Encoding and Computer Analysis of Macroseismic Effects

G. Vannucci¹, P. Gasperini², G. Ferrari³ and E. Guidoboni³

¹Dipartimento di Scienze della Terra, Universita di Firenze, Via G. La Pira 4, 50121 Firenze, Italy.
E-mail: gfranco@steno.geo.unifi.it
²Dipartimento di Fisica, Universita di Bologna, Viale B. Pichat 8, 40127 Bologna, Italy.
E-mail: paolo@ibogfs.df.unibo.it
³S.G.A. Storia Geofisica Ambiente, Via Bellombra 24/2°, 40126 Bologna, Italy.
E-mail: ferrari@sga-storiageo.it, guidoboni@sga-storiageo.it

Received 08 June 1998; accepted 28 September 1998

Abstract. We propose a method for the encoding and the computer analysis of the macroseismic effects deduced from historical sources allowing the complete formalization of the process of seismic intensity assessment. In the framework of historical seismology we make use of a multi-criteria decisions-support algorithm, based on the theory of the Fuzzy Sets. By analyzing the texts of the available sources for the 1919 Mugello and 1920 Garfagnana earthquakes, we followed a classification criterion which is independent of any macroseismic scale: we "disarrange" each sentence reported on the sources into 5 syntactic elementary components and represent it by a set of alphanumeric codes. This allows us to retain the maximum adherence to the original sources and to avoid forced interpretations and losses of information due to the need of fitting a given description to each observed effect. Moreover this scheme also allows to gather equivalent effects by reassigning them the same code, and to use this new classification in further processing. This procedure could even be seen as an attempt to define a new macroseismic scale on the basis of a statistical counting of different effects occurrences.

© 1999 Elsevier Science Ltd. All rights reserved

1 Introduction

Intensity scales were originally compiled to classify the effects of earthquakes occurring synchronously to the observers. The scales are then tools which have been thought to be applied on the basis of direct observation and had not been really conformed to an efficient usage with historical sources. Often arises the paradox of a very detailed source that furnishes a framework of effects which cannot be clearly addressed to the descriptions of the degrees of the scale.

All the macroseismic scales (even the more recent updates) have been formulated and improved without a real statistical analysis of real data but only on the basis of a qualitative comparison of the different effects based on "expert" experience. The only work (Brazee, 1979) which approached the problem more systematically, seems to be completely forgotten by the following literature. Moreover, the intensity assessment is usually not based only on the canonical definitions (written on the scale) but often also relies on other implicitly assumed (non written) criteria which depend on the personal experiences and beliefs of the investigator and then may not be homogeneously applied to different sites and to different times by different investigators (i.e. every macroseismic operator assigns more or less consciously different weights or priorities to the various effects).

To investigate on these inadequacies and contradictions, in this work we propose a formalization of the intensity assessment procedure which is able to keep trace of the entire process, and makes use of an intensity assessment algorithm which gives the same result every time the base observations are the same. To achieve this objective we faced the problem of intensity assessment in the framework of the decision making support methods. Since the experience both on the field and with historical sources demonstrated that the effects of seismic events are gathered less hard (or in a more fuzzy way) than stated in the scales, we adopted a multi-criteria decision making model (MCDM) (Xiang et al., 1987), based on the Fuzzy Sets Logic (see Appendix A), which was already applied in the past to different fields, ranging from the landscape planning to the evaluation of natural hazards (see in Appendix B an outline of the MCDM algorithm).

We took advantage of the presence, among the authors, of two experts of historical seismology which assigned all the "expert" intensities and the weights of the different sources. In particular this last aspect, the only one which needs to be defined a priori by an expert, is an important element for the scientific approach of the historical information on single earthquakes. Only an expert of historical seismology can be able to define the value of the basic testimonies. Such a value is established on the basis of rigorous disciplinary criteria, is defined, in other words, by the hermeneutic rules of the historical research. This means that the various testimonies (sources) must be
considered with regard to their contemporaneity and
authoritativeness, and evaluated in their relative historical
context.

2 Decision making and macroseismic data

Decision making is a complex, human activity which can be
defined as the choosing, usually on the basis of many
criteria, of a course of action among alternatives to
accomplish one or several objectives. Decision making in
the real world mostly takes place in an environment in
which the objectives, the constraints and the consequences
of possible actions are not known precisely. Before the
introduction of the fuzzy approach (Zadeh, 1965) the only
source of this imprecision was considered the randomness,
while, after that, many authors have argued that the major
source of imprecision is fuzziness, i.e. the real impossibility
in many cases to attribute precise properties to different
subjects. The Fuzzy Sets logic tries to reproduce the mental
processes of the human brain and in particular its ability,
taking advantages from the tolerance of the imprecision, to
obtain a result even in case of a lack of complete and
precise data. Under the Fuzzy "philosophical approach" the
ambiguous evidences which can be encountered in the
application of a macroseismic scale are not due to the
randomness in the appearance of certain effects but to the
uncertainty (or fuzziness) in recognizing them as belonging
to different grades of the scale. So that the belonging, or
better the "membership", of a given effect to the set of
effects associated to a certain intensity degree, may be
better defined by a real number in the interval [0,1] whose
lower limits stands for not belonging at all and the higher
for full belonging (see Appendix A for more details about
the Fuzzy Sets).

In our intensity assessment problem the different
alternatives are the degrees of a macroseismic scale while
the effects actually observed at each given locality are the
attributes. Two possible approaches to obtain the
membership function are the subjective approach and the
empirical approach. In the first case, the membership
function values are freely assigned by the macroseismic
expert on the basis of his proper belief while in the second
one it is derived somehow from data. The debate is open in
the literature on the effectiveness of these two options (for
example: we assigned to direct contemporary sources
larger weights than to later indirect sources and we
assigned to macroseismic sources larger weights than to
newspapers).

3 Data analysis

We analyzed two Italian earthquakes which occurred
relatively close in time and space: the Mugello earthquake
(Lat=43°56', Lon=11°27') of June 29, 1919 (Ms=6.2
Imax=IX) and the Garfagnana earthquake (Lat=44°15',
Lon=10°17') of Sept. 7, 1920 (Ms=6.5, Imax=X). Due to
their proximity in space, time and energy they give us an
excellent chance to make crossed tests on the results of this
method of analysis.

In order to keep trace of the entire process, from the
reading of the sources to the intensity estimate, we
developed a computer procedure consisting of four steps:

1. Initial encoding of the database of observed effects.
2. Application of re-encoding rules in order to
equivalence effects previously kept distinct.
3. Selection of most significant effects and computation
of membership functions and weights.
4. Intensity evaluation with the fuzzy algorithm.

The initial encoding step is based on the decomposition
of each useful sentence which can be found in the sources into
two main syntactic components. They are:
- Object/subject of the phrase.
- Quantifier of the object/subject.
- Predicate.
- Modifier of the predicate.
- Specification of the object/subject or of the predicate.

| Table 1: Encoding phase: disarrangement of sentences into 5
| syntactic components. |
|-----------------------|-----------------------|
| Quantifier | Object/subject | Specification | Predicate | Modifier |
| most | people | at rest | left |
| many | houses | stone | cracks | light |
| cw | glasses | broken | |

After their disarrangement, the set of descriptions forms a
matrix (see Table 1) where each row represents a single
macroseismic effect and each column a different syntactic
component. By assigning a two-character code to each
different word which can be found in each column, it is
possible to assign a ten-character code to each effect which
thus can be analyzed later by computer techniques. In order
to better explain this passage, an example can help: the
phrase: "light cracks in many stone houses" include an
"object/subject" which is "houses" and a "predicate" which
can be represented by the word "cracks" (notwithstanding
the latter is not really a verb it however express the action
of cracking), the term "light" represents a "modifier" of the
"predicate", "many" the "quantifier" of the
"object/subject" and " stone " the "specification" (in Table
The highlighted row shows the result of this decomposition.

Now the first element of the phrase can be compared with the list of the words previously found in other sentences for the corresponding syntactic component (see Table 2). If the same word is already present in the list, the corresponding two-character code is assigned, otherwise a new code is allocated and the new word is inserted in the list with the given code. After repeating this procedure for all the five elements, our phrase “light cracks in many stone houses” will be represented by the ten-character code “d4-62-51-42-26”.

This procedure is not always obvious and sometimes cannot be unique but in most of the cases can be carried out easily with the help of the computer code we developed. Through this encoding scheme, it is possible to maintain an almost complete and accurate recording of the information contained in the sources without any adaptation to the descriptions of the events.

During the re-encoding step it is possible to modify the classification defined in the previous step in order to make equivalent two or more effects which formerly had kept distinct. This can be done by equating different codes of the same column: for example the subject/object “houses” (code 62) can be made equivalent to “buildings” (code 63) or the predicate “to break” (code 41) can be made equivalent to “to crack” (code 42). The re-encoding can also be done using combinations of codes belonging to different columns for example: “railway tracks” (object/subject) “bent” (predicate) can be made equivalent to “railway line” (object/subject) “closed” (predicate). After these “re-encoding rules” are compiled, a computer program automatically makes the changes and builds a new database of observed effects. The main advantage of this procedure is that everybody can apply their own “rules” and change them at will without modifying the original database.

In the selection step all the effects that are rarely observed are discarded and not furthermore processed. In the following computation we will only consider the effects with at least 5 occurrences at different sites, but this threshold could even be increased to improve the reliability of the results. In fact this selection guarantees that the computed empirical membership functions are less biased by possible anomalous cases and also reduces the danger of overfit (see discussion below).

In the intensity evaluation step, the MCDM algorithm described in Appendix B is applied and the “fuzzy” intensity is computed at each site. This point can be repeated independently by using different membership schemes. The algorithm also allows multiple memberships and weighting schemes to be used simultaneously.

A graphical sketch on how the intensity assessment procedure works is shown in Figure 1 where the shapes of the empirical membership functions of the effects of the Garfagnana earthquake observed in Florence are reproduced. In Figure 2 the decision function obtained by taking the minimum membership value for each intensity among all of the functions is plotted. The intensity degree chosen by the algorithm as the “least objectionable” solution is the one corresponding to the maximum of the decision function.
4 Results and discussion

In order to check the efficiency and the reliability of our methodology we tested its ability to correctly reproduce the macroseismic expert decisions. We thus computed the coefficient of variation $R^2$ of the regression of the intensity estimated by the expert with the fuzzy intensity, which represent the amount of the variance of the former estimate that can be explained by the latter, and the average absolute difference $r$ between the expert ($I_e$) and the fuzzy intensity ($I_f$) over the entire data set of evaluated localities. This last is given by:

$$r = \frac{1}{N_{\text{total}}} \sum |I_e - I_f|$$

(1)

where $N_{\text{total}}$ is the total number of evaluated localities. We also computed the average signed difference $r_{\text{sign}}$ between expert and fuzzy intensities which represents a sort of “offset” indicator between the expert and “fuzzy” intensity estimates:

$$r_{\text{sign}} = \frac{1}{N_{\text{total}}} \sum (I_e - I_f)$$

(2)

Furthermore, in order to evaluate the overall efficiency of the algorithm in determining the intensity, we also report on the following tables the total number of evaluated localities ($N_{\text{total}}$), the number of univocal intensity determinations ($N_{\text{single}}$), and the number of intensity determination ($N_{\text{multiple}}$) which are uncertain between two or more grades.

In Table 3 the summary results for the two earthquakes, on the basis of the empirical membership functions and weights, separately estimated from the data of each event, are shown. For both sets the small values of $r$ and the high $R^2$ indicate that the algorithm satisfactorily reproduces the expert intensities (within half of a degree on average). The values of $r_{\text{sign}}$, which are positive for the Garfagnana event and slightly negative for the Mugello one, correspond to an underestimation of the “fuzzy” intensity with respect to the expert for the former earthquake and an overestimation for the latter. These differences, which nevertheless lay both largely below the average residuals, might be caused by different statistical distribution of various intensities for the two earthquakes.

In order to investigate on this problem in Table 4 we show the result of the same comparison when the fuzzy membership function and weights are computed using, as learning set, the data coming from both earthquakes merged together. The scores do not vary very much with respect to the previous table but you can note a slight improvement of the fit for the Garfagnana earthquake and a worsening for the Mugello event. The average signed differences $r_{\text{sign}}$ confirms the tendency of the fuzzy algorithm to overestimate the Mugello intensities while it shows an almost perfect coincidence (on average) of the two estimates for the Garfagnana data.

In Table 5 Swapped membership functions

Since in both previous computations we used the data of each earthquake to determine the algorithm parameters
used for the same event, it is possible that the goodness of the agreement might be due to overfit (that means that the algorithm fits not only the average tendencies of the data but also their statistical fluctuations). We then performed additional computations (see Table 5) where for each event, the membership functions and weights are derived from the data of the other event. The marked decrease of the $R^2$ and the increase of $r$ (especially for the Mugello earthquake) clearly shows that a certain amount of overfit is certainly present in previous computations. However the fit remains quite acceptable for both earthquakes notwithstanding the complete independence of the learning and testing sets. The opposite signs of $r_{\text{sign}}$ for the two events confirm the tendency indicated by previous cases with a remarkable increase of the offsets (however still largely below the average absolute deviations). A possible explanation of this behaviour could be that, even in presence of the same effects, the expert is more confident to assign higher degrees in the framework of a strong earthquake like the Garfagnana event rather than the weaker Mugello one.

<table>
<thead>
<tr>
<th>Table 6: Macroseismic bulletins only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Garfagnana</td>
</tr>
<tr>
<td>Mugello</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 7: Newspaper sources only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Garfagnana</td>
</tr>
<tr>
<td>Mugello</td>
</tr>
</tbody>
</table>

To analyze the influence of types of sources on the resulting intensity determinations we performed additional separate computations by only using macroseismic bulletins (Table 6) or only newspapers (Table 7). The average residuals' slight increase is obviously caused by the worsening of the definitions of membership functions due to the smaller number of data used, while the marked decreases of the $R^2$ values with respect to Table 4 may also be linked to a reduction of the overall variance of the intensity. In fact the intensity ranges differ between the two type of sources: while macroseismic bulletins mainly refers to low intensities, newspaper are usually the main source of information for more severely damaged localities.

5 Conclusions

The method of analysis of macroseismic effects we developed allows to describe in detail the process of intensity assessment that in many cases is performed by the macroseismic expert without a trace of the assumptions done and of the decision made and then sometimes could not be reproducible even by the same expert. In particular, the reliability of the different sources and the weights of the different effects, established by historical seismologists, can be accounted for explicitly.

Our approach can be useful to reduce the arbitrariness of the intensity assignment process, and could practically cancel all possible mistakes as far as the encoded data correctly interpret the text. It may even be a useful support tool for the macroseismic expert himself who, from the comparison with the algorithm results, can improve the understanding of his own choices (sometimes not fully rationalized) and decisional processes.

The ability of the multi-criteria decision making algorithm to combine different membership and weighting schemes (determined, for example, from the intensities assigned by different experts) could be used to obtain "objective" estimates in debated cases.

Our method of encoding effects, being independent of a particular macroseismic scale, could be used in the future, with a large enough database, to define the characteristics of a new macroseismic scale which can be more useful for historical testimonies.

Appendix A: Fuzzy Sets

In the ordinary sets algebra the membership of an object $X_i$ (belonging to an universe of objects $X=\{X_1, X_2, \ldots, X_n\}$) in a set $A$ can be defined as a characteristic function $U$ from $X$ to a valuation set $[0,1]$ such that:

\begin{align*}
U_A = 1 & \text{ if } X_i \in A & (A1) \\
U_A = 0 & \text{ if } X_i \notin A & (A2)
\end{align*}

In the fuzzy algebra the membership function is not limited to be only two-valued ($1=$belonging, $0=$not belonging) but it can assume all the real values in the interval $[0,1]$. From this definition, the theory of the Fuzzy Sets can be developed defining operations among sets in a way similar to the ordinary (hard) sets theory.
operations between fuzzy sets can be seen as an extension of the classic sets theory. We can define, on two fuzzy sets \( A \) and \( B \), the fuzzy operations of union (\( A \cup B \)), intersection (\( A \cap B \)) and complementation (\( A' \)) in terms of membership function as:

\[
\forall \, X_i \in X, \quad U_{A \cup B} (X_i) = \max \{ U_A (X_i), U_B (X_i) \}; \quad (A3)
\]

\[
\forall \, X_i \in X, \quad U_{A \cap B} (X_i) = \min \{ U_A (X_i), U_B (X_i) \}; \quad (A4)
\]

\[
\forall \, X_i \in X, \quad U_{A'} (X_i) = 1 - U_A (X_i) \quad (A5)
\]

Note that these definitions also hold for the hard (non-fuzzy) sets using the appropriate (two-valued) definition of membership function.

Appendix B: Decision making algorithm

The Fuzzy Sets logic can be applied to the field of decision making and in particular to represent the relationship between the possible alternative decisions and the objectives or attributes which must be taken into account for making the decision. In the classic decision theory, one considers a system of variables with a set of constraints which limits the choice among alternatives and some attributes or objectives, which sort the alternatives on the basis of certain criteria. Bellman and Zadeh (1970) suggested that each attribute can be represented as a fuzzy subset over the set of alternatives \( X = \{ X_1, X_2, ..., X_i, ..., X_n \} \). Thus, if \( A_j \) indicates the \( j \)-th attribute, then the grade of membership of alternative \( X_i \) in \( A_j, U_{Aj} (X_i) \) indicates the grade to which \( X_i \) satisfies this attribute. In order to combine multiple attributes and to form the decision making function, the “minimax” decision making procedure has been proposed by Bellman and Zadeh (1970). The decision function \( D(X_i) \) that satisfies all of the attributes is obtainable as an intersection of the fuzzy sets corresponding to all the attributes \( A_j \). This corresponds to get the minimum of the membership function for each alternative:

\[
\forall \, X_i \in X, \quad D(X_i) = \min_j \{ U_{Aj} (X_i) \} \quad (B1)
\]

To find the “least objectionable” solution \( X^* \), the maximum value over the alternatives in \( D \) must be calculated so that the whole procedure can be summarized by:

\[
D(X^*) = \max_i \{ \min_j \{ U_{Aj} (X_i) \} \} \quad \forall \, X_i \in X \quad (B2)
\]

This procedure does not allow to consider attributes that may differ in importance but appropriate weighting of attributes can effectively do this.

The model we adopted is a slightly modified version of the one used for assessment of landscape planning by Xiang et al. (1987). It allows different membership functions (coming for example from different data sets or experts criteria) to be used at the same time. This can be done by the aggregation of the different membership \( U_{K,Aj} \) and weighting \( W_{kj} \) functions relative to each \( k \)-th data set/criterion. The aggregate membership is given by:

\[
U_{Aj} (X_i) = \min_k \{ U_{K,Aj} (X_i) \}, \quad \forall \, k=1,...,q \quad (B3)
\]

being \( q \) the number of different memberships and \( U_{K,Aj} \) the grade of membership of alternative \( X_i \) to the \( j \)-th attribute and relatively to \( k \)-th criterion. A similar expression is given for aggregation of weights:

\[
W_j = \min_k \{ W^k_j \}, \quad \forall \, k=1,...,q \quad (B4)
\]

where \( W^k_j \) is the weight of \( j \)-th attribute on the basis of \( k \)-th data set/criterion. The least objectionable solution \( X^* \) is obtained via the decision rule:

\[
D(X^*) = \max_i \{ \min_j \{ W^k_j U_{Aj} (X_i) \} \}, \quad \forall i=1,...n, \forall i=1,...m \quad (B5)
\]

being \( n \) the total number of alternatives and \( m \) the total number of attributes.

References


