SADI Journal of Interdisciplinary Research
ISSN: 2837-1909| Impact Factor: 6.08
Volume. 10, Number 1; January-March, 2023;
Published By: Scientific and Academic Development Institute (SADI)
8933 Willis Ave Los Angeles, California
https://sadipub.com/Journals/index.php/SJIR/index|

RELIABILITY ASSESSMENT OF HISTORICAL SOURCES USING THE HISTORICAL DOCUMENT QUALITY METHOD (HDQM)

Anthony Emmanuelle and Frank Roberto
Saint-Venant Hydraulics Laboratory (LHSV), Ecole des Ponts ParisTech

Abstract: Historical sources and data play an important role in studies related to extreme events, natural hazards, and environmental sciences. However, it is crucial to ensure the reliability of these sources and the data they contain, in order to avoid errors of interpretation, misleading information, and incorrect conclusions. In this paper, we present an innovative method to assess the quality of historical documents called the Historical Document Quality Method (HDQM). The HDQM works in three steps: historical critical analysis, decision tree and evaluation of four criteria, and assigning a final score using expert systems. This method is user-friendly, easy-learning, operational, and functional. It has been developed through an interdisciplinary work between historians, engineers, and mathematicians. The results obtained using the HDQM are very consistent, with models developed to assign a final score to historical documents having a coefficient of determination close to 1. This method can be particularly useful for studies on extreme events, where the reliability of historical sources and data is essential.

Keywords: Historical sources, Data reliability, Extreme events, Qualitative data, Historical critical analysis, Decision tree, Expert systems, HDQM.

INTRODUCTION

Located on the path of mid-latitude storms, Europe is often hit by these events that can cause coastal flooding, major damage and death (e.g: the storm of January 31thFebruary 1st 1953; the storm of October 15th 1987; Lothar and Martin, end of December 1999). Due to strong impacts on humans, economy, ecology and industry, the necessity arose to implement an efficient historical reconstruction and analysis of past storms, coastal flooding, and surges over a longer period to better assess these risks. The value of using historical sources and data to improve the knowledge of these past extremes has been identified since decades and is, nowadays, no longer to be demonstrated (Abadie et al., 2018; Athimon, 2021; Baart et al., 2011; Brazdil and Kotyza, 2004; Breilh et al., 2014; Camuffo, 1993; Chaumillon et al., 2017; De Vries and Winsemius, 1970; De Kraker, 2006; Garnier et al., 2018; Gottschalk, 1971-1977; Gram-Jensen, 1985; Haigh et al., 2016; Hickey, 1997; Kempe, 2006; Lamb and Frydendahl, 1991; Pfister et al., 2010; Soens, 2009; Sweeney, 2000)

In fact, studies on extreme events increasingly use historical sources and data. Useful information can either come from quantitative or qualitative historical data. Quantitative data are numerical data, that can be found in different kind of documents such as account registers, weather records, invoices, tidal ledgers, seismograms… They are often used to improve statistical analyses on extreme events. Qualitative data are descriptive and narrative data. The content of this paper will only focus on qualitative historical sources and data which mainly come from primary and / or secondary historical sources such as chronicles, diaries, newspapers, letters, account / parish / notarial or city council registers, post-disaster investigations, engineers' reports, etc. A primary source is a document written by a person who is contemporaneous with the historical event. It contains first-hand information and descriptions made by the author who experienced the events. On the other hand, a secondary source is written by a non-contemporary author who copies or draws inspiration from some primary sources. The author of the secondary source, who has not experienced the event of which he/she is speaking, produces a speech about it. Both, primary and secondary sources need a historical critical approach in order to avoid misunderstandings, errors of interpretation, misleading findings, false conclusions on the use of historical documents (Van Bavel et al., 2019). In addition, the research in history benefits both from historical sources and from the work of colleagues (living or dead) carried out from ancient documents. These technical or scientific productions are part of the literature and should therefore not be confounded with secondary historical sources. They are written by an expert and contain an analysis and interpretation of the data, they benefit to the topic they address (Bonnechere, 2008).

The question of the reliability of historical sources and data is of highest importance when we deal with extreme events. In particular, the historical sources and data used to estimate past extreme sea levels and quantify skew surge – which is the difference between the maximum observed water level and the maximum predicted water level during one tidal cycle (Haigh et al., 2015) – must be trustworthy and of good quality. Otherwise, some of the possible risks would be to impact statistical analyses using over or underestimate reconstructed water levels and skew surges or taking into account storms that did not occur at that time (error of date) or did not impact that specific area (error of location). The consequences of such misunderstandings, errors of interpretation, misleading information can be very important for the population, the environmental or town planning, engineering, the economics, industries, insurance companies, etc. For example, the skew surges computed with historical sources (Athimon et al., 2021; Giloy et al., 2019) are used in studies that proposes new statistical methods for the protection of coastal nuclear power plants from the risk of coastal flooding (Frau, 2018; Frau et al., 2018; Hamdi et al., 2015; Hamdi et al., 2018).

Both in France and abroad, researchers and research groups on extremes events such as river floods, earthquakes, avalanches, coastal flooding, who use historical documents for their studies are aware of the necessity of the quality control to work with reliable historical sources and data (Benito et al., 2004; Fradet, 2016; Frau, 2018; Giacona et al., 2017; Glaser, 1996; Idier et al, 2020; Lambert, 1986; Mangeon et al., 2020; Molinari et al., 2017; Porfido et al., 2009; Veale et al., 2017). Some of them (e.g: SSHAC, SISFrance, BDHI) have tried to build a quantified rating system in order to define the reliability of historical documents and data they provide (Albini et al., 2013; Arnoux et al., 2021; Hamdi et al., 2018; Lang et al., 2017; Scotti et al., 2004; Torres-Vera, 2010). Still, when their method is detailed and explained, important lacks remain regarding the evaluation of these documents. Typical errors are, for example, the confusion of old scientific literature with historical sources or the combination of two points that must be clearly distinguished: the authenticity of a historical document and the reliability of the data it contains. They can also assign *de facto* a high level of reliability to a primary document when a primary source may not be authentic, contain errors, exaggerations, biases, etc. Moreover, none has truly implemented an interdisciplinary approach by starting as close as possible to the method of historical critical analysis.

This paper aims to introduce and discuss a method to assess the quality of historical documents. This interdisciplinary work has required significant skill sharing between historians, engineers and mathematicians. It was first developed within the framework of historical research on storms, coastal flooding and skew surge in France (Athimon et al., 2021). However, the issues surrounding the assessment of the reliability of historical sources and data extend well beyond research into past meteo-marine events. For this reason, the methodology has been designed and created to be used for different fields of application using primary and secondary historical sources.

The paper is structured as follow: first we explain the methodology we created to evaluate a historical document quality's (HDQM). Secondly, we present the results through the comparison of the four expert systems developed. Finally, we discuss the historical document quality method (HDQM), its interests, its strengths, its limitations, and the results obtained by the different expert systems.

METHODOLOGY TO ASSESS THE HISTORICAL DOCUMENTS AND DATA QUALITY

Starting with the historical critical methodology

As suggested earlier, researches using historical documents and data often do not consider the historical critical analysis method. While using historical sources, it should however be the starting point. This method is specific to the history field and has been developed and enhanced by historians over decades since the end of the 19th century (Charland, 1948; Cellier and Cocaud, 2001; Halkin, 1951; Langlois and Seignobos, 1898; Le Goff and Nora, 1974; Lemercier and Zalc, 2008; Veyne, 1971).

To reduce the risk of misunderstanding, over or under interpretation, the historical critical approach requires strong scientific rigor. This approach aims to 1) to assure the authenticity of historical sources by ensuring its origin and 2) to define the reliability of data by making certain they do not contain errors of date, place, event, etc. The historical critical analysis method then works in two stages:

At first, historians use the "external criticism" or "authenticity criticism" which amounts to the validation process and allows to identify the authenticity of each historical source. The appearance (paper / parchment / papyrus, handwritten / printed / typescript, etc), the language used, the physical state (good state / rip / mold / scorch / deletion), the type (primary or secondary source) and nature (letter / parish register / engineer report / newspaper, etc) of the document are evaluated to ensure that the source is not falsified or a counterfeit. The historical environment of the document in terms of social, political, economic and scientific context is also questioned. Authenticity is therefore essential to establish, but it is not enough.

Secondly, historians use the "internal criticism", also known as "value criticism", which makes it possible to assess the intrinsic quality of a historical document. The internal criticism examines the consistency of the text, questions its compatibility with what we know about the event by documentary cross-checking and confrontation of various testimonies. This stage analyses the author, and seeks to determine what he meant, what he refused to say, what he said despite him. The date, the context, the motivation of the production of the document, etc. are also taken into account. It aims to establish the extent to which the data reflects the observations of the time and also makes it possible to give an opinion on the reliability of the data contained within this historical document.

At the end, the critical analysis of historical sources and their content should implicitly answer the five main questions (and their sub-questions):

- What? (type, nature, language, physical appearance, contents, etc. of the document)
- Who? (information about the author: name, birth and death date, job, skills/expertise, social origins, place he/she lived, contemporary or not, eyewitness or not, originality of the story or not, etc.)
- When? (date of document production, date of the events, consistency of the date with the content of the historical source and with the historical context…)
- Where? (location of production of the document, site in which the events occurred, consistency between the location of the production and the content, circumstances of the production, location of conservation of the historical document, etc.)
- Why? (motivations, intentions and interests of the author, reasons to produce the document, sponsorship or not…)

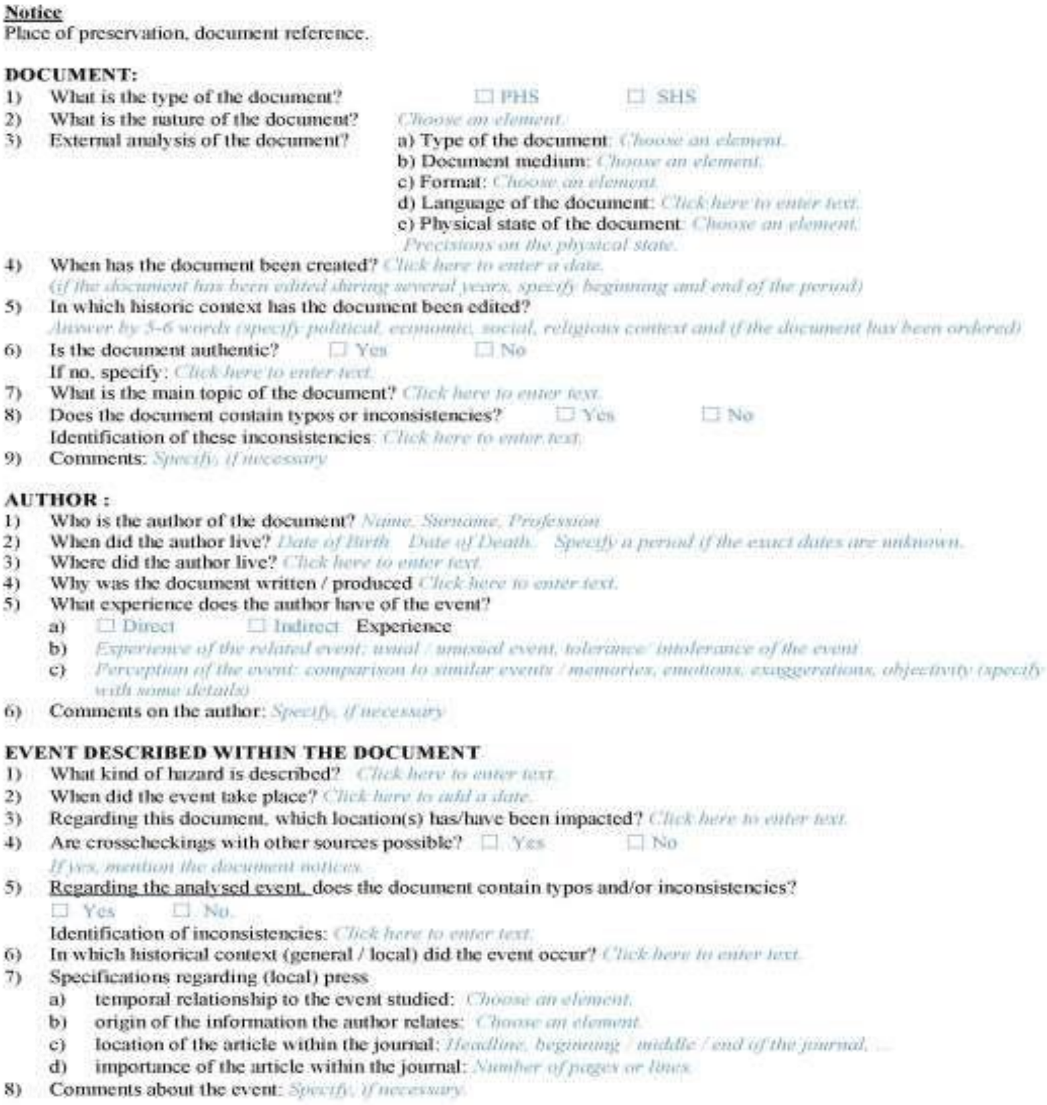
Based on the historical critical method we developed a method to evaluate historical documents quality. As we are eager to stick to the historical critical method as closely as possible, the aim is to identify the criteria on which historians rely on to define a level of reliability of historical sources and data they contain.

Operation of the historical document quality method (HDQM)

The HDQM combines history and expert systems based on mathematics. It works in three steps.

Step 1: Historical critical analysis

A complete and precise critical analysis for each primary and secondary historical source studied must be written by a historian. This analysis is based on 20 questions and three open comments organized into three sections: Document, Author, Event(s) (figure 1). This critical analysis thus approaches the external and the internal criticism of a historical source and allows to answer the key questions of what, who, when, where and why. It precisely analyses the historical document to be evaluated, without neglecting the uniqueness of each source. The historical critical analysis is supplemented by a "system of filiation" (figure 2), which allows to highlight the relationships that may exist between primary and secondary historical sources (copy, inspiration, etc.), as well as the identification of well-known sources used in the literature or novel historical sources. The system of filiation must be dated, updated and completed as documentary discoveries are made (Fradet, 2016). On figure 2, a red square indicates a bibliographic reference or a secondary historical source for which no parentage could be established. The success of steps 2 and 3 depends on the quality of the work done in step 1.



Notice
Place of preservation, document reference.

DOCUMENT:

1) What is the type of the document? ☐ PHS ☐ SHS
2) What is the nature of the document? *Choose an element.*
3) External analysis of the document? a) Type of the document: *Choose an element.*
b) Document medium: *Choose an element.*
c) Format: *Choose an element.*
d) Language of the document: *Click here to enter text.*
e) Physical state of the document: *Choose an element.*
Precisions on the physical state.
4) When has the document been created? *Click here to enter a date.*
(if the document has been edited during several years, specify beginning and end of the period)
5) In which historic context has the document been edited?
Answer by 5-6 words (specify political, economic, social, religious context and if the document has been ordered)
6) Is the document authentic? ☐ Yes ☐ No
If no, specify: *Click here to enter text.*
7) What is the main topic of the document? *Click here to enter text.*
8) Does the document contain typos or inconsistencies? ☐ Yes ☐ No
Identification of these inconsistencies: *Click here to enter text.*
9) Comments: *Specify, if necessary*

AUTHOR :

1) Who is the author of the document? *Name, Surname, Profession*
2) When did the author live? *Date of Birth Date of Death. Specify a period if the exact dates are unknown.*
3) Where did the author live? *Click here to enter text.*
4) Why was the document written / produced *Click here to enter text.*
5) What experience does the author have of the event?
a) ☐ Direct ☐ Indirect Experience
b) *Experience of the related event: usual / unusual event, tolerance / intolerance of the event*
c) *Perception of the event: comparison to similar events / memories, emotions, exaggerations, objectivity (specify with some details)*
6) Comments on the author: *Specify, if necessary*

EVENT DESCRIBED WITHIN THE DOCUMENT

1) What kind of hazard is described? *Click here to enter text.*
2) When did the event take place? *Click here to add a date.*
3) Regarding this document, which location(s) has/have been impacted? *Click here to enter text.*
4) Are crosscheckings with other sources possible? ☐ Yes ☐ No
If yes, mention the document notices.
5) Regarding the analysed event, does the document contain typos and/or inconsistencies?
☐ Yes ☐ No.
Identification of inconsistencies: *Click here to enter text.*
6) In which historical context (general / local) did the event occur? *Click here to enter text.*
7) Specifications regarding (local) press
a) temporal relationship to the event studied: *Choose an element.*
b) origin of the information the author relates: *Choose an element.*
c) location of the article within the journal: *Headline, beginning / middle / end of the journal, ...*
d) importance of the article within the journal: *Number of pages or lines.*
8) Comments about the event: *Specify, if necessary.*

Figure 1: Analytical and critical notice to be written for each historical source to be evaluated. It contains 20 questions and 3 open comments organized into 3 sections: Document, Author, Event(s).

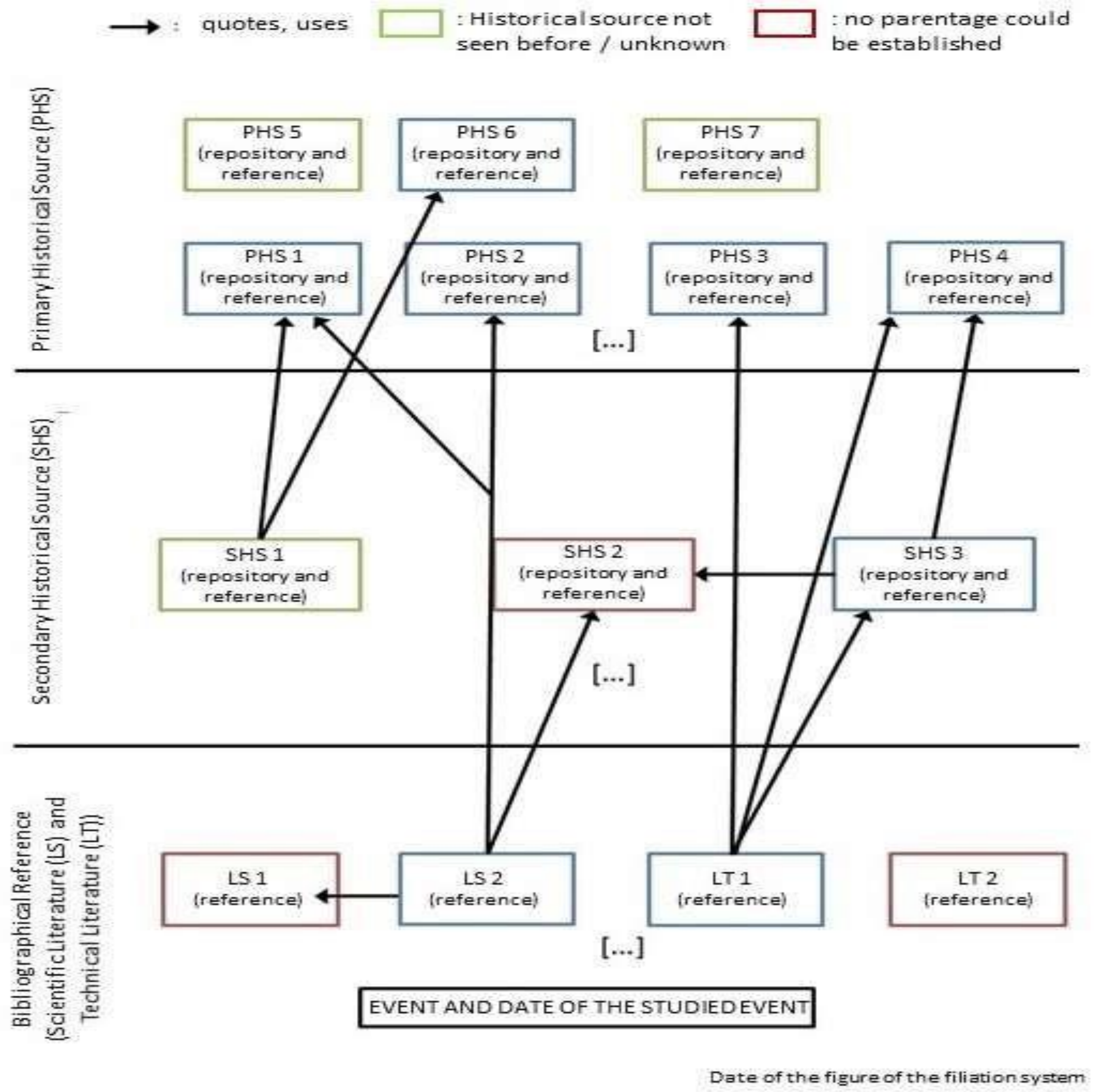

: quotes, uses
: Historical source not seen before / unknown
: no parentage could be established
Primary Historical Source (PHS)
PHS 5 (repository and reference)
PHS 6 (repository and reference)
PHS 7 (repository and reference)
PHS 1 (repository and reference)
PHS 2 (repository and reference)
PHS 3 (repository and reference)
PHS 4 (repository and reference)
[...]
Secondary Historical Source (SHS)
SHS 1 (repository and reference)
SHS 2 (repository and reference)
SHS 3 (repository and reference)
[...]
Bibliographical Reference (Scientific Literature (LS) and Technical Literature (LT))
LS 1 (reference)
LS 2 (reference)
LT 1 (reference)
LT 2 (reference)
[...]
EVENT AND DATE OF THE STUDIED EVENT
Date of the figure of the filiation system

Figure 2: Generic representation of the system of filiation of primary and secondary historical sources, and bibliographical references on an event (inspired by Fradet, 2016, p. 593).

Step 2: Decision tree and evaluation of criteria

Based on the historical critical analysis written by a historian (step 1), a decision tree has been developed. Four criteria that can systematically be applied to any content and specificity of all historical sources, primary or secondary, have been identified. According to historians and their historical critical methodology, these four criteria are necessary to establish the level of reliability of a historical document and its content. These four criteria are:

1) The type of the document (C1). In C1, the decision tree first questions the type of the document, whether it is a primary or a secondary historical source. It then ensures its authenticity. Finally, it defines if the historical source to be evaluated is an original (i.e. produced by its original author), if it is a complete copy or publication made by someone else than the author, an incomplete copy or publication made by someone else than the author;
2) The author's link to the testimony of an event (C2). In C2, the decision tree first questions whether the author is contemporary or not of the event he/she is writing about. It then examines the relationship the author has regarding the event (eyewitness, indirect witness, drawing on primary or secondary historical sources, untraceable origins of the information, etc);
3) The cross-checking (C3). In C3, the decision tree first defines whether crosschecks with other historical documents have been made. It then questions the type of documents on which the

crosschecks are based (primary or secondary historical documents) and their number. Finally, it examines whether the crosschecks confirm / refute the content of the historical source to be evaluated;

4) The consistency of the source contents (C4). In C4, the decision tree questions the consistency of the content of the historical document to be evaluated, and the importance of the errors identified. The inconsistencies can be related to contextual anachronisms, errors in the dating and/or localization of the event, inventions, nonsense, etc. They can be identified by the critical analysis of the document (step 1), as well as by crosschecking with other historical sources (step 2, C3).

Each criterion is a branch of the decision tree and consists of closed questions (yes/no) to ensure a rigorous approach (figures 3, 4, 5 and 6). In addition, feedback loops have been introduced to complement the strength of the method. After following each branch, the user is left with four marks, one per criterion, which gives 1372 possible combinations. It is important to note that based on the critical analysis established in step 1, it is possible for everyone to give a score for each criterion in step 2.

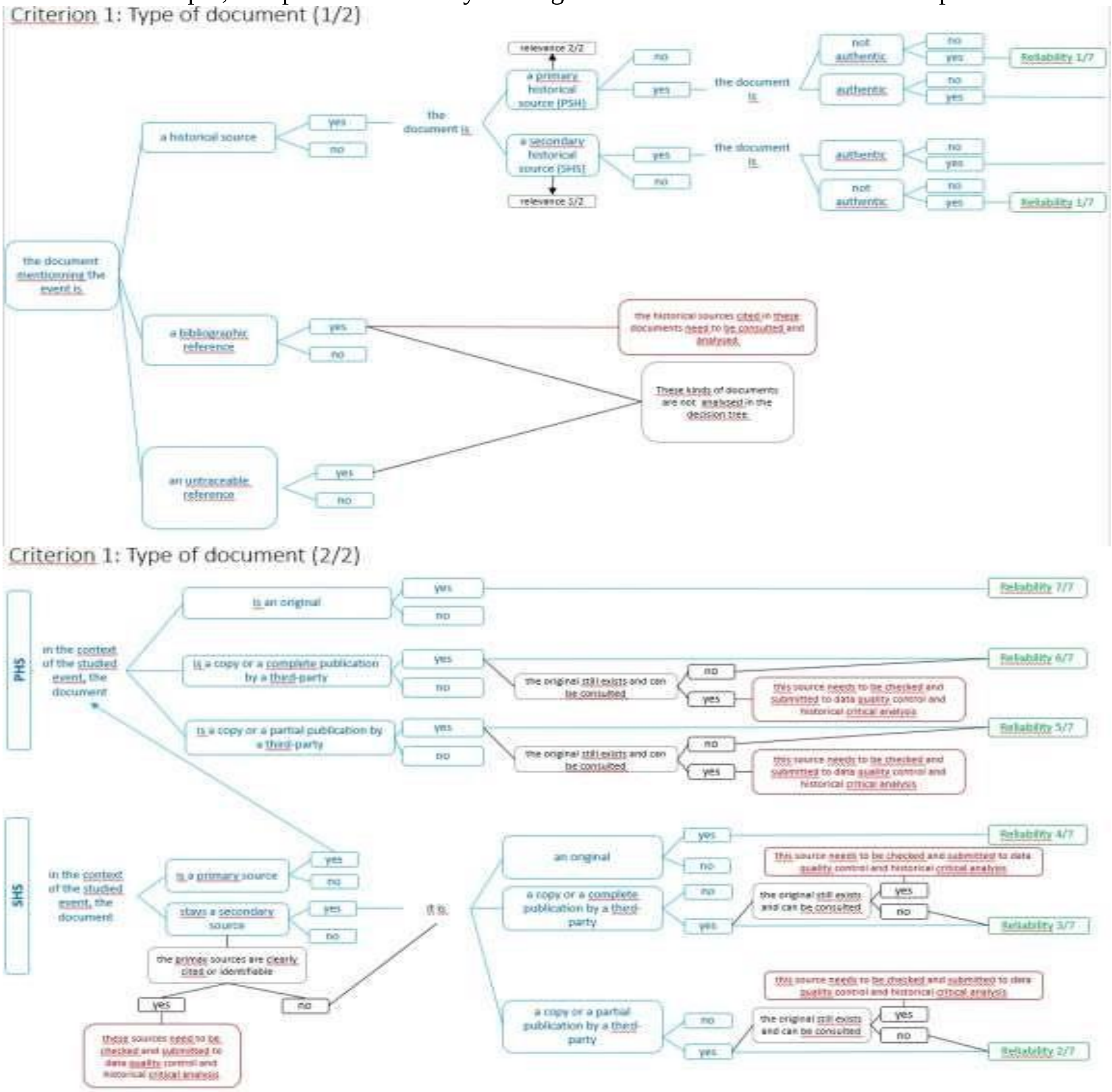

Criterion 1: Type of document (1/2)
the document mentionning the event is
a historical source
yes
no
the document is
a primary historical source (PSH)
relevance 2/2
a secondary historical source (SHS)
relevance 5/2
the document is
not authentic
authentic
Reliability 1/7
a bibliographic reference
the historical sources cited in these documents need to be consulted and analysed.
These kinds of documents are not analysed in the decision tree.
an untraceable reference
Criterion 1: Type of document (2/2)
PHS
in the context of the studied event, the document
is an original
Reliability 7/7
is a copy or a complete publication by a third-party
the original still exists and can be consulted
Reliability 6/7
this source needs to be checked and submitted to data quality control and historical critical analysis
is a copy or a partial publication by a third-party
Reliability 5/7
SHS
is a primary source
stays a secondary source
the primay sources are clearly cited or identifiable
these sources need to be checked and submitted to data quality control and historical critical analysis
it is
an original
Reliability 4/7
a copy or a complete publication by a third-party
Reliability 3/7
a copy or a partial publication by a third-party
Reliability 2/7

Figure 3: Schematic representation of the tree structure of criterion 1 on the type of documents analyzed (page 1 and 2).

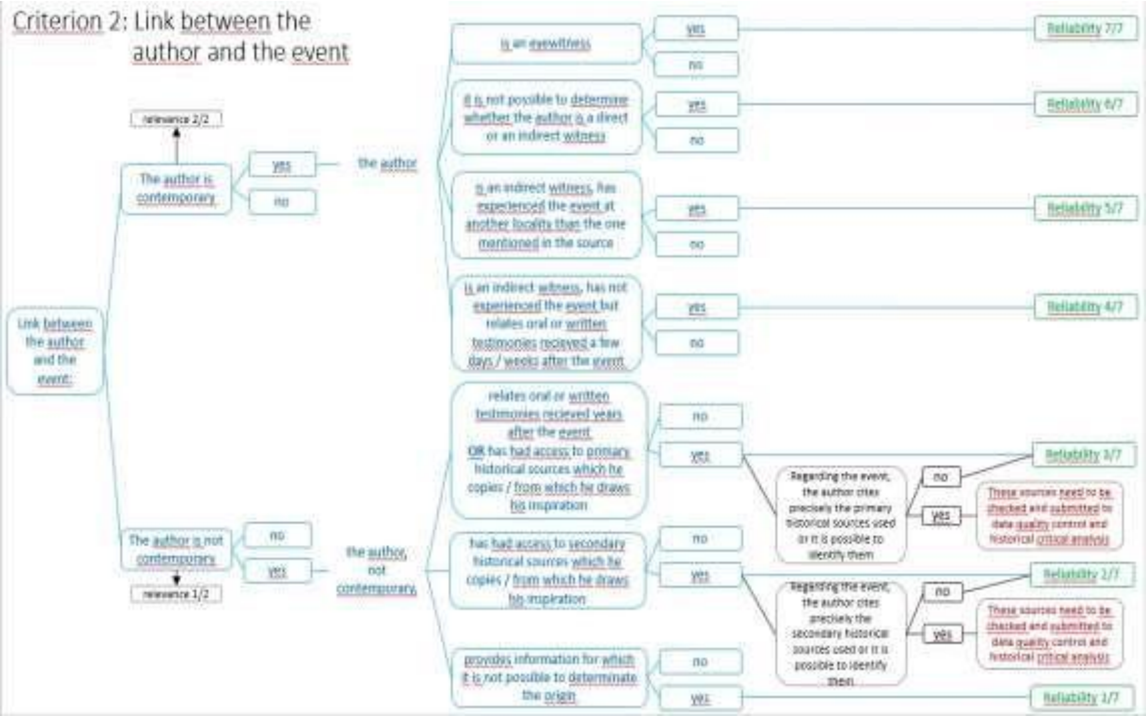
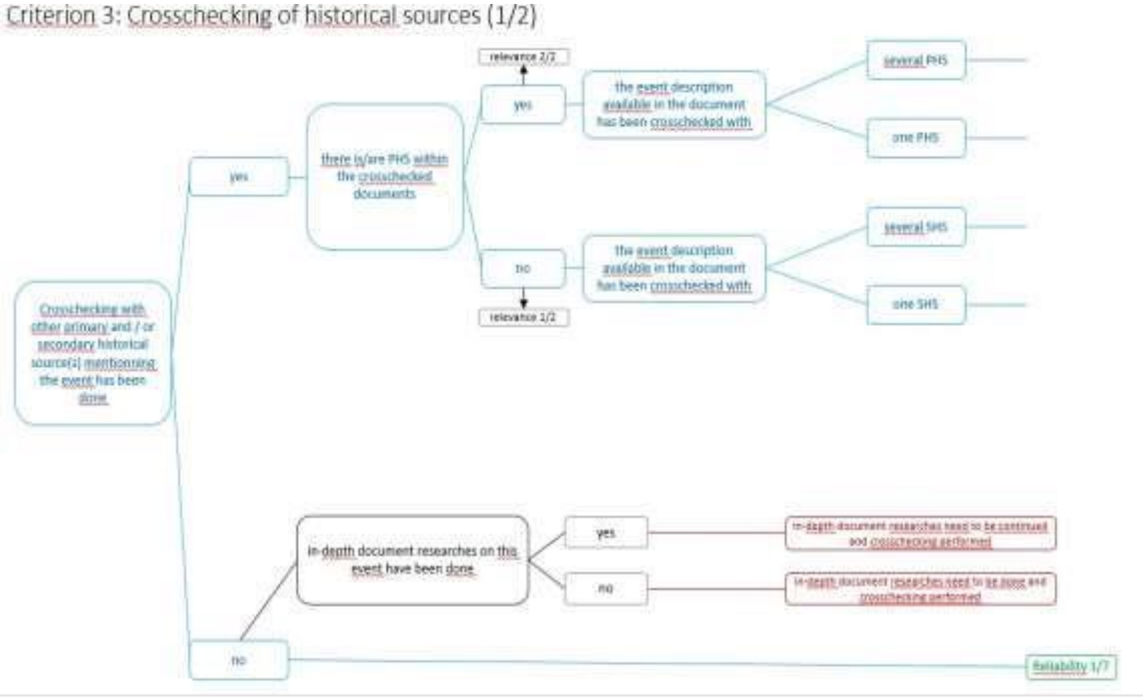

Criterion 2: Link between the author and the event
Link between the author and the event:
relevance 2/2
The author is contemporary
yes
no
the author
is an eyewitness
yes
no
Reliability 7/7
it is not possible to determine whether the author is a direct or an indirect witness
yes
no
Reliability 6/7
is an indirect witness, has experienced the event at another locality than the one mentioned in the source
yes
no
Reliability 5/7
is an indirect witness, has not experienced the event but relates oral or written testimonies recieved a few days / weeks after the event
yes
no
Reliability 4/7
The author is not contemporary
no
yes
relevance 1/2
the author, not contemporary,
relates oral or written testimonies recieved years after the event OR has had access to primary historical sources which he copies / from which he draws his inspiration
no
yes
Reliability 3/7
Regarding the event, the author cites precisely the primary historical sources used or it is possible to identify them
no
yes
These sources need to be checked and submitted to data quality control and historical critical analysis
has had access to secondary historical sources which he copies / from which he draws his inspiration
no
yes
Reliability 2/7
Regarding the event, the author cites precisely the secondary historical sources used or it is possible to identify them
no
yes
These sources need to be checked and submitted to data quality control and historical critical analysis
provides information for which it is not possible to determinate the origin
no
yes
Reliability 1/7

Figure 4: Schematic representation of the tree structure of criterion 2 on the link between the author of the historical sources analyzed and the event.


Criterion 3: Crosschecking of historical sources (1/2)
Crosschecking with other primary and / or secondary historical source(s) mentionning the event has been done
yes
there is/are PHS within the crosschecked documents
yes
relevance 2/2
the event description available in the document has been crosschecked with
several PHS
one PHS
no
relevance 1/2
the event description available in the document has been crosschecked with
several SHS
one SHS
no
in-depth document researches on this event have been done
yes
in-depth document researches need to be continued and crosschecking performed
no
in-depth document researches need to be done and crosschecking performed
Reliability 1/7

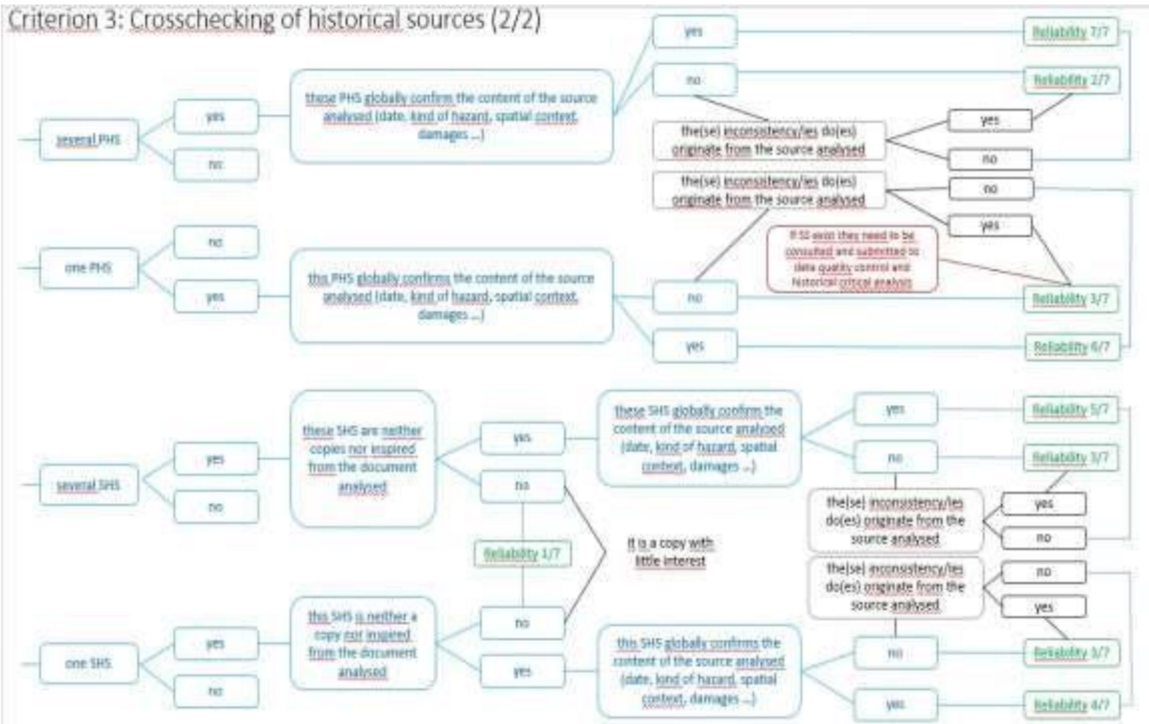

Criterion 3: Crosschecking of historical sources (2/2)
several PHS
yes
no
these PHS globally confirm the content of the source analysed (date, kind of hazard, spatial context, damages ...)
yes
no
Reliability 7/7
Reliability 2/7
the(se) inconsistency/ies do(es) originate from the source analysed
yes
no
the(se) inconsistency/ies do(es) originate from the source analysed
no
yes
If SS exist they need to be consulted and submitted to data quality control and historical critical analysis
one PHS
no
yes
this PHS globally confirms the content of the source analysed (date, kind of hazard, spatial context, damages ...)
no
yes
Reliability 3/7
Reliability 6/7
several SHS
yes
no
these SHS are neither copies nor inspired from the document analysed
yes
no
these SHS globally confirm the content of the source analysed (date, kind of hazard, spatial context, damages ...)
yes
no
Reliability 5/7
Reliability 3/7
Reliability 1/7
It is a copy with little interest
the(se) inconsistency/ies do(es) originate from the source analysed
yes
no
the(se) inconsistency/ies do(es) originate from the source analysed
no
yes
one SHS
yes
no
this SHS is neither a copy nor inspired from the document analysed
no
yes
this SHS globally confirms the content of the source analysed (date, kind of hazard, spatial context, damages ...)
no
yes
Reliability 3/7
Reliability 4/7

Figure 5: Schematic representation of the tree structure of criterion 3 on the crosschecking of historical sources (page 1 and 2).

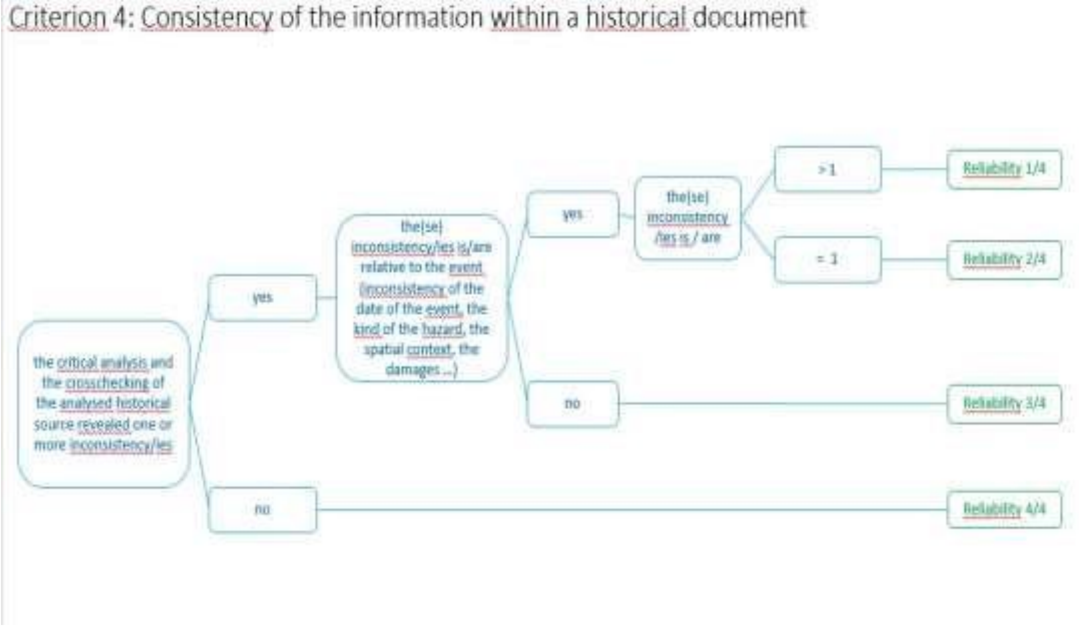

Criterion 4: Consistency of the information within a historical document
the critical analysis and the crosschecking of the analysed historical source revealed one or more inconsistency/ies
yes
no
the(se) inconsistency/ies is/are relative to the event (inconsistency of the date of the event, the kind of the hazard, the spatial context, the damages ...)
yes
no
the(se) inconsistency /ies is / are
>1
= 1
Reliability 1/4
Reliability 2/4
Reliability 3/4
Reliability 4/4

Figure 6: Schematic representation of the tree structure of criterion 4 on the consistency of the information and data within a historical document.

Step 3: Assigning a final score

To be able to compare different sources, a final score is assigned to each historical source based on the four marks obtained in step 2. A weight was assigned to each mark in each criterion (tables 1, 2, 3 and 4).

CR TERION 1 (C1)						
1	2	3	4	5	6	7

Table 1: Table of weightings (weight functions) applied for criterion 1 (C1) in the first and the second expert system.

Light grey: low weight; Grey: mean weight; Dark grey: high weight; white: weight changes depending on the scores obtained in the other criteria.

Indeed, when the score of 1 (white) is given in C1, the weighting changes from low to mean depending on the scores obtained in the other criteria. This is due to their nominal characteristics. So, if:

- **C1 = 1 (non-authentic historical source), but C3 ≥ 4 (the cross-checks corroborate the data within the document to be evaluated) and C4 ≥ 3 (no inconsistency), then the weight applied on the score of 1 will be mean weight.**
- **C1 = 1 (non-authentic historical source), but C3 ≤ 3 (no cross-check or the cross-checks did not confirm the content of the document to be evaluated) and C4 ≤ 2 (inconsistency), then the weight applied on the score of 1 will be low weight.**

CR TERION 2 (C2)						
1	2	3	4	5	6	7

Table 2: Table of weightings (weight functions) applied for the criterion 2 (C2) in the first and the second expert system.

Light grey: low weight; Grey: mean weight; Dark grey: high weight.

CR TERION 3 (C3)						
1	2	3	4	5	6	7

Table 3: Table of weightings (weight functions) applied for the criterion 3 (C3) in the first and the second expert system.

Light grey: low weight; Grey: mean weight; Dark grey: high weight; white: weight changes depending on the scores obtained in the other criteria.

Indeed, when the score of 1 (white) is given in C3, the weighting changes from low to mean depending on the scores obtained in the other criteria. This is due to their nominal characteristics. So, if:

- **C3 = 1 (no cross-check), but C2 ≥ 4 (contemporary author) and C4 ≥ 3 (no inconsistency), so the weight applied on the score of 1 will be mean weight.**
- **C3 = 1 (no cross-check), but C2 ≤ 3 (non-contemporary author) and C4 ≤ 2 (inconsistency), so the weight applied on the score of 1 will be low weight.**

CRITERION 4 (C4)			
1	2	3	4

Table 4: Table of weightings (weight functions) applied for the criterion 4 (C4) in the first and the second expert system.

Light grey: low weight; Grey: mean weight; Dark grey: high weight.

Weightings have been assigned according to the historical critical methodology and common critical judgement. Following this methodology, a primary historical source is more reliable than a secondary historical source, which draws on primary sources and risks distorting reality. Also, an original document (i.e. any document produced by its original author, whether it is a manuscript, a print, a publication by the author, a typescript…) is more reliable than a copy or a publication made by someone else than the original author as copy errors, modifications, deletions of the text could have occurred. If an author is contemporary with the event – even being in the emotion –, his testimony is considered more reliable than a non-contemporary author, who lived decades or centuries later and can imagine, may confuse date or place. While being contemporary with the event, an eyewitness is considered more reliable than someone who writes about others' experiences. A historical source crosschecked with primary sources which confirms its content is more reliable than one whose content is confirmed by only one secondary historical source. When crosschecking with several primary sources that are coherent with each other, but which do not confirm the content of the source to be evaluated the level of reliability is low. The weighting of mark 1 of C3 is a special case. In fact, an absence of crosscheck can have various causes such as an incomplete research, the loss of archives, an invention by the author… Therefore, C2 and C4 are used to assign a value to mark 1 of C3. Similarly, for the weighting of mark 1 of C1, as an unauthentic historical document can still contain reliable information.

Based on these marks and weightings, four different expert systems (ES) have been established and tested to estimate the final relevance of a document. This part of the method combines historical and mathematical approaches.

Expert system 1 (ES1) is a rating proposed by historians based on the obtained marks and associated weightings (tables 1, 2, 3 and 4), and on their expertise, knowledge and experience of historical documents. These final marks, between 0 and 100, are considered as consensus level. They are the target value to reach as closely as possible by the other ES, so the other systems, which give predicted scores (L_{pred}), will be compared to this system hereafter (figure 7).

Expert system 2 (ES2) is a linear function. At first, each mark is multiplied with a factor depending on its weight w available after step 2 (tables 1, 2, 3 and 4):

- Low weight: multiplication by 1
- Mean weight: multiplication by 2 - High weight: multiplication by 3.

In addition, C4 is awarded double weight as the consistency criteria is of highest importance to define the reliability of a historical document and the data it contains. The final mark is then a sum of all values.

$$L_{PredES2} = C1.w_{C1} + C2.w_{C2} + C3.w_{C3} + 2C4.w_{C4}$$

To be able to compare these ES, absolute marks have been translated into relative marks, i.e. percentage of reliability. The lowest mark possible is 5 which equals 0% of reliability and 100% of reliability corresponds to 87, the highest possible mark.

$$L_{PredES2\%} = \frac{L_{PredES2} - 5}{87 - 5} * 100$$

Expert system 3 (ES3) is based on algorithms to define the weightings and find the final mark. Two correlations (ES3a and ES3b) have been tested to rate the final predicted score of the historical document evaluated according to the four marks of the criteria. These marks were first normalized by their maximum possible score, so for criteria 1, 2 and 3, the marks are divided by 7 and for criterion 4 it is divided by 4 (tables 1, 2, 3 and 4 ; figures 3, 4, 5 and 6). For the first correlation (ES3a) a linear relation is proposed:

$$L_{PredES3a} = \frac{C1}{7}.a_1 + \frac{C2}{7}.a_2 + \frac{C3}{7}.a_3 + \frac{C4}{4}.a_4 + b$$

The values a_{Ci} and b found are given in table 5. They represent the coefficients of the correlation.

b	aC1	aC2	aC3	aC4
-28,3	27,6	27,2	31,4	35,9

Table 5: Weightings of each criterion defined through the first correlation of the expert system 3. Note that C3 and C4, which are associated with a higher weight after the optimization, are therefore considered a bit more important in this case.

The second correlation of the expert system 3 (ES3b) rests on the weightings proposed by historians (tables 1, 2, 3 and 4) and used on the ES1 and ES2. It is linked to the finding that heavy weights are associated with good marks. A linear relation is then sought:

$$L_{PredES3b} = \left(\frac{C1}{7}\right)^{\alpha}.a_1 + \left(\frac{C2}{7}\right)^{\alpha}.a_2 + \left(\frac{C3}{7}\right)^{\alpha}.a_3 + \left(\frac{C4}{4}\right)^{\alpha}.a_4 + b$$

On this equation, α is a parameter to optimize. α is equal to 2,2. It allows to better consider the highest marks of each criterion. The values a_{Ci} and b found are given in table 6. They represent the coefficients of the correlation to be optimized.

b	aC1	aC2	aC3	aC4
1,32	27,5	19,2	26,8	26,8

Table 6: Weightings of each criterion defined through the second correlation of the expert system 3. Note that C1, which is associated with a higher weight during the optimization, is therefore considered a bit more important in this case.

Expert system 4 (ES4) is a neural network. Neural networks learn by processing examples, each of which contains a known “entry” as well as a known “result” forming weighted associations whose weights are stored in the data structure of the network itself. The learning performed by a neural network from a given example is usually done by determining the difference, i.e the error between the output of the network and the target value. The network then automatically adjusts its weighted associations according to a learning rule using the value of the error. When the network is useful, successive adjustments will cause it to produce an output that is increasingly similar to the target value. After a sufficient number of these adjustments, the training can be stopped according to certain criteria. Neural networks are conventionally used for classification problems, which is the case here. Several types of neural networks have been tested to have the closest results to those of ES1. Numerous sensitivity tests were carried out on the parameters, namely the number of layers, the number of neurons as well as the activation functions. Finally, a network with 4 inputs, then two layers of 21 neurons each and one output showed a very good prediction capacity (which means a very good correlation with ES1). The activation function is the reLU (Rectified Linear Unit) function:

$f(x) = \max(0, x)$

It allows or not to transmit the information if the stimulation threshold is reached. Concretely, its role will be to decide whether or not to activate a neuron response. These are the most popular functions nowadays. They allow a faster training compared to the sigmoid and tanh functions (Glorot et al., 2011).

Moreover, no matter the ES used, the application of this method in three steps gives a final score that matches the credit of the historical source and of the data it contains. Final marks given by all the ES are relative marks and are therefore considered as percentage of reliability. This scale has then been divided in five groups:

[0-20[%: not reliable at all;

[20-40[%: unreliable;

[40-60[%: moderately reliable; [60-80[%: reliable;

[80-100]%: very reliable.

RESULTS: COMPARISON OF THE EXPERT SYSTEMS

To test the three steps of the HDQM, we created 147 fictive historical sources defined by their criteria values, which represent a large panel of possible characteristics. The methodology has also been tested and applied on real historical documents (Athimon et al., 2021).

As aforementioned, ES1 is a rating proposed by historians. The final consensus marks of ES1 are the target, so results of ES2, ES3a, ES3b and ES4 will be compared to this system (figure 7).

Comparing ES1 and ES2, it appears that the minimum and maximum differences between the consensus score of the ES1 and the predicted score of the ES2 are -9 and +10. 79% of the difference between the consensus and the predicted scores are smaller than 5, which is very consistent and results in a very good correlation: coefficient of determination $R^2 = 0.975$ (figure 7, on the top corner left).

Comparing ES1 and ES3a, it appears that the minimum and maximum differences between the consensus score of the ES1 and the predicted score of the ES3a are -3 and +20. The model predicts the results well as the coefficient of determination $R^2 = 0.93$

(figure 7, on the top corner right). 56% of the difference between the consensus and the predicted scores are smaller than 5. Nevertheless, the model is less efficient than ES2.

In the second case of ES3 (ES3b), differences with ES1 range from -14 and +14. 71% of the difference between the consensus and the predicted scores are smaller than 5. The coefficient of determination is even better, as $R^2 = 0.96$ (figure 7, at the bottom left).

Comparing ES1 and ES4, it appears that the minimum and maximum differences between the consensus score of the ES1 and the predicted score of the ES2 are -12 and +13. 77% of the difference between the consensus and the predicted scores are smaller than 5, which is very consistent and results in a very good correlation: coefficient of determination $R^2 = 0.968$ (figure 7, at the bottom right).

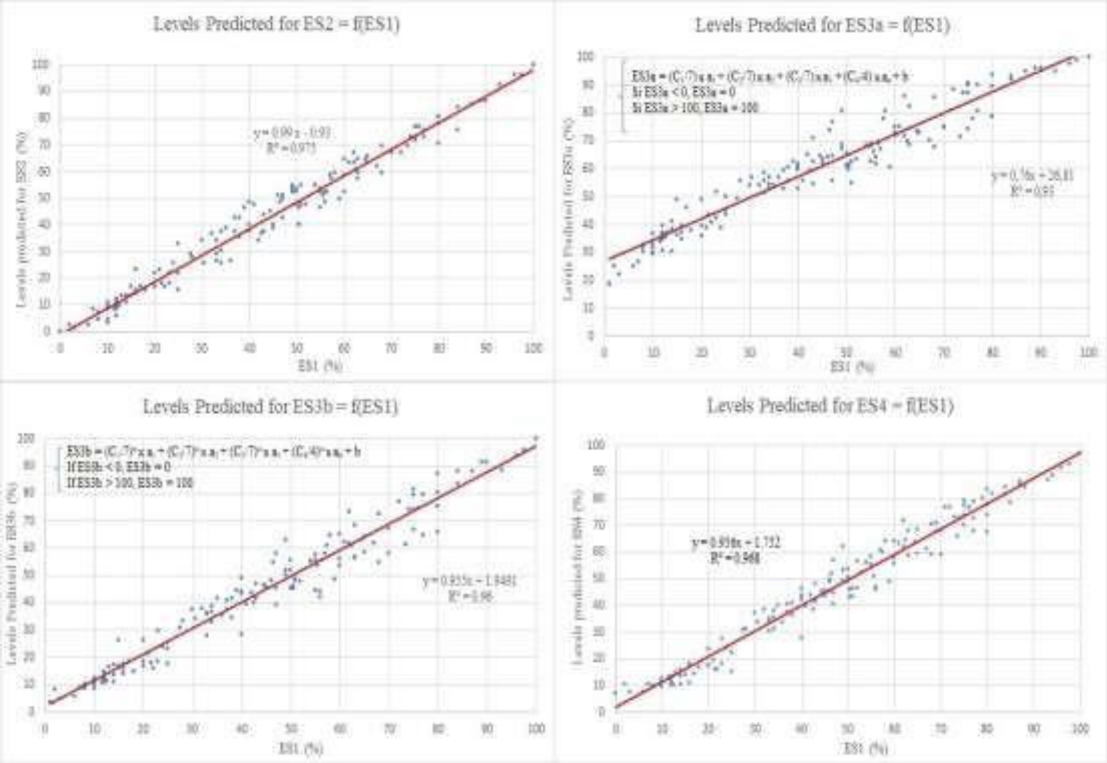


Figure 7: Correlation between the predicted scores of the ES2 (on the top corner left), ES3a (on the top corner right), ES3b (at the bottom left), ES4 (at the bottom right) and the consensus scores (ES1).

DISCUSSION

The HDQM has been developed for a wide range of studies using historical documents. The main advantage of the HDQM is the strictness of its structure. In particular, the historical critical analysis (step 1) is based on elementary questions and comments and is performed by historians, who are experts of historical sources. The decision tree (step 2) is based on yes/no questions and the final mark is established using an expert system (step 3). So far, no choice has been made concerning the expert systems.

Although the step 1 pays attention to a lot of details and document specificities (e. g. the historical context, the authors position / job as priest, engineer, journalist, scholar, unknown, etc., or the temporal distance between the experience of the event by an eyewitness and its writing since the memory is person-dependent), the assessment of singularity of each historical source is a challenging task (step 2). However, the selected criteria need to be standardized to define the quality of one document and further compare different historical sources. The consideration into a decision tree of other criteria such as the historical context or the authors position / job has been tested but it was questionable, and the definition of related reliability levels was highly subjective. Moreover, the decision tree offers 1372 possible combinations, which makes this criticism irrelevant.

So far, no crosschecks neither with scientific literature nor other kind of data (e.g: geosciences, simulation data) are made. There are several reasons for it: 1) on the framework of extreme events, crosschecks with

other kind of data are still too rare to be standardized and in other fields they may not exist; 2) defining the reliability of crosschecks with other data than historical ones is difficult due to associated residual uncertainties such as imprecise dating; 3) crosschecks with scientific publications do not say much about the reliability of a historical source. Still, these crosschecks are pointed out in step 1, both in the critical analysis written by a historian and in the
"system of filiation" (figure 2).

No feedback loops have been implemented in C4. Indeed, the consistency or inconsistency of the data contained in a historical source is identified in relation to "something". Nevertheless, feedback loops are difficult to establish in C4 and they cannot be standardized. Moreover, they become negligible since the logic of the decision tree in C4 is done by comparison of historical documents and thanks to the answers obtained in step 1 and the score of C3.

Concerning step 3, assigning a final score, the weightings proposed by historians (tables 1, 2, 3 and 4) are the result of several versions of the weightings which have been made and tested before getting to this final one. This final version is considered as functional and operational, and it overcomes the limits and problems found on the other previous versions. Here, the weightings are associated to the nominal characteristics of the branch of each criterion of the decision tree and to their reliability level. It follows the logics and the way historians proceed. This explains why the weighting changes from low to mean when the score of 1 is given in C1 and C3 (tables 1 and 3), as same as this explains why in C1 the most important weighting is granted to mark 7/7 and 4/7: an original document, either primary or secondary, carries more weight than a copy or a publication by someone else than the author.

Another interesting point to discuss are the ES, how they work and what they shed light on the expertise of historians:

In ES1 and ES2, the importance is not given to the criteria, but to the score of each criterion. The four scores rated in step 2 allow to estimate the level of reliability of the historical source to be evaluated. The higher these scores are, the more reliable is the document. Following the historian's expertise, C3 (crosschecks between historical documents) and C4 (consistency of the data contained) are crucial. C1 (authenticity or not, primary or secondary historical source, original document or not) is also essential. C2 appears a little less important. The coefficient of determination between the ES2 and the ES1 is 0.975 (figure 7, on the top corner left). The ES2 predicts so well because it relies directly on the weightings established by historians in tables 1, 2, 3 and 4 and used to set the final consensus scores of the ES1. Indeed, historians being the specialists of historical sources, it is legitimate to rely on the weightings they fixed to define the importance to be given to the mark of a criterion while assigning the final score.

ES3 gives more importance to the criterion itself, then to the score obtained in each criterion. In ES3a, C4 is the most important, followed by C3, while C1 and C2 are of equal importance and have a lower weight (table 5). In ES3b, the final score assigned mainly rests on C1, and then C3 and C4 (table 6). Again, C2 has the lowest weight. ES3b has a better coefficient of determination (0.96, figure 7 at the bottom left) than ES3a (0.93, figure 7 on the top corner right). This is explained by the fact that the second correlation (ES3b) uses the power function for each score C_i of each criterion (tables 1, 2, 3 and 4). In doing so, it makes a relation between the highest score and the most important weight, which brings it closer to the logic used in ES1 and ES2.

Regarding the ES4, its performances are very close to ES2. The correlation coefficient obtained, close to 0.97 (figure 7 at the bottom right), indicates that the model predicts very well, like ES2. Further work on neural networks would probably make it possible to further improve this model. Nevertheless, for a set of first tests, the results obtained seem satisfying.

The models developed to assign a final score to historical documents have all a coefficient of determination close to 1. These results are very consistent, and we assume that they are predictive and allow a good estimation of the final mark. So far, no choice has been made to support one specific ES, as it seemed more relevant to keep all of them to continue their use on more and more historical sources.

CONCLUSION

Historical sources and data are more and more taken into account in various studies, in particular those relating on extreme events, natural hazards or environmental sciences. When dealing with these kinds of documents, it is very important to certify the relevance of historical sources and the data they contain. For this reason, the HDQM has been developed.

The HDQM is an operational and functional method that allows the evaluation of the quality of historical documents. This method works in three steps: 1) historical critical analysis, 2) decision tree and evaluation of four criteria, 3) assigning a final score. Except for the step 1 which must imperatively be performed by a historian, otherwise this method has been developed to be used by everyone. It is a user-friendly and easy learning method and the results obtained in this study are very convincing.

Therefore, the HDQM allows the detection of inconsistencies (e.g. confusion, invention, error in date, location, context, etc.) present in historical sources. An interesting perspective would be to try to quantify the uncertainties linked (exaggeration, attenuation, interpretation, etc.) of each historical document and to take this quantification into account in the final score. A major advantage of this method is that it can be applied to a wide variety of research fields and case studies using historical sources.

ACKNOWLEDGMENT

Authors wish to acknowledge Jean-Paul Bahman for his work on the neural network (ES4) during his internship at EDF R&D. Authors also wish to thank the other members of the Working Group “Historical Storms and Floodings”. This research has been funded through EDF R&D research project MOISE.

REFERENCES

Abadie, S., Beauvivre, M., Egurrola, E., Bouisset, C., Degrémont, I. and Arnoux, F. (2018) *A Database of Recent Historical Storm Impact on the French Basque Coast*, in Journal of Coastal Research, 85 721-725.

Albini, P., Musson, R., Gomez Capera, A., Locati, M., Rovida, A., Stucchi, M. and Vigano, D. (2013) Global Historical Earthquake Archive and Catalogue (1000-1903), GEM Technical Report, [online].

Arnoux, F., Abadie, S., Bertin, X. and Kojadinovic, I. (2021) Coastal flooding event definition based on damages: Case study of Biarritz Grande Plage on the French Basque coast, Coastal Engineering, 166-3 103873.

Athimon, E. (2021) Tempêtes et submersions marines sur les territoires de la côte atlantique (XIV^{e}XVIII^{e} siècle), Les Indes Savantes, Paris.

Athimon, E., Giloy, N., Sauzeau, T., Andreevsky, M. and Frau, R. (2021) *Quantification of historical skew surges: challenges and methods*, in Advances in Hydroinformatics, Special Issue: SimHydro 6th International Conference, Springer (accepted).

Baart, F., Bakker, M. A. J., van Dongeren, A., den Heijer, C., van Heteren, S., Smit, M.W.J., van Koningsveld, M. and Pool, A. (2011) *Using 18th century storm-surge data from the Dutch Coast to improve the confidence in flood-risk estimates*, Natural Hazards Earth System Science, 11 2791–2801.

Benito, G., Lang, M., Barriendos, M., Llasat, C., Francès, F., Ouarda, R., Thorndycraft, V., Enzel, Y., Bardossy, A., Coeur, D. and Bobée, B. (2004) *Use of Systematic, Palaeoflood and Historical Data for the Improvement of Flood Risk Estimation. Review of Scientific Methods*, Natural Hazards, 31 623-643.

Breilh, J-F., Bertin, X., Chaumillon, E., Giloy, N. and Sauzeau, T. (2014) *How frequent is storminduced flooding in the central part of the Bay of Biscay?*, Global and Planetary Change, 122 161-175.

Bonnechere, P. (2008) Profession : historien, POM, Montréal.

Brazdil, R. and Kotyza, O. (eds) (2004) History of weather and climate in the Czech lands, vol. 6 : Strong winds, Masaryk University, Brno.

Bulteau, T., Idier, D., Lambert, J. and Garcin, M. (2015) How historical information can improve estimation and prediction of extreme coastal water levels: application to the Xynthia event at La Rochelle (France), Natural Hazards and Earth System Sciences, 15-6 1135-1147.

Camuffo, D. (1993) *An analysis of the sea surges at Venice from A.D 782 to 1990*, Theoretical and Applied Climatology, 47 1-14.

Cellier, J. and Cocaud, M. (2001) Traiter des données historiques. Méthodes statistiques / Techniques informatiques, PUR, Rennes.

Charland, T. (1948) *La critique d'authenticité*, Revue d'histoire de l'Amérique française, 1-4 483-494.

Chaumillon, E., Bertin, X., Fortunato, A. b., Bajo, M., Schneider, J-L., Dezileau, L., Walsh, J.P.,

Michelot, A., Chauveau, E., Créach, A., Hénaff, A., Sauzeau, T., Waeles, B., Gervais, B., Jan, G., Baumann, J., Breilh, J-F. and Pedreros, R. (2017) *Storm-induced marine flooding: Lessons from a multidisciplinary approach*, Earth-Science Reviews, 165 151-184.

De Kraker, A. (2006) *Flood events in the southwestern Netherlands and coastal Belgium, 1400-1953*, Hydrological Sciences, Special Issue: Historical Hydrology, 51-5 913-929.

De Vries K. and Winsemius, J.P. (1970) De Allerheiligenvloed van 1570, Miedema Pers., Leeuwarden.

Fradet, T. (2016) Vulnérabilité et perception face aux tremblements de terre en France, 1650-1850, PhD, Saint-Quentin University, Saint-Quentin.

Frau, R. (2018) Utilisation des données historiques dans l'analyse régionale des aléas marins extrêmes : la méthode FAB, PhD, Paris-Est University, Paris.

Frau, R., Andreevsky, M. and Bernardara, P. (2018) *The use of historical information for regional frequency analysis of extreme skew surge*, Natural Hazards and Earth System Science, 18 949962.

Garnier, E., Ciavola, P., Spencer, T., Ferreira, O., Ar-maroli, C. and McIvor, A. (2018) *Historical analysis of storm events:case studies in france, england, portugal and Italy*, Coastal Engineering, 134 10–23.

Giacona, F., Eckert, N. and Martin, B. (2017) *A 240-year history of avalanche risk in the Vosges Mountains basedon non-conventional (re)sources*, Natural Hazards and Earth System Science, 17 887-904.

Giloy, N., Hamdi, Y., Bardet, L., Garnier, E. and Duluc, C-M. (2019) *Quantifying historic skew surges : an example for the Dunkirk Area, France*, Natural Hazards, 98 869–893.

Glaser, R. (1996) Data and methods of climatological evaluation in historical climatology, Historical Social Research, 21 56–88.

Glorot, X., Bordes, A. and Bengio, Y. (2011) *Deep sparse rectifier neural networks. Appearing*, in

Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) 2011, Volume 15 of JMLR: W&CP 15, Fort Lauderdale, USA.

Gottschalk, E. (1971-1977) Stormvloeden en rivieroverstromingen in Nederland, 3 volume, Van Gorcum, Assen.

Gram-Jensen, I.B. (1985) Sea Floods. Contributions to the climatic history of Denmark, Danish Meteorological Institute, Copenhague.

Haigh, I., Wadey, M., Wahl, T., Ozsoy, O., Nicholls, R., Brown, J. and Gouldby, B. (2016) *Spatial and temporal analysis of extreme sea level and storm surge events around the coastline of the UK*, Scientific Data, 3 160107.

Haigh, I., Wadey, M., Gallop, S., Loehr, H., Nicholls, R. J., Horsburgh, K., Brown, J. and Bradshaw, E. (2015) *A user-friendly database of coastal flooding in the United Kingdom from 1915–2014*, Scientific data, 2-1 1-13.

Halkin, L. (1951) Initiation à la critique historique, Armand Colin, Paris.

Hamdi, Y., Bardet, L., Duluc, C.-M. and Rebour, V. (2015) Use of historical information in extremesurge frequency estimation. The case of marine flooding on the La Rochelle site in France, Natural Hazards and Earth System Sciences, 15-7 1515-1531.

Hamdi, Y., Garnier, E., Giloy, N., Duluc, C-M. and Rebour, V. (2018) *Analysis of the risk associated to coastal flooding hazards: A new historical extreme storm surges dataset for Dunkirk, France*, Natural Hazards and Earth System Sciences, 18-12 3383-3402.

Hickey, K. (1997) Documentary records of coastal storms in Scotland, 1500-1991 A.D, 2 volume, PhD, Coventry University, Coventry.

Idier, D., Rohmer, J., Pedreros, R., Le Roy, S., Lambert, J., Louisor, J., Le Cozannet, G. and Le Cornec, E. (2020), Coastal flood: a composite method for past events characterisation providing insights in past, present and future hazards joining historical, statistical and modelling approaches, Natural Hazards, 101 465–501.

Kempe M., (2006), *'Mind the Next Flood!' Memories of Natural Disasters in Northern Germany from the Sixteenth Century to the Present,* The Medieval History Journal, 10(1-2) 327-354.

Lamb, H. and Frydendahl, K. (1991) Historic storms of the North sea, British Isles and Northwestern Europe, Cambridge University Press, Cambridge.

Lambert, J. (1986) Actualisation et interprétation des données de sismicité historique relatives au Bassin Aquitain et au Quercy, technical report, BRGM.

Lang, M., Cœur, D., Audouard, A., Villanova Oliver, M. and Pene, J-P. (2017) *BDHI: a french national database on historical floods,* in FLOODrisk 2016, Lyon.

Langlois, Ch-V. and Seignobos, Ch. (1898) Introduction aux études historiques, Hachette, Paris.

Le Goff, J. and Nora, P. (eds) (1974) Faire de l'histoire, I. Nouveaux problèmes, II. Nouvelles approches, III. Nouveaux objets, Gallimard, Paris. Lemercier, C. and Zalc, C. (2008) Méthodes quantitatives pour l'historien, La Découverte, Paris.

Mangeon, M., Duluc, C-M., Bardet, L., Giloy, N., Hamdi, Y. and Chanton, O. (2020) Opportunités,

limites et frontières de l'évaluation des aléas extrêmes : Plongée au cœur du travail des experts scientifiques de l'IRSN, in Paralia, Journées Nationales Génie Côtier-Génie Civil, JNGCGC, June 2020, Le Havre, France.

Molinari, D., Menoni, S. and Ballio, F. (eds) (2017) Flood Damage Survey and Assessment: New Insights from Research and Practice, John Wiley & Sons and The American Geophysical Union, Hoboken.

Pfister, C., Garnier, E., Alcoforado, M-J., Wheeler, D., Luterbacher, J., Nunes, M-F. and Taborda, J-P. (2010) *The meteorological framework and the cultural memory of three severe winter-storms in early eighteenth century Europe*, Climatic Change, 101/1-2 281-310.

Porfido, S., Esposito, E., Alaia, F., Molisso, F. and Sacchi, M. (2009) *The use of documentary sources for reconstructing flood chronologies on the Amalfi rocky coast (southern Italy)*, in Geohazard in Rocky coastal areas (Eds Violante C.), The Geological Society, London.

Scotti, O., Baumont, D., Quenet, G. and Levret, A. (2004) *The France macroseismic database SISFRANCE: objectives, results and perspectives*, Annals of Geophysics, 47 571-581.

Soens, T. (2009) De spade in de dijk. Waterbeheer en rurale samenleving in de Vlaamse kustvlakte (1280-1580), Directory of Open Access Books, ID : 10670/1.pd356d

Sweeney, J.C. (2000) A three-century storm climatology for Dublin (1715-2000), Irish Geography, 331 1-14.

Torres-Vera M.A. (2010) Historical seismicity in Mexico during 1568-1837: intensity evaluation and data reliability, Natural Hazards, 54 863-878.

Van Bavel, B., Curtis, D., Hannaford, M., Moatsos, M., Roosen, J. and Soens, T. (2019) *Climate and society in long-term perspective: opportunities and pitfalls in the use of historical datasets*, Wires CC, 10-6 e611.

Veale, L., Endfield, G., Davies, S., Macdonald, N., Naylor, S., Royer, M-J., Bowen, J., Tyler-Jones, R.

and Jones, C. (2017) *Dealing with the deluge of historical weather data: the example of the TEMPEST database*, Geo: Geography and Environment, 4-2 e00039. Veyne, P. (1971) Comment on écrit l'histoire, Seuil, Paris.