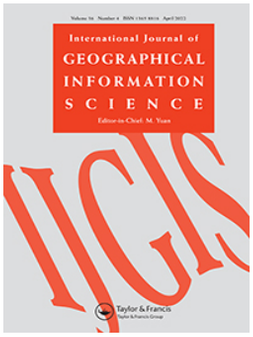


International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Visualising post-disaster damage on maps: a user study

Candela Thomas, Péroche Matthieu, Sallaberry Arnaud, Rodriguez Nancy, Lavergne Christian & Leone Frédéric

To cite this article: Candela Thomas, Péroche Matthieu, Sallaberry Arnaud, Rodriguez Nancy, Lavergne Christian & Leone Frédéric (2022): Visualising post-disaster damage on maps: a user study, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2022.2063872

To link to this article: https://doi.org/10.1080/13658816.2022.2063872

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2022.2063872

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Visualising post-disaster damage on maps: a user study

Candela Thomas[a], Péroche Matthieu[a], Sallaberry Arnaud[b,c], Rodriguez Nancy[b], Lavergne Christian[c,d] and Leone Frédéric[a]

[a]LAGAM - Université Paul-Valéry, Montpellier 3, France; [b]LIRMM - Université de Montpellier - CNRS, Montpellier, France; [c]AMIS - Université Paul-Valéry, Montpellier 3, France; [d]IMAG - Université de Montpellier - CNRS, Montpellier, France

ABSTRACT

The mapping of the damage caused by natural disasters is a crucial step in deciding on the actions to take at the international, national, and local levels. The large variety of representations that we have observed leads to problems of transfer and variations in analysis. In this article, we propose a representation, Regular Dot map (RD), and we compare it to 4 others routinely used to visualise post-disaster damage. Our comparison is based on a user study in which a set of participants carried out various tasks on multiple datasets using the various visualisations. We then analysed the behaviour during the experiment using three approaches: (1) quantitative analysis of user answers according to the reality on the ground, (2) quantitative analysis of user preferences in terms of perceived effectiveness and appearance, and (3) qualitative analysis of the data collected using an eye tracker. The results of this study lead us to believe that RD is the best compromise in terms of effectiveness among the various representations studied.

ARTICLE HISTORY
Received 4 June 2021
Accepted 5 April 2022

KEYWORDS
Semiology; geovisualisation; maps; post-disaster damage; user study; eye tracking

1. Introduction

Mapping is used extensively in the field of risk management. Maps are used at various stages of the risk management cycle, from the study of phenomena to the monitoring of post-crisis territorial reconstruction (Girres *et al.* 2018). During phases of crisis management, in an emergency situation, a map is traditionally used to spatially visualise the physical phenomena and their territorial consequences, whether future (impact scenarios) or past (damage assessment), and thus inform the decisions of institutional and local players and emergency services.

Today, the availability and accuracy of data have been improved by Earth observation technologies, in particular by satellite data. On the international scale, initiatives such as the International Charter 'Space and Major Disasters'[1] and the European Copernicus mechanism[2] use satellite data to map the consequences of a disaster in short times (several hours). These maps exhibit significant cartographic diversity due to the large number of producers involved and the absence of an international

CONTACT Candela Thomas thomas.candela@univ-montp3.fr

consensus. Researchers have already begun asking questions about the consequences of this diversity of cartographic representations (Griffin and Fabrikant 2012), in particular in the context of post-disaster maps (Kerle and Hoffman 2013). Such maps have not, however, been subject to an analysis of the cognitive and visual impact of the various methods of representing the damage in the process of reading and interpretation by the final users. Moreover, these processes can be disturbed by the fast pace and extraordinary character of the working environment, as is the case in emergency management. In this context, the effectiveness and the quality of interpretation are crucial to guaranteeing fast and appropriate decision-making in response to a crisis.

In this article, we study user behaviour by comparing a representative sample of existing visualisation methods routinely used in rapid post-disaster mapping presenting the same information (quantitative and qualitative). To these standard methods, we have added a visualisation, not used in this product, constructed with size and colour visual variables applied to a regular tessellation according to Bertin's rules of semiology[3] (Bertin 1967). Our hypothesis is that this technique optimises the identification of post-disaster damage by improving the visual salience of the information. In order to make this comparison, we conducted an experiment in which participants had to carry out a list of identification tasks using the various visualisations. We then quantitatively analysed the results in terms of accuracy (i.e. conformity of user responses to the reality on the ground), response time, and user preference. In addition to this statistical analysis, we performed a qualitative analysis of the data collected using an eye tracker. These analyses shed light on the processes of reading of the various types of maps by the participants.

2. Background and motivation

In this section, we describe the context of our study, and more particularly how maps constitute tools for abstraction of real phenomena that can be used for decision support in emergency situations. We then explain our comparative approach by introducing elements relative to the theory of map reading.

2.1. Maps as tools for abstraction of real phenomena

For a long time, maps have been perceived as a tool for consultation, communication and decision support. They provide an abstraction of reality via the graphical representation of geographic information. They illustrate both tangible phenomena such as land occupation and intangible phenomena such as investment flows (Kraak and Fabrikant 2017). Spatial analysis via maps makes it possible to study territorialised processes in order to understand the dynamics thereof over time. In this respect, maps are used in numerous fields for various purposes. Examples include topographic maps for orientation and direction; statistical maps for synthesising complex data; predictive maps for anticipating natural phenomena; regulatory maps for governing territorial development; and operational maps for guiding risk and crisis management.

This short typology highlights the great diversity of the maps produced. At the same time, the tools used in their design have become extremely widespread. As a

result, maps are no longer the carefully crafted product of specialised cartographers. They have become a participative and collaborative tool far removed from the vision of the traditional schools of cartography. Today, new forms of representation have emerged through the development of GIS (Geographic Information Systems), the geo-visualisation tools that have flourished through the 'GeoWeb', and data visualisation (Herring 1994, Roche *et al.* 2013). However, despite the significant advances in map design tools and diffusion media, the main paradigms that govern this discipline have remained unchanged, as evidenced by the continuing fundamental importance of Bertin's work on the semiology of graphics (Bertin 1967). These semiological rules are a response to the necessity for the designer to communicate an effective piece of information while reducing, by previous convention, the potential interpretation biases (Chauvin 2005). The semiology of graphics forms a reference frame for the relationship between the effectiveness, appearance, and usefulness of a map (Kuveždić Divjak and Lapaine 2018). In particular, it provides a framework for the creation of maps that satisfy a specific need. This need takes the form of a certain number of tasks that the user must be able to accomplish with the map, according to the problem they are seeking to solve and the available data. In other words, the semiology of graphics allows the definition of a map (*how?*) according to a problem stated by a user and the creation of a list of what the user wants to visualise (*what?*) and why they want to visualise it (*why?*) (Munzner 2014). When used correctly, the semiology of graphics helps the user concentrate on the data essential to carrying out a task.

2.2. Operational use of a map in emergency situations

Risk management is traditionally characterised by several phases before and after a crisis. In each phase, maps are routinely used for interpretation and/or anticipation before a phenomenon occurs and as a tool for communication and planning of actions in the short term. During or just after a crisis, they are used in the emergency phase to identify zones to be evacuated, get rescue teams on the ground in the areas hit the hardest, and deploy vital resources according to needs (Coyle and Meier 2009, Roche *et al.* 2013). Once the emergency phase is over, they are used for monitoring reconstruction and as spatial reference frames for scientific feedback (Rey *et al.* 2019, Cécé *et al.* 2021).

Among the various types of maps, post-crisis damage maps play an essential role in decision-making by allowing the size of the operational response to be optimised. They are produced with Earth-observation (satellite) technologies by rapid mapping services, which provide the international community with information on the magnitude of a disaster for free. These actions are inscribed in the International Charter 'Space and Major Disasters'. When it is activated, rapid mapping services such as SERTIT[4] and DLR ZKI[5] are deployed in order to carry out missions of damage detection and inventory. They transform the raw spatial imaging data into geographic information usable by the crisis management operations. At present, the damage maps are mainly created via manual photointerpretation. The comparison of the satellite images acquired before and immediately after a disaster reveal the changes in the state of buildings caused by varying amounts of damage and allow the hardest-hit areas to be

mapped. This work results in the production of maps and vector data that satisfies the informational needs of decision support, with the main goal of rapidly optimising the organisation and deployment of rescue forces (Chiroiu 2004). However, the analysis of several of these maps on the international scale shows that there has been no concerted reflection on the choice of semiology, the levels of analysis and the techniques used in these post-disaster damage maps.

This diversity was particularly notable during the earthquake in Haiti (2010). The plethora of maps produced, rapidly available and using varied modes of representation, raises questions about the consequences of such production on the reader's understanding and the risks of confusion and uncertainty caused by a large influx of different maps. Such situations are considered to be counterproductive, in particular for rapid decision-making with limited time (Staněk *et al.* 2010), and especially when the damage is mapped differently using the same datasets. However, this diversity is counterbalanced by the standards found in risk and crisis mapping (Konecny and Bandrova 2006, Stachoň *et al.* 2016). These standards can be found more particularly in the colours used to visualise the consequences of the event or the symbologies used to locate points of interest in crisis management. It is therefore essential to take into account the rules of cartography and semiology, as well as user feedback, in order to optimise the capacity of maps to provide information and thus participate in effective crisis communication.

2.3. Theory of map reading: how to analyse the reading of a map

When reading a map, there is a difference between the information mapped by the cartographer and the perception that the target group of users has of this information (Popelka *et al.* 2012). The difference between these input and output variables can be used to measure the effectiveness and the readability of a map, which depend on several factors. These include the compromise between the cognitive capacity of the reader, their knowledge of cartography, and semiological characteristics of the map (Fairbairn 2006, Jégou and Deblonde 2012). Another method involves analysing the map-reading strategy via anatomical processes of the body. To understand the cognitive processes that condition this reading, we must understand its main sensory pathway: vision (Board and Taylor 1977). This involves, *inter alia*, the movements of the eye, and more particularly the path of the foveal zone over a visual scene. This field of study has given rise to theoretical approaches, namely in psychology and psychophysics, the latter of which links the intensity of the stimulus to the force of sensation on which perception thresholds depend (Stevens 1975). Perception is thus implicitly related to the combination of factors that drive our senses and favours pre-attentive and attentive vision processes, which involve detecting salient regions (pre-attentive) and then focusing visual attention on these regions (attentive) in order to describe the information (Itti and Koch 2001). In this respect, Jégou (2013) mentions bottom-up and top-down processes corresponding, respectively, to the acquisition and the visual decoding of a scene.

Modern studies in this field are carried out with the technologies of eye tracking (Popelka *et al.* 2019). Based on the recording and analysis of the movements of the

eye, and more particularly of the fovea, this technique allows an objective assessment (Goldberg and Kotval 1999) of the impact of the visual on the processes of reading and interpreting a map. These studies mainly focus on the ease of use and on the semiology of the display of web interfaces.

The present study focuses on the visualisations used in damage maps that can be reproduced using data gathered on the building scale. It also integrates new visualisations based on rules of semiology. The goal of this approach is to analyse the visual capabilities of this range of visual variables in tasks of locating damage and their ability to transmit the information that provides the correct results. For this, we used eye-tracking technology on a homogenous group of users. A protocol was put in place in order to collect qualitative and quantitative data on the strategies of the eye, the responses provided and the time required to complete a series of tasks.

3. Visual techniques

The goal of this study is to analyse several techniques of cartographic representation of the density of damaged buildings and their level of damage after a disaster. We compare 6 visualisation techniques, which can be classified as shown in Figure 1.

Harmonised maps were produced using open data for several territories. The map surround consists of a background with identical colours in which maritime zones are represented with a light blue, zones of study with a medium grey, and the other land zones with a darker grey. Grey was chosen because it promotes similar degrees of saturation and acts as a neutral link among the colours while limiting the discrimination thereof (Itten 1967, Sève 2009). The legends of the maps were also harmonised.

Besides the map background, the visualisations can contain one or two layers. In the former group, the two variables studied, building density and damage level, are represented using the same graphic elements. In the latter group, a first layer contains a dot-based layout of the buildings and a second, semi-transparent layer shows the damage levels. Figure 2 shows examples of the 6 techniques used. The damage level

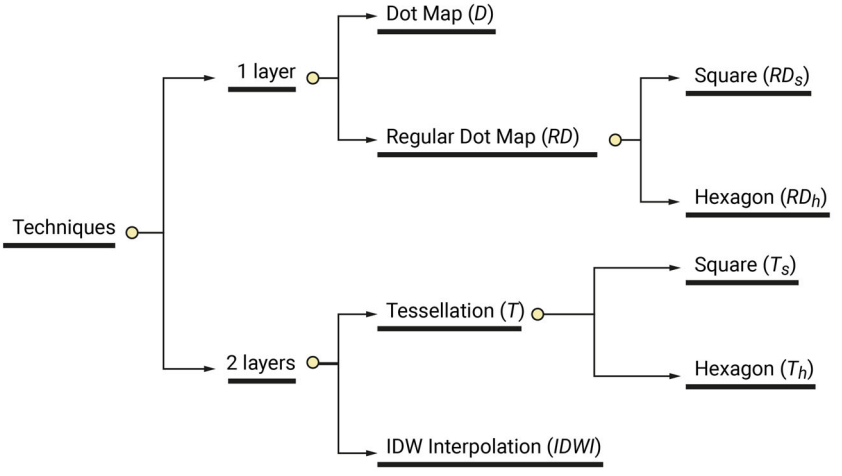


Figure 1. List of techniques evaluated.

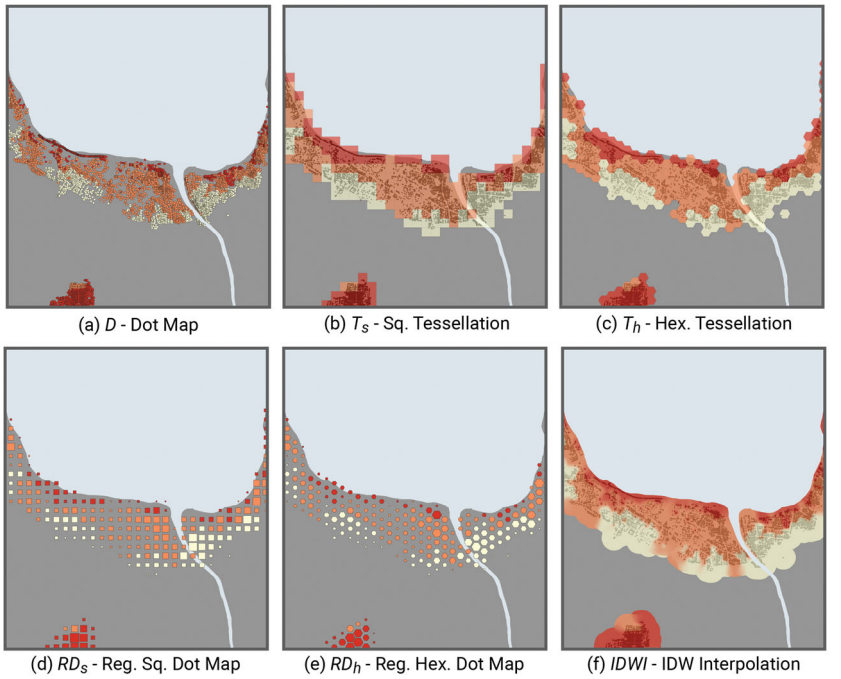


Figure 2. Visual techniques compared in the study.

is always shown using the same colours: light yellow for a low level of damage, orange for an intermediate level, and red for a high level. To choose these colours, we used the Color Brewer[6] diagnostic tool to generate a palette of optimal colours for our 3 classes of damage. This palette preserves the ordinal differentiation of the levels of damage regardless of the type of screen, printing process and colour-blindness of the participant (Brewer *et al.* 2003). We will now describe the specificities of each of these visualisations in detail.

3.1. Dot map (D)

The principle of this technique is to represent each building with a dot, the colour of which indicates the level of damage (see Figure 2(a)). This is one of the most commonly used techniques in rapid post-disaster mapping. In general, dot maps are used to visualise a spatial distribution of a qualitative variable. The dispersion or clustering of the units visually evokes a certain density of the number of objects. This visual technique has the advantage, according to the amount of superposition, of displaying the actual location and layout of the object (Hey and Bill 2014). Nevertheless, it requires a certain number of criteria in order to be readable: (1) effects of superposition that could hide certain values must be avoided, (2) it is sensitive to visual overload caused by data in the background, (3) the dot must have a diameter of 2 mm to be perceived correctly (Bertin 1967). For this representation, the major limitation lies in the loss of readability when the maps are made on the smallest scales: overlapping of the dots (1) or dots too small to be perceived correctly (3).

3.2. Tessellation (T)

Tessellation is a visual technique frequently used in numerous fields, for example in the socio-demographic studies of the British Census and the INSEE (200-meter grid) on the national scale and with the data of GRUMP (10-km grid) on the international scale, as well as in archaeology (Moreau and Locht 2016) and biodiversity (Krebs 1989, Birch *et al.* 2007, Gros-Desormeaux *et al.* 2015). It is also one of the techniques used to create choropleth maps.

It involves constructing a regular tiling (called grid) of identical polygons (called cells), the dimensions of which can vary according to the scale, the dataset and the desired level of precision to be displayed. By way of analogy, tessellation is comparable to the pixels that compose a raster image. This capacity to refine the resolution of the grid allows the map to be representative of the source datasets. With the tools of GIS, this method is relatively simple to implement. Its regular grid facilitates comparison and allows the change in one or more pieces of data to be assessed via an unchanging and easily shared support. However, These regular geometric shapes, of varying dimensions, smooth out the information located at the boundaries of geographical entities, particularly when natural boundaries (island boundaries, land use) or administrative boundaries are involved (Lajoie 1992). Moreover, the choice of the optimal dimension of the cells is difficult to determine and there are no real predefined rules. The final result and the statistical smoothing of the data during the aggregation are thus largely dependent on these dimensions.

In our context, the colour of the polygons of the tessellation represents the most frequent value of the damage levels observed on the scale of the buildings that they contain. As indicated in the introduction, the tessellation is carried out on a semi-transparent layer that shows the buildings as dots in order to allow the visualisation of their distribution.

- *Square Tessellation* (T_s) This technique involves constructing a regular grid of square polygons (see Figure 2(d)) that displays qualitative information (damage level). In our case, each polygon has a dimension of 150m $\times$ 150m. This parameter ensures a fineness compatible with the scale of the maps to be produced (1: 25,000).
- *Hexagonal Tessellation* (T_h) The juxtaposition of hexagons, also called a honeycomb (Kienberger *et al.* 2009), allows the formation of a regular tiling that has numerous advantages (see Figure 2(e)): (1) this structure allows the sampling deformations to be reduced by substantially approaching the circular shape while ensuring the constitution of a graphic view without discontinuities (Krebs 1989), (2) the perimeter/surface area ratio, close to that of a circle, reduces the edge effects (Blondel and Ferris 1995) and involves a more natural abstraction of the data, (3) the hexagonal shape, less rectilinear and thus less rigid for the eye (Birch *et al.* 2007), optimises readability and thus facilitates the interpretation of the data (Carr *et al.* 1992). In our context, the dimensions used are a height of 150m and a width of 173m.

3.3. Regular dot map (RD)

The two variables that must be visualised in damage maps – the quantity of affected buildings and the levels of damage – are quantitative and ordinal qualitative variables, respectively. According to the rules of semiology first set out by Bertin (1967) and then refined in various works in visualisation and cognitive sciences (Munzner 2014, Ware 2019), these two types of variables are called 'ordered' and can be represented, in the same dot, using two visual magnitude variables. In our context and in order to remain coherent with the other visualisations, the colour represents the damage level. According to the degrees of effectiveness proposed by Munzner (2014), and given that the position cannot be used (it already represents the geolocation of the dots), we propose visualising the number of buildings via the size of the dots. We thus propose a quantitative and qualitative visualisation of the damage via a regular distribution of dots[7] (see Figures 2(b,c)). This method, borrowed from Bertin (1967), uses the centroid of the polygons of the tilings described in the previous visualisation and varies the colour and surface area visual variables to represent the two dimensions of the data. More precisely, the damage level most often observed for the buildings contained in a polygon of the tiling is represented by the colour of the polygon. The number of buildings is represented by its surface area, the dimensions of which are calculated by Flannery compensation (Flannery 1971) in order to facilitate the reading of the proportionality. In this case, the polygons thus no longer form a tiling, but their centroids remain uniformly distributed on the map.

Like for the previous visualisation, we use two types of polygons, the square and the hexagon, which produces two visualisations, the *Regular square Dot map*, (RD_s, Figure 2(b)) and the *Regular hexagonal Dot map*, (RD_h, Figure 2(c)).

3.4. IDW interpolation (IDWI)

The last technique studied is called spatial interpolation. This method is well known in the world of geographic information because it allows complex data to be displayed graphically and makes it intelligible (Garnero and Godone 2014). In our context, it allows us to display the levels of damage semi-transparently, with the buildings being represented on a layer placed below, as mentioned in the introduction of this section. A colour is assigned to each pixel according to a numerical value calculated on the basis of the spatial data in the pixel's neighbourhood. The technique exists in two forms: deterministic and stochastic. Their implementation is identical, and the difference lies in the type of weighting that is applied. The general procedure involves defining a search area (or search neighbourhood) around a location and assigning a weighting according to the data observed in this area.

For our study, we chose the deterministic technique of IDW (Inverse Distance Weighting) spatial interpolation, which is routinely used in damage maps (see Figure 2(f)). This automatic technique has the advantage of being relatively easy to implement due to its limited number of input parameters. Indeed, the operation requires a neighbourhood search parameter, an exponent, and optionally a smoothing factor (Hessl *et al.* 2007). The method is based on the theory that the value of a location is a weighted average of the data available in the zone surrounding this location

(Mitas and Mitasova 1999). In our context, we used the damage data for the buildings in order to calculate this weighted average, which is based on a search (neighbourhood) radius of 200 meters, the same as in the techniques used by SERTIT[8] for the maps produced in Haiti after the earthquake of 2010.

4. User study

In our experiment, we asked participants to observe several maps created with the techniques presented above and to perform several tasks routinely carried out in crisis situations. This section describes the experimental protocol: tasks to be performed by the participants, datasets used, working hypotheses, participants, materials, and experimental procedure.

4.1. Tasks

The tasks to be performed by the participants were defined on the basis of scientific feedback recently obtained (February 2020) from operational players (prefecture, firefighters, and private companies) and a questionnaire distributed to 85 people at a scientific and technical conference. This event on the theme of geographic information in risk and crisis management brought together emergency services, risk managers and scientists. The results of the survey clearly show that the two main tasks expected in relation to post-disaster damage maps are the evaluation of the damage (40/85) and coordination and intervention (27/85) (see Figure 3).

Experts first seek to locate the zones with high levels of damage. Then, they try to estimate the quantity of buildings affected (and/or human risks if this information is available) in order to optimise the resources deployed. The priority is not necessarily to intervene in the zones with the highest density. For example, if a road to a group of buildings that are few in number but isolated is cut off, the consequences can be substantial and it is therefore necessary to intervene there as fast as possible. Based on these observations, we proposed the 4 tasks described in Figure 4. The first two involve locating the regions that sustained a high level of damage and contain a large (*T*1.1) and a small (*T*1.2) quantity of buildings. The next two tasks involve evaluating the level of damage (*T*2.1) and the quantity of buildings affected (*T*2.2) for a given region.

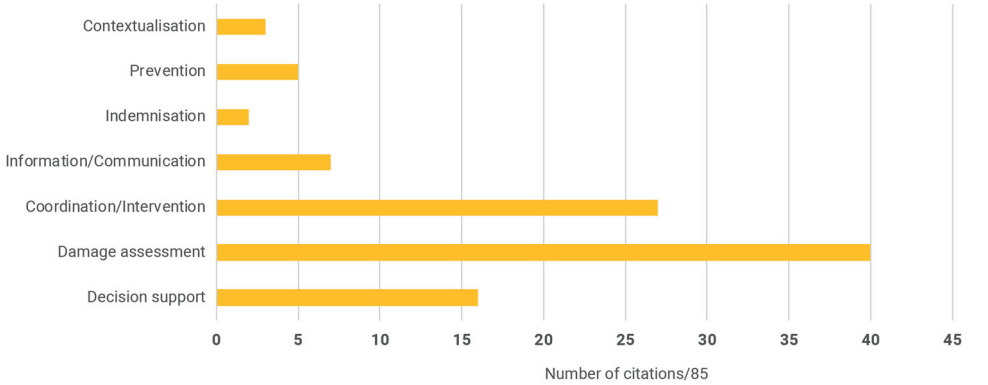


Figure 3. Number of conference participants citing each task.

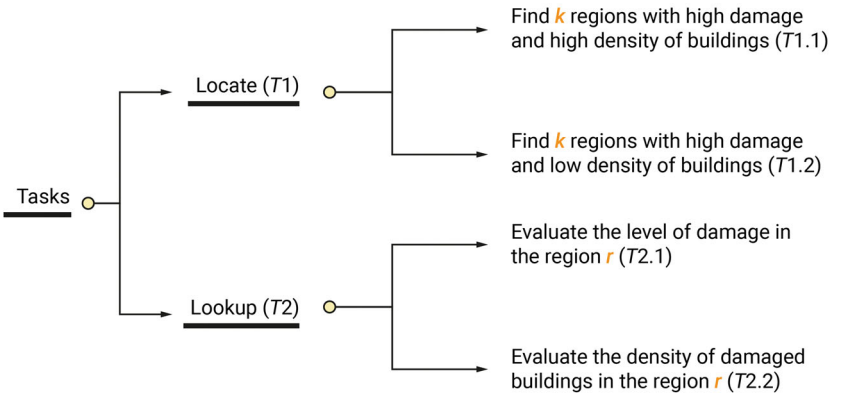


Figure 4. List of tasks used in this study.

Figure 5. Example of labelling of a visualisation of the RD_h type in order for the user to be able to identify the various zones and perform the tasks.

According to the search-task taxonomy described by Munzner (2014), the tasks *T*1 are classified in the category *locate* because the user knows what they are looking for (*target known*) but does not know the location thereof (*location unknown*). The tasks *T*2 are classified in the category *lookup* because the user knows what they are looking for (*target known*) and knows the location thereof (*location known*).

Letters were placed on each of the maps in order to allow the participants to name the highly damaged regions for the tasks *T*1 and identify the zone to be observed for the tasks *T*2 (see Figure 5). For each of the tasks and for each of the datasets, two experts in damage mapping determined the correct answers, in particular via GIS tools. These answers were used as references for analysing the validity of the answers given by the participants via an accuracy score.

4.2. Datasets

We used 4 datasets each relating to a different geographic area (see Table 1). Different datasets were used so as to vary the sectors analysed in order to limit

Table 1. Datasets used for the experiment.

Dataset	Disaster type	Data source	Year
DS_1 - Indonesia (Palu)	Earthquake and Tsunami	Copernicus EMSR 317[14]	2018
DS_2 - Dominica (Saint-Joseph)	Hurricane (Maria)	Copernicus EMSR 246[15]	2017
DS_3 - Saint-Martin (Marigot - Nettlé Bay)	Hurricane (Irma)	Copernicus EMSR 232[16]	2017
DS_4 - Haiti (Port-au-Prince - Carrefour)	Earthquake	UNITAR[17]	2010

cognitive redundancy, a phenomenon in which learning processes may have the effect of making the tasks progressively easier during the study. Each dataset consists of information on land occupation, limited to the terrestrial and maritime zones and rivers, and on the location of buildings, with 3 levels of damage: low, medium, and high. The datasets were selected on the basis of the availability of open data and the magnitude of the disasters in terms of damage to buildings.

These 4 datasets were used to create $4 \times 6 = 24$ maps using the 6 visualisation techniques described in section 3. As mentioned in the previous section, the toponomy of the maps was simplified with letters in order to facilitate location when performing the tasks (see Figure 5). Each map was accompanied by an inset for the title, the legend, the scale and north. These elements were essential in order for our visuals to be considered maps.

4.3. Hypotheses

An effective map must provide readable, accurate and easily understandable information, limiting interpretation bias. These conditions are even more important in crisis situations, where decisions must be made quickly, and participants are under intense stress. Rapid post-disaster mapping still shows a diversity of damage representations. The ex-post study of these visualisations implies higher visual and cognitive charges during the reading processes. These charges are related to the aggregation methods and the addition of layers required to permit a quantitative and qualitative analysis of damage. We propose a representation based on the rules of graphical semiology to take advantage of size and colour variables to provide this dual information on a single layer. Furthermore, the literature shows the visual advantages of the hexagonal shape over the square. In order to test the potential interpretation bias, the contribution of the graphical semiotics rules and the hexagonal shape, we formulated 6 hypotheses from several tasks performed using rapid maps for post-disaster assessment.

- H_1**:** The reading of the maps via the tessellation techniques (RD_s; RD_h; T_s; T_h) is more accurate overall than with the Dot map (D) and the technique of IDW interpolation ($IDWI$).
- H_2**:** Regular Dot (RD_s; RD_h) should give more accurate overall results than Tessellation (T_s; T_h).
- H_3**:** In general, the Regular Dot (RD_h) and Tessellation (T_h) techniques that use a hexagonal shape to aggregate the data provide better accuracy than the same techniques using the square shape (RD_s; T_s).

- H_4**:** Regular Dot hex (RD_h) is the representation that provides the best overall accuracy in the analysis of the maps.
- H_5**:** The techniques that use two layers to present the quantity and the quality of the damage levels (T_s, T_h, *IDWI*) take more time to analyse than techniques that use two visual variables in a single layer to present the same information (D, RD_s, RD_h).
- H_6**:** Given the previous works mentioned in section 4.2, the users should prefer, in terms of use and appearance, the Tessellations and the Regular Dot maps with hexagons (T_h and RD_h) over those with squares (T_s and RD_s).

4.4. Participants

The experiment was carried out on a sample of 24 participants, all students. They come from a master's program specialised in disaster and natural risk management at the Université Paul-Valéry Montpellier 3 with a strong foundation in the cartographic approach and GIS tools. 9 were female and 15 male. The average age was 24.6 years old, the youngest being 22 and the oldest 39.

In order to favour proper assimilation of the protocol, participants were welcomed with a presentation of the approach and the goals of the project. In this phase, we systematically carried out a test of colour perception. This allowed any possible cases of dyschromatopsia to be identified and guaranteed the reliability of the results. We used the Ishihara test (Ishihara 1918), which is simple and fast and allows detection of all forms of colour blindness. It consists of 38 plates divided into 7 sections. Each section concentrates on the levels of colorimetric deficiency (protanopia, deuteranopia, etc.), except for the first, the results of which should be identical for all people who take the test, even those with deficiencies. This particular section is used to identify potential cheaters. Each section was presented successively to the participants. The results did not bring to light any problems of colour perception for any of the sample.

4.5. Apparatus

The techniques of eye tracking measure the movements of the eyes of the user to find out which zones of the support image they looked at and when, the duration of the gaze, and the path followed by the eyes while carrying out a specific task. The data was collected by the Tobii X2-60 eye-tracking system with a frequency of 60 Hz, i.e. a measurement of the position of the eye was taken every 16.67 milliseconds. The eye tracker was connected to a Windows laptop with an Intel Core i7-4710MQ Quad Core 2.5 GHz processor, 16GB RAM, and an NVIDIA Quadro K3100M video card. The 17.3-inch screen had a resolution of 1920 x 1080. The various maps to be analysed by the users were displayed at screen resolution. We used the program Tobii Studio Professional, version 3.4.8, to implement the test protocol, collect the data and export it for processing. Once the user was in place facing the computer, a visualisation of the position of the eyes was used to adjust the distance and the orientation of the screen. Then, the calibration was carried out. Once this step was completed, the sequence (see section 4.6) was launched and the recording of the eye movements began.

4.6. Procedure

The trials of the study involved a *TECH* $\in$ [D,T_s,T_h,RD_s,RD_h,*IDWI*], a *TASK* $\in$ [$T1.1$,$T1.2$,$T2.1$,$T2.2$] and a *DS* $\in$ [DS_1,DS_2,DS_3,DS_4]. In total, there were 96 trials (*TECH* $\times$ *TASK* $\times$ *DS*), to be performed the same number of times. Each trial provided two measurements, a completion time and an accuracy score. For 'Locate' tasks ($T1$), this score was a real number between 0 and 1, corresponding to the percentage of correct regions found. For 'Lookup' tasks ($T2$), it was 1 or 0, depending on whether the participant evaluated the level of damage/density of buildings correctly or not.

In order to maintain the participants' concentration, we decided to set the number of trials for each participant to 36 since a pilot study performed with 4 additional participants showed us that 36 trials were completed in about 20–25 minutes. We thus employed a [$3 \times 4 \times 3$] within-subject design with the 3 primary variables of *TECH*, *TASK* and *DS*.

Given the two constraints (same number of trials and $3 \times 4 \times 3 = 36$ trials per participant), we followed a procedure inspired by Lobo *et al.* (2015) to define a set of 24 sessions (a session being an ordered set of trials for a participant), ensuring proper counterbalancing (see Figure 6). Indeed, as highlighted by Lobo *et al.* (2015), proper counterbalancing is of utmost importance when dealing with real-world data because the complexity of the trials cannot be controlled.

The trials of each session were grouped into 3 blocks, one per *TECH*. Since each session involved 3 out of the 6 available *TECH*, we selected and ordered the *TECH* for each session as follows: A Latin square gave us 6 orders for the 6 *TECH*, which we repeated once, leading to 12 orders. Then, each order was split in two to produce the 24 counterbalanced *TECH* orders, one per session. This way, pairs of consecutive participants completed trials on the 6 *TECH*, 3 per participant.

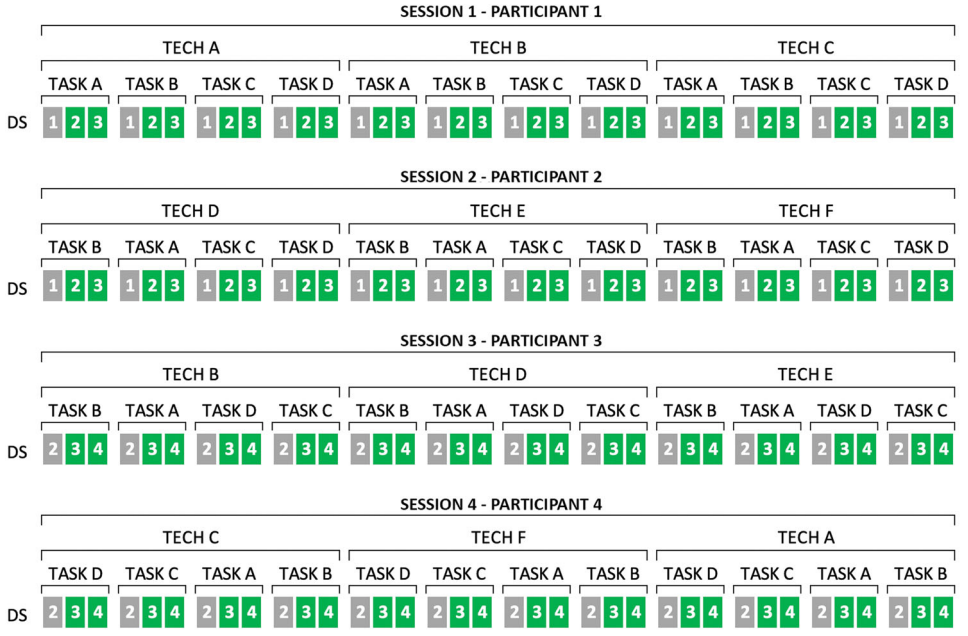


Figure 6. First four sessions of the study.

One of the $4! = 24$ TASK orders was assigned to each session. Thus, the order of the tasks was the same for all the *TECH* in a session, but different from one session to another. The *DS* were grouped into batches of 3 out of the 4 available. Pairs of consecutive sessions had the same *DS* to ensure counterbalancing. Thus, each of the $\binom{4}{3} = 4$ batches was used 6 times. The first trial of each *TECH* × *TASK* pair was considered training (grey background in Figure 6), It allowed us to check that the participant had understood the visual encoding and the task at hand. The accuracy and completion time of the second and third trial were recorded (green background in Figure 6). The order of the *DS* in each batch varied so that a different *DS* was first (DS_1,DS_2,DS_3 - DS_2,DS_3,DS_4 - DS_3,DS_4,DS_1 - DS_4,DS_1,DS_2); this also ensured counterbalancing. This procedure produced $24 \times 3 \times 4 \times 3 = 864$ trials, 576 of which were recorded to perform the analysis. The counterbalancing ensured that the 96 combinations of *TECH* × *TASK* × *DS* were performed 9 times by different participants and recorded 6 times.

A trial started with a screen showing the task question. The participant then had to click on the screen to display the map and answer the question. When done, they had to click again to switch to the next trial. There was no time limit and the recorded time corresponded to the duration of display of the map. An experimenter was available to answer questions about the task if there was any confusion. When the 36 trials were completed, the experimenter asked the participant which *TECH* they found the most useful to perform the tasks, and which one was the most aesthetically pleasing. The participant was then asked to rate their degree of visual fatigue after the session on a Likert scale. The average time of a session was 20.5 minutes (min: 14.9, max: 31.5).

5. Results

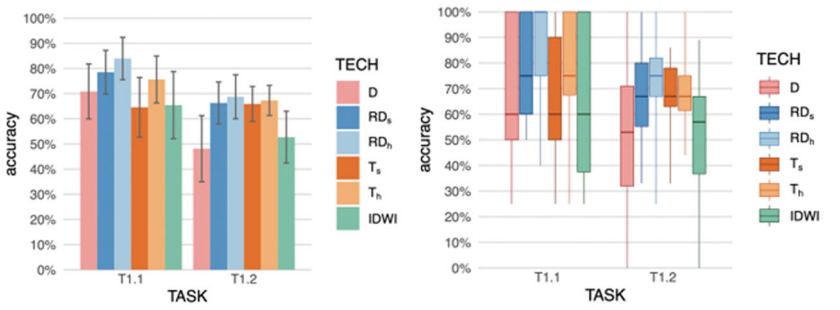
In this section, we start by statistically analysing the results obtained by the participants in terms of accuracy and response time. Then, we present the feedback from the participants in terms of perceived effectiveness and appearance. Finally, we discuss the results in relation to the hypotheses formulated above, the analysis of the maps, and the data collected by the eye tracker.

5.1. Accuracy

The tasks of the *T*1 type involved locating several sectors on a map. The value recorded for each trial corresponds to a percentage of correct answers. For example, for a particular trial, the participant may have found 4 out of the 10 sectors having a high level of damage and a high building density. On the contrary, the tasks of the *T*2 type involved identifying a level of damage or a building density for a given sector. A participant's answer was thus either true or false. The statistical analysis of these two types of tasks must thus be based on different approaches given the nature of the values observed.

5.1.1. Locate tasks

Figure 7 shows the results obtained for the tasks of the *T*1 type ('Locate', see Figure 4). For each task and technique, we show the average and the confidence interval



(a) Mean accuracy for $T1$: error bars represent the 95% confidence interval

(b) Box plots of the accuracy for $T1$

Figure 7. Accuracy for $T1$ per *TECH*.

(Figure 7(a)), as well as the minimum, the first quartile, the median, the third quartile, and the maximum (Figure 7(b)) of the accuracy rates of participant answers.

For task $T1.1$, the techniques T_s and *IDWI* give the worst results, with an average of 65%. Technique *D* is next with an average of 71%. Although the results with this technique are slightly better on average than with the previous two, the median is the same (60%). The techniques T_h and RD_s give better results, with an average rate of correct answers of 76% and 79% and a median at 75%. The technique RD_h gives the best results, with an average rate of correct answers of 84% and 50% of participants answering everything correctly (i.e. the participant identified all the sectors in question, the median is 100%). A statistical analysis[9] shows that the technique RD_h is significantly better than the techniques T_s ($p = 0.0076$) and *IDWI* ($p = 0.0107$). The other p-values are all greater than 0.07. If the techniques are grouped together by type (*T* and *RD*), we can see that the techniques of the *RD* type are significantly better that the techniques of the *T* type ($p = 0.0299$) and *IDWI* ($p = 0.0124$). With regard to the number of layers, the techniques that use one are better than those that use two ($p = 0.0271$). The techniques that use tiling (*RD* and *T*) do not obtain better results than the others ($p = 0.0903$). However, Figure 8(a) shows that the type of tiling used seems to have an effect (hexagons appear to give better results than squares), but this difference is not significant ($p = 0.088$).

For task $T1.2$, the techniques *D* and *IDWI* give the worst results (average: 48% and 52%, median: 53% and 57%). The difference between these techniques and all the others is in all cases significant ($p < 0.04$). Consequently, the techniques that use tiling (the other four) give better results ($p = 2.48e-05$). With regard to these four, the results are very close for the techniques RD_s, T_s and T_h (averages between 66% and 67%, median at 67%), and are slightly better for RD_h (average: 69%, median: 75%). However, the difference is not significant ($p \geq 0.66$). Contrary to the observations for the first task, the number of layers used by the technique does not seem to have an influence on the results here ($p = 0.813$). The same is true for the type of polygon used for the tiling of the *RD* and *T* techniques ($p = 0.603$).

In conclusion, the techniques that use tiling (*RD* and *T*) are in most cases more effective than the others (*D* and *IDWI*) for performing the tasks of the $T1$ type.

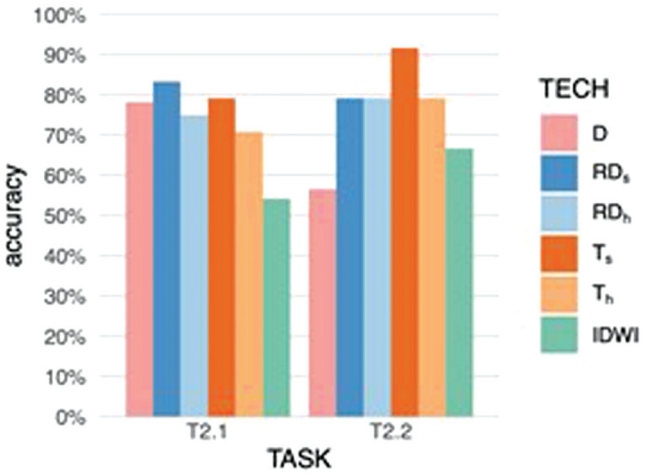


Figure 8. Mean accuracy of *T*2 per *TECH*.

Therefore, this confirms the hypothesis H_1 for the tasks *T*1. Among the techniques that use tiling, given an identical type of tiling (hexagonal or square), the techniques of the *RD* type give better results than those of the *T* type, which confirms the hypothesis H_2. With regard to the type of tiling, our results suggest that a hexagonal division is preferable, which confirms the hypothesis H_3. Finally, the technique RD_h gives better results than the others. RD it seems that this technique should be preferred for the *T*1 tasks, which confirms the hypothesis H_4.

5.1.2. Lookup tasks

Figure 8 shows the percentage of tasks for which the user correctly identified the level of damage (task *T*2.1) and the density of damaged buildings (task *T*2.2) for a given sector.

For task *T*2.1, the participants obtained worse results with the technique *IDWI* (54% correct answers). This difference is significant[10] only for RD_s ($p = 0.0356$) and more generally for both techniques of the *RD* type ($p = 0.0313$). Next come the techniques T_h (71%), RD_h (75%), *D* (78%), T_s (79%) and finally RD_s (83%). No difference is observed between the techniques that use one layer (*D*, *RDs*) or two (*T*, *IDWI*) or between the presence (*RD*, *T*) or absence (*D*, *IDWI*) of tiling. It should be noted that contrary to the tasks *T*1.1 and *T*1.2, the participants performed better with the techniques that use square tiling (RD_s and T_s) than with those that use hexagonal tiling (RD_h and T_h). Still, the difference is not significant ($p = 0.33355$).

For task *T*2.2, the participants obtained significantly worse results ($p = 0.00839$) when they used the techniques that do not feature tiling: *D* and *IDWI* (57% and 67% correct answers). The next 3 techniques, RD_s, RD_h and T_h, obtained similar results (79%), and one technique stood out (92%). However, the difference is not significant ($p = 0.23413$). No difference is observed between the techniques that use one (*D*, *RD*) or two (*T*, *IDWI*) layers. The type of tiling, square or hexagonal, does not have an effect on the results for the techniques of the *RD* type, but it does for the techniques of the *T* type, for which participants obtained better results when squares were used.

In conclusion, the techniques that use tiling are not more effective for the task *T*2.1, but they are for the task *T*2.2, which partially confirms the hypothesis H_1. Among the techniques that use tiling, given an identical type of tiling, the techniques of the *RD* type give better results for task *T*2.1 but not for task *T*2.2. The hypothesis H_2 is thus only partially confirmed. With regard to the type of tiling, the squares always give results that are better than or equal to the hexagons, which invalidates the hypothesis H_3. In general, the technique T_s seems to be the most effective for the tasks of the *T*2 type since it provides the second-best percentage for *T*2.1 and the best percentage for *T*2.2. The hypothesis H_4 is thus invalidated here.

5.2. Time

We used the data from the eye tracker to calculate the time spent by each participant on each of the maps. Figure 9 shows the averages and the confidence intervals for this time for each of the tasks and each technique.

The analysis was carried out after eliminating the trials in which the participants took an abnormally long time. These anomalies were mainly due to questions of understanding the techniques. For task *T*1.1, two trials for *RD* and three trials for *T* were thus eliminated. For task *T*1.2, two trials for *RD* and two trials for *T* were eliminated. For task *T*2.1, one trial for *RD* was eliminated. For task *T*2.2, one trial for *D* was eliminated. This corresponds to what we observed during the sessions: the techniques that use tiling require more time to learn. Since we eliminated these anomalies in the following analysis, the results are more representative of the time required after learning.

For task *T*1.1, the response time is slightly shorter for the maps that use the technique RD_h (average: 15.2) and, to a lesser extent, *Ts* (average: 17.17). The other techniques have an average between 18 and 20. The differences observed are not statistically significant ($p \geq 0.1$ in all cases).[11] As is clear in Figure 10, they are also not

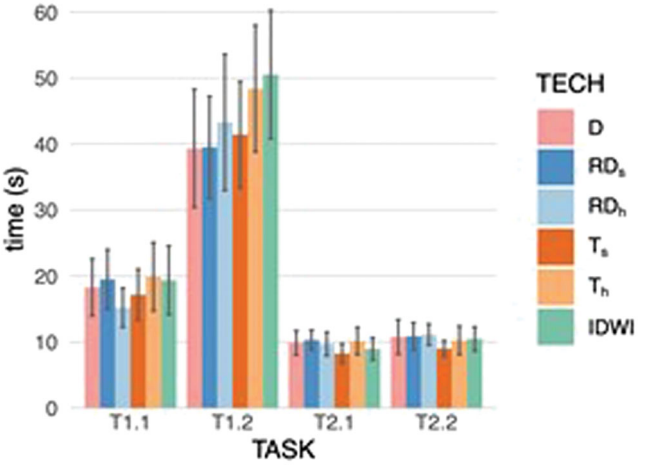


Figure 9. Mean completion time per *TECH* × *TASK*: error bars represent the 95% confidence interval.

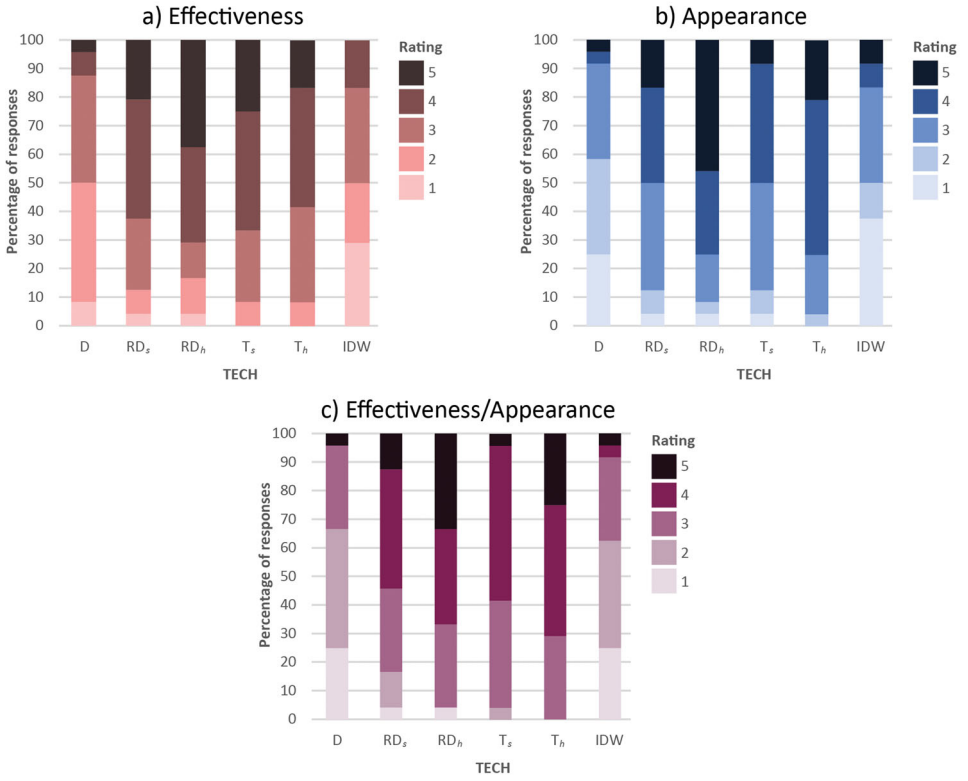


Figure 10. User preferences rated from 1 (very poor) to 5 (very good): (a) percentage of responses for the effectiveness of each technique, (b) percentage of responses for the appearance of each technique, (c) percentage of responses for the effectiveness/appearance ratio of each technique.

significant for the number of layers (1 or 2), the use of tiling (*RD* and *T* with respect to *D* and *IDWI*), and the type of polygon used for the tiling (square or hexagon).

For task *T*1.2, Figure 9 shows that the techniques *D* and RD_s have similar results (average: 39.33 and 39.5). The techniques T_s and RD_h are next (average: 41.42 and 43.29), while the techniques T_h and *IDWI* have the longest times (average: 48.41 and 50.54). Figure 10 also shows that the techniques that use square tiling, RD_s and T_s, allow the tasks to be carried out faster on average than the techniques that use hexagonal tiling, RD_h and T_h. Likewise, the techniques that only use a single layer (*D*, RD_s and RD_h) on average allow the tasks to be carried out faster than the techniques that use two layers (T_s, T_h and *IDWI*). However, like for task *T*1.1, the generalised regression model does not reveal any significant differences ($p \geq 0.08$ in all cases).

For task *T*2.1, Figure 9 shows that the tasks were carried out faster with the technique T_s and, to a lesser extent, with the technique *IDWI*. While the average time is 8.25 and 8.96 seconds for these two techniques, the average times for the other techniques vary between 9.7 and 10.67. This observation is partially confirmed by the generalised regression model, which shows that T_s allows the task *T*2.1 to be caried out significantly faster than the techniques *D* ($p = 0.040$), RD_s ($p = 0.049$) and T_h ($p = 0.035$). There are not, however, any significant differences for the number of layers, the use of tiling, and the type of polygon used for the tiling.

For task *T*2.2, Figure 9 shows that once again, the tasks were carried out faster with the technique T_s (average: 9). The other techniques have similar times, with averages that vary from 10.25 to 11.12. However, this difference between T_s and the other techniques is not confirmed by the generalised regression model ($p \geq 0.07$). As above, the number of layers in the techniques, the use of tiling, and the type of polygon used for the tiling do not have any effect on the response time.

In conclusion, even though the generalised regression model does not show any significant differences in most cases, it appears that T_s is a good compromise in terms of execution time for carrying out the tasks of the *T*1 type. Moreover, this technique allows the tasks of the *T*2 type to be carried out faster. These observations seem surprising given the strong resemblance to the technique T_h. This particularity will be discussed in section 5.4. The hypothesis H_5, which states that the techniques that use a single layer (*D*, RD_s and RD_h) allow the tasks to be carried out faster than those that use two layers (T_s, T_h and *IDWI*), can be validated for task *T*1.2 but not for the others. A substantial difference is also observed between the execution times for the task *T*1.1 and the task *T*1.2. This can be explained by an atypical operation of interpretation (identifying the regions with significant damage to few buildings) of this type of data, which requires a more detailed reading of the map.

5.3. User feedback

At the end of each session, the participants were shown all 6 techniques for the same geographic sector (taken from DS_2). They were asked to rate the techniques according to three criteria: the first was the sensation of effectiveness with regard to the execution of the tasks, the second was appearance, and the third was the perceived ratio between effectiveness and appearance. The ratings ranged from 1 to 5, with 1 being 'very poor' and 5 being 'very good'.

The results show that the techniques RD_h and T_h, followed by RD_s and T_s, are judged to be the most effective (Figure 10(a)). For these techniques, 70%, 67%, 63% and 58% of the ratings are greater than or equal to 4. These results are rather similar to the ratings for the assessment of appearance (Figure 10(b)). RD_h and T_h both have ratings greater than or equal to 4 in 75% of cases, with an average of 4.1 and 3.9, respectively. RD_s and T_s have substantially lower results, with 50% of ratings greater than or equal to 4 (average of 3.50 for RD_s and 3.42 for T_s). This observation allows the hypothesis H_6 to be validated. The techniques *IDW* and *D* obtained the lowest ratings for these two criteria (Figures 10(a,b)). The average ratings for *IDW* and *D* were 2.38 and 2.44, respectively.

Consequently, whether in terms of effectiveness or appearance, the participants seem to prefer the techniques that use tiling (Regular Dot map and Tessellation) over the others (Dot map and IDW Interpolation). This is confirmed by the third criterion – the ratio between effectiveness and appearance (Figure 10(c)). The techniques RD_h and T_h received the best ratings, with one being an average of 3.96/5 (i.e. approximately 70% of participants gave a rating greater than or equal to 4). T_s comes third with an average rating of 3.58, with more than 50% of participants giving it a rating greater than or equal to 4. RD_s had an average rating of 3.46, while *IDWI* and *D* were

Figure 11. Comparison of the average accuracy for all tasks (a) and the perceived effectiveness by participants (b) for each technique.

the techniques with the poorest ratings, with averages of 2.25/5 and 2.17/5, respectively. Over 63% of participants gave them ratings less than or equal to 2/5.

According to these first results, appearance is defined in terms of effectiveness just as much as in terms of subjective preference, which is a reminder of the fact that behind the map, there is a designer and an audience. In the literature, this relationship between the appearance and effectiveness of a map is often described as a corollary. The map must be pleasant to look at and appealing (Cauvin *et al.* 2007) in terms of certain aesthetic properties (Pouivet 2010, Jégou 2013). These visual aspects are important for the effectiveness of the cartographic communication (Kent 2012, Jégou 2016), the exploration of the map, and, *in fine,* the interpretation of its message (Keates 1964). However, the aesthetic quality of maps is not limited to the subjective assessment of those viewing them (Kent 2012). It must be possible to compare it to the exactness of the message that the map transmits. To examine this idea more closely, we compared the average value of the accuracy of the techniques for all tasks (Figure 11(a)) to the perception of effectiveness by the participants (Figure 11(b)). The results obtained show the existence of a correlation between these two variables. For example, RD_h obtained both the best results for the tasks and the best average rating for perceived effectiveness. RD_s seems to be less well perceived in comparison to the results obtained. The perception of effectiveness for T_h is also correlated to the average accuracy percentage. The techniques *D* and *IDWI*, which had worse results, were considered to be the least effective on average. Nevertheless, *D* appears to be undervalued with respect to *IDWI* in comparison to the average number of correct answers.

For each participant, we also noted the technique for which they were the least effective and the most effective. This information was compared to their perception of the effectiveness and appearance of the technique in question. For the poor results, this comparison showed that approximately 35% of the participants overrated the effectiveness of the techniques with respect to the results obtained (Figure 12). Among them, 13% also overrated the appearance, while 22% overrated the effectiveness but assigned a rating for appearance that corresponded to their

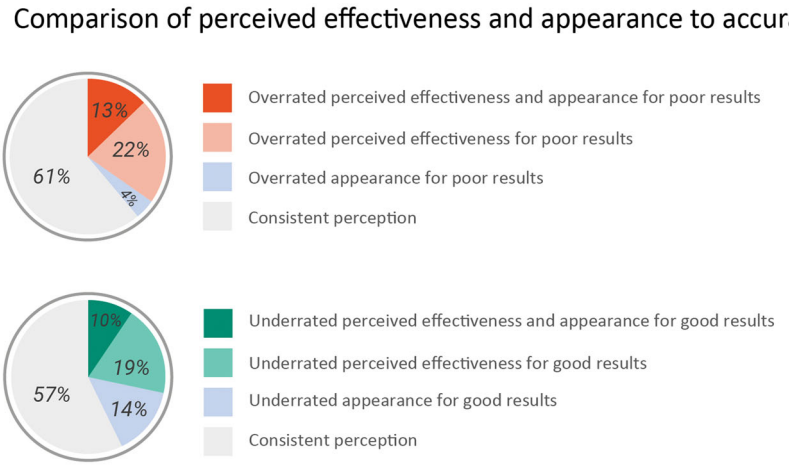


Figure 12. Comparison of perceived effectiveness and appearance to accuracy.

results. Only 4% of participants overrated just the appearance. When the analysis is carried out for the best results, it shows that approximately 30% of the participants underrated the effectiveness of the techniques for which they obtained good results. Among these 30%, 10% also underrated the appearance. In 19% of cases, the techniques were perceived as better in appearance than effectiveness, as opposed to 14% in which participants underrated the appearance but not the effectiveness.

In general, it would appear that there is a correlation between the criteria of appearance and perceived effectiveness for approximately 70% of cases. When the perceived effectiveness is consistent with the results or overrated, appearance is as well. Approximately 60% of participants provided an assessment consistent with the accuracy of their answers. These results thus appear to confirm that for over half of participants, the effectiveness, and thus the appearance, which is correlated thereto, do appear to be influenced by the ability to analyse and interpret the map. A technique with a less pleasing appearance, and which is thus less effective, for the viewer can be linked to a greater difficulty in carrying out the tasks, and vice versa. *In fine*, this comparison corroborates the relationship between the effectiveness and the average percentage of correct answers in Figure 11.

5.4. Discussion

5.4.1. Hypotheses

The analysis of the statistical results of the protocol of our study shows that certain working hypotheses were validated and others not (see Table 2). Our experiment confirmed all the hypotheses related to accuracy (H_{1_4}) for the tasks of the *T*1 type ('Locate'). As a result, the technique RD_h appears to be the most effective in this context. However, the same hypotheses were not validated (with some exceptions) for the tasks of the *T*2 type ('Lookup'). In this case, it is more difficult to determine which

Table 2. Report of the hypotheses validated according to the task.

		*T*1.1	*T*1.2	*T*2.1	*T*2.2
H_1	(accuracy): *RD,T* > *D,IDWI*	✓	✓	×	✓
H_2	(accuracy): *RD* > *T*	✓	✓	✓	×
H_3	(accuracy): *h* > *s*	✓	✓	×	×
H_4	(accuracy): RD_h > other techniques	✓	✓	×	×
H_5	(time): 1 layer > 2 layers	×	✓	×	×

technique is the best. It is noted, however, that RD_s is a good compromise. In general, the techniques of the *RD* type seem to be the best compromises in terms of accuracy if a single visualisation must be used to perform all the types of tasks. With regard to time, the hypothesis H_5 is only validated for the task *T*1.2. The times obtained with the *RD* techniques are comparable or better to those obtained with the other techniques, which supports the choice of these techniques. Finally, the hypothesis H_6, which does not appear in the table, is also validated. From the point of view of effectiveness and appearance, the participants prefer the regular distributions of dots and hexagonal tiling. In most cases, the results show a correlation between the results obtained and the perception of the participants.

5.4.2. Accuracy of T_s for T2

The statistical results for *T*2 (see Figure 8) show a significant difference between the percentages of correct answers obtained by the square tessellation (T_s) and the hexagonal one (T_h). We suspect a bias related to the offset produced by the difference in geometry between the square and hexagonal shapes. This difference can lead to a shift in the results of the aggregation of the damage during the generation of cells in Qgis (3.10).

5.4.3. Further remarks based on the eye-tracker logs

The data from the eye tracker allowed us to extract and analyse the time necessary for the participants to complete the various tasks demanded of them. The eye tracker also provided important data for better understanding the processes of reading a map, thus contributing to the interpretation of the differences in the results for the various representations. A first exploration of the results is performed on the cumulative number of fixations obtained for each pixel of the image. This accumulation is calculated from all the records selected for T1.1 for the selected time interval. This addition of fixation values allows the determination of the highest fixation points in the image. From these fixation points, colour values were added to all pixels in the image according to their distance from these points. This process ensures a smoother colour distribution to generate a relevant heat map that are superimposed on the visual studied. In order to provide some material for discussion on the largest statistical differences observed, we propose analysing the impact of the distribution of the data for *T*1.1 for a territory with dispersed damage (Figure 13) and another with clustered data (Figure 14), for the techniques RD_h and *IDWI*, which obtained the best and the worst scores for accuracy across all tasks, respectively. On the maps of the two Figures 13 and 14, a dotted circle shows the location of the correct answers. The black dots with a white exclamation point in the centre indicate the zones of high concentration outside of the correct answers.

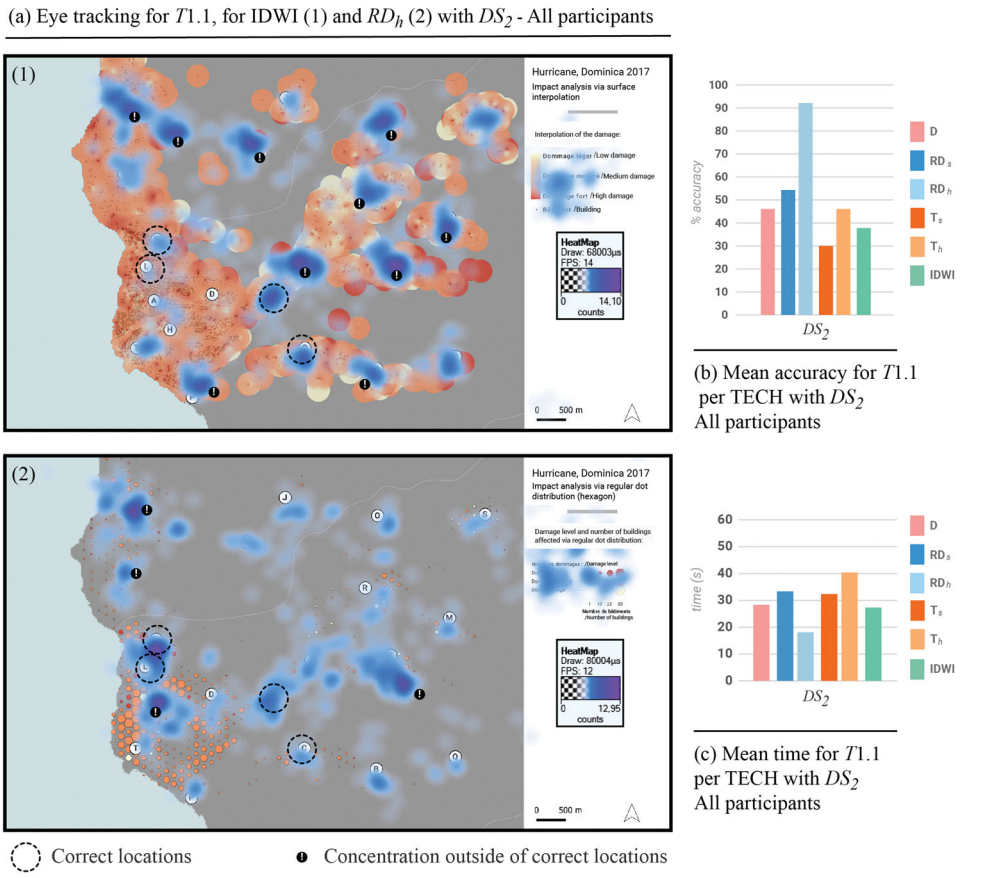


Figure 13. Interpolation vs RD_h for *T*1.1 with DS_2 *for all participants*.

When the data is dispersed (DS_2), the technique *IDWI* is less effective than RD_h. For the former, only two out of four sectors were correctly identified by the participants. On the contrary, the four expected sectors were the object of particular attention by the participants for RD_h. Moreover, there are more zones of concentration visible outside of the correct answers for *IDWI* (10) than for RD_h (4). These observations are correlated to the percentage of correct responses and the times for completion of task *T*1.1. RD_h was both faster (18 s) and more effective (91) on average than *IDWI* (average: 27 s and 37.5). When the data is more clustered (DS_3), the eye-tracker data shows that the two sectors to be identified for *T*1.1 attracted visual attention regardless of the technique (see Figure 14). RD_h is once again more effective in attracting attention onto the zones most affected, with a single zone outside of the correct answers versus five for *IDWI*. Nevertheless, for this dataset, the percentage of correct answers and the completion times were similar for both techniques.

Beyond the percentage of correct answers and the times required for completing the task, it is noted that the technique *IDWI* is capable of causing more visual confusion. This is potentially caused by the calculation method (radius of 200 metres), which creates greater red surfaces at the edges of the interpolated surfaces, in particular

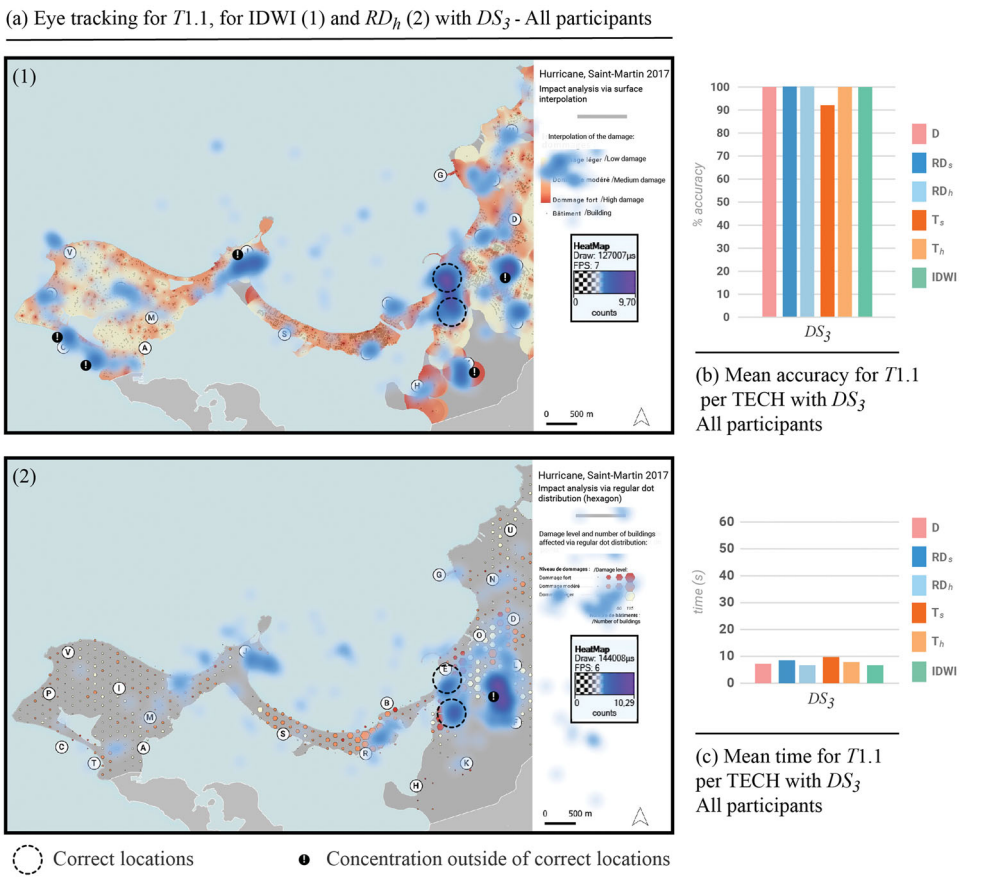


Figure 14. Interpolation vs RD_h for *T*1.1 with DS_3 *for all participants*.

when the latter are isolated or more dispersed. The technique RD_h, which uses the visual variables of colour and size, is more effective in attracting the eye to the qualitative and quantitative information requested.

These preliminary observations are promising. They allow the visual salience of the various cartographic techniques used by the EMS (Emergency Mapping Service) to be interpreted and their effectiveness to be estimated, in particular when the salience is compared to the percentages of correct answers. To take this a step further, it would be possible to analyse the trajectories of the gaze of each participant in order to determine the scanning strategies used to find the information necessary for the completion of a task. This scanning would allow the jumps between the legend and the graphic image to be interpreted. A correlation similar to the above analysis (percentage of correct responses and response time) would reveal, for example, whether certain techniques are predisposed to providing correct answers without systematic viewing of the legend.

5.4.4. Focus on RD_h

Overall, these results seem to confirm the general hypothesis that the main advantage of the methods for aggregating damage on a system of regular cells (*RD*) is to simplify

the spatial processes (Pumain and Saint-Julien 2010) and make the visual identification of information faster. When such a method is used to generate a regular distribution of proportional points, using the visual variables of size and colour, the map delivers a more effective message in comparison to the dot (*D*) and interpolation (*IDWI*) techniques. Moreover, the preliminary results from the eye tracker seem to demonstrate that these visual variables would also be more effective regardless of the geographic distribution of the data. The cognitive and visual performance influence the perception of the participants, which allows us to define an optimal ratio between the actual and perceived effectiveness with the technique RD_h.

6. Conclusion and future work

Satellite rapid mapping services form an integral part of the tools used to assist in emergency and disaster management. These services provide support for risk management via cartographic media dedicated to the evaluation of damaged areas in the first few hours after a disaster. Despite their operational value and their importance for making the right decisions in a post-disaster situation, the variety of the techniques used to represent the damage can lead to errors in interpretation by the user.

In this article, we proposed a new visualisation of post-disaster damage based on Bertin's regular distribution of points (*RD*) (Bertin 1967). In order to test the capabilities of this technique with respect to a sample representative of the techniques usually used by rapid mapping services, we carried out a user study in which a set of participants was asked to carry out various tasks on several datasets using the various visualisations. We then analysed the behaviour during the experiment through three approaches: (1) quantitative analysis of the answers given by the users, (2) quantitative analysis of preferences in terms of perceived effectiveness and appearance, and (3) qualitative analysis via the data collected using an eye tracker. The main results revealed the benefits of the technique *RD* in completing the tasks tested. This technique is effective both in terms of the results obtained and according to the preliminary analysis of the eye-tracker data. These results are promising because the effectiveness of this method of aggregation is clear even though all of the maps tested provide two types of information (quantitative and qualitative). Beyond these measurements of effectiveness, we observed that this technique was well perceived by users. Finally, it provides the best compromise between actual effectiveness and effectiveness perceived by the users.

These results support the use of this technique in the development of a multiscale cartographic visualisation tool. By allowing a dynamic aggregation of the cells, the dimensions and the positions of which vary according to the zoom, *RD* can be used to create a cartographic tool based on a system of standardised scales suitable for damage processing (e.g. Web Mercator Projection, EPSG: 3857,[12] Stefanakis 2017). We were thus able to create a dynamic mapping prototype[13] to visualise the damage caused by Hurricane Irma (September 2017) recorded on the French part of Saint-Martin.

Notes

1. https://disasterscharter.org/web/guest/home (accessed on April 13, 2021)

2. https://www.copernicus.eu/en (accessed on April 13, 2021)
3. In the sense of the observations and rules governing the rational design of graphics. In this context, semiology integrates the natural structure and the properties of visual perception into the construction of a graphics, namely through a set of visual variables (position, colour, value, size, grain, orientation, shape).
4. https://sertit.unistra.fr/en/ (accessed on March 02, 2021)
5. https://www.dlr.de/eoc/en/ (accessed on March 02, 2021)
6. https://colorbrewer2.org/#type=sequential&scheme=OrRd&n=3 (accessed on March 02, 2021)
7. After our observation of decisionmakers in the field, we did not find any examples of the use of this type of visualisation.
8. https://sertit.unistra.fr/ (accessed on May 11,2020)
9. The accuracy distributions measured for each task and for the response time are not considered *a priori* to be Gaussian, and can also be exponential or gamma for the continuous case and binomial or Poisson for the discrete case. Therefore, we use statistical models based on the Generalised Linear Model (GLM) in order to identify the significant differences between the visualisation techniques (*glm* function in *R*). This model is similar to *ANOVA* but takes the various distributions into account. For the tasks of the *T*1 type, we use the Gaussian family (normal distribution) that seems to be the most suitable. We will see below that this is not the case for the tasks of the *T*2 type and the time.
10. Like above, we use a model of the GLM type to identify the significant differences between the visualisation techniques for the tasks of the *T*2 type. However, since the responses of the participants are dichotomous here (either correct or wrong), we use the binomial family.
11. We first used a GLM model with the gamma family (exponential distribution). Then, we introduced a random effect on the participants in order to take into account the information that the same individual carries out each of the tasks (glmer function in R). This allows the correlation between 2 answers from the same individual to be modelled. For each task, we observed a decrease in the AIC between the two approaches (which was not the case for the analysis of accuracy), which allowed us to select the second model. For example, for task T1.1, the AIC decreased from 1019 to 990 when adding the random effect.
12. The Web Mercator coordinate system, also known as Spherical Mercator and Pseudo-Mercator, is a standard used for web mapping (Google, Bing, etc.) and online services (Géoportail, ArcGIS Online).
13. https://worldismap.com/dommages-irma/ (accessed on April 15, 2021)
14. https://emergency.copernicus.eu/mapping/list-of-components/EMSR317 (accessed on May 14, 2020)
15. https://emergency.copernicus.eu/mapping/list-of-components/EMSR246 (accessed on May 14, 2020)
16. https://emergency.copernicus.eu/mapping/list-of-components/EMSR232 (accessed on May 14, 2020)
17. https://unitar.org/sustainable-development-goals/satellite-analysis-and-applied-research (accessed on May 14, 2020)

Acknowledgements

The authors thank the IUT de Montpellier-Sète for providing the eye-tracking device and software used in the experiment. The authors also thank Catherine Trottier, statistician at IMAG - Université de Montpellier and AMIS - Université Paul-Valéry Montpellier 3, for her advice on the statistical analysis of the results, and the Master 2 GCRN class of 2019–2020 at the Université Paul-Valéry Montpellier 3 for their participation in this experiment. Finally, the authors thank RisCrises for providing support and 'guinea pigs' for the calibration of the experiment.

Ethical approval

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institution and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The research has been conducted in an ethical and responsible manner. The protocol was clearly presented to the participants as well as the scientific publication purpose. The participants were voluntary and gave their consent. The tasks performed by the participants were anonymous and no personal data were collected during the study. The research ethics committee of the university Paul-Valéry Montpellier 3 was not yet in place at the time of the experiment. It started functioning after the data collection (September 2020).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data and codes availability statement

The data that support the findings of this study are openly available in figshare - Code_and_data at 10.6084/m9.figshare.14724879.

Funding

This work was supported by the CFP HURRICANES 2017: DISASTER, RISK AND RESILIENCE and the ANR's TIREX project (Sharing learning from post-disaster research for strengthening individual and collective response and adaptation capacities in the context of climate change (Leeward Islands – 2017 hurricane season)), 2018–2022: https://anr.fr/Projet-ANR-18-OURA-0002.

Notes on contributors

Candela Thomas is a PhD at Paul-Valéry Université de Montpellier and LAGAM (Laboratory of Geography and Territorial Planning in Montpellier). He is also Research and Development Manager at RisCrises SARL, Alès, France. His research interests include the development of cartographic and geovisualisation tools for the analysis of vulnerabilities and consequences of natural hazards at the national and international levels.

Péroche Matthieu is an Assistant Professor at the Paul-Valéry Université de Montpellier and researcher at LAGAM (Laboratory of Geography and Territorial Planning in Montpellier). His research focuses on the use of spatial information to help in the management of natural risks and disasters.

Sallaberry Arnaud is an Assistant Professor at the Paul-Valéry Université de Montpellier. He is the head of the applied mathematics and computer science research team (AMIS). He is also a member of the data mining and visualisation research team (ADVANSE) of the Laboratory of Computer Science, Robotics and Microelectronics of Montpellier (LIRMM). His research interests include information visualisation and visual analytics.

Rodriguez Nancy is an Assistant Professor at the Université de Montpellier and researcher at LIRMM (Laboratory of Computer Science, Robotics and Microelectronics of Montpellier) at the ADVANSE team. Her research interests include virtual and augmented/mixed reality, interaction, accessibility and visualisation.

Lavergne Christian is a Professor at Paul-Valéry Université de Montpellier with the applied mathematics and computer science research team (AMIS). He is also a member of the

probability and statistics research team (EPS) at the Alexander Grothendieck Institute of Montpellier (IMAG). He conducts research in statistics and data science.

Leone Frédéric is a Professor at Paul-Valéry Université de Montpellier. He is the head of the Laboratory of Geography and Territorial Planning in Montpellier (LAGAM). His research focuses on the spatial reconstruction of natural disasters and the integrated assessment of associated risks.

References

Bertin, J., 1967. *Sémiologie graphique, Les diagrammes, Les réseaux, Les cartes. Sémiologie graphique, Les diagrammes, Les réseaux, Les cartes*. Paris: Gauthiers-Villars.

Birch, C.P., Oom, S.P., and Beecham, J.A., 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, 206 (3-4), 347–359.

Blondel, J., and Ferris, R., 1995. *Biogéographie: approche écologique et évolutive*. Paris: Masson.

Board, C., and Taylor, R., 1977. Perception and maps: Human factors in map design and interpretation. *Transactions of the Institute of British Geographers*, 2 (1), 19–36.

Brewer, C.A., Hatchard, G.W., and Harrower, M.A., 2003. ColorBrewer in print: a catalog of color schemes for maps. *Cartography and Geographic Information Science*, 30 (1), 5–32.

Carr, D.B., Olsen, A.R., and White, D., 1992. Hexagon mosaic maps for display of univariate and bivariate geographical data. *Cartography and Geographic Information Systems*, 19 (4), 228–236.

Cauvin, C., Escobar, F., and Serradj, A., 2007. *Cartographie thématique 1–Une nouvelle démarche*. Paris: Hermès-Lavoisier.

Cécé, R., *et al.*, 2021. A 30 m scale modeling of extreme gusts during Hurricane Irma (2017) landfall on very small mountainous islands in the Lesser Antilles. *Natural Hazards and Earth System Sciences*, 21 (1), 129–145.

Chauvin, S., 2005. Visualisations heuristiques pour la recherche et l'exploration de données dynamiques: l'art informationnel en tant que révélateur de sens. Doctoral dissertation. Université Paris VIII, Vincennes-Saint Denis.

Chiroiu, L., 2004. Modélisation de dommages consécutifs aux séismes. Extension à d'autres risques naturels. Thesis (PhD). Université Paris-Diderot - Paris VII.

Coyle, D., and Meier, P., 2009. *New technologies in emergencies and conflicts: The role of information and Social Networks*, 54–54. Washington, D.C. and London, UK: UN Foundation-Vodafone Foundation Partnership.

Fairbairn, D., 2006. Measuring map complexity. *The Cartographic Journal*, 43 (3), 224–238.

Flannery, J.J., 1971. The relative effectiveness of some common graduated point symbols in the presentation of quantitative data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 8 (2), 96–109.

Garnero, G., and Godone, D., 2014. Comparisons between different interpolation techniques. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5/W3, 139–144.

Girres, J.-F., *et al.*, 2018. Analysis of tsunami evacuation maps for a consensual symbolization rules proposal. *International Journal of Cartography*, 4 (1), 4–24.

Goldberg, J.H., and Kotval, X.P., 1999. Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24 (6), 631–645.

Griffin, A.L., and Fabrikant, S.I., 2012. More maps, more users, more devices means more cartographic challenges. *Cartographic Journal*, 49 (4), 298–301.

Gros-Desormeaux, J.-R., *et al.*, 2015. Cartographie des tensions spatiales au sein des Znieff à la Martinique: modèle d'accessibilité à la connaissance botaniste.

Herring, C., 1994. *An architecture of cyberspace: Spatialization of the internet*. Champaign, IL: US Army Construction Engineering Research Laboratory.

Hessl, A., *et al.*, 2007. Mapping paleo-fire boundaries from binary point data: comparing interpolation methods. *The Professional Geographer*, 59 (1), 87–104.

Hey, A., and Bill, R., 2014. Placing dots in dot maps. *International Journal of Geographical Information Science*, 28 (12), 2417–2434.

Itten, J., 1967. *Art de la couleur. in 1961: Kunst der Farbe, translated in French in 1967, then 2001*. Dessain and Tolra, 155 p.

Itti, L., and Koch, C., 2001. Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging*, 10 (1), 161–169.

Jégou, L., 2013. Vers une nouvelle prise en compte de l'esthétique dans la composition de la carte thématique: propositions de méthodes et d'outils. These de doctorat en Géographie. Université Toulouse le Mirail - Toulouse II.

Jégou, L., 2016. L'imagination esthétique dans la conception graphique des cartes: proposition de typologie illustrée, 22.

Jégou, L., and Deblonde, J.-P., 2012. Vers une visualisation de la complexité de l'image cartographique. *Cybergeo: European Journal of Geography*. Article 600.

Keates, J., 1964. Cartographic communication. Presented at the 20th International Geographical Congress, 20–28.

Kent, A.J., 2012. From a dry statement of facts to a thing of beauty: understanding aesthetics in the mapping and counter-mapping of place. *Cartographic Perspectives*, (73), 37–60.

Kerle, N., and Hoffman, R.R., 2013. Collaborative damage mapping for emergency response: the role of Cognitive Systems Engineering. *Natural Hazards and Earth System Sciences*, 13 (1), 97–113.

Kienberger, S., Lang, S., and Zeil, P., 2009. Spatial vulnerability units–expert-based spatial modelling of socio-economic vulnerability in the Salzach catchment. *Natural Hazards and Earth System Sciences*, 9 (3), 767–778.

Konecny, M., and Bandrova, T., 2006. Proposal for a standard in cartographic visualization of natural risks and disasters. *International Journal of Urban Sciences*, 10 (2), 130–139.

Kraak, M.-J., and Fabrikant, S.I., 2017. Of maps, cartography and the geography of the International Cartographic Association. *International Journal of Cartography*, 3 (Sup1), 9–31.

Krebs, C.J., 1989. *Ecological methodology*. Harper & Row New York.

Kuveždić Divjak, A., and Lapaine, M., 2018. Crisis Maps—Observed Shortcomings and Recommendations for Improvement. *ISPRS International Journal of Geo-Information*, 7 (11), 436.

Lajoie, G., 1992. *Le carroyage des informations urbaines: une nouvelle forme de banque de données sur l'environnement du Grand Rouen*. Publication Univ Rouen Havre.

Lobo, M.-J., Pietriga, E., and Appert, C., 2015. An evaluation of interactive map comparison techniques. Presented at the Proceedings of the 33rd annual ACM conference on human factors in computing systems, 3573–3582.

Mitas, L., and Mitasova, H., 1999. Spatial interpolation. *Geographical Information Systems: principles, Techniques, Management and Applications*, 1 (2), 18.

Moreau, G., and Locht, J.-L., 2016. Importance de l'utilisation du Système d'Information Géographique pour les vastes sites de plein air du Paléolithique moyen en France septentrionale. L'Exemple de Caours (Somme, France) et Beauvais (Oise, France).

Munzner, T., 2014. *Visualization analysis and design*. CRC press.

Popelka, S., *et al.*, 2012. Advanced map optimalization based on eye-tracking. *Cartography–A Tool for Spatial Analysis*, 99–118.

Popelka, S., Vondrakova, A., and Hujnakova, P., 2019. Eye-tracking Evaluation of Weather Web Maps. *ISPRS International Journal of Geo-Information*, 8 (6), 256.

Pouivet, R., 2010. L'ontologie de l'oeuvre d'art. Vrin.

Pumain, D., and Saint-Julien, T., 2010. *Analyse spatiale: Les localisations*. Armand Colin.

Rey, T., *et al.*, 2019. Coastal processes and influence on damage to urban structures during Hurricane Irma (St-Martin & St-Barthélemy, French West Indies). *Journal of Marine Science and Engineering*, 7 (7), 215.

Roche, S., Propeck-Zimmermann, E., and Mericskay, B., 2013. GeoWeb and crisis management: Issues and perspectives of volunteered geographic information. *GeoJournal*, 78 (1), 21–40.

Sève, R., 2009. *Science de la couleur: Aspects physiques et perceptifs*. Chalagam.

Stachoň, Z., *et al.*, 2016. Cartographic principles for standardized cartographic visualization for crisis management community. Presented at the Proceedings, 6th international conference on cartography and GIS, 781–788.

Staněk, K., *et al.*, 2010. Selected issues of cartographic communication optimization for emergency centers. *International Journal of Digital Earth*, 3 (4), 316–339.

Stefanakis, E., 2017. Web Mercator and raster tile maps: two cornerstones of online map service providers. *GEOMATICA*, 71 (2), 100–109.

Stevens, S., 1975. *Psychophysics: Introduction to its perceptual, neural, and social prospects*. 2000 ed. New York: John Willey and Sons.

Ware, C., 2019. *Information visualization: perception for design*. Morgan Kaufmann.