these questions, such as the use of interactivity to explore the pairs {type of imperfection, dimension of information} one after the other. We also think that graphic visualization next to the map could be of interest to show the level of imperfection for each point of the map. Using charts involve to transform qualitative imperfection into quantitative data. Figure 3 shows two examples of charts, that we draw by giving a mark out of ten to each subtype of imperfection, according to its supposed impact on decision making from the map. Thus we turn temporarily the qualitative data into an ordered one, that provides information on the seriousness of imperfection towards the analysis. Of course, the sequence of ranking of the subtypes of imperfection depends on the context of use. This order should be objectivized thanks to a survey of users.

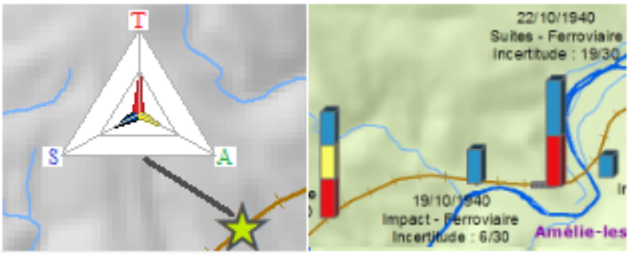

Fig. 3. Visualization of the imperfection of data for the three dimensions of information (time, space, attribute) thanks to graphs

5 Conclusion

Through several examples from a concrete use case, we have tried to show that the construction of a historical GIS raises issues, the answers of which are largely influenced by the context of use of data

With the aim to present the impacts of a flood, we stored imperfections in a qualitative format. We assigned a single date and location to each phenomenon of the flood, even if these parameters were imprecise or uncertain, in order to simplify the spatio-temporal visualization. This simplification should not prevent from including the dimension of imperfection in the visualization.

Nevertheless, there are still open issues. Some of these are inherent to the historical aspect of data and cartographers have to adapt. Others are relative to the cartographic visualization. In front of the limitations of GIS to deal with the temporal and uncertainty dimensions, we use empirical approaches, which do not follow the rules of graphic semiology.

6 References

1. A. Arnaud, P.A. Davoine. Temporal geovisualization in risk area. International Cartographic Conference, November 15-21, 2009, Santiago de Chile.
2. J. Bertin. La sémiologie graphique. Paris, Gauthiers-Villars, 431 p., 1967.
3. F. Bessadok, C. Domingués. Automatic evaluation and improvement of map readability. 25th International Cartographic Conference (ICC'11), Paris (France), July 3-8, 2011.
4. B. Bouchon-Meunier. La logique floue. Presses universitaires de France, Paris, "Que sais-je?" collection,127 p., 2007.
5. D.C. Cliburn, J.J. Feddema, J.R. Miller, T.A. Slocum. Design and evaluation of a decision support system in a water balance application. Computers and Graphics. 26(6):931-949, 2002.
6. P-A. Davoine, A. Arnaud, J. Gensel. A Tool for historical cartography about natural hazards, International Cartographic Conference, A Coruna ICC 2005, July 2005.
7. P-A. Davoine, B. Moisuc, J. Gensel, H. Martin. SIHREN Conception de Systèmes d'Information Spatio-Temporelle dédiés aux Risques Naturels, Revue Internationale de Géomatique, N° spécial, 16(3), April 2006.
8. S. Dragicevic, D. Marceau. A Fuzzy Set Approach for Modelling Time in GIS. International Journal of Geographical Information Science, 14:3, pp.225-245, 2000.
9. Theresia Gschwandtner, Johannes Gärtner, Wolfgang Aigner, and Silvia Miksch: A Taxonomy of Dirty Time-Oriented Data. In Lecture Notes in Computer Science (LNCS 7465): Multidisciplinary Research and Practice for Information Systems, 58 - 72, 2012.
10. H.D. Kraft, G. Pasi Disco, G. Bordogna. Vagueness and Uncertainty in Information Retrieval: How can Fuzzy Sets Help? IWRIDL, India, 2006.
11. S. Konecny, S. Zlatanova, T.-L. Bandrova. Geographic Information and Cartography for Risk and Crisis Management. Towards Better Solutions. Springer, 2010.
12. A.M. MacEachren, A. Robinson, S. Hopper, S. Gerdner, R. Murray, M. Gahegan, E. Hetzler. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. Cartography and Geographic Information Science, Vol. 32, No. 3, pp.139-160, 2005.
13. A.M. MacEachren, R.E. Roth, J. O'Brien, B. Li, D. Swingley, H. Gahegan. Visual Semiotic and Uncertainty Visualization: An Empirical Study. IEEE Transactions on Visualization and Computer Graphics, 2012.
14. B. Moisuc, J. Gensel, P.-A. Davoine, H. Martin. Designing Adaptive Spatio-temporal Information Systems for Natural Hazard Risks with ASTIS. Proceedings of W2GIS 2006, pp.146-157, 2006.
15. L. Sanders. Système d'information géographique (S.I.G), in Hypergeo electronic encyclopaedia, 2004. Available online: http://www.hypergeo.eu/spip.php?article63
16. Scaife, M., and Y. Rogers. 1996. External cognition: How do graphical representations work? International Journal of Human-Computer Studies 45:185-213.

17. Senaratne, H., Gerharz, L., Pebesma, E., & Schwering, A. (2012). Usability of Spatio-Temporal Uncertainty Visualization Methods. In Bridging the Geographic Information Sciences (pp. 3-23). Springer Berlin Heidelberg.
18. P. Smets. Imperfect Information: Imprecision – Uncertainty. UMIS, 1999.
19. Snoussi M., Gensel J., Davoine P.A., Extending TimeML and SpatialML languages to handle imperfect spatio-temporal information in the context of natural hazards studies, AGILE 2012.
20. M. Snoussi, P.-A. Davoine. Methodological proposals to handle imperfect spatial and temporal information in the context of natural hazard studies. International Journal of Geomatics and Spatial Analysis 3-4, 23, pp. 495-517, 2014.
21. G. Soutadé. Quand la terre s'est ouverte en Roussillon, l'Aiguat-octobre 1940, l'Olivier, Perpignan, 171 p., 2010.
22. J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, M. Pavel. A typology for visualizing uncertainty. Proceedings of SPIE & IS&T Conference, Electronic Imaging, Vol. 5669: Visualization and Data Analysis, pp. 146-157, 2005.
23. A. Tversky, D. Kahneman. Judgement under uncertainty: Heuristics and biases. Science 185:1124-31.
24. L.A. Zadeh. Fuzzy sets, Information Control, vol 8, pp.338-353, 1965