The SISFRANCE database of historical seismicity. State of the art and perspectives

Herve Jomard, Oona Scotti, Samuel Auclair, Pascal Dominique, Kevin Manchuel, Déborah Sicilia

INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus

Géoscience

Sciences de la Planète

Hervé Jomard, Oona Scotti, Samuel Auclair, Pascal Dominique, Kévin Manchuel and Déborah Sicilia

The SISFRANCE database of historical seismicity. State of the art and perspectives

Online first, 8th December 2021

<https://doi.org/10.5802/crgeos.91>

Part of the Special Issue: Seismicity in France

Guest editors: Carole Petit (Université de Nice & CNRS, France), Stéphane Mazzotti (Univ. Montpellier & CNRS, France) and Frédéric Masson (Université de Strasbourg & CNRS, France)

CENTRE MERSENNE

Les Comptes Rendus. Géoscience — Sciences de la Planète *sont membres du Centre Mersenne pour l'édition scientifique ouverte*
www.centre-mersenne.org

Comptes Rendus
Géoscience — Sciences de la Planète
Online first, 8th December 2021
https://doi.org/10.5802/crgeos.91

Seismicity in France / *Sismicité en France*

The SISFRANCE database of historical seismicity. State of the art and perspectives

Hervé Jomard [a, d]*, **Oona Scotti** [a, d], **Samuel Auclair** [b, d], **Pascal Dominique** [b, d], **Kévin Manchuel** [c, d] and **Déborah Sicilia** [c, d]

[a] Institut de Radioprotection et Sûreté Nucléaire, BERSSIN - 92262 Fontenay-aux-Roses, France

[b] BRGM, Risk and Prevention Division, 3, avenue Claude-Guillemin, BP 36009, 45060, Orléans Cedex 2, France

[c] EDF-DIPNN-CEIDRE-TEGG, 905 avenue du Camp de Menthe, 13097, Aix en Provence Cedex 02, France

[d] GT SISFRANCE (SISFRANCE working group)

E-mails: herve.jomard@irsn.fr (H. Jomard), oona.scotti@irsn.fr (O. Scotti), s.auclair@brgm.fr (S. Auclair), p.dominique@brgm.fr (P. Dominique), kevin.manchuel@edf.fr (K. Manchuel), deborah.sicilia@edf.fr (D. Sicilia)

Abstract. The SISFRANCE database aims to collect and interpret archives relating the historical seismicity in metropolitan France. The database presently contains some 12,500 documents that allow to define 108,832 macroseismic data points, for a total of 6427 events, of which 5743 can be qualified as real earthquakes. Although half of these events are and will probably remain poorly known, completing our knowledge of historical earthquakes remains a fundamental element for the definition of seismic hazard and the understanding of seismicity in France and low to moderate seismicity regions in general. The first aim of this paper is to explain the state of the art of the database and remind end users how the limits introduced by the necessary interpretation of archival data are translated into quality factors that should be carefully considered when using such data for seismotectonic and seismic hazard purposes.

Originally built as part of the necessary datasets to define the seismic hazard for nuclear facilities, the SISFRANCE database is the result of 40 years of work within a consortium bringing together the French civil nuclear operator (EDF), the French geological survey (BRGM), and the French institute of nuclear safety (IRSN). SISFRANCE is also the heir to archival research carried out since the second half of the 19th century, thus constituting the richest collection of data related to historical seismicity in France. The second aim of this paper is to explain to end users the current state of archival research within the SISFRANCE consortium, present avenues for future research strategies, and list potential improvements of the database structure.

Archival research within the SISFRANCE consortium, no longer carried out by a dedicated historian since 2018, is presently being pursued by engaging individual researchers or academics, focusing research on specific archives and regions, as well as by developing data-mining techniques to exploit the increasing wealth of accessible numerical archives. Given the new impetus of French governmental research agencies that encourage transdisciplinarity research, SISFRANCE is also pursuing the idea

* Corresponding author.

of further engaging the academic community of historians and seismologists, presently greatly underrepresented in the field of historical seismicity in spite of the importance of this data, fundamental for any seismotectonic and seismic hazard study in France.

Keywords. Historical seismicity, Macroseismic database, SISFRANCE, Seismic hazard assessment, Earthquakes, France.

Online first, 8th December 2021

1. Introduction

SISFRANCE (BRGM, EDF, IRSN) is the French macroseismic database of historical seismicity, a collection of written archives containing mentions of earthquakes that occurred in metropolitan France and surrounding regions and that have been interpreted into values of macroseismic intensity following the MSK-64 intensity scale [Medvedev et al., 1967]. The database incorporates contemporaneous macroseismic data from BCSF (Bureau Central Sismologique Français), between 1921 (birth of BCSF) and 2007, and from BRGM for the period from 1978 to 1987 [see Sira et al., 2021, Roger, 2021].

The aim and origins of the SISFRANCE research program are well described by two contributions in a compendium dedicated to the French historical seismologist Jean Vogt, who passed away in 2005, (https://emidius.mi.ingv.it/vogt/):

Dieter Mayer-Rosa from the Swiss seismological survey wrote "*He devoted his scientific life to the investigation of the past, well aware that this is the main key to understanding the present*". If he could also have added "to prepare for the future", such sentence, valid for many fields of interest, is obviously true when it comes to estimating seismic hazard [Mayer-Rosa and Schwarz-Zanetti, 2004]. Indeed, the French electronuclear program, at its peak in the 1970's, imposed a better knowledge of the seismicity of the metropolitan territory [Levret and Mohammadioun, 1984], which gave a new impetus to historical seismicity studies [Levret, 1991]. Hence, in the framework of the seismotectonic map of France [Vogt, 1981] started in 1975, a major part of the project was devoted to the reassessment of the historical seismicity [Vogt, 1979, see also Roger, 2021 for more details], which gave rise to the French computerized macroseismic data base "SIRENE" [Godefroy et al., 1980, Thirion, 1983, Godefroy and Levret, 1992], later known as SISFRANCE.

Agnès Levret from the French institute for nuclear Safety (IRSN) and as President of the French group dedicated to archeoseismology (APS group), wrote "*With him disappears the person who initiated the updating of the historical earthquake investigation in France; in some way, the father of modern historical seismology. He started the way back to the investigation of original sources, coeval to the event; moreover, he put this investigation in the historical, economical and social context of the event, for a better understanding of what happened and getting close, as far as possible, to the reality. He found out date and place mistakes, fake quakes and the very many confusions in the current catalogs. Some earthquakes disappeared, some new ones appeared*". The challenge in the 1970s was indeed to move from the Rothé's catalog toward a truly "historical catalog" where each event was critically assessed after a systematic historical approach. For the events preceding the 19th century, Rothé's catalog indeed represented the state of the art at the time [Roger, 2021], however mainly founded on the numerous publications and interpretations from Perrey [e.g. Perrey, 1845]. Subsequently, Vogt [1979] and then Lambert [Lambert et al., 1996] patiently revisited these catalogs during their career at BRGM and afterward, hence developing and enriching the SISFRANCE database.

It is important to remind that, even today, any change in the interpretation of historical earthquakes could have a direct impact on the definition of the seismic hazard of nuclear installations in France, where the deterministic nature of the employed method is still in force [RFS 2001-01 guidelines]. The method relies on the definition of the "Maximum Historically Probable Earthquake" selected from macroseismic or instrumental databases and considered to be the most penalizing earthquakes liable to occur [Levret and Mohammadioun, 1984, Scotti et al., 2014]. In practice, the so-called "reference earthquakes" selected to define the ground accelerations to be taken into account for the design or the verification of the sizing of nuclear installations are largely dating prior to the 20th century. In addition, not only the strongest events considered

in deterministic seismic hazard assessment (DSHA), but the overall database has an increasing role to play since probabilistic hazard assessment (PSHA) is more and more applied to calculate seismic hazard in France. This arises from the fact that earthquake catalogs, derived from both instrumental and historical datasets [Manchuel et al., 2018], contribute to the definition of earthquake rates of occurrence at the heart of the PSHA calculation. SISFRANCE has therefore continued this work of collecting and interpreting archives relating the effect of earthquakes for 40 years.

This paper aims to first (i) propose an updated inventory of the SISFRANCE database since the last synthesis proposed in 2004 by Scotti et al. [2004] and (ii) enlighten end users on the way historical seismicity data should be interpreted by presenting how such data were gathered and what their limits are.

In a second part, we will take stock of the state of archival research in France for historical seismicity in the light of our experience acquired over the years. Then we will discuss the future of SISFRANCE by addressing some possible adjustments of the database and archival research strategies in a context where, for the first time since the 1970s, there is no longer a researcher specifically dedicated to this task.

2. The database: architecture and statistics

In this section, we provide an overview of the SISFRANCE database as released in 2017 at its last update. The general structure of SISFRANCE is described and key information and statistics regarding the data composing the database as well as their related uncertainties will be discussed. Some aspects regarding the history, development, and content of the database are described in Scotti et al. [2004]. The reader will then be referred to this publication when necessary for complementary information.

2.1. *The SISFRANCE database and associated website*

The design of the SISFRANCE database, originally named SIRENE (for "système informatique de rassemblement des événements naturels existants", i.e., computerized system for collecting existing natural events), dates back to 1979. It was initially composed of a sequence of indexes related by common keys, exploitable through a specific data extraction program [Thirion, 1983]. In 1986, all these indexes were gathered in a single relational database designed with the ORACLE software, compatible with personal computers [Levret, 1991]. The architecture and content of the database is described in Godefroy and Levret [1992]. In 2000, when it came to making the database public, which was not the case before, SIRENE was renamed SISFRANCE. The structure of the database has been lightened and now contains nine individual tables (out of 11 originally) linked together through relational keys (Figure 1).

SISFRANCE is designed to be a self-supporting database, avoiding data coming from external databases. Thus, in addition to the documentary tables collecting the written archives (BIBLIO and DOCUMENTS) and to the interpretative tables of SISFRANCE (EVTSIRENE, OBSIRENE, EPCSIRENE, ISOSEISMAL MAPS), geographical tables are included, serving as inputs to report the location of the individual macroseismic data points and seismic events (LOCALITIES, COUNTRIES). In addition to these tables, an ARCHIVE table acts as a memory of what has been modified from a SISFRANCE release version to another.

The principal tables (in blue, Figure 1) composing the SISFRANCE database are hereafter shortly described:

- The event table EVTSIRENE is at the heart of the database, it brings together all the true, false, and doubtful earthquakes interpreted in SISFRANCE and contains information mainly related to the nature and date of occurrence of the events. Both false and doubtful earthquakes are kept within the database in order not to lose information and hence duplicate future archival researches. A unique identity number is assigned to each event of the table (Numevt), allowing to link the EVTSIRENE table with the other main tables of the database (Figure 1). A true event only exists when at least one individual observation (IDP or Felt) is contained in the OBSIRENE table. Note that an IDP is defined as a quantified Intensity Data Point. A "felt" is defined as unquantified intensity data point and the addition of the number of "felt" and IDP provides the total number of Macroseismic Data Point (MDP);

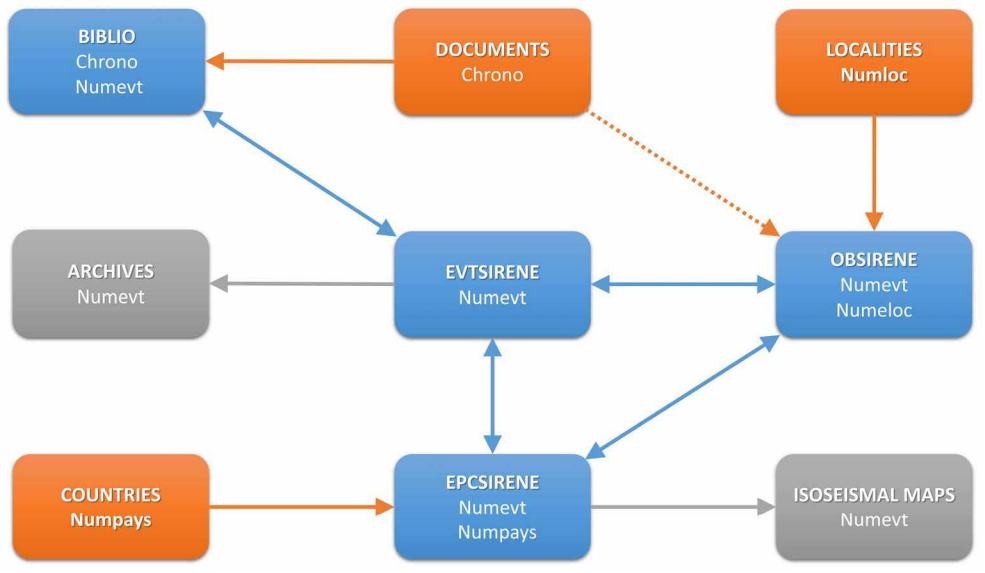

Figure 1. General structure of SISFRANCE and relationships between individual tables. One may individualize basic data (databank, in orange) from the main relational tables of the database (in blue) and the output tables in gray. A dashed arrow indicates that data are used to feed a table without any relational index. Updated from Scotti et al. [2004].

- The OBSIRENE table contains the individual observations (MDP) related to true, false, and doubtful earthquakes. When the historical information allows quantifying an intensity level, the IOBS field is completed, otherwise it is left empty (i.e., "felt"). Each single MDP, quantified (IDP) or not (felt), comes together with a location corresponding to actual municipalities. When it can be defined, the intensity is quantified according to the MSK-64 macroseismic scale;
- The EVTSIRENE and OBSIRENE tables are implemented from the interpretation of the archives, which are referenced in the BIBLIO table and indexed in the DOCUMENT table;
- The epicenter table EPCSIRENE provides the geographic component (location) of each true event, as well as an estimation of their epicentral intensity (IEPC or Io) following the MSK-64 scale, when attributable (see Section 2.2 for further information related to the definition of epicentral parameters).

Even though the concept of intensity is an integer classification [Musson et al., 2010], half-intensity values are sometimes reported in SISFRANCE for both individual and epicentral intensities. These half degree values indicate end users that either the upper or the lower value can be used (e.g., VI or VII for VI–VII), given that the historical information is sometimes not precise enough.

At each level of interpretation of the archives, a quality index is produced in order to provide a state of uncertainty related to each parameter finally implemented in the database. Table 1 lists the fields of each SISFRANCE table that correspond to a quantified uncertainty. The uncertainty relative to epicentral information, QPOS, and QIE, for example, are presented in Table EPCSIRENE and can vary from A to K underlining very different states of knowledge. This is further explained in Scotti et al. [2004]. In addition to these uncertainties, specific information relative to the nature of the seismic events are also provided and summarized in Table 1. For instance, the RELATION field in the EVTSIRENE table allows distinguishing foreshocks and aftershocks activity from mainshocks, but also to pinpoint events belonging to an earthquake swarm.

The SISFRANCE database was made public at the beginning of the 2000's with the launch of the website (http://www.sisfrance.net). Most of the useful information attached to the seismic events reported in the database are provided, concerning the location and

Table 1. Descriptions of the uncertainties (in bold) and additional information (in italic) implemented in the different tables of SISFRANCE

Table	Field	Meaning	Code (meaning)
EVTSIRENE	**CODDATE**	**Qualify the date**	**"LE" (date is complete), "EN" (date is incomplete), "OU" (uncertainty between 2 dates), "ENTRE" (duration between two dates)**
	QD	**Date uncertainty**	**"A" (certain), "B" (pretty sure), "C" (uncertain)**
	CODHEURE	**Qualify the hour**	**"A" (very precise hour), "VERS" (pretty precise hour), "ENTRE" (interval between two hours), "DEBUT" (starting at)**
	QH	**Hour uncertainty**	**"A" (certain), "B" (pretty sure), "C" (uncertain)**
	NATURE	*Nature of event*	*"VS" (true earthquake), "FS" (False earthquake), "SD" (doubtful earthquake)*
	RELATION	*Relates an event to others*	*Empty for mainshock, "R" (aftershock), "P" (foreshock), "E" (individualized event of a swarm), "Z" (swarm)*
	SECOUSSES	*Number of shocks*	*Exact number when known, "QQ" (some), "BCP" (numerous)*
EPCSIRENE	**QPOS**	**Location uncertainty**	**"A" (few kilometers), "B" (~10 km), "C" (from 10 to 20 km), "D" (up to 50 km), "E" (more than 50 km), I (Isolated, located at the only observation point available)**
	QIE	**Epicentral intensity uncertainty**	**"A" (reliable, MDP close to each other, Iobs are certain), "B" (pretty reliable, MDP less close, Iobs certain), "C" (Uncertain, MDP sparse and Iobs uncertain), "K" (pretty reliable and calculated from an attenuation law), "E" (arbitrary, when MDP are sparse and distant from each other), "I" (deduced from a unique MDP)**
OBSIRENE	**QOBS**	**Intensity**	**"A" (certain), "B" (pretty sure), "C" (uncertain)**
	EFFET	*Environmental effects*	*"MT" (gravitational movements), "RZ" (tsunami), "EE" (hydrogeological effect), "PL" (lighting effect), "ES" (site effect), "EH" (topographic effect)*
ISOSEISTE	**Q**	**Quality of isoseismal parameters**	**"A" (certain), "B" (pretty sure), "C" (uncertain)**
DOCUMENT	**QTXT**	**Quality of archive sources**	**"A" (certain), "B" (pretty sure), "C" (uncertain)**
	COMMENTAIRE	*Comment on archive sources*	*Short textual description*
LOCALITE	*NOMLOC*	*Former name of locality, district, hamlet*	*Former name of a locality that may change through time*

intensity parameters of individual observations and earthquakes, as well as their associated uncertainties. Likewise, the available bibliographic information is provided. However, the website suffers from some limitations:

- Solely written archives that are free of rights can be displayed;
- Epicentral locations of the most poorly known events, for which uncertainties are the greatest, are not provided. Note that this limitation will be lifted in 2022 with the release of all new website design.

However, information's that are not available on the website can be perused digitally with SISFRANCE partners and physically at BRGM. In addition, the overall database can be shared upon request to the SISFRANCE consortium and the establishment of a collaboration agreement. Finally, SISFRANCE data are also used and disseminated at the European level through the AHEAD portal of the European Platform Observing System EPOS (https://www.emidius.eu/AHEAD/) for the period up to year 1900.

2.2. *The latest release of the SISFRANCE database*

The latest version of the SISFRANCE database was released in 2017. We first present in Figure 2a, a seismicity map centered on metropolitan France as seen from written archives, from year 463 (first "true event" reported in the database), to year 2007 when BCSF data were no longer included in the database. Note that the database contains some remote earthquakes occurring outside of the map perimeter. In total, 5743 earthquakes are contained within the EPCSIRENE table. A timeline representing the extent of our knowledge while going back in time is provided in Figure 2b. For better readability of the data, the epicenters have been displayed in Figure 2 using a simplified scale based on the MSK-64 macroseismic scale:

- Weakly and/or locally felt events (II ≤ Io ≤ III–IV) correspond to events felt by few people under favorable conditions. 566 events are reported, the majority within the most recent periods, which could mainly be explained by the fact that there are very little chances to find archives relating to weak events (i.e., without consequences) when going back in time;
- Widely and/or strongly felt events (IV ≤ Io ≤ V–VI) correspond to events felt by many people over wider areas and where non-structural damages may occur in some buildings. 1960 events are reported over a longer period of time. This intensity category is the most represented in the database. Indeed, these events are more recurrent than damaging events and important enough to be noticed by the population and reported within the archives;
- Damaging events (VI ≤ Io ≤ VII) correspond to frightening events, inducing structural damages to vulnerable buildings in the epicentral area, in larger and larger areas as the shaking increases. Being rarer than the above-mentioned categories, only 516 events are reported over the entire history,
- Strongly damaging events (Io ≥ VII–VIII). Heavy structural damages and first collapses of vulnerable buildings are starting to be observed in the epicentral area and in larger areas as the shaking increases. Solely 89 strongly damaging events are known throughout the history, especially because of their long return periods;
- Events with undefined epicentral intensity (Io is unknown) correspond to events for which an earthquake is reported in the archives without referring to their detailed effects on people and buildings. 2612 of these are reported in the database, which is by far the largest category.

It is worth mentioning that this representation (Figure 2) does not account for uncertainties reported in Tables 1 and 2. However, it allows to raise first order qualitative observations, especially on the spatial repartition of the historical seismicity. As previously noticed in Mazzotti et al. [2020], the general pattern of historical seismicity is consistent with the instrumental one in SI-Hex [Cara et al., 2015], with the exception of local discrepancies such as in Provence and in the Pas-de-Calais, where significant historical events occurred, and where the instrumental seismicity is on the contrary low [see also regional synthesis contained in this issue by Beucler et al., 2021, Doubre et al., 2021, Larroque et al., 2021, Sylvander et al., 2021]. In addition, while damaging earthquakes occurred in almost all regions

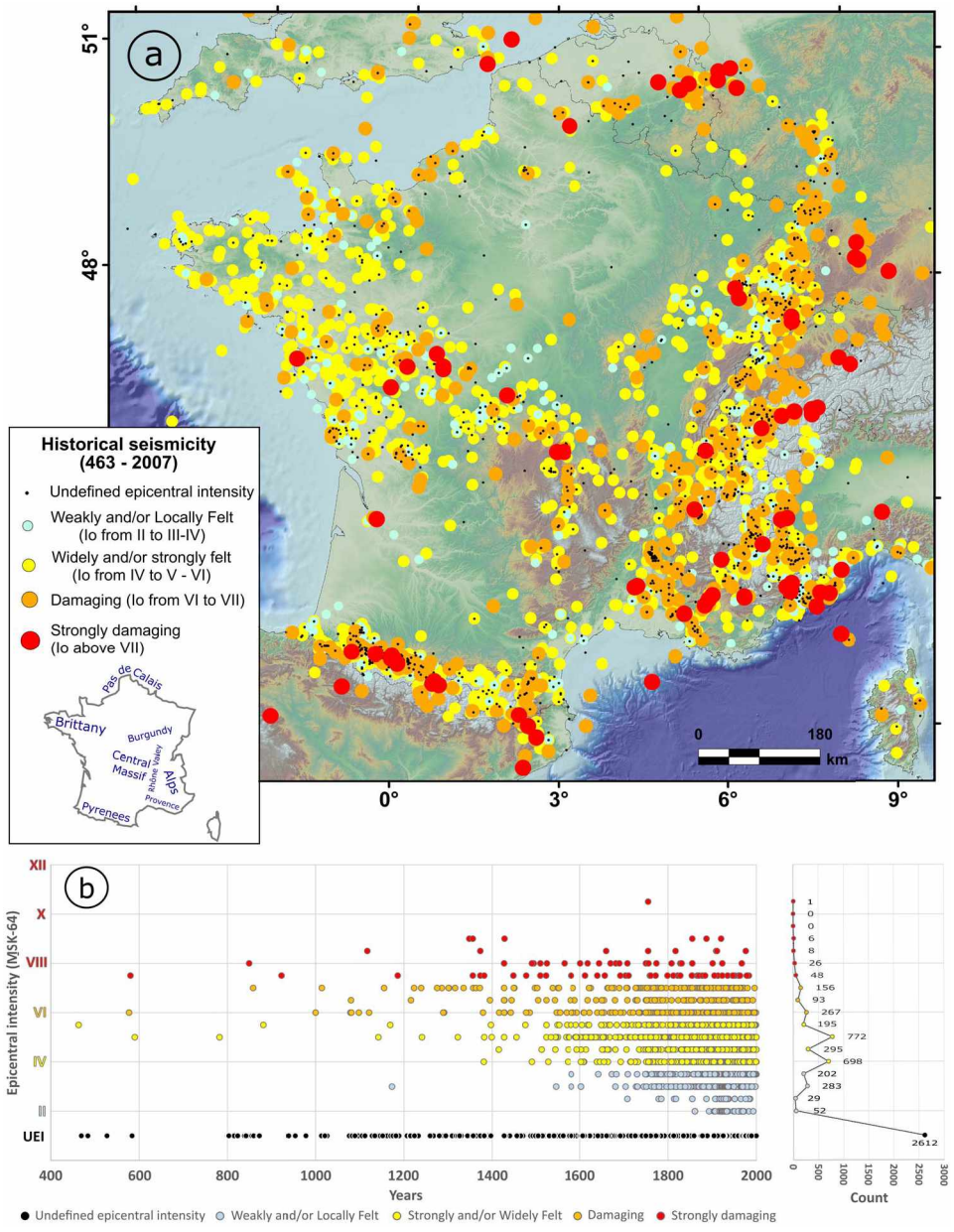

Figure 2. Known historical seismicity from macroseismic data. (a) Seismicity map centered on metropolitan France (SISFRANCE database, 2017 release). Earthquakes are plotted according to five distinct epicentral intensity classes. Regions mentioned throughout the text are reported in the caption; (b) temporal plot of the seismicity, including the 5743 events of the EPCSIRENE table, from year 463 to year 2007. The number of events for each intensity level is presented on the right. Undefined epicentral intensities (UEI) are arbitrarily plotted along the Io = I axis. Note that the event showing an epicentral intensity value of X–XI is the 1755 Lisbon earthquake, not reported in (a).

Table 2. Number of events, macroseismic data points (MDP), and documents contained in the principal tables of the SISFRANCE database (2014 and 2017 releases), detailed according to their associated uncertainties (2017 release)

Table	SISFRANCE release		Quality factor							
	2004	2017								
EVTSIRENE			VS (True event)–SD (Doubtful event)–FS (False event)							
Number of events	**5960**	**6427**	**5743**				**487**			**197**
			A	B	C	D	E	I	K	N/A
EVTSIRENE			Related to intensity (QIE—bold) Related to position (QPOS—italic)							
Number of epicenter	**5351**	**5743**	—	—	—	—	—	**1465**	—	**2027**
			9	*12*	*5*	*710*	*235*	*2521*	—	—
→ with epicentral intensities	**2792**	**3131**	**209**	**976**	**831**	—	**224**	**880**	**11**	—
			209	*342*	*268*	*1208*	*224*	*880*	—	—
→ with Io ≥ VI	**576**	**605**	**69**	**185**	**198**	—	**102**	**41**	**10**	—
			67	*75*	*83*	*239*	*100*	*41*	—	—
OBSIRENE			Related to intensity (QIOBS)							
Number of MDP (false and doubtful events)	**1113**	**1240**	692	215	333	—	—	—	—	—
Number of MDP (True events)	**89448**	**107592**	36885	65141	5566	—	—	—	—	—
→ with quantified Iobs	**70008**	**86346**	17167	64409	4770	—	—	—	—	—
→ with Iobs ≥ VI	**3467**	**3684**	1191	1717	776	—	—	—	—	—
DOCUMENTS			Related to intensity (QTXT)							
Number of archives	**9765**	**12532**	1762	3849	75	—	—	—	—	5771

highlighting seismicity, this is not the case, for example, in northern Brittany, southern central Massif, northern Rhône Valley, and Burgundy that have not suffered any known strongly damaging events in the history. Finally, a significant part of these strongly damaging events occurred near the boundaries of neighboring countries, highlighting the need for cross-boundaries exchanges and studies.

2.2.1. *General statistics*

A synthesis of the principal data and associated uncertainties contained in the database is presented in Table 2. Among these data, one should essentially notice:

- 6427 events are contained in the EVTSIRENE table, 5743 of them are qualified as "true events". Others are considered either doubtful or false events;
- among the 5743 true events reported in the EPCSIRENE table, only 3131 of them come along with a quantified epicentral intensity "Io" (54.5%), of which 880 are known through a single IDP (Qpos = I "Isolated"). This means that solely 2251 events (39.2%) of the database have been characterized with more than 1 IDP;
- 605 true events highlight epicentral intensities equal or higher than VI MSK (i.e., events having potentially produced structural disorder on buildings);
- 12532 documents (written archives) allowed determining a total of 108832 MDP. 107592 of them are associated with true events and 86346 have quantified intensities (IDP);
- In addition to the information contained in Table 2, one may note that a total of 1235 earthquakes are qualified as foreshocks (256 events) or aftershocks (979 events).

As a consequence, one should be aware that at least half the database is composed of events that could be qualified as poorly characterized. On the other hand, taking into account an uncertainty of A and B (see Table 1 for details) onto the estimation of the epicentral intensity, around 1100 events may be

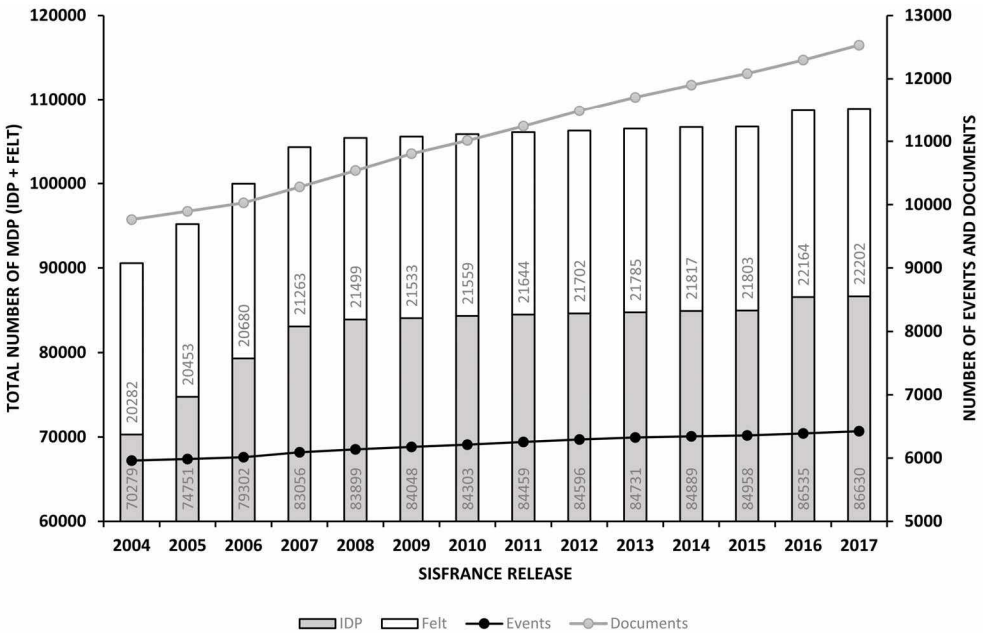

Figure 3. Evolution of the number of MDP (IDP + felt), number of events, and number of documents contained in SISFRANCE starting from the 2004 release until the 2017 release.

qualified as properly known.

The evolution of the database since 2004 [Scotti et al., 2004] is presented in Figure 3 (and Table 2). While the number of new documents (~213) and new events (~36) discovered per year is relatively constant over the time period, it is not the case concerning the number of MDP (IDP and Felt). From 2004 to 2007, a mean number of 4586 MDP/yr were added to the database, and 451 MDP/yr afterward and until 2017. This difference could be explained by the end of the integration within SISFRANCE of the BCSF surveys in 2007. Since then, the rate of new MDP/yr is that corresponding to historical events strictly speaking and remains globally stable over time. This provides a relatively clear idea of what we may expect to find annually with the work of an historian dedicated to this project (note that the person in charge at BRGM was dedicated 30% of his time working on this task).

2.2.2. *Focus on epicentral parameters*

An important point when it comes to feed a database is that the methodology should be as homogeneous and reproducible as possible and when subjected to expert opinion, that such interpretation can be traced in order to exploit them in a seismotectonic framework. Indeed, as already pointed out by Ambraseys [1983], a seismotectonic approach to hazard assessment requires good quality epicentral locations and precise epicentral intensity determinations. The methods used to determine epicentral parameters and their use to define seismogenic sources (zones or faults) are a matter of debate [Richter, 1958, Cecic et al., 1996]. Cecic et al. [1996] present a review of these issues across Europe.

In SISFRANCE, the epicentral location of earthquakes is determined prior to their epicentral intensity which depends on it. Few simple rules are in general followed to establish the location of events:

- When a single MDP is available, the epicenter of the event corresponds to the location of this data point, hence qualified as isolated (QPOS = I; 2521 events, Figure 4). The epicentral intensity is that of the single available MDP, often unknown (felt—1641 events), but it happens that events having only one MDP have a quantified Io (i.e., 880 events, 41 of them having an Io ≥ VI). This is the simplest case, corresponding to a significant part of

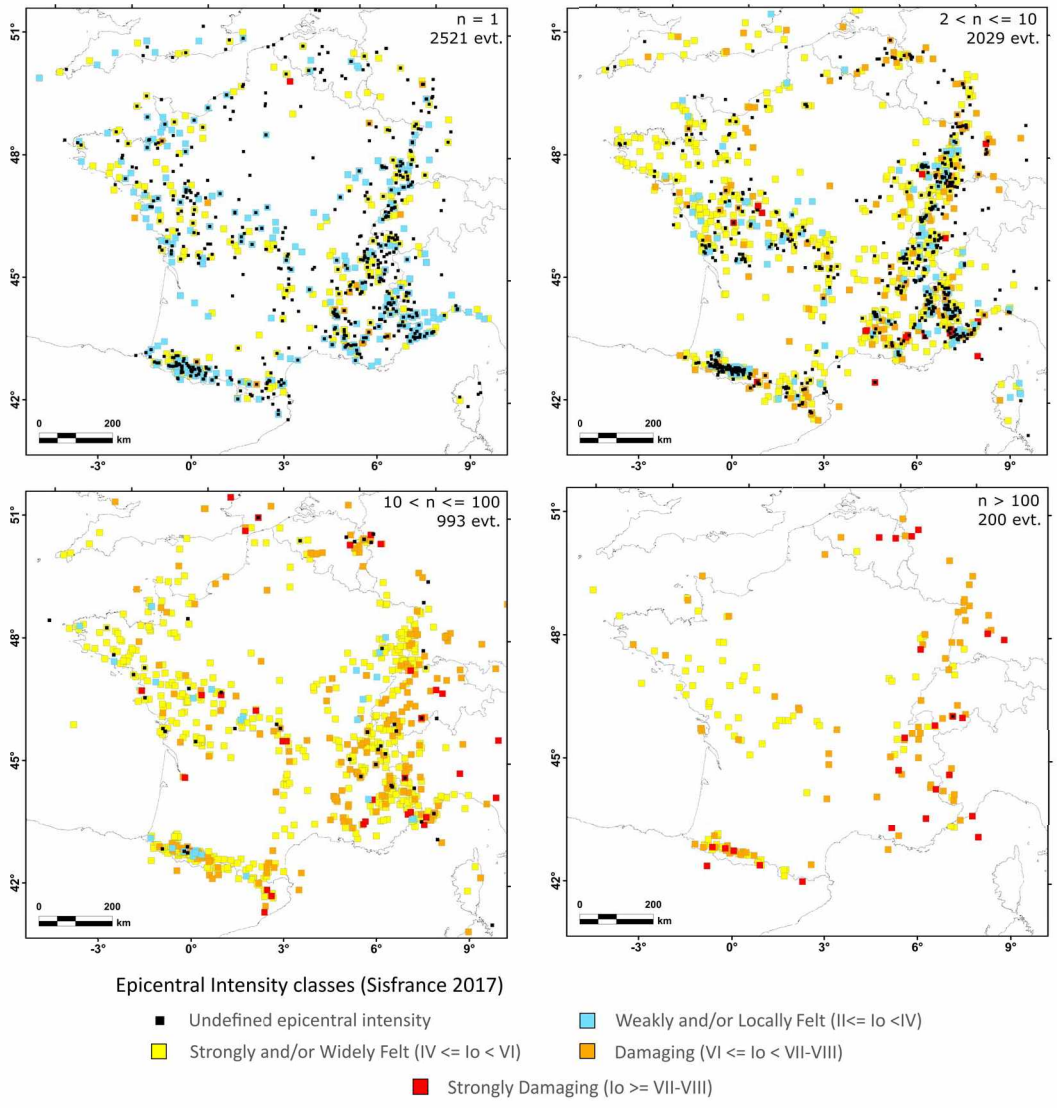

Figure 4. Number (n) of macroseismic data points (MDP) for each single true earthquake of the EPC-SIRENE table of SISFRANCE plotted in four distinctive maps (from $n = 1$ MDP to $n > 100$ MDP). The number of corresponding events (evt.) is reported on each map as well as the epicentral intensity classes defined in Figure 1.

the database;

- When a few MDP are available, the location of the earthquake may correspond either to (1) the location of maximal intensity when characterized by a single IDP; (2) the barycenter of the strongest IDPs, weighted by their quality (Qiobs); (3) the barycenter of weaker IDP's, when they are better represented in number and spatially; (4) the barycenter of felt data points when no IDP is available;
- When a significant number of MDP are available, allowing to draw isoseismal lines, it enables the determination of the macroseismic epicenter as the barycenter of the highest intensity area [see Levret et al., 1994]. In such

cases, the epicentral intensity is not always equal to the maximum intensity observed locally [owing to site effects due to local amplification, Levret et al., 1994].

However, additional information contained in the archives also led the contributors of SISFRANCE to constrain the epicentral locations with more subjective criteria. A recurring criteria being, for example, to consider the occurrence of foreshocks or aftershocks felt very locally in order to guide the choice of an epicentral location. Considering that part of these epicenters was proposed on such subjective criteria, it was decided at the beginning of the 2000s that any modification of an earthquake of intensity greater than VI (i.e., impacting most the seismic hazard) would be accompanied by an internal justification note. This approach having been lately established, a given number of earthquake epicentral parameters might need reinterpretation. In this regard, a future improvement of the database may consist in establishing and applying a more systematic definition and application of criteria leading to the definition of the epicentral position of historical earthquakes, even though it can be questionable in a seismotectonic perspective.

Once the epicentral position of an event is defined, the epicentral intensity can then be deduced. In most cases, the epicentral intensity is equal to IOBSmax, especially when the epicenter is defined at the location of a punctual observation, which is the case with isolated events, but not only. The intensity Io also corresponds to IOBSmax when the macroseismic observations of maximum intensity are close in space to each other. When this is not the case (i.e., the observation points are spaced from each other and the epicentral location is at the barycenter of this distribution), a calculated intensity (QIE = K) is established on the basis of an attenuation law describing the decrease of intensity with distance following a Sponheuer [1960] relationship [Levret et al., 1994], or an arbitrarily intensity (QIE = E) is defined when the data points are too scattered and/or too far from each other.

A cartographic illustration of the number of MDP that allowed establishing epicentral parameters is presented in Figure 4. Similarly, the evolution over time of the number of MDP collected for each event is presented in Figure 5, while the temporal distribution of seismic events by intensity class is presented in Figure 2b. These different representations allow appreciating the representativeness and completeness of historical data:

- Most of the historical database contains events for which the number of MDP is equal or lower than 10 (4550 event over the 5743 of the true events, Figure 4). They obviously mainly correspond to events of unknown epicentral intensity (Figure 5), but also to non-damaging events and even some damaging events, going back through time (Figures 4 and 5). Indeed, a simple linear regression on the different defined classes of events (Figure 5) shows that the number of MDP we may expect for a given earthquake decreases as a function of time and epicentral intensity;
- Figure 2b suggest that the chances of finding the mention of an earthquake within the historical archives increases with increasing intensities going back in time. It somehow provide hints concerning the completeness of the dataset for each epicentral intensity class (discussed hereafter), and rather clearly shows that our knowledge significantly increases starting from the 15th century;
- Looking at the spatial repartition of the best-known events (>10 MDP), we observe that the number of low intensity events (in blue and yellow, Figure 4) in the northwestern part of France is over-represented in comparison to what is observed in the Pyrenees and in the Alps, where observed seismicity rates are higher [Mazzotti et al., 2020]. This is difficult to interpret directly because different factors may come into play, such as the completeness of the historical archives, the habitat distribution in plains versus mountainous regions, or even regional differences of intensity attenuations with distance [Bakun and Scotti, 2006, Baumont et al., 2018]. As an example, the Provence region highlights a number of significant events for which few MDP have been collected. One explanation may come from the fact that in this region of written/Roman law (i.e., archived documents are strongly regulated), where ancient narrative sources are rarer than elsewhere in France [Giordanengo, 1988], annotations relating natural phenomena are rare

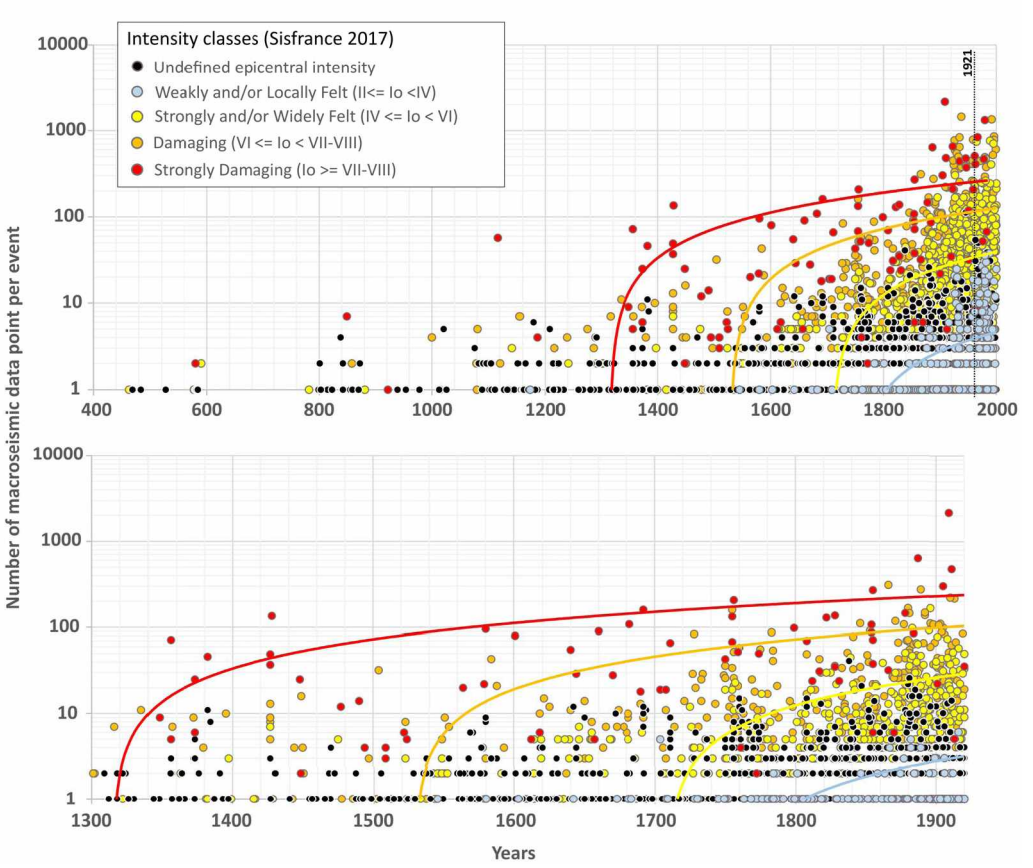

Figure 5. Number of macroseismic data points (MDP) for each event of the EPCSIRENE table of SISFRANCE. Colors corresponds to the five intensity classes defined in Figure 2. On top: for all the database (year 463 to year 2007). Below: close-up for the period between year 1300 and year 1921, corresponding to the most complete part of the database before the birth of the BCSF. Linear regressions associated with each intensity class are the same on both graphs, established for the overall period covered by the database. Figures with one plot per intensity range are provided in Supplementary materials.

throughout the middle ages.

- Concerning events with more than 100 MDP (Figure 4), they mostly correspond to earthquakes occurring since the 19th century, with a strong increase starting from 1921 (birth of BCSF, Figure 5);
- A significant number of events for which an epicentral intensity has not been defined are characterized by more than one MDP (Figures 4 and 5). If this is normal for the oldest events, because few descriptions are provided in ancient archives (i.e., MDP often qualified as "felt"); it is more surprising concerning the younger ones for which an effort to precise their epicentral characteristics should be performed;
- By construction, all the events characterized by a single MDP are located onshore (Figure 4), which does not mean that part of these could be situated offshore. More generally, the location of earthquakes in the marine domain is associated to strong uncertainties [e.g. Larroque et al., 2012];

However, such a global overview of the database does not allow retrieving enough information that would allow at the national or more regional scales to directly address the notion of completeness as needed for the definition of seismicity rates [Rydelek and Sacks, 1989, Stucchi et al., 2004]. Indeed, his-

torical knowledge varies and evolves regionally due to geographic, socio-economic, and political conditions, which implies that the completeness of historical data must be assessed specifically for each single region, before it can even be extrapolated to the entire country. Such an observation is all the more true concerning "small" earthquakes for which the impacted area is small, but it is also true when it comes to focusing on major events that could be missing, missed, or underrated [Valensise and Guidoboni, 2000].

The take-home message for any end user interested in using the epicentral dataset is to be aware that epicentral intensity and location are the two most expert-opinion-driven parameters in historical databases, derived from a second level of interpretation [Scotti et al., 2004]. While these data are crucial in order to plot the seismicity, perform seismotectonic, and seismic hazard analyses, they must be handled with care and consciousness, keeping in mind the historical and geographical context of the data they derive from. The uncertainties provided in SISFRANCE can definitely help exploiting these data [e.g. Provost and Scotti, 2020], but they may not to be always sufficient for specific studies. As an example, the great number of events associated with an undefined epicentral intensity, even within the most recent times, recalls that the knowledge provided by archives is intrinsically limited.

2.2.3. *Focus on damaging earthquakes*

A focus on damaging events reported in SISFRANCE is provided in this section. 605 earthquakes have an epicentral intensity compatible with the occurrence of structural damage on buildings (Io ≥ VI, Table 2), of which 336 are located within the French territory. The others are mainly distributed near the French border, but some of the strongest events occurring at a great distance and having been felt in France also appear in the database, such as the great Lisbon earthquake of 1755.

A cartographic representation of these events, associated with their uncertainties is presented in Figure 6. A distinction is made between events qualified as "fairly well known" (QPOS = A, B, C and QIEPC = A, B, K, see Table 1) and those for which the uncertainty about the location or the epicentral intensity is greater (QPOS = D, E, I; QIEPC = C, E, I). Of the 412 events qualified as "uncertain", practically half are located in France (202 events). Likewise, out of the 193 well-known events, a much larger part of them is located in France (134 events). We then notice that a majority of events located abroad, but also at sea are of "uncertain" quality. This is obvious concerning offshore events, for which fishes, unfortunately, cannot witness. But for those located beyond the borders, this is explained by the fact that research was carried out with the objective of prioritizing the collection of MDP in France and also that the data from foreign databases are not integrated into SISFRANCE. Again, it underlines the benefits from future cross-border collaborations on past events characterization.

Beyond this spatial aspect, the birth of the BCSF 100 years ago (i.e., in 1921) allowed an important leap forward in terms of knowledge concerning, in particular, the stronger events. Indeed, one may notice that, of the 134 best-known events reported in France (Io ≥ VI), 75 took place after the creation of the BCSF, and therefore since the establishment of almost systematic macroseismic surveys. Surprisingly, since the onset of macroseismic inquiries, 14 events located in France are still qualified as uncertain, the last of them dating back to 1982.

Of the slightly more than 100,000 MDP associated with real events in SISFRANCE, ~50% of them correspond to events of epicentral intensity greater than or equal to VI (51003 MDP), of which 15045 were acquired before 1921. Between 1921 and 2007, SISFRANCE data corresponds primarily (70% of events) to a reinterpretation by SISFRANCE of the macroseismic surveys and questionnaires from BCSF and BRGM. Macroseismic data for events after 2007 are only available at the BCSF website (www.franceseisme.fr).

The number of IDP for each level of intensity exceeding VI is presented in Figure 7. The highest intensity points (I = IX–X to X–XI) are only found abroad. In France, the strongest IDP are 2 points of intensity IX recorded following the Lambesc earthquakes in 1909 (Io = VIII–IX) and the Ligurian earthquake of 1887 (Io = IX). Regarding intensity levels VII–VIII to VIII–IX, we note that they were mostly recorded before 1921 (73 out of 87 IDP and 68 out of 71 IDP, respectively), with the occurrence of few strong earthquakes since then. The 14 points of intensity VII–VIII and the 3 points of intensity VIII observed since

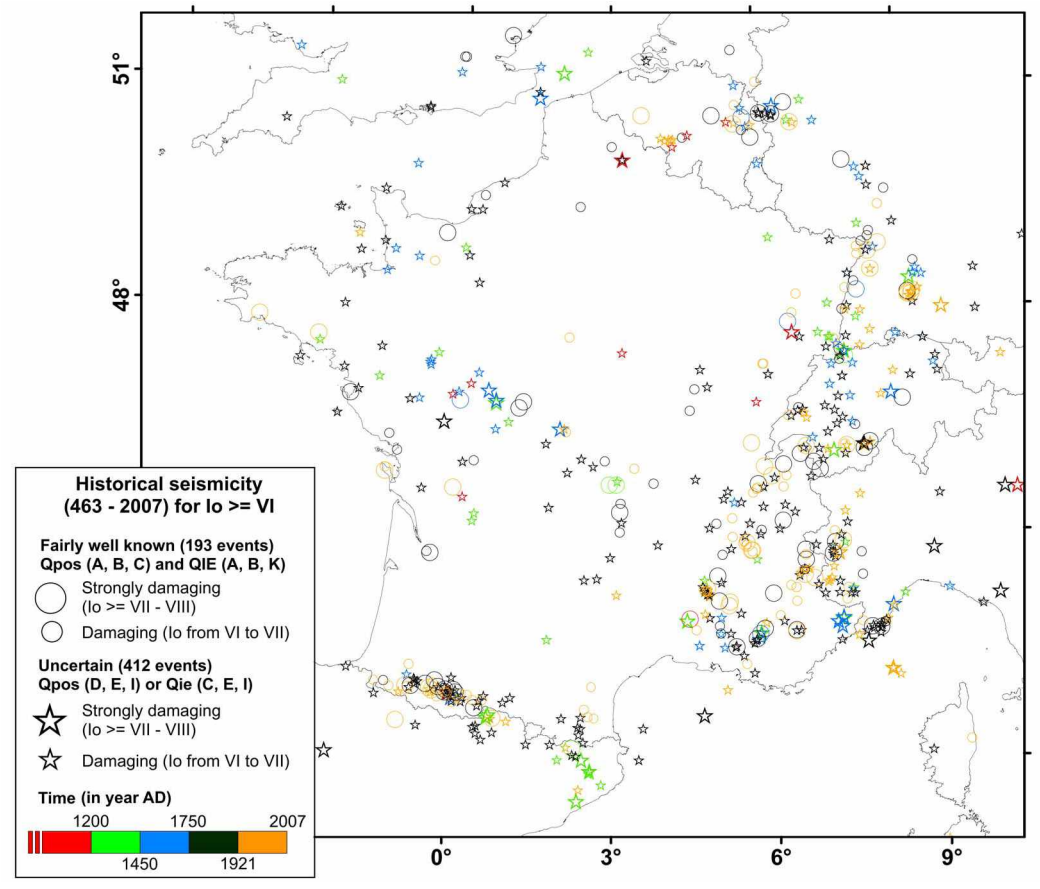

Figure 6. Earthquakes of the EPCSIRENE table of SISFRANCE, with epicentral intensities equal or higher than VI (MSK-64). Two classes of uncertainties are presented (fairly well known and uncertain), following criteria reported in Table 1. The color scale corresponds to key historical periods reported in Figure 8. Modified and updated from Scotti et al. [2004].

1921 are associated with the earthquakes of Viella in Spain (1923, Io = VIII), Chasteuil (1951, Io = VII–VIII), St Paul sur Ubaye (1959, Io = VII–VIII), Corrençon en Vercors (1962, Io = VII–VIII), Arette (1967, Io = VIII) and Arudy (1980, Io = VII–VIII). Since 2007, only the Le Teil earthquake reached such levels of intensity [see Schlupp et al., 2021]. Concerning IDP's below VII–VIII in France, we note a significant increase in the number of municipalities affected by these intensity levels since 1921. This is of course not related to an increasing rate of damaging events in recent times, but to a more complete view of the impact of earthquakes provided by the macroseismic surveys of BCSF. On the contrary, the number of municipalities impacted by intensities above VIII drops drastically in France, occurring only before the birth of BCSF (7 IDP of intensity VIII–IX, 2 IDP of intensity IX). This could be explained (1) by a return period of strong earthquakes equal or longer than the period of time covered by the database, or (2) by the incomplete knowledge we have concerning the stronger events, especially for the older ages. Both hypotheses are only valid when considering stable seismicity rates over time. We favor the second hypothesis considering the poor historical knowledge we have for the older events. For instance, the Basel earthquake of year 1356 is an interesting example. This event of epicentral intensity Io = IX, is characterized by only two IDPs = VIII–IX, and 42 IDP = VIII, coming from the fact that we most probably miss information

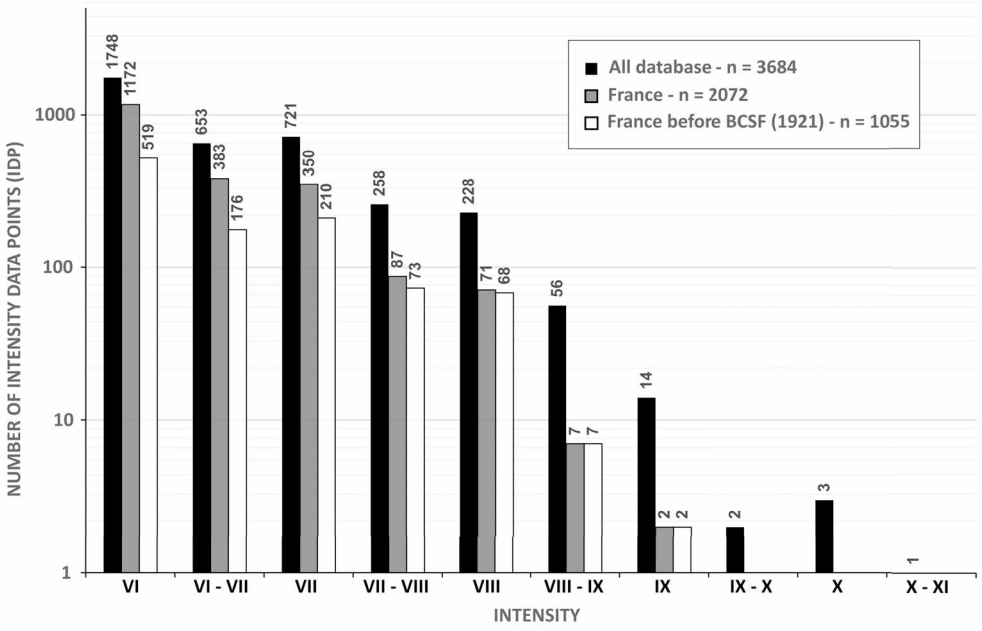

Figure 7. Number of intensity data points (IDP) corresponding to each intensity level equal or higher than VI (MSK-64 scale) for the entire SISFRANCE database (in black), for France (in gray), and for France before 1921 (in white).

mentioning cities or villages located the closest to the epicenter.

3. Archival researches and perspectives

The nature of archival research dedicated to historical seismicity has changed little over time. Of course, the form has evolved with the evolution of archives centers and their organization, with the development of consultation tools from microfilm to the Internet; but if the work has been greatly facilitated by these developments, it still fundamentally resembles a "Benedictine work" [Lambert et al., 2012]. In this section, we offer a synthesis of our feedback concerning the archival research carried out to date, then present and discuss some avenues for research and improvement of SISFRANCE for the years to come.

3.1. *Archival researches and expected progresses in SISFRANCE*

In general, annotations of earthquakes are hidden in archives of all kinds. Any document, whether accountant or notarial, being likely to conceal it (Figure 8). Sometimes, it only consists of half a line, without direct relationship to the document bearing it (e.g., parish registers). Their collection involves the examination of a large number of archives kept in archive centers, as well as locally and abroad, and undertaken with an effort to prioritize documents potentially carrying information relating to the occurrence of an earthquake. Indeed, the examination of one corpus or another is made in the light of a cost-benefit ratio, evaluated taking into account the difficulty of access to the document (i.e., time of consultation, transcription ...) on the one hand, and the probability of finding interesting information on the other hand. Overall, it is considered that around 10% of the consulted archives provide usable information, of course meaning that one has to go through the other 90% [Lambert et al., 2012]. Once new documents come to light, many errors of interpretation of these documents would remain without a stage of interpretation and contextualization of historical sources [Ambraseys, 1983, Fradet, 2016]. The first of these aiming to establish the genealogy of archival sources, which encourages the search for first-hand

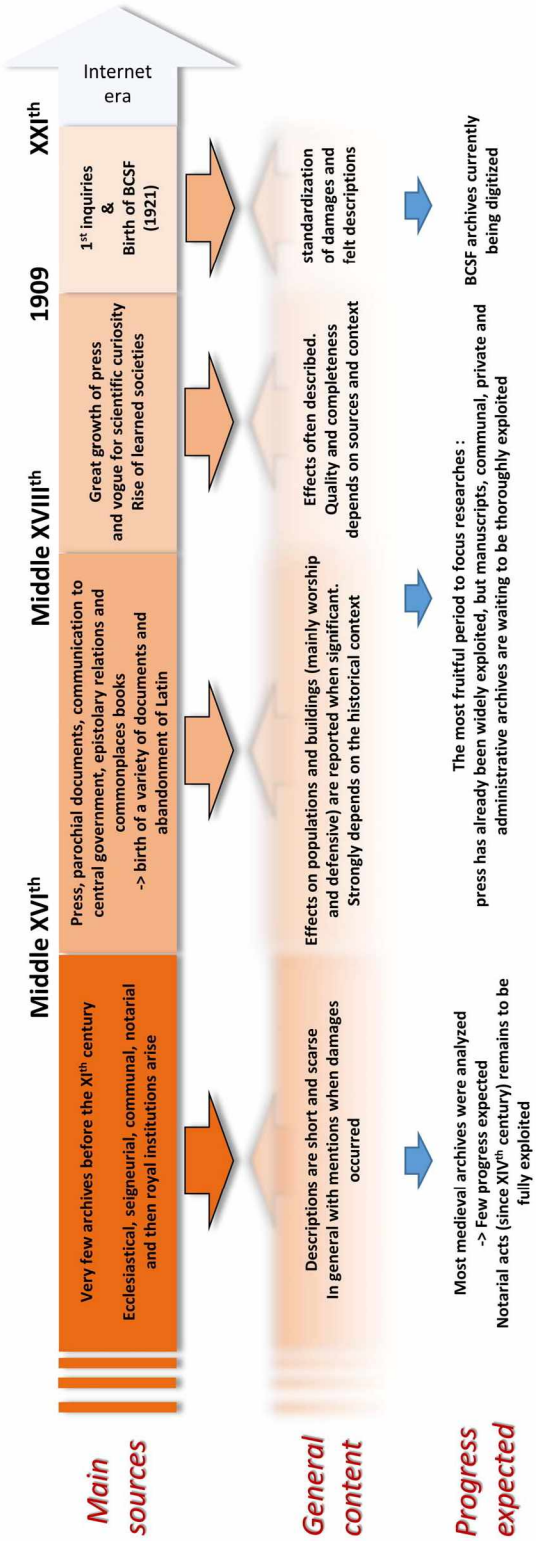

Figure 8. Timeline presenting a synthesis of the main archives where mentions of earthquakes are reported, their general content, and the expected progress in pursuing their exploration.

documents (i.e., primary sources) and if possible, the most contemporary to each event. These very general statements may, however, not be equally applicable to all the historical periods and locations, the research carried out on the Middle Ages is, for instance, completely different from that concerning more recent periods.

3.1.1. *Before the renaissance (XVIth century)*

In France, until the Middle Ages, historical sources have an eminently religious character (annals), while civil chronicles are rarer. Thanks to the increasing centralization of powers, accompanied by an increasing documentary richness over time, civil chronicles also became better and better archived and preserved (Figure 8). Given the relatively limited amount of ancient archives available, it is now possible to believe that the work carried out for years has enabled the vast majority of earthquakes that have been reported for this period to be identified. In particular, Alexandre [1990] established a critical catalog of the narrative sources at a European scale, thus discarding many documents whose little historical value had been demonstrated. For example, following the "pneuma theory" of Aristotle [Poirier, 2008], confusion could often be made with the occurrence of storms which, until the 16th century were often referred to as "terrae motus", hence classified as "false earthquake" within SISFRANCE. The earthquake annotations recorded during the Middle Ages are, with the exception of major events such as the Basel earthquake in 1356 [Lambert et al., 2005], short and lacking in detail, even when damages are mentioned. Then, uncertainties regarding both the location and intensity of ancient earthquakes remain most often strong. However, notarial records, which appear from the beginning of the 14th century, and which have been little explored to date, could still provide indirect details concerning a potential impact of earthquakes on people and properties. These handwritten archives are, however, difficult to access, although their exploration yielded interesting results for a more recent period [i.e., for the Manosque earthquake in 1708, Quenet, 2001]. Finally, exceptional discoveries can still be made on the occasion of research in archival records accessible only to a few historians, as was the case in the Vatican archives concerning the earthquake of 1186 near Uzès [Castelli et al., 2012]. Overall, the period of the Middle Age is marked in SISFRANCE by a significant increase in the number of MDP for the strongest events and from the 12th century, with a steady increase afterward (Figure 5).

3.1.2. *From the renaissance until the mid-XVIIIth century*

From the middle of the 16th century, the renaissance in general and more specifically the development of printing press (e.g., birth of the Gazette de France by T. Renaudot in 1631 and other Gazettes across Europe), and the keeping of parish registers made compulsory by the ordinance of Villers-Cotterêts of August 1539 [Foyer, 1989], contributed to considerably increase the number and the nature of archives, thus increasing the likelihood of finding records, which could mention earthquakes (Figure 8). Also corresponding to the appearance of the notion of "disaster", this period saw the appearance of the quantification of human and material tolls caused by disasters (including earthquakes), in connection with the development of the figures of the victim and of the hero [Labbé, 2017]. Even though the effects of earthquakes on people and buildings are then frequently reported, they are unequally detailed depending on the authors and the purpose of their writings, which do not yet have an established scientific approach. Moreover, the historical context, whether national or more local, is also largely influencing the quantity and quality of the information available. Indeed, long periods of "silence" are observed in the archives depending on whether periods of war, famine or other major events naturally take precedence over the occurrence of "minor" seismic events. We may, however, consider that the strongest events are identified for that period and, with a few exceptions, relatively well characterized. On the other hand, the hope of enriching knowledge for this period is widely permitted, in particular, because private and administrative archives (e.g., municipal deliberations, notarial archives, etc.), and in general, handwritten archives have not been explored much in comparison with the printed archives. As the increase in the number of MDP per event noted in SISFRANCE since the end of the Middle Ages continues, events of lower epicentral intensity are beginning to be more systematically reported (Figure 5).

3.1.3. *From the mid-XVIIIth century until the modern era (XXth century)*

From the second half of the 18th century, the development experienced during the Renaissance period continued and intensified. Press organs are multiplying and the general interest to science is taking off with the multiplication of the number of learned societies. Concerning earthquakes, in particular, the Lisbon earthquake of 1755 (Io = X–XI), widely felt in France, deeply marked the people during this period of enlightenment [Dynes, 2000], and marks a real turning point in terms of available earthquake related information (Figure 5). However, as quantity is not always a guarantee of quality, the circulating information is often repeated and sometimes quickly interpreted, distorted, amplified, leading to a necessary effort to interpret and cross-reference important data. This is all the more true in the 19th century when the press exists at all levels of society and, as it does today, often relays second hand information. In parallel, the scientific literature is developing with the aim of being readable by the general public [Metzger, 1934] and the first syntheses of historical and contemporary seismicity emerge [e.g. Pallassou, 1815, Perrey, 1845, Montessus de Ballore, 1906], as well as the first surveys aimed at locally quantifying the impact of earthquakes [see Vogt, 2003, Camelbeeck et al., 2021 and reference therein]. A very marked peak in the number of events and associated MDP is observable following the Lisbon earthquake of 1755, then followed with a strong and steady increase during the 19th century. In this very rich period, the quantity of archives available is such that their exploitation is today incomplete and that many new information can still be discovered, with the possible exception of the strongest events for which in-depth research has often already been carried out.

3.1.4. *Modern era*

Just like the earthquakes of Basel (1356, Io = IX) and Lisbon (1755) in their time, the earthquake that occurred near Lambesc in 1909 (Io = VIII–IX) marks a new turning point. This is indeed the first major earthquake for several generations [Vogt, 2003] in the French territory. The Central Meteorological Bureau took matters in hand with a detailed macroseismic survey published by Angot and Lemoine [1910], which led, a few years later, to the creation of the BCSF in 1921 [Fréchet, 2008, Roger, 2021] and to a systematization of the macroseismic surveys during the 20th century. Resorting to other sources of information remains, however, necessary and can still clarify our knowledge, in particular, because the action of the BCSF has not been continuous and homogeneous over the century [Roger, 2021, Sira et al., 2021]. The period was marked by a high and almost constant number of MDP recorded in SISFRANCE, until 2007 (Figure 5).

3.2. *SISFRANCE in the near future*

3.2.1. *Archival researches: lessons from past and present*

For each aforementioned historical period, new avenues of archival research can be considered to enrich the database (Figure 8). However, the experience gained so far shows that the depth of investigation in the archives, and hence the necessary time dedicated to it, becomes increasingly more important as the most accessible documents have already been searched for and analyzed. Between 2004 and 2017, the pace of unpublished archive discoveries (~213 per year; see Section 2.2) was maintained by continuous research and systematic monitoring of the archives by a dedicated historian, largely facilitated by the continuous online posting of new archives on the Internet. On the other hand, research carried out on more specific events, targeted considering their potential impact on the seismic hazard for nuclear installations (i.e., in general, strong and poorly known events), have often proved to be more complicated, each new piece of data implying a substantial investment. It is important to underline again that the efforts carried out in SISFRANCE were largely oriented by the needs linked to the development of the nuclear industry in France [Roger, 2021]. As a result, some regions in general and specific events, in particular, have been the subject of more in-depth study than others. This does not mean that the rest of the historical seismicity in France has been left out, continuous research all over the country allowing the knowledge to evolve regularly, but it is important to note that many earthquakes have not been reviewed since their implementation in SISFRANCE in the 1980's.

To summarize, the constitution of a database of historical seismicity relies first and foremost on the ability to find, contextualize, and analyze original documentary sources. This work requires combining the wide range of skills of historians (knowledge of the organization of and access to archives, translation into modern language, filiation of sources, historical contextualization ...) with those of historical seismologist for the translation of testimonies into values of macroseismic intensity. The retirement of J. Lambert in 2017 and the non-renewal of his position led the SISFRANCE consortium to subcontract historical researches to academics or research firms.

This approach allowed diversifying and modernizing the search methodology for original documents, with the participation of new people bringing a fresh look and new research skills. However, in order to guarantee the homogeneity and continuity of interpretations with the work carried out in the past, this phase is now completely devoted to the SISFRANCE working group. After three years of operation, this way of conducting new researches offers the possibility to target on specific earthquakes, archives or periods of interest, with requirement specifications defined upstream. However, it has the main drawback of no longer allowing wide-ranging researches and therefore precludes the continuous enrichment of the database for events not considered as a priority. In addition, the research carried out by professional historians is limited to their disciplinary field, and the training of people capable of having a critical look between history and seismology is no longer guaranteed.

In order to complement the continuous enrichment of the database performed with classical archival researches, new avenues of research are under investigation such as "data-mining techniques" aimed at automating the collection of new data from the huge amount of digitalized archives [Nayman et al., 2020], or participatory science. For the latter, the ongoing modernization of the www.sisfrance.net website is an opportunity to collect information by allowing the public, whether from academia, associations or individuals, to report new documents still unexploited. In addition, in the framework of EPOS-RESIF [Masson et al., 2021], macroseismic data from the Sisfrance database will soon be available together with data from BCSF in a unique website (www.franceseisme.fr), offering the opportunity of an enhanced visibility. It is nowadays difficult to draw consolidated feedbacks concerning these actions as we do not have the necessary hindsight. But it however appears, as shown by the last update of the database in 2017, that historical seismicity research will no longer be as effective as in the past as long as academic researches and hence the training of researchers skilled to the use of historical data related to natural events (earthquakes or other) will not be carried out in parallel to SISFRANCE.

3.2.2. *Improving the database*

The SISFRANCE database relies on a structure developed at the beginning of the 1980s. If it has proven its efficiency, remaining fully usable after 40 years, it is indeed not flawless.

There is a clear need to improve the hierarchization of the archives in SISFRANCE as pointed out by Quenet [2001] by setting up hierarchical trees explicitly presenting the relation between primary and secondary sources [Quenet et al., 2004]. He also recommended an enhanced contribution of historians to the critical analysis of historical data. Since fifteen years, the systematic production of justification notes tracing and explaining the modification of earthquakes of epicentral intensity greater than or equal to VI allowed, at least in part, to tackle this problem. However, a retrospective review of older modifications, for which specific notes were not systematically produced, constitutes a difficult task needing the involvement of historians or people trained to the critical analysis of archives and to the specificities of macroseismic analysis.

The experience gathered within the SISFRANCE working group also makes it possible to identify a certain number of areas for improvement, which are more achievable today:

- In the SISFRANCE database, documents are related to earthquakes (through Chrono and Numevt field, Figure 1). The initial architecture of the database, conceived in the early eighties did not anticipate the possibility to link each single MDP to the documents that made it possible to establish them (see the dotted arrow in Figure 1). If this point does not pose a real problem in the case of earthquakes for which few documents and few MDP are available (i.e., the major part of

the database), great difficulties arise as soon as they are more numerous. Indeed, when a modification has to be made for such or such well-known event, it is difficult today to recover the document(s) at the base of the definition of an MDP, and therefore to discuss its possible modification on the basis of a new document. OCR (*Optical Character Recognition*) techniques available today should help overcome this problem in the coming years by offering the possibility to relate the OBSIRENE and DOCUMENTS tables together. However, while a systematic ocerization (defined as the digitization of the texts contained in the archives) of printed documents is feasible, it is still complicated for handwritten documents and a manual work is still required. The ongoing "data-mining" project [Nayman et al., 2020], which requires the ocerization of SISFRANCE, will constitute an essential step forward, offering the possibility to overcome this problem in the coming years;

- Each year, the SISFRANCE working group has to establish a research program for the following one. Such a yearly exercise revealed the difficulty to know precisely which archives have already been investigated (where and when), and also those which gave negative results (therefore not referenced in SISFRANCE). A cartography of the investigated archives is partly possible starting from the referenced documents (feasibility is actually under evaluation), but it is clear that a complete vision is not achievable and that the consultation of archives that have already been visited in the past cannot be systematically excluded;
- The increasing availability of archives on the Internet is a major element allowing archival research to be carried out more and more efficiently. As part of these archives are also ocerized, their exploration by keywords are fast and comfortable. However, this implies the constitution of an extensive lexicon related to the phenomenology and impact of earthquakes. The simple term earthquake is, for example, not used systematically over the history. Such work has been carried out by Fradet [2016] for a limited time period and is also underway in order to supervise computerized searches during the "data-mining" process [Nayman et al., 2020, see next section].

3.2.3. *The promising data-mining approach*

The principle of data mining is to extract automatically useful information from various numerical databases available online. Applied to seismicity, it is an interesting methodology to improve knowledge of historical earthquakes by finding new testimonies in written archives available on the web and, therefore, to enrich the SISFRANCE database.

A dedicated method for historical documents investigation has been developed and applied to the digital library of the Bibliothèque Nationale de France (BNF), named Gallica [Nayman et al., 2020]. Given the large amount of harvested texts (3.8 millions) on this website, algorithms were designed to build and implement an efficient workflow strategy: (i) Definition of seismological ontology (6 concepts: seismic, damage, assembly, behavior, noise, divine) from SISFRANCE database which is used as dedicated dictionary to extract relevant information from the Gallica collection of documents; (ii) Semantic and knowledge enrichment of harvested documents which constitute the knowledge base; (iii) Use of advanced techniques of data mining, as for example the use of a similarity process that dramatically helps to find relevant text through the background "noise".

After one year of investigations in the Gallica Corpus, the work led to more than 1600 documents dealing with earthquakes felt in mainland France, among which 62% are not listed in the SISFRANCE database. All this new information will be analyzed in the frame of the SISFRANCE consortium for updating the database. The approach is now being improved, mainly concerning data-mining techniques, but given the success of the first application, the exploration of other websites is already planned (RetroNews, LecturaPlus). Although the proposed methodology aims at facilitating source findings, it does not preclude the expertise of the historian for analyzing and interpreting new sources.

4. Conclusion

In this paper, we presented the current state of knowledge reported in the SISFRANCE database of historical seismicity. We also illustrated the need for a careful use of SISFRANCE, in understanding the limits introduced by the interpretation of archival data, even if SISFRANCE provides several quantified elements allowing to better appreciate the quality related to each basic or interpreted data. For instance, we showed that during the 40 years of data collection and interpretation, the homogeneity of treatment of the data, especially concerning the epicentral parameters, cannot be systemically justified and that there is often a need to go back to the primary data to understand what has been done. As such, some avenue for improving the database are enlightened, in particular the need to apply a more systematic procedure to define epicentral locations and intensities of earthquakes.

Out of the 5743 real earthquakes inventoried in 2017, half of them are and will remain poorly or very poorly known. However, clarifying our knowledge for some of them as well as for some better but still insufficiently characterized ones, in particular those indicating an Io $\geq$ VI, remains a fundamental element in understanding the seismicity of metropolitan France [Mazzotti et al., 2020] and its related seismic hazard. The hereby envisaged avenues of archival research, especially for the periods from the 15th century, are compatible with such an objective. Since 2017 however, archival research is no longer carried out by a dedicated historian within SISFRANCE, hence leading the SISFRANCE working group to promote historical seismicity research and explore new strategies to continue to enrich and improve the database. Given the new impetus of French governmental research agencies that encourage transdisciplinarity research, SISFRANCE is also pursuing the idea of further engaging the academic community of historians and seismologists, presently greatly underrepresented in the field of historical seismicity in spite of the importance of this data, fundamental for any seismotectonic and seismic hazard study in France.

Acknowledgments

We sincerely thank our colleague Jérôme Lambert who patiently collected and interpreted archives for SISFRANCE during almost 40 years. He shared his passion and strong opinions over the years, helping us familiarize ourselves with historical data and their often subtle interpretation.

We also acknowledge our predecessors within the SISFRANCE working group, especially Agnès Levret who greatly contributed to enrich and develop the database. We also thank the members of the SISFRANCE coordination committee as well as our institutions which have enabled SISFRANCE to continue living after more than 40 years.

Finally, we wish a happy 100th Birthday to the Bureau Central Sismologique Français (BCSF) and thank them for their crucial contribution to the SISFRANCE project.

This publication is dedicated to the memory of our colleague and friend Christophe Clément.

Supplementary data

Supporting information for this article is available on the journal's website under https://doi.org/10.5802/crgeos.91 or from the author.

References

Alexandre, P. (1990). Les séismes en Europe occidentale de 394 à 1259. In *Nouveau catalogue critique*. Publications de l'Observatoire Royal de Belgique, Série Géophysique, Bruxelles.

Ambraseys, N. (1983). Notes on historical seismicity. *Bull. Seismol. Soc. Am.*, 73(6A), 1917–1920.

Angot, A. and Lemoine, P. (1910). Le tremblement de terre du 11 juin 1909 dans le sud-est de la France. *Ann. Géogr.*, 19(103), 8–25.

Bakun, W. H. and Scotti, O. (2006). Regional intensity attenuation models for France and the estimation of magnitude and location of historical earthquakes. *Geophys. J. Int.*, 164(3), 596–610.

Baumont, D., Manchuel, K., Traversa, P., Durouchoux, C., Nayman, E., and Ameri, G. (2018). Intensity predictive attenuation models calibrated in Mw for metropolitan France. *Bull. Earthq. Eng.*, 16(6), 2285–2310.

Beucler, É. et al. (2021). Characteristics and possible origins of the seismicity in northwestern France. *C. R. Géosci.*, 353(S1), 1–25. (in this issue).

Camelbeeck, T., Knuts, E., Alexandre, P., Lecocq, T., and Van Noten, K. (2021). The 23 February 1828

Belgian earthquake: a destructive moderate event typical of the seismic activity in Western Europe. *J. Seismol.*, 25(2), 369–391.

Cara, M., Cansi, Y., Schlupp, A., Arroucau, P., Béthoux, N., Beucler, E., et al. (2015). SI-Hex: a new catalogue of instrumental seismicity for metropolitan France. *Bull. Soc. Géol. Fr.*, 186(1), 3–19.

Castelli, V., Camassi, R., and Molin, D. (2012). The Uzège (Southeastern France) 22 March 1186 Earthquake Reappraised. *Seismol. Res. Lett.*, 83(3), 604–614.

Cecic, I., Musson, R. M., and Stucchi, M. (1996). Do seismologists agree upon epicentre determination from macroseismic data? A survey of ESC Working Group "Macroseismology". *Ann. Geophys.*, 39(5), 1013–1027.

Doubre, C., Meghraoui, M., Masson, F., Lambotte, S., Jund, H., de Berc, M. B., and Grunberg, M. (2021). Seismotectonics in Northeastern France and neighboring regions. *C. R. Géosci.*, 353(S1), 1–33.

Dynes, R. R. (2000). The dialogue between voltaire and Rousseau on the lisbon earthquake: the emergence of a social science view. *Int. J. Mass Emerg. Disasters*, 18(1), 97–115.

Foyer, J. (1989). L'Ordonnance de Villers-Cotterêts. *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, 133(3), 636–646.

Fradet, T. (2016). *Vulnérabilité et perception face aux tremblements de terre en France, 1650–1850.* PhD thesis, Université Paris-Saclay.

Fréchet, J. (2008). Past and Future of historical seismicity studies in France. In *Historical Seismology*, pages 131–145. Springer, Dordrecht.

Giordanengo, G. (1988). Le droit féodal dans les pays de droit écrit. In *L'exemple de la Provence et du Dauphiné. XIIe-début XIVe siècle*, pages 5–331. Ecole française de Rome, Rome. (Bibliothèque des Écoles françaises d'Athènes et de Rome, 266).

Godefroy, P. and Levret, A. (1992). Creation and management of a data base dedicated to historical and contemporary macroseismicity in France: various applications in earthquake engineering. In *Recent Advances in Earthquake Engineering and Structural Dynamics*, pages 89–116.

Godefroy, P., Thirion, S., Lambert, J., and Cadiot, B. (1980). Informatisation du patrimoine de sismicité historique de la France. *Bull. BRGM*, pages 139–145.

Labbé, T. (2017). *Les catastrophes naturelles au Moyen Âge.* CNRS Editions, Paris.

Lambert, J. et al. (2012). Reconstituer la séismicité historique : un travail de bénédictin nécessaire pour tirer les enseignements des anciens séismes. In *Revue Géologues*, volume 169, pages 68–71. Société Géologique de France, Paris.

Lambert, J., Levret-Albaret, A., Cushing, M., and Durouchoux, C. (1996). In *Mille ans de séismes en France (catalogue d'épicentres, paramètres et références)*, page 70. Ouest Éditions, Nantes.

Lambert, J., Winter, T., Dewez, T. J., and Sabourault, P. (2005). New hypotheses on the maximum damage area of the 1356 Basel earthquake (Switzerland). *Quat. Sci. Rev.*, 24(3–4), 381–399.

Larroque, C., Baize, S., Albaric, J., Jomard, H., Trévisan, J., Godano, M., et al. (2021). Seismotectonics of southeast France: from the Jura mountains to Corsica. *C. R. Géosci.*, 353(S1), 1–47.

Larroque, C., Scotti, O., and Ioualalen, M. (2012). Reappraisal of the 1887 Ligurian earthquake (western Mediterranean) from macroseismicity, active tectonics and tsunami modelling. *Geophys. J. Int.*, 190(1), 87–104.

Levret, A. (1991). Management and processing of the french macroseismic data base Sirene (No. CEA-DAS–776). CEA Centre d'Etudes Nucleaires de Fontenay-aux-Roses.

Levret, A., Backe, J. C., and Cushing, M. (1994). Atlas of macroseismic maps for French earthquakes with their principal characteristics. *Nat. Hazard.*, 10(1–2), 19–46.

Levret, A. and Mohammadioun, B. (1984). Determination of seismic reference motion for nuclear sites in France. *Eng. Geol.*, 20(1–2), 25–38.

Manchuel, K., Traversa, P., Baumont, D., Cara, M., Nayman, E., and Durouchoux, C. (2018). The French seismic CATalogue (FCAT-17). *Bull. Earthq. Eng.*, 16(6), 2227–2251.

Masson, F., Auclair, S., Bertil, D., Grunberg, M., Hernandez, B., Lambotte, S., et al. (2021). The transversal seismicity action RESIF: a tool to improve the distribution of the French seismicity products. *Seismol. Soc. Am.*, 92(3), 1623–1641.

Mayer-Rosa, D. and Schwarz-Zanetti, G. (2004). On historical earthquakes in Switzerland: summary of compilations and investigations. *Ann. Geophys.*, 47(2–3), 621–630.

Mazzotti, S., Jomard, H., and Masson, F. (2020). Pro-

cesses and deformation rates generating seismicity in metropolitan France and conterminous Western EuropeProcessus et taux de déformation générant la sismicité en France métropolitaine et en Europe occidentale proche. *Bull. Soc. Géol. Fr.*, 191(1), 1–20.

Medvedev, S. V., Sponheuer, W., and Kárník, V. (1967). *Seismic Intensity Scale, Version 1964 (MSK-64)*, volume 48. Institute of Geodynamics, Jena, Germany, Publ. page 13.

Metzger, H. (1934). La littérature scientifique française au XVIIIe siècle. *Archeion*, 16(1), 1–17.

Montessus de Ballore, F. (1906). *Les tremblements de terre: Géographie séismologique.* Libr. Armand Colin, Paris. page 475.

Musson, R. M., Grünthal, G., and Stucchi, M. (2010). The comparison of macroseismic intensity scales. *J. Seismol.*, 14(2), 413–428.

Nayman, E., Bothua, M., Mayor, J., and Testut, N. (2020). On the use of data mining to improve the knowledge of historical earthquakes. Deliverable SIGMA-2-2019-D2-039, 68 pp., SIGMA2 Research Programm (www.sigma-2.net).

Pallassou, P. B. (1815). Liste de plusieurs tremblements de terre ressentis dans les Pyrénées. In *I.D., Mémoires pour Servir à l'Histoire Naturelle des Pyrénées et des Pays Adjacents*, pages 260–284. T1 Pau, Vignancour.

Perrey, A. (1845). *Mémoire sur les tremblements de terre ressentis en France, en Belgique et en Hollande, depuis le quatrième siècle de l'ère chrétienne jusqu'à nos jours (1843 inclusiv.).* M. Hayez, Bruxelles.

Poirier, J. P. (2008). Histoire de la sismologie. *Travaux du Comité français d'Histoire de la Géologie, 3e série*, 22(2), 27–35.

Provost, L. and Scotti, O. (2020). QUake-MD: Open-source code to quantify uncertainties in magnitude–depth estimates of earthquakes from macroseismic intensities. *Seismol. Res. Lett.*, 91(5), 2520–2530.

Quenet, G. (2001). *Les tremblements de terre en France aux XVIIe et XVIIIe siècles: une histoire sociale du risque.* PhD thesis, Paris 1.

Quenet, G., Baumont, D., Scotti, O., and Levret, A. (2004). The 14 August 1708 Manosque, France earthquake: new constraints on the damage area from in-depth historical studies. *Ann. Geophys.*, 47(2–3), 583–595.

RFS 2001-01 (2001). French Safety Rule, published by the French Nuclear Safety Authority. http://www.asn.fr/index.php/Divers/Autres-RFS/RFS-2001-01.

Richter, C. F. (1958). *Elementaty Seismology (No. BOOK).* W. H. Freeman and Co., San Francisco, California. page 768.

Roger, M. (2021). Rothé's legacy to the French Central Seismological Bureau (BCSF): a history of hegemony in French seismicity. *C. R. Géosci.*, 353(S1), 1–18. (in this issue).

Rydelek, P. A. and Sacks, I. S. (1989). Testing the completeness of earthquake catalogues and the hypothesis of self-similarity. *Nature*, 337(6204), 251–253.

Schlupp, A. et al. (2021). EMS98 intensities distribution of the "Le Teil" earthquake, France, 11 November 2019 (Mw 4.9) based on macroseismic surveys and field investigations. *C. R. Géosci.*, 353(S1), 1–28. (in this issue).

Scotti, O., Baumont, D., Quenet, G., and Levret, A. (2004). The French macroseismic database SISFRANCE: objectives, resultsand perspectives. *Ann. Geophys.*, 47(2/3), 571–581.

Scotti, O., Clément, C., and Baumont, D. (2014). Seismic hazard for design and verification of nuclear installations in France: regulatory context, debated issues and ongoing developments. *Boll. Geofis. Teor. Appl.*, 55(1), 135–148.

Sira, C. et al. (2021). 1921–2021: 100 years of macroseismic studies at BCSF. *C. R. Géosci.*, 353(S1), 1–29. (in this issue).

Sponheuer, W. (1960). Methoden zur Herdtiefenbestimmung in der Makroseismik. In *Freiberger Forschungsh C. 88.* Akademie Verlag, Berlin. page 120.

Stucchi, M., Albini, P., Mirto, M., and Rebez, A. (2004). Assessing the completeness of Italian historical earthquake data. *Ann. Geophys.*, 47(2/3), 659–673.

Sylvander, M., Rigo, A., Sénéchal, G., Battaglia, J., Benahmed, S., Calvet, M., et al. (2021). Seismicity patterns in southwestern France. *C. R. Géosci.*, 353(S1), 1–26.

Thirion, S. (1983). SIRENE: Système informatique de rassemblement des événements naturels existants. Application à la sismicité historique de la France. Rapport BRGM 83SGN623GEG, 108 pp.

Valensise, G. and Guidoboni, E. (2000). Towards new research strategies: silent seismogenic areas or silent sources? *Ann. Geophys.*, 43(4), 797–812.

Vogt, J. (1979). *Les tremblements de terre en France.* Mémoire du BRGM, No 96. BRGM, Orléans.

Vogt, J. (1981). Problèmes de séismicité historique de France.

Vogt, J. (2003). Bref historique des enquêtes macrosismiques en France, dans leur contexte, avec quelques exemples de pays voisins. *Travaux du Comité français d'Histoire de la Géologie*, 3(tome 17), 175–192.