Knowledge-Based Systems 101 (2016) 15–30

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction

G. Asencio-Cortés [a], F. Martínez-Álvarez [a,*], A. Morales-Esteban [b], J. Reyes [c]

[a] Division of Computer Science, Universidad Pablo de Olavide, ES-41013 Seville, Spain
[b] Department of Building Structures and Geotechnical Engineering, University of Seville, Spain
[c] TGT-NT2 Labs, Santiago, Chile

ARTICLE INFO

Article history:
Received 22 June 2015
Revised 16 November 2015
Accepted 20 February 2016
Available online 16 March 2016

Keywords:
Sensitivity analysis
Earthquake prediction
Seismicity indicators
Supervised learning

ABSTRACT

The use of different seismicity indicators as input for systems to predict earthquakes is becoming increasingly popular. Nevertheless, the values of these indicators have not been systematically obtained so far. This is mainly due to the gap of knowledge existing between seismologists and data mining experts. In this work, the effect of using different parameterizations for inputs in supervised learning algorithms has been thoroughly analyzed by means of a new methodology. Five different analyses have been conducted, mainly related to the shape of training and test sets, to the calculation of the *b*-value, and to the adjustment of most collected indicators. Outputs sensitivity has been determined when any of these factors is not properly taken into consideration. The methodology has been applied to four Chilean zones. Given its general-purpose design, it can be extended to any location. Similar conclusions have been drawn for all the cases: a proper selection of the sets length and a careful parameterization of certain indicators leads to significantly better results, in terms of prediction accuracy.

1. Introduction

The problem of predicting earthquakes has fascinated the human being. Although this problem seems to be irresolvable, recent works have proposed new paradigms of prediction that should be taken into consideration [1]. In particular, the use of data mining techniques has emerged in this field as a powerful tool with undeniable benefits [2–5].

This work is focused on the analysis of the inputs used in several supervised machine learning classifiers in order to improve earthquake prediction accuracy. In particular, some studies recently conducted propose the use of various seismicity indicators (or attributes containing geophysical information associated with earthquake occurrence) for earthquake prediction [6–8].

The correlation of such indicators with the binary class (either an earthquake is coming or not) was analyzed in [9], showing that some of them exhibited information gain close to zero. This work goes one step ahead because all of these indicators have been used with a baseline configuration. This is, none of the works above referenced considered that most of the indicators are functions of certain variables. These works just used standard values omitting the fact that different configurations may lead to different results and, in some cases, to better results. And this is the main goal of this research: to conduct an exhaustive analysis on how an adequate adjustment of the seismicity indicators may improve the accuracy of the classifiers.

In particular, in [8] a new set of seismicity indicators was proposed as inputs for earthquake prediction. Later, in [9], such set of indicators were combined with those published in [7] and applied feature selection methods to discover that some of the indicators proposed in both [7] and [8] exhibit null information gain with the class. In this work, it should be highlighted that some indicators are highly dependent with their initial parameterization. A sensitivity study is performed to show that results can be highly improved if an adequate initialization is done.

A new methodology is thus proposed and the following issues have been explored. First, the size of the most adequate training sets and whether training and test sets must be contiguous or not. Second, how the *b*-value (a key predictive value [10]) must be calculated. Finally, how certain attributes introduced in [7,8] must be configured so that the best possible prediction is achieved. In other words, it provides some guidelines in order to properly parameterize the seismicity indicators proposed to date. Also, the best training set selection is performed.

Four zones of Chile, the country with the highest seismic activity [11], have been analyzed to validate the applicability of this

* Corresponding author. Tel.: +34 954 977 370.
E-mail addresses: guaasecor@upo.es (G. Asencio-Cortés), fmaralv@upo.es (F. Martínez-Álvarez), ame@us.es (A. Morales-Esteban), daneel@geofisica.cl (J. Reyes).

http://dx.doi.org/10.1016/j.knosys.2016.02.014
0950-7051/

methodology. Nevertheless, it has been defined so that it can also be applied to any zone in the world.

The rest of the paper is structured as follows. Section 2 provides a general overview on the state-of-the-art. Section 3 describes the new methodology proposed in order to find the set of seismicity indicators with the most adequate initialization (when used as input in supervised classifiers). All the results of applying the methodology to four cites in Chile have been presented in Section 4. Finally, the conclusions drawn from this study have been summarized in Section 5.

2. Related works

The possibility of predicting earthquakes has been questioned and answered in various ways, from denial to optimism, including the contribution of mathematical proofs and empirical support for each hypothesis [8,12–15].

To ensure that statements related to earthquake prediction are rigorous, the following information must be simultaneously provided according to [16]:

1. A specific location or zone.
2. A specific span of time.
3. A specific magnitude range.
4. A specific probability of occurrence.

Additionally, the U.S. Geological Survey (USGS) founded the Collaboratory for the Study of Earthquake Predictability (CSEP) in 2007 [17]. The goal of this organization is to develop a virtual and distributed laboratory that can support a wide range of scientific prediction experiments in multiple regional or global natural laboratories. This earthquake system science approach seeks to provide answers to the questions:

1. How should scientific prediction experiments be conducted and evaluated?
2. What is the intrinsic predictability of the earthquake rupture process?

In this context, several methods have been proposed to predict any of the features detailed by Allen [16]. According to the Accelerating Moment Release (AMR) method, the rate of seismic moment release for magnitude is rapidly increased before a large event occurs [18,19].

Variations of *b*-value have also been analyzed. For a large magnitude earthquake to occur, it is necessary a prior elastic potential energy accumulation. This fact causes a deficit of small and moderate earthquakes. This leads to an abnormal alteration of the Gutenberg–Richter law's *b*-value [10,20].

M8 algorithms study the occurrence of earthquakes of magnitude larger than 8.0. They are based on the evolution of several time series composed of earthquakes of moderate magnitude. The goal is to decide if a time of increased probability (TIP) exists for an event of larger magnitude [21,22].

Region–Time–Length (RTL) is an algorithm that analyzes temporal sequences of earthquakes. It only takes into consideration location, time, magnitude, and detects anomalies in seismicity prior to large events [23,24].

It is thought that for a large earthquake to occur, it is necessary that more energy is released during the loading period than during the unloading one. Based on this assumption, Load-Unload Response Ratio (LURR) uses the ratio of energy released as a potential precursor to make predictions [25,26].

Another widely used method is Every Earthquake is a Precursor According to Scale (EEPAS). This method is based on the observation of an increment of small earthquakes, as this is considered a precursory phenomenon for larger earthquakes [27,28].

Epidemic-Type Aftershock Sequence (ETAS) considers that every earthquake is a simultaneously potential aftershock, main shock or foreshock, with its own aftershock sequence. This way, anomalous configurations for temporal and spatial seismicity can be found [29,30].

The Simple Smoothed Seismicity model, or simply TripleS, provides space-rate-magnitude forecasts based on a spatial clustering of seismicity. To get this done, a Gaussian smoothed is applied to the seismic catalogue, which estimates the amount of foreseen earthquakes in particular zones for particular periods of time [31].

Increased attention is being payed to algorithms based on machine learning nowadays. These algorithms include a vast variety of solutions ranging from unsupervised learning [10,32] to supervised one [4,9]. It must be noted that in [10] clustering techniques were used to obtain patterns that model the behavior of seismic temporal data and can help to predict medium-large earthquakes. It is true that in [9] the seismicity indicators used were based on [10] and others. But, in [10], the results were patterns and in [9] the results were a probability of an earthquake to happen after the hit of every earthquake of magnitude larger than 3.0. Moreover, the results in [9] were improved thanks to feature selection techniques. In supervised learning, every earthquake is modeled by means of certain attributes that Panakkat and Adeli [7] named seismicity indicators. From its initial application, several works have proposed new indicators. Such is the case of [8] or [6], where the authors also added Bath and Omori–Utsu laws, as well as variations of *b*-value, to the set of proposed seismicity indicators. The model was assessed by artificial neural networks, a method also used in [2,5,33].

Recently, Ikram and Qamar [34] introduced an expert system for earthquake prediction, which extended [35]. They considered the historic record of earthquakes and divided the Earth into four zones. Then, association rules were applied to predict earthquakes in each of the four zones with a horizon of prediction equals to one day.

Nonetheless, some of the seismicity indicators proposed exhibit parametrical dependence, this is, there is a need of an initial setup so that they can properly work with supervised classifiers. Moreover, the original studies do not explicitly propose a specific tuning for them. In this context this work has been carried out: to determine the influence of either an adequate or wrong adjustment for all the existing seismicity indicators reported in the literature.

3. Methodology

This section introduces a methodology to systematically identify those values for certain parameters, somehow hidden in a set of seismicity indicators, that generate better results in terms of average accuracy when predicting earthquakes. In this sense, a set of parameters that may deeply influence the accuracy for predictions has been first identified. Later, a sensitivity analysis over such parameters has been conducted in order to determine how a wrong setup may lead to the occurrence of a major loss of accuracy in predictions.

Note that the proposed methodology must be applied to every geographical zone. In this work, the four Chilean zones studied in [8] and [9] have been considered.

Section 3.1 explains how supervised learning is applied to predict earthquakes. Once this strategy is defined, Section 3.2 details the five proposed studies to analyze the sensitivity of the parameters involved in the prediction.

3.1. Procedure for earthquake prediction

Generally speaking, the prediction of earthquakes is carried out in the context of supervised learning by means of well-known

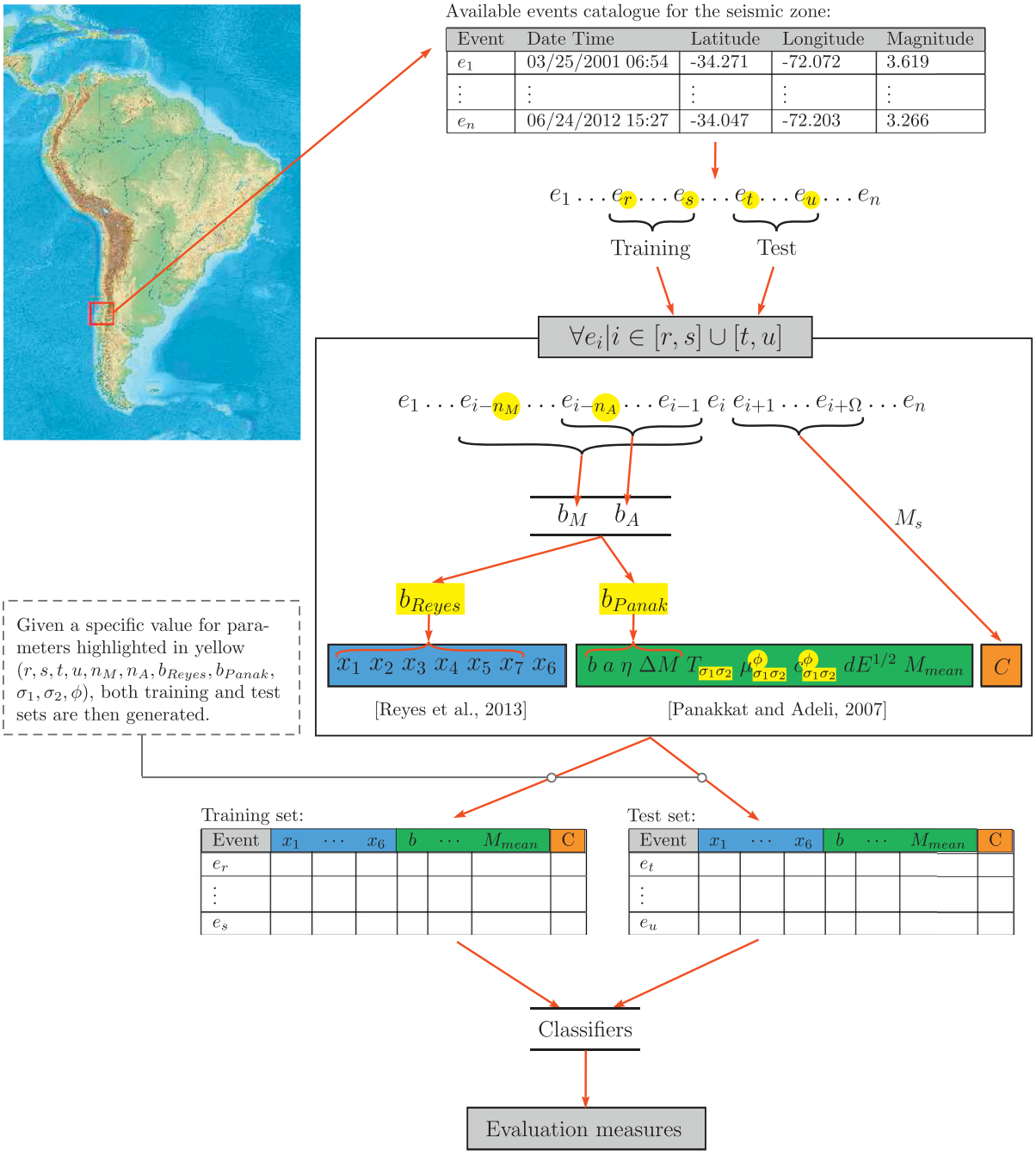


Fig. 1. Global procedure of the methodology for a given seismic zone. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

classifiers, using *hold-out* as validation method. These approaches have already been used in the literature [6–9] and consist in predicting the class (or label) for a test set, given a training set.

Fig. 1 depicts the methodology proposed in this work. First, the target zone has been selected. The city of Pichilemu (Chile) and surroundings have been picked for illustrative purposes (red frame). The information is retrieved from seismic catalogues, which feed the system with data of earthquakes occurred in the past. From this data, both training and test sets must be created (highlighted in blue and green in Fig. 1, respectively), as well as the class (in orange) so that the classifiers can perform the predictions. Subsequent sections detail how every task is carried out.

All elements used to define the methodology are shown in Tables 1–4. First, the complete set of attributes which can be considered for predictions are indicated in Table 1. Note that these attributes were firstly introduced in [8] and [7]. However, some of these attributes need parameterization and, in previous works, this has not been done. In this sense, the initial parameters taken into consideration in the methodology are summarized in Table 2. In Table 3 the nomenclature used in the assessment of the Gutenberg–Richter law's *b*-value [36] is indicated. Finally, some required auxiliary definitions are presented in Table 4.

The methodology developed in this work allows the comparison of effectiveness when predictions are made using some attributes as input and some parameterization for them. Five studies have been performed changing the input attributes and their parameterization. The first study changes four temporal parameters (*r, s, t, u*) which affect to all attributes. The second study changes

Table 1
Attributes considered for predictions.

Name	Description
b	Gutenberg–Richter law's b-value
x_1	Increment of b between the events i and $i-4$
x_2	Increment of b between the events $i-4$ and $i-8$
x_3	Increment of b between the events $i-8$ and $i-12$
x_4	Increment of b between the events $i-12$ and $i-16$
x_5	Increment of b between the events $i-16$ and $i-20$
x_6	Maximum magnitude from the events recorded during the last week (OU's law)
x_7	Probability of recording an event with magnitude larger or equal to 6.0 using a probability density function
a	Gutenberg–Richter law's a-value
η	Mean square deviation
ΔM	Magnitude deficit
T	Elapsed time
μ	Mean time
c	Coefficient of variation
$dE^{1/2}$	Rate of square root of seismic energy
M_{mean}	Mean magnitude
$T_{\sigma_1\sigma_2}$	Tuple of T values using magnitude thresholds in the interval $[M_s-\sigma_1, M_s+\sigma_2]$
$\mu^{\phi}_{\sigma_1\sigma_2}$	Tuple of μ values using magnitude thresholds in the interval $[M_s-\sigma_1, M_s+\sigma_2]$ and ϕ as characteristic interval amplitude
$c^{\phi}_{\sigma_1\sigma_2}$	Tuple of c values using magnitude thresholds in the interval $[M_s-\sigma_1, M_s+\sigma_2]$ and ϕ as characteristic interval amplitude

Table 2
Initial parameters to tune the prediction methodology.

Name	Description
r, s	Indices that point to start and end events to define the training set from a catalogue
t, u	Indices that point to start and end events to define the test set from a catalogue
n_M, n_A	Number of previous events used to assess b_M and b_A
M_0	Reference magnitude of a geographic zone
M_s	Magnitude threshold used to predict next events in μ and c
dtp	Number of next days considered to perform predictions
σ_1	Lower bound for interval of magnitude thresholds considered to assess T, μ and c
σ_2	Upper bound for interval of magnitude thresholds considered to assess T, μ and c
ϕ	Amplitude of the interval of magnitudes that defines characteristic events
λ	Step value between values of θ (fixed to 0.1 in this work)

Table 3
Definitions for calculation of the b value.

Name	Description
b_M	Calculation of b value based on maximum likelihood approach (proposed in [8])
b_A	Calculation of b value based on least squares fitting (proposed in [7])
b_{Reyes}	Represents the b value used to assess the attributes x_1, x_2, x_3, x_4, x_5 and x_7
b_{Panak}	Represents the b value used to assess the attributes a, η and ΔM

Table 4
Auxiliary elements to define the methodology.

Name	Description
M_i	Magnitude of the event i
N_i	Number of events with magnitude greater than or equal to M_i among the n_A events previous to the event i
θ	Magnitude threshold for T, μ and c
m	Number of values for θ
$h_{\theta i}$	Ordered sequence of times for events previous to e_i with magnitude greater than or equal to θ
$ch^{\phi}_{\theta i}$	Ordered sequence of times for events previous to e_i with magnitude centered in θ with maximum deviation of ϕ
$T_{\theta i}$	Elapsed time between the first and (n_A)th event previous to e_i with magnitude greater than or equal to θ
$\mu^{\phi}_{\theta i}$	Average time differences between consecutive elements in sequence $ch^{\phi}_{\theta i}$
$c^{\phi}_{\theta i}$	Variation coefficient of time differences between consecutive elements in sequence $ch^{\phi}_{\theta i}$

between two sets of attributes (previously introduced in [7] and [8]) and changes a parameter b_M/b_A (the b-value assessment type). The third study changes two parameters (σ_1 and σ_2) which affect to three attributes (T, μ and c) of the set of attributes. The fourth study changes one parameter (ϕ) which affects to two attributes (μ and c) of the set of attributes. The fifth study changes equally two parameters (n_M and n_A) which affect to all attributes. The detailed description of these studies is presented in Section 3.2 while their results are discussed in Section 4.4.

3.1.1. Catalogue retrieval and selection of training and test sets

The data for every seismic zone is obtained from official catalogues that are usually available on-line. The information provided by such catalogues may vary from one zone to another, but they generally contain: date, time, latitude, longitude and magnitude for every event. Spatial coordinates will be used to filter data and to make predictions in restricted zones. In this work, the dimension of the zones will be $1^\circ \times 1^\circ$, which is another interesting aspect of this work.

Then, training and test sets have been selected from the catalogue. Let $e_1 \ldots e_n$ be the events (or earthquakes) in the catalogue. Training and test sets have been identified as $e_r \ldots e_s$ and $e_t \ldots e_u$, respectively, with $1 \leq r \leq s < t \leq u \leq n$.

A very common strategy is to use 70% of data to train and 30% to test: $\frac{s-r+1}{(s-r+1)+(u-t+1)} \simeq 0.7$. Note that, usually, $e_{s+1} = e_t$ although variations in values r, s, t and u have been considered to perform one of the sensitivity analysis proposed in this work (see Section 3.2.1).

3.1.2. Generation of new seismic features

From every event e_i, where $i \in [r, s] \cup [t, u]$, two set of features (or seismicity indicators) have been generated. The first set was introduced in [7] and consists of features b, a, η, ΔM, T, μ, c, $dE^{1/2}$ and M_{mean}. The second one was proposed in [8] and is highlighted in blue in Fig. 1. It is composed of the following attributes: x_1, x_2, x_3, x_4, x_5, x_6 and x_7.

Features T, μ y c have been extended by adding three new parameters, σ_1, σ_2 y ϕ, which will be explained in detail in subsequent sections. This has been done in order to analyze the predictive ability of such attributes regarding the introduced parameters. This information is shown in the studies carried out in Sections 3.2.3 and 3.2.4. Therefore, the set of attributes from [7] has been used but modifying their calculation accordingly. Thus, the used attributes have been b, a, η, ΔM, $T_{\sigma_1\sigma_2}$, $\mu^{\phi}_{\sigma_1\sigma_2}$, $c^{\phi}_{\sigma_1\sigma_2}$, $dE^{1/2}$ and M_{mean}, which have been highlighted in green in Fig. 1.

The calculation of attributes x_1, x_2, x_3, x_4, x_5, x_7, b, a, η and ΔM is based on the b-value from the Gutenberg–Ritcher's law [36]. The b-value is the size distribution factor. It reflects the tectonics of the underlying zone and it is related to the geophysical properties of the zone.

There are two different ways to calculate the b-value given an event e_i. The first one, denoted as b_M, is used in [8] and is shown in Eq. (1). The parameters involved are: number of considered events prior to e_i, n_M; magnitude of the event e_{i-j}, M_{i-j}; and the cutoff magnitude for the seismic zone, M_0.

$$b_M = \frac{\log e}{(1/n_M)\sum_{j=0}^{n_M-1} M_{i-j} - M_0} \tag{1}$$

The second way to calculate b, hereafter denoted as b_A, is the one used in [7] and is shown in Eq. (2), where N_{i-j} is the number of events with magnitude equal or larger than M_{i-j} in the n_A events prior to e_i.

$$b_A = \frac{n_A \sum_{j=1}^{n_A}(M_{i-j}\log N_{i-j}) - \sum_{j=1}^{n_A} M_{i-j} \sum_{j=1}^{n_A} \log N_{i-j}}{(\sum_{j=1}^{n_A} M_{i-j})^2 - n_A \sum_{j=1}^{n_A} M_{i-j}^2} \tag{2}$$

In Fig. 1, b_{Reyes} and b_{Panak} indicate the b-values used for the calculation of attributes x_1, x_2, x_3, x_4, x_5, x_7 (set proposed in [8]) and b, a, η, ΔM (set proposed in [7]), respectively. Note that in [6,8,9] the b_{Reyes} value used for the calculation of x_1, x_2, x_3, x_4, x_5 y x_7 was obtained with Eq. (1). Analogously, in works [7,9] the b_{Panak} used for the calculation of b, a, η and ΔM was obtained with Eq. (2).

It is necessary to introduce five new elements $h_{\theta i}$, $ch^{\phi}_{\theta i}$, $T_{\theta i}$, $\mu^{\phi}_{\theta i}$ and $c^{\phi}_{\theta i}$ to define attributes $T_{\sigma_1\sigma_2}$, $\mu^{\phi}_{\sigma_1\sigma_2}$ and $c^{\phi}_{\sigma_1\sigma_2}$. The expression $h_{\theta i}$ represents the ordered series of times for the events prior to e_i with magnitude equal or larger than a threshold θ, as shown in Eq. (3), where t_j identifies the date and time of event e_j.

$$h_{\theta i} = \{t_j : M_j \geq \theta \wedge j < i\} \tag{3}$$

Let $ch^{\phi}_{\theta i}$ be an ordered sequence of time for the characteristic events occurred before e_i with magnitude θ and maximum deviation ϕ, as shown in Eq. (4).

$$ch^{\phi}_{\theta i} = \{t_j : M_j \in [\theta - \phi, \theta + \phi] \wedge j < i\} \tag{4}$$

Let $T_{\theta i}$ be the time elapsed between the first and the n_Ath event, prior to e_i with magnitude equal or larger than θ, as defined in Eq. (5).

$$T_{\theta i} = h_{\theta i}(1) - h_{\theta i}(n_A) \tag{5}$$

Also, $\mu^{\phi}_{\theta i}$ is the average elapsed time for n_A consecutive characteristic events prior to e_i, as can be seen in Eq. (6).

$$\mu^{\phi}_{\theta i} = \frac{\sum_{j=1}^{n_A} ch^{\phi}_{\theta i}(j) - ch^{\phi}_{\theta i}(j-1)}{n_A} \tag{6}$$

Let $c^{\phi}_{\theta i}$ the variation coefficient for the elapsed times for n_A consecutive characteristic events prior to e_i, as shown in Eq. (7).

$$c^{\phi}_{\theta i} = \frac{\sqrt{\frac{1}{n_A}\sum_{j=1}^{n_A}(ch^{\phi}_{\theta i}(j) - ch^{\phi}_{\theta i}(j-1) - \mu^{\phi}_{\theta i})^2}}{\mu^{\phi}_{\theta i}} \tag{7}$$

Once these five elements have been introduced, $T_{\sigma_1\sigma_2}$, $\mu^{\phi}_{\sigma_1\sigma_2}$ and $c^{\phi}_{\sigma_1\sigma_2}$ can be already calculated for event e_i. Associated formulas can be found in Eqs. (8)–(10), respectively. Note that M_s is the threshold magnitude to make predictions (predictions have been made for magnitudes exceeding such threshold) as detailed in Section 3.1.3. Parameter λ is the step for which θ is incremented in each iteration and has been set to 0.1. Finally, $m = 1 + (\sigma_1 + \sigma_2)/\lambda$ indicates the number of values that θ reaches according to σ_1, σ_2 and λ.

$$T_{\sigma_1\sigma_2 i} = \{T_{\theta i} \ : \ \theta = M_s - \sigma_1 + k\lambda \ \wedge \ k = \{0 \ldots m-1\} \in \mathbb{N}\} \in \mathbb{R}^m \tag{8}$$

$$\mu^{\phi}_{\sigma_1\sigma_2 i} = \{\mu^{\phi}_{\theta i} \ : \ \theta = M_s - \sigma_1 + k\lambda \ \wedge \ k = \{0 \ldots m-1\} \in \mathbb{N}\} \in \mathbb{R}^m \tag{9}$$

$$c^{\phi}_{\sigma_1\sigma_2 i} = \{c^{\phi}_{\theta i} \ : \ \theta = M_s - \sigma_1 + k\lambda \ \wedge \ k = \{0 \ldots m-1\} \in \mathbb{N}\} \in \mathbb{R}^m \tag{10}$$

It is worth noting that $T_{\sigma_1\sigma_2}$, $\mu^{\phi}_{\sigma_1\sigma_2}$ and $c^{\phi}_{\sigma_1\sigma_2}$ are vectors in $\mathbb{R}^m$ including values for attributes T, μ and c from [7] that have been calculated for different M_s, which will eventually define the class to be predicted.

3.1.3. Class generation

The class (or label) C for every event is defined as a logical value, $C = \{0, 1\}$. The class C is equals to 1 if the maximum magnitude for events occurred in the horizon of prediction (dtp days after e_i) is larger than M_s; $C = 0$ otherwise. Formally, the class C_i associated with the event e_i is defined in Eq. (11), where Ω is the number of events encountered in the catalogue for dtp days after the event e_i.

$$C_i = \begin{cases} 1, & \text{if } max\{M_j \ : \ i < j \leq i + \Omega\} \geq M_s \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

3.1.4. Classification and evaluation

Training and test sets have been obtained after the generation of the features and class for events e_i, where $i \in [r, s] \cup [t, u]$, as described in previous sections. For doing so, the parameters highlighted in yellow in Fig. 1 (r, s, t, u, n_M, n_A, b_{Reyes}, b_{Panak}, σ_1, σ_2, ϕ) need a preset value, since they all participate in the obtention of such sets.

Once sets have been generated, a battery of well-known supervised classifiers have been used in order to predict the class of the current event, which involves predicting if during the dtp next days an event with magnitude larger than M_s is occurring or not.

Then, differences between predicted and actual classes must be assessed. There is a wide range of quality measures and its use

Table 5
Reference catalogues and configuration from [8] and [9].

Catalogue	M_s	r'	s'	t'	u'	t_r	t_s	t_t	t_u
Pichilemu	4.6	19	140	141	262	08/10/05	03/31/10	04/01/10	10/08/10
Santiago	3.4	141	262	263	384	05/13/03	06/02/04	06/23/04	01/16/06
Talca	4.4	19	140	141	185	06/19/03	03/21/10	03/24/10	01/04/11
Valparaíso	3.8	691	812	813	918	01/31/06	12/19/08	12/19/08	02/10/11

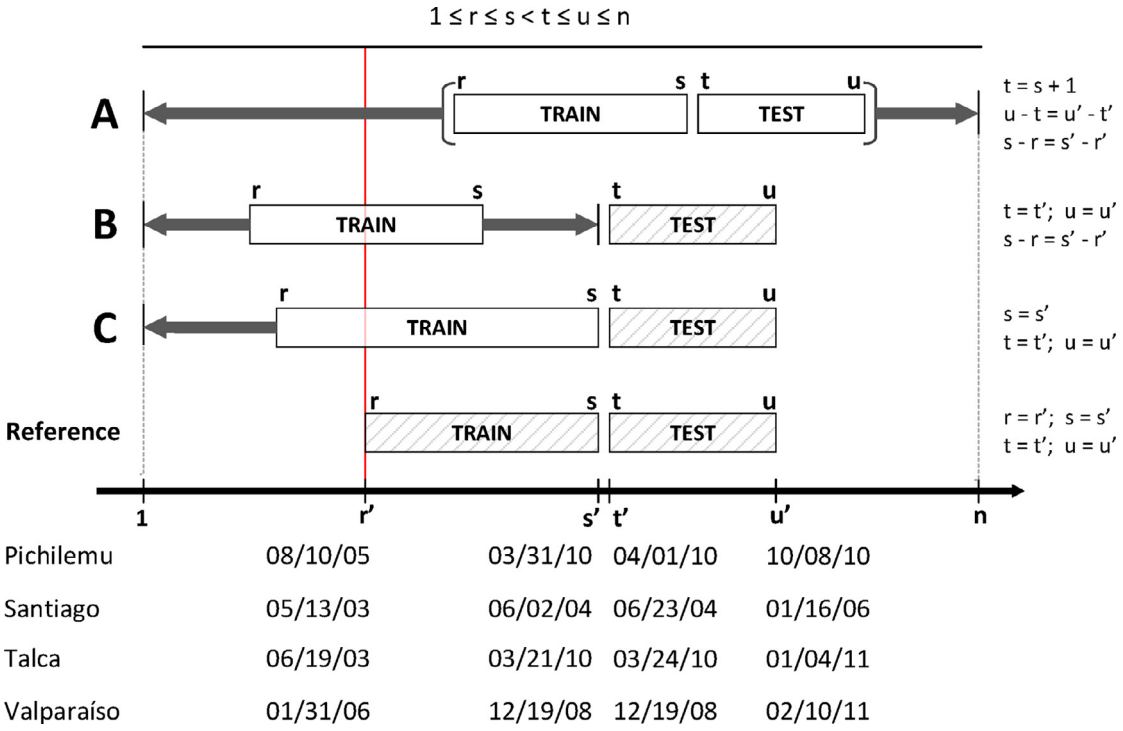


Fig. 2. Temporal variation of parameters r, s, t and u in the three analyses (A, B, C) of the Study #1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

may vary from one researcher to another, depending on what it is pretended to be evaluated. Section 4.3 defines the set of measures proposed in this work to evaluate the performance.

3.2. Sensitivity analysis for seismicity indicators

From the reading of the previous section, it can be easily concluded that there are many parameters that need to be adjusted before making any prediction. Five studies have been conducted to analyze the sensitive of some parameters. In particular: r, s, t and u have been analyzed in Section 3.2.1. b_{Reyes} and b_{Panak} have been studied in Section 3.2.2. Additionally, the influence of σ_1 and σ_2 is evaluated in Section 3.2.3. Section 3.2.4 analyzes ϕ. Finally, a sensitivity analysis of n_M and n_A is proposed in Section 3.2.5.

All results stemmed from these sensitivity analyses can be compared in terms of accuracy to those published in [8] and [9], since the same catalogues and thresholds M_s, M_0 and dtp have been used. These values, as well as temporal information (in format mm/dd/yy) about training and test sets have been shown in Table 5. In that table, r', s', t' and u' reference values have been indicated, and t_r, t_s, t_t, t_u stand for data and time of events $e_{r'}$, $e_{s'}$, $e_{t'}$ y $e_{u'}$. Note that $M_0 = 3.0$, $dtp = 5$, $n_M = 50$ and $n_A = 4$ for all the four datasets and $b_{Reyes} = b_M$, $b_{Panak} = b_A$.

3.2.1. Study #1: temporal distribution in training and test sets

The main goal of this study is to determine the influence of choosing different sets for training and testing, in terms of temporal distribution and size. Three different analyses (A, B C) have been performed when parameters r, s, t y u vary, since they directly affect both the training and test sets distribution.

Fig. 2 schematically shows the variation of temporal parameters r, s, t and u performed in each analysis (A, B, C) of the Study #1. The red vertical line indicates the origin of the training set of reference. The results of this study (shown in Section 4.4.1) refer to the red vertical line defined here. Each value of effectiveness shown in Figs. 4–6 was achieved varying only the parameter r and satisfying the restrictions shown in the right margin of the Fig. 2.

The first analysis (A) jointly slides training and test sets over time, by modifying the reference values r', s', t' and u' published in [8] and [9]. These values have been summarized in Table 5. For this purpose, sizes $s - r$, $u - t$ remain constant and t is set to $t = s + 1$. Therefore, the only parameter identifying this study is r (since all of them depend on it) and its values vary from 1 to $n - u + r'$.

The second analysis (B) consists in the addition of a new degree of freedom to both training and test sets: it is allowed that sets are not contiguous (there might exist events between training and test sets). The sizes $s - r$, $u - t$ remain unaltered but, this time t is not set to $t = s + 1$. The test set defined by (t, u) is fixed to the reference values shown in Table 5 (columns t' and u').

Finally, the size of the training set is also allowed to change in the third analysis (C). The size of the test set is not altered though. For this last study contiguity between sets is also required: $t = s + 1$. The size ratios between training and test sets vary from 30% to (100s/u)%, including thus the original ratios published in [8] and [9]. The test set is fixed to the reference values shown in Table 5 (columns t' and u').

Table 6
Results for study #2: *b*-value assessment type.

Dataset	Method	Conf. 1	Conf. 2	Conf. 3	Conf. 4	Conf. 5	Conf. 6	Conf. 7	Conf. 8	Conf. 9
Pichilemu	KNN	0.597	0.567	0.543	0.638	0.685	0.609	0.409	0.448	0.655
Pichilemu	NB	0.725	0.621	0.615	0.488	0.615	0.532	0.659	0.504	0.623
Pichilemu	ANN	0.623	0.653	0.514	0.651	0.582	0.575	0.412	0.509	0.599
Pichilemu	J48	0.309	0.668	0.74	0.713	0.692	0.713	0.713	0.669	0.713
Pichilemu	SVM	0.441	0.554	0.522	0.683	0.701	0.721	0.572	0.699	0.726
Pichilemu	(AVG)	0.539	0.613	0.587	0.635	0.655	0.63	0.553	0.566	0.663
Santiago	KNN	0.576	0.427	0.489	0.549	0.59	0.579	0.524	0.5	0.503
Santiago	NB	0.618	0.463	0.479	0.556	0.616	0.551	0.493	0.488	0.53
Santiago	ANN	0.54	0.516	0.482	0.447	0.58	0.514	0.543	0.56	0.616
Santiago	J48	0.62	0.521	0.561	0.441	0.496	0.432	0.537	0.449	0.56
Santiago	SVM	0.471	0.471	0.471	0.471	0.471	0.471	0.471	0.471	0.471
Santiago	(AVG)	0.565	0.48	0.496	0.493	0.551	0.509	0.514	0.494	0.536
Talca	KNN	0.502	0.596	0.298	0.556	0.512	0.523	0.487	0.498	0.378
Talca	NB	0.289	0.503	0.289	0.693	0.289	0.289	0.554	0.679	0.289
Talca	ANN	0.289	0.62	0.551	0.601	0.458	0.491	0.444	0.521	0.558
Talca	J48	0.641	0.542	0.404	0.7	0.702	0.653	0.702	0.702	0.653
Talca	SVM	0.289	0.461	0.48	0.48	0.289	0.289	0.533	0.533	0.289
Talca	(AVG)	0.402	0.544	0.404	0.606	0.45	0.449	0.544	0.587	0.433
Valparaíso	KNN	0.755	0.698	0.47	0.594	0.727	0.758	0.58	0.686	0.78
Valparaíso	NB	0.744	0.618	0.34	0.674	0.514	0.697	0.332	0.724	0.539
Valparaíso	ANN	0.726	0.765	0.709	0.695	0.6	0.609	0.663	0.695	0.727
Valparaíso	J48	0.753	0.784	0.676	0.506	0.733	0.733	0.683	0.494	0.716
Valparaíso	SVM	0.804	0.591	0.681	0.733	0.797	0.749	0.784	0.714	0.797
Valparaíso	(AVG)	0.756	0.691	0.575	0.64	0.674	0.709	0.608	0.663	0.712

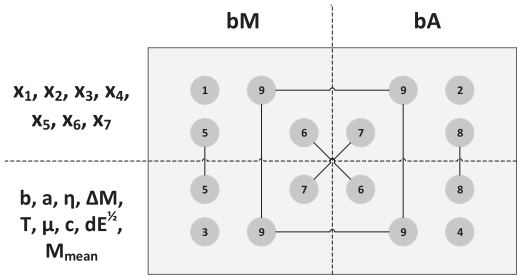


Fig. 3. The nine configurations of the Study #2 according to the set of attributes used and the *b* calculation type.

3.2.2. Study #2: b-value assessment type

Since the use of b_M and b_A for a given event e_i does not generate same values, the influence of using either one or the other is analyzed when attributes depending on b have been calculated: x_1, x_2, x_3, x_4, x_5, x_7, b, a, η y ΔM.

Hence, in this study the b-value assessment types, b_M or b_A, to obtain b_{Reyes} and b_{Panak} have been analyzed by using, jointly or separately, the sets of attributes proposed in [8] and [7].

Nine different configurations were designed and they are shown schematically in Fig. 3, that indicates the attribute sets used and the type of b-value calculation in each configuration.

Specifically, Configurations 1 and 2 contain the results achieved using the attributes [8] assessed with b_M and b_A, respectively. Configurations 3 and 4 contain the results achieved using the attributes [7] assessed with b_M and b_A, respectively. Configuration 5 uses attributes in [7] and [8] both assessed with b_M. Configuration 6 uses attributes in [8] with b_M plus attributes in [7] with b_A. Configuration 7 attributes in [8] with b_A plus attributes in [7] with b_M. Configuration 8 uses attributes in [7] and [8] both assessed with b_A. Finally, Configuration 9 uses all possible combinations: attributes in [8] with b_M and b_A and attributes in [7] with b_M and b_A.

The Table 6 in Section 4.4.2 shows the effectiveness achieved for each dataset and method according to the nine different configurations introduced here.

The relationship between the results achieved in [8,9] and those presented in this study is the following. The work in [8] presented results only for the Configuration 1. By contrast, the results presented in [9] were both for Configuration 1 and 4. Moreover, a feature selection over Configuration 6 was performed in [9]. In this work the study of the rest of combinations have been extended in order to clarify what are the possible improvement choosing the attribute set and the type of b-value calculation.

3.2.3. Study #3: on the sensitivity of parameters σ_1 and σ_2 in T, μ and c

Parameters σ_1 and σ_2 indicate the range of thresholds for θ in T, μ and c, as explained in Section 3.1.2. In particular, σ_1 affects the number of attributes with $\theta \leq M_s$. On the contrary, σ_2 affects the number of attributes with $\theta > M_s$.

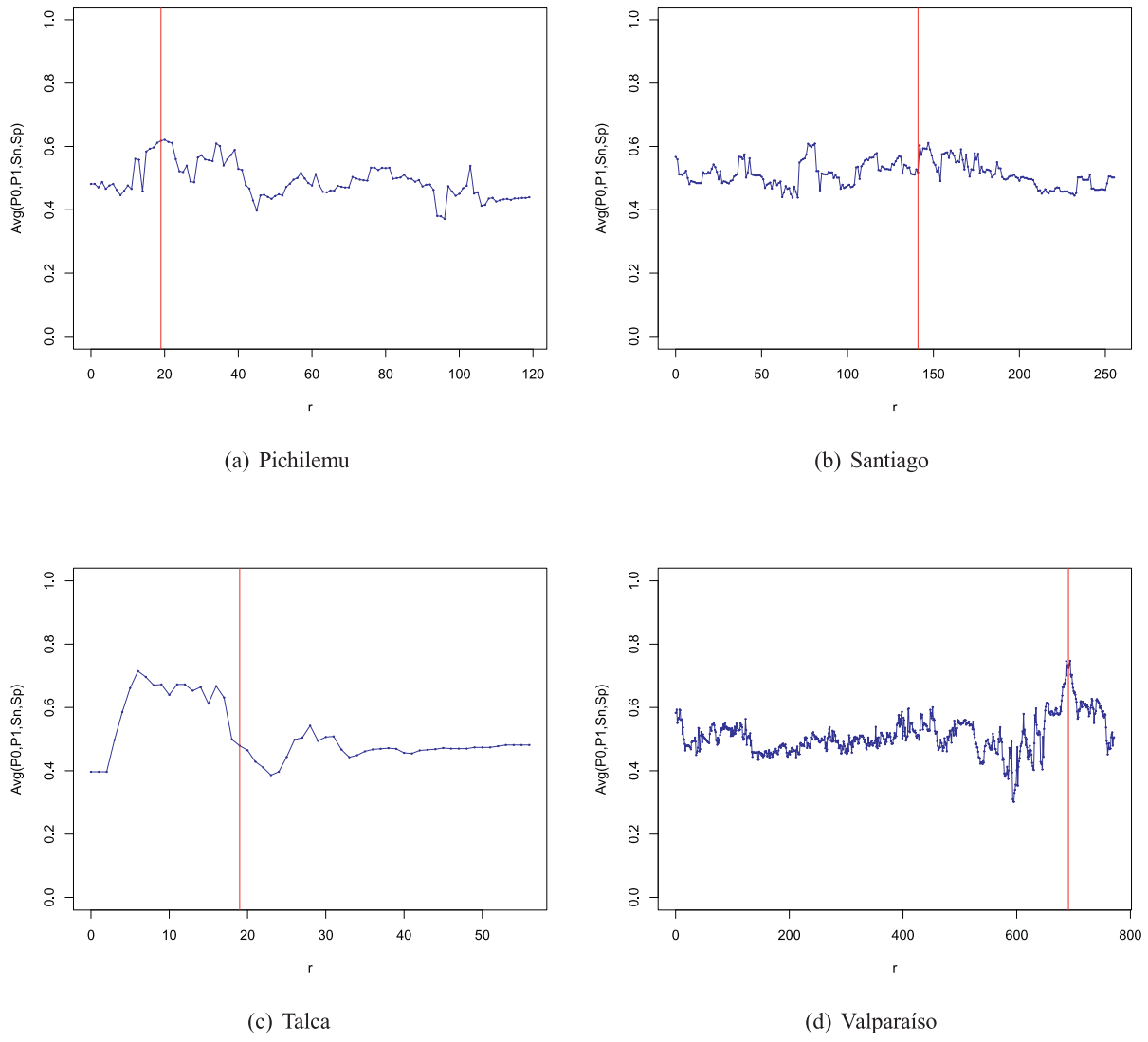
Reference values in [9] refer to the calculation of attributes T, μ and c with no threshold. Consequently, in this study differences for both preset thresholds $\theta = M_s$ ($\sigma_1 = \sigma_2 = 0$) as well as varying thresholds ($\sigma_1 \neq 0 \vee \sigma_2 \neq 0$) have been analyzed.

3.2.4. Study #4: on the sensitivity of parameter ϕ in μ and c

Parameter ϕ determines the range of approximation of magnitudes that identifies a characteristic event within a series. Generally speaking, given a fixed distribution of magnitudes, if ϕ decreases to values close to 0 then the number of characteristic events will probably decrease too and, therefore, the average elapsed time (used in attributes μ and c) will increase. By contrast, a greater number of characteristic events will be encountered if ϕ increases, decreasing thus the average elapsed time. Parameter ϕ is then varied in this study to determine its influence in terms of accuracy for the classifiers.

3.2.5. Study #5: on the sensitivity of the number of previous events considered, n_M and n_A

Last, a sensitivity analysis for parameters n_M and n_A involved in the calculation of b_M (along with x_1, x_2, x_3, x_4, x_5, x_6 and x_7) and b_A (along with b, a, η, ΔM, $T_{\sigma_1\sigma_2}$, $\mu^{\phi}_{\sigma_1\sigma_2}$, $c^{\phi}_{\sigma_1\sigma_2}$, $dE^{1/2}$ and M_{mean}) has been performed.



(a) Pichilemu (b) Santiago

(c) Talca (d) Valparaíso

Fig. 4. Results for study #1A: sliding training+test window. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Results

This section gathers the results obtained from the studies introduced in the previous section. First, a description on how datasets have been retrieved and preprocessed can be found in Section 4.1. Later, in Section 4.2 the data and settings have been described. Section 4.3 introduces the set of quality parameters used to assess the system performance. Finally, Section 4.4 presents and discusses all the results for the experiments that have been proposed in Section 3.2.

4.1. Datasets description

The catalogue of earthquakes used in this paper has been obtained from the Chile's National Seismological Service. This database was evaluated in [8]. It was determined that 2001 is the year of completeness of the catalogue, for earthquakes of magnitude larger than or equal to 3.0 for the four zones.

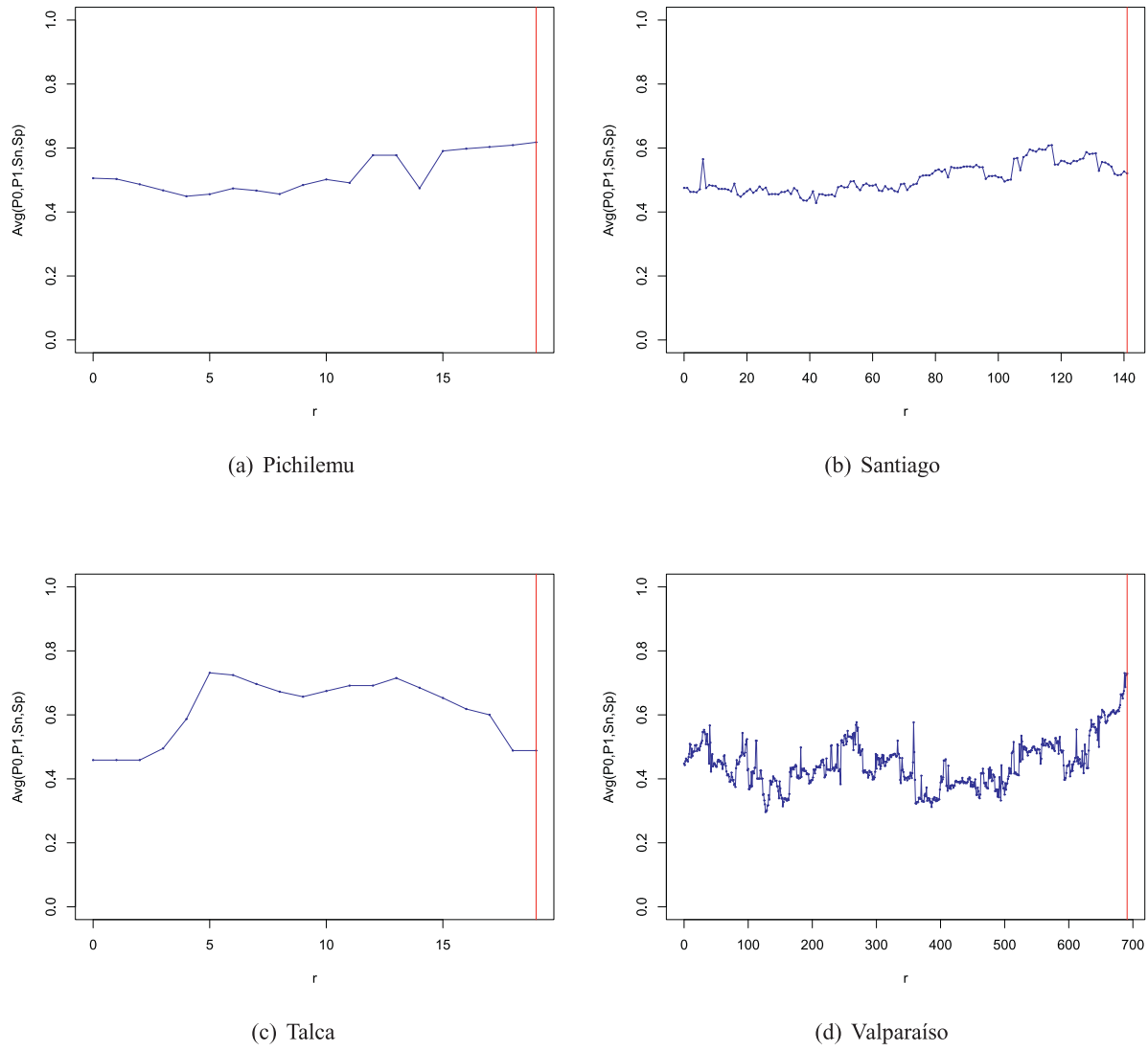
4.2. Configuration of the experimentation

In this section, data and parameter settings used in the five studies described previously are provided. A baseline configuration has been introduced, common to all experiments. Each experiment only introduces one change in these data and parameters from the baseline configuration, in order to analyze the impact of such change in prediction accuracy.

The baseline configuration corresponds to experimental conditions used in [8,9] and indicated in Section 3.2. The four geographic zones shown in Table 5 have been considered, using the specified temporary decomposition for training and test (t_r, t_s, t_t, t_u).

The attributes used have been taken from both groups [8] and [7] described in the methodology section. Specifically, the attributes used have been x_1, x_2, x_3, x_4, x_5, x_6, x_7, b, a, η, ΔM, T, μ, c, $dE^{1/2}$ and M_{mean}. To assess these attributes, in the baseline configuration $b_{Reyes} = b_M$, $b_{Panak} = b_A$, $M_0 = 3$, $n_M = 50$ and $n_A = 4$ have been used. To assess the class, dtp is equal to 5 and M_s values have been indicated in Table 5.

The applied classifiers have been: Nearest Neighbors (KNN) with $k = 1$, Naive Bayes (NB), Support Vector Machines (SVM) with $c = 1$ and polynomial kernel of degree 1, C4.5 decision trees (C45) with a confidence factor of 0.25 and 2 minimum instances per leaf and Artificial Neuronal Networks (ANN) with backpropagation, sigmoid nodes, one hidden layer, as many inputs as attributes (16 attributes in baseline configuration) and two



(a) Pichilemu (b) Santiago

(c) Talca (d) Valparaíso

Fig. 5. Results for study #1B: sliding training window (constant reference test). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

outputs (corresponding to the two values of the class specified in Section 3.1.3).

4.3. Quality parameters

To assess the performance of any classifier, several parameters have been used. Here, those used in the works in which this analysis has been inspired [7,8] have been used. In particular:

1. True positives (TP). The number of times that an upcoming earthquake was properly predicted.
2. True negatives (TN). The number of times that neither a classifier triggered an alarm nor an earthquake occurred.
3. False positives (FP). The number of times that a classifier erroneously predicted the occurrence of an earthquake.
4. False negatives (FN). The number of times that a classifier did not trigger an alarm but an earthquake did occur.

The combination of these parameters leads to the calculation of:

$$P_0 = \frac{TN}{TN + FN} \tag{12}$$

$$P_1 = \frac{TP}{TP + FP} \tag{13}$$

where P_0 denotes the well-known negative predictive value, and P_1 the well-known positive predictive value.

Additionally, two more parameters corresponding to widely used statistical measures for supervised classifiers have been used to evaluate their performance. These two parameters, sensitivity or rate of actual positives correctly identified as such (denoted by S_n) and specificity or rate of actual negatives correctly identified (denoted by S_p), have been defined as:

$$S_n = \frac{TP}{TP + FN} \tag{14}$$

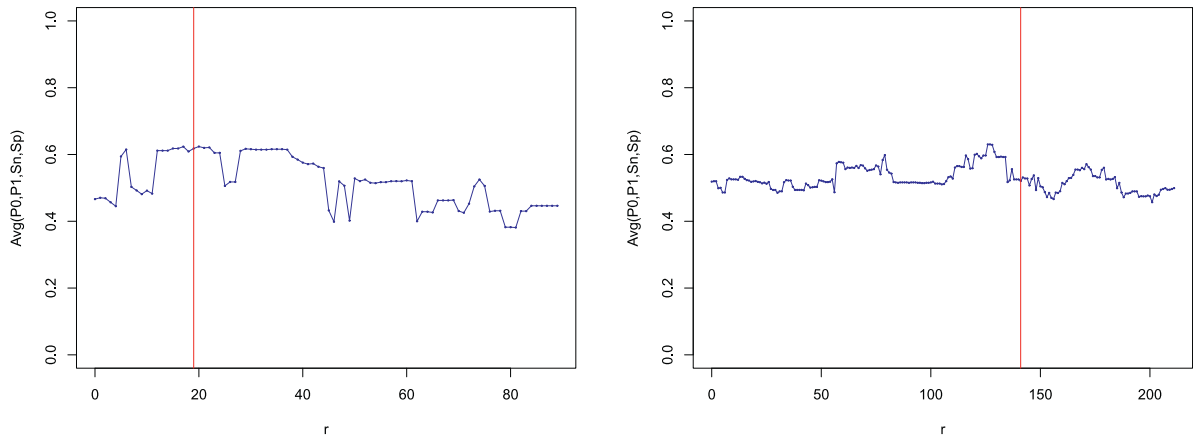
$$S_p = \frac{TN}{TN + FP} \tag{15}$$

Finally, the sensitivity analysis carried out in the following section has been done on the basis of the average value of all these parameters. This is:

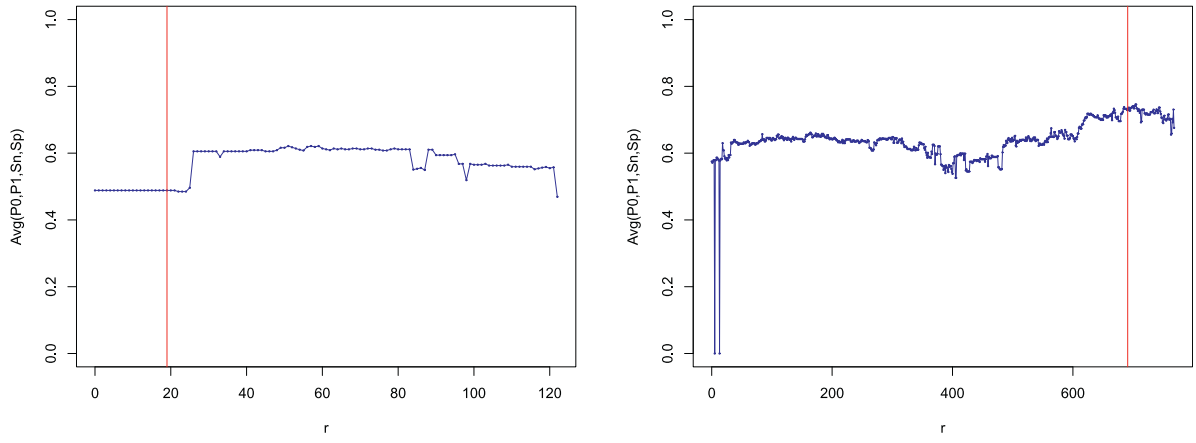
$$Avg(P_0, P_1, S_n, S_p) = 0.25(P_0 + P_1 + S_n + S_p) \tag{16}$$

4.4. Sensitivity analysis

Five studies of sensitivity have been presented. The baseline configuration shown in Section 4.2 has been kept and only some



(a) Pichilemu (b) Santiago



(c) Talca (d) Valparaíso

Fig. 6. Results for study #1C: modifying the training size (constant reference test). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

values of the model have been modified. In these studies the values for $Avg(P_0, P_1, S_n, S_p)$ have been graphically depicted. These have been obtained as the average of the five classifiers indicated in Section 4.2 (KNN, NB, SVM, C45, ANN).

4.4.1. Results for study #1. Temporal distribution in training and test sets.

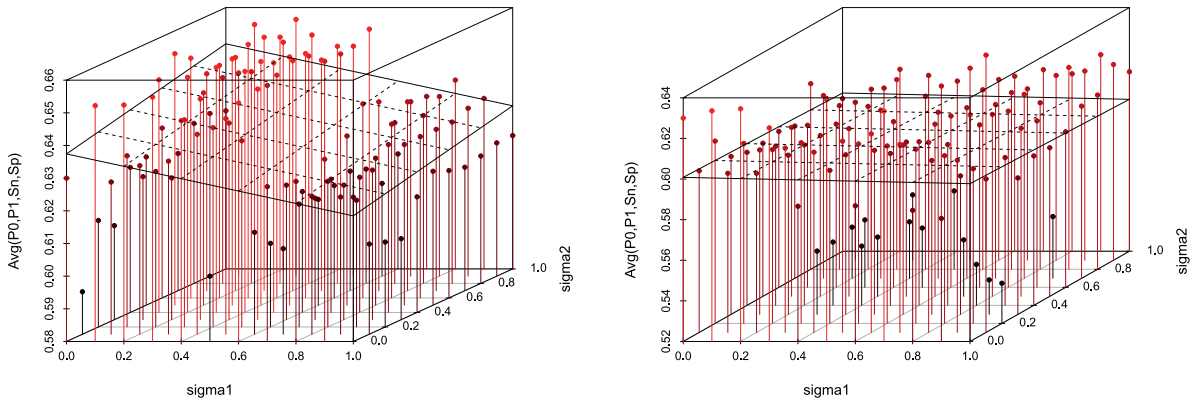
Three kind of temporary studies have been described. The values of *r, s, t* and *u*, described in the methodology, have been modified. The values of $Avg(P_0, P_1, S_n, S_p)$ in function of *r* have been graphically depicted. Aforementioned is that *r* decides *s, t* and *u*. In the graphics shown in this research a vertical red line has been included. It points out the reference value r' specified in Table 5.

4.4.1.1. Sliding training+test window. Fig. 4 plots the effectiveness distribution when the set of adjacent training and test have been shifted in time. Pichilemu and Santiago have a similar behavior with an effectiveness value of 0.4–0.6. The deviation of the values has also been also similar, even considering that the catalogue of Santiago is twice the length of Pichilemu's. The range of values for Talca and Valparaíso has been wider: 0.3–0.7. This makes that the choice of the study set within the available catalogue has a great importance in the predictions effectiveness. This has been even more relevant for Valparaíso, due to the large length of the catalogue, which increases the choice possibilities within the study-set (from $r = 0$ to $r = 771$ choice possibilities).

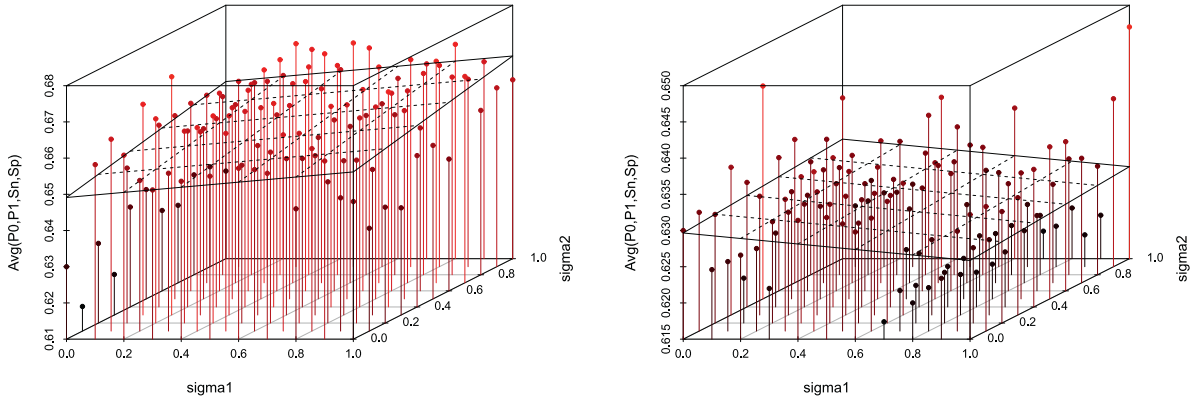
The lineal trend of the effectiveness in the four zones is almost horizontal. The values of the slope of the regression line have been close to zero. This demonstrates, according to the data of the studied zones, that, a priori, there is not a better origin that produces more quality predictions.

4.4.1.2. Sliding training window (constant reference test). Fig. 5 depicts the effectiveness distribution when the training test has been moved backwards in time from its reference value (*r* values less than the reference value r'). The test set has been constant with regard to the reference value ($t = t'$ and $u = u'$). For this reason, the vertical red line has always been plotted in the last *r* value shown, which corresponds to the reference value r'. This study allows evaluating the effect in the effectiveness that produces the insertion of gaps of incremental size between the training and the test sets.

The figure shows that the effectiveness distribution for *r* values minor than the reference one behaves approximately the same than the previous study, for the same values of *r*. This result was



(a) ALL (b) T



(c) μ (d) c

Fig. 7. Study #3: on the sensitivity of parameters σ_1 and σ_2 in T, μ and c for Pichilemu.

expected because the training sets have been exactly the same, although the test sets have been different.

Nevertheless, except for Talca, a positive slope for the linear regression has been observed. This means that the r values that move away from the reference value decrease the effectiveness. In other words, it has been proved that introducing gaps between the training and the test deteriorates the predictions quality proportionally to the gap size.

4.4.1.3. Modifying the training size (constant reference test). Fig. 6 presents the effectiveness distribution when the training set size has been modified and therefore the r value has been varied. The values of s, t and u have been kept constant with regard to their reference values. The size of the training set has been modified from a proportion of 30% to 70% to the maximum training size allowed by the available catalogue. This study allows testing the impact in the effectiveness produced by the size of the training set.

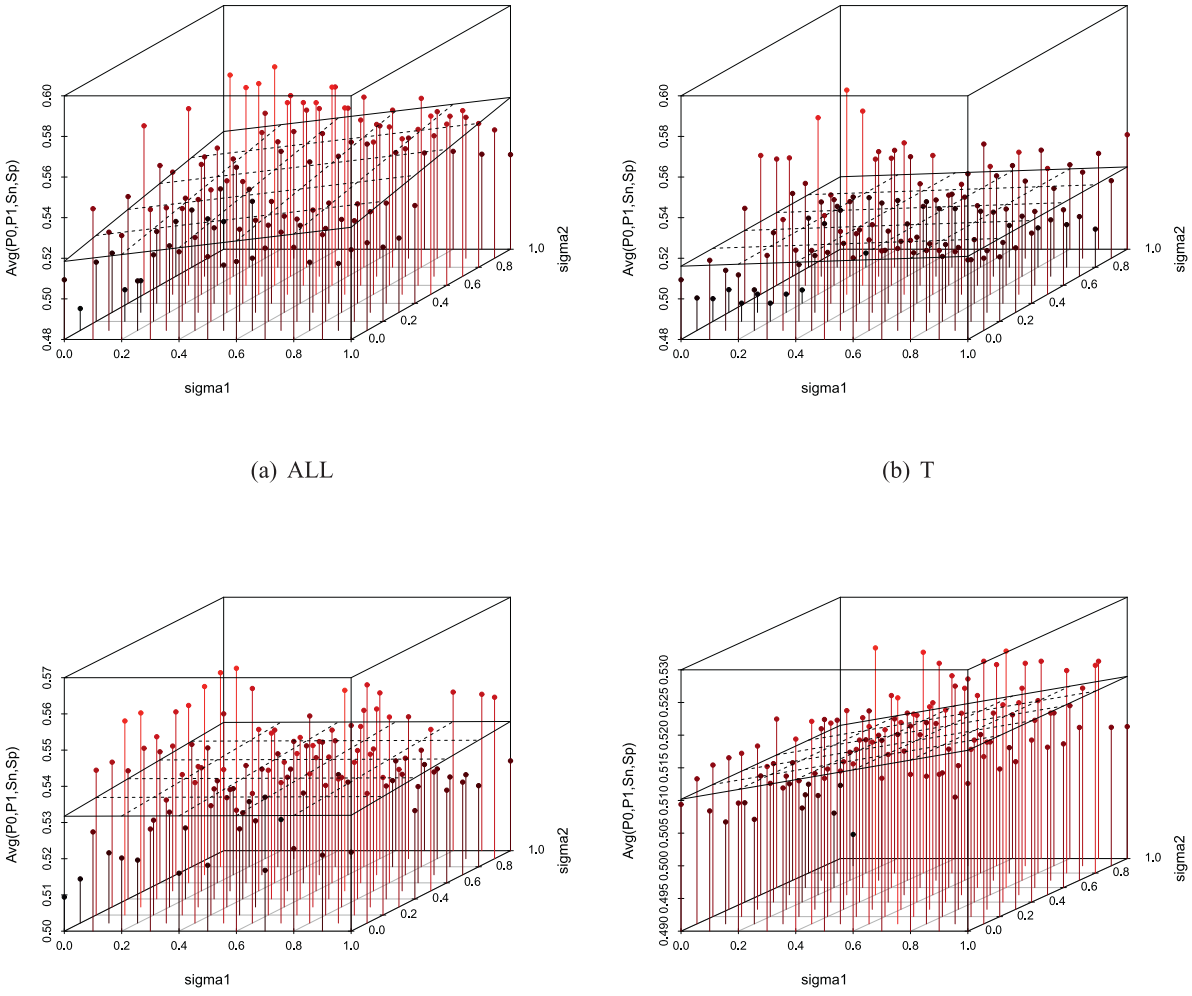
Excepting Pichilemu, there has not been a clear trend in the predictions quality due to the reduction of the training set, as it could be a priori supposed in supervised learning methods. These results point the possibility of increasing the efficiency (execution time and memory usage) of prediction methods without losing their effectiveness, because smaller training sets could be used.

4.4.2. Results for study #2: b*-value assessment type*

In this study the assessment of the b-value has been analyzed in order to check a possible impact to the prediction effectiveness. In order to perform a fair comparison between the two types of b-value assessments, b_M and b_A, their dependent parameters n_M and n_A have been set to the same value, $n_M = n_A = 50$, which has been the value used in the reference works [8,9]. This study includes comparisons between b_M and b_A for [8] attribute set, [7] attribute set and both attribute sets together in different combinations.

Table 6 shows the effectiveness, $Avg(P_0, P_1, S_n, S_p)$, achieved for each dataset and method according to the nine different configurations presented in Fig. 3.

No significant variations of effectiveness has been found in general terms over all methods and studied datasets. The average difference of effectiveness, $Avg(S_n, S_p, P_0, P_1)$, between b_M and b_A has been 0.014 with a standard deviation of 0.111. Some isolated cases of an effectiveness difference higher than 0.2 have been though found. Specifically, the algorithm J48 (a decision tree-based classifier), when it has been applied to the catalogue of Talca, b_A produces better results for attributes in [8] (together with attributes in [7]) than the b_M, with a difference of 0.33. As for the ANN algorithm in Pichilemu, b_M produces better results for attributes in [7] (together or not with attributes in [8]) than b_A, with a difference of 0.36.



(a) ALL (b) T

(c) μ (d) c

Fig. 8. Study #3: on the sensitivity of parameters σ_1 and σ_2 in *T*, μ and *c* for Santiago.

Regarding to the improvement achieved using b_M or b_A as calculation type of the *b*-value, the results of the Configurations 5 and 8 can be compared; also the Configurations 1 versus 2, and 3 versus 4. With respect to Configurations 5 and 8, it can be observed that, on average, there are two cities (Pichilemu and Talca) with a difference higher than 0.1 in effectiveness: best b_M for Pichilemu, 0.655 versus 0.553, and best b_A for Talca, 0.587 versus 0.45. Contrarily, the cities Santiago and Valparaíso do not present important variations in effectiveness. It is remarkable that some algorithms are specially sensitive to the *b* calculation. For example, in Pichilemu, KNN gains 0.24 of effectiveness with b_M, or NB improves 0.39 with b_A in Talca.

The comparison between results of Configurations 1 and 2 reveals that while the city of Talca, on average, improves using b_A (0.14) as before, the city of Pichilemu improves using b_A (0.07) for the first set of attributes ($x_1, \ldots, x_7$). With respect to Configurations 3 and 4, Talca presents the higher difference, on average, with improvement of 0.2 of effectiveness using b_A, mainly due to the algorithms NB and J48.

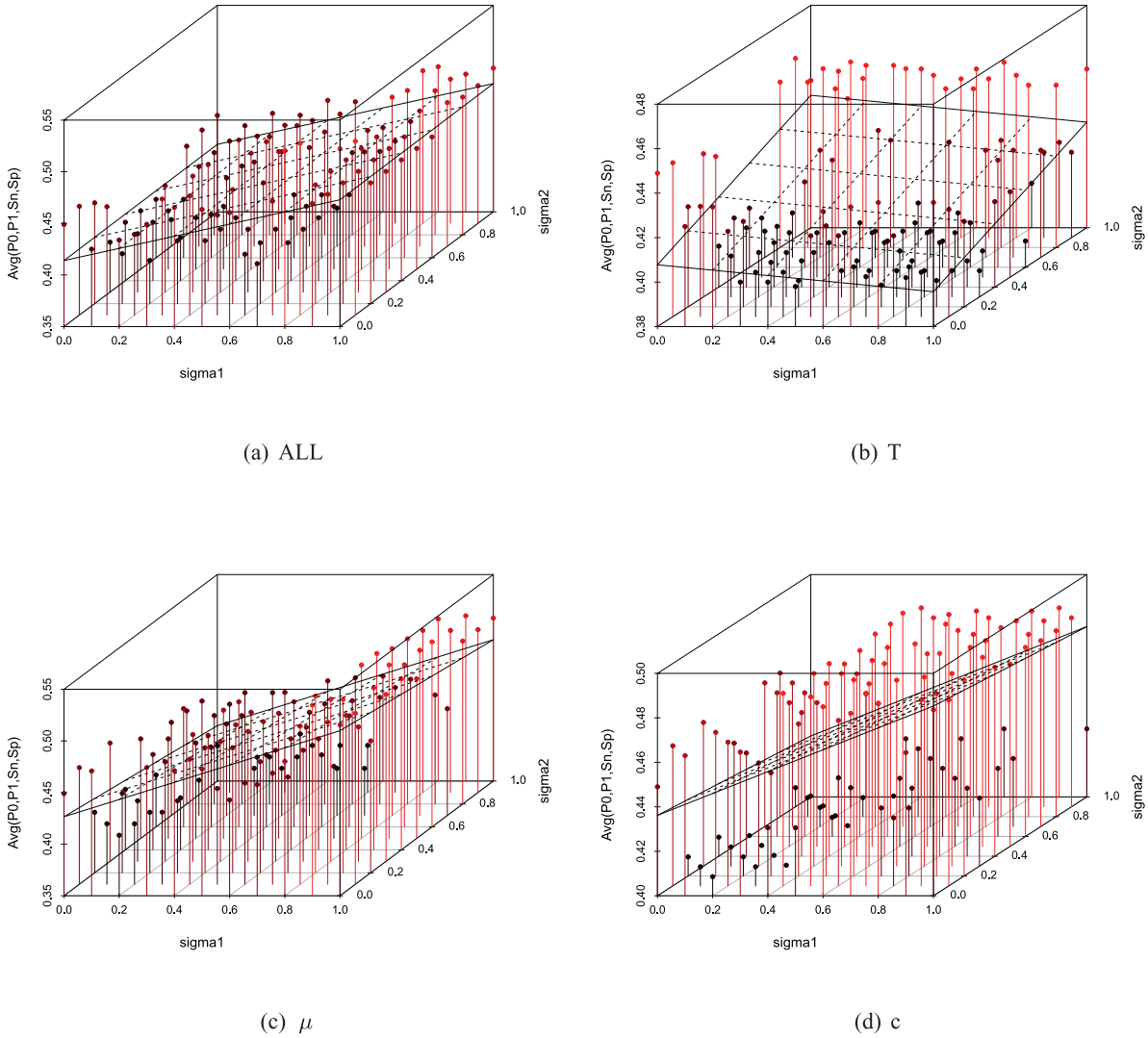
4.4.3. Results for study #3: on the sensitivity of parameters σ_1 and σ_2 in T, μ and c

In this study, the effectiveness distribution in function of the variation of σ_1 and σ_2 has been analyzed. Parameters σ_1 and σ_2 affect on *T*, μ and *c*, within the set of values $[0, 0.1, 0.2, \ldots, 1]$. 3D-graphics illustrate four possibilities: *ALL*, where *T*, μ and *c* have been modified with the same values of σ_1 and σ_2; *T*, where σ_1 and σ_2 just affects on *T*, keeping the *c* and μ values constant; μ, analogously for μ; and *c*, modifying only *c*. Figs. 7–10 summarize these results.

With regard to *ALL*, from the regression planes of the graphics, high values of σ_2 produce better rates of effectiveness for Pichilemu and Santiago, with particularly low values of σ_1 for Pichilemu. By contrast, high σ_2 values mean considering earthquakes of magnitude noticeably larger than the reference threshold, as it was explained in the methodology. High values of σ_1 produce better results in effectiveness regardless of the σ_2 values for Talca. High values of σ_1 implies considering shakes of magnitude noticeably minor than the reference threshold. No clear trend in the values of σ_1 and σ_2 has been found for Valparaíso.

Regarding the changes in *T*, μ and *c*, in the four zones analyzed, similar patterns to the ones found for *ALL* have been found. In general, a wider range between the minimum and the maximum value of effectiveness have been found for Talca and Valparaíso, 0.35–0.55 and 0.55–0.80, respectively, compared to Pichilemu and Santiago (0.52–0.68 and 0.48–0.60, respectively).

These noticeable variations of the effectiveness in function of σ_1 and σ_2, even considering average values among five different



(a) ALL (b) T

(c) μ (d) c

Fig. 9. Study #3: on the sensitivity of parameters σ_1 and σ_2 in T, μ and c for Talca.

methods, point to the possibility of automatically optimizing the values of σ_1 and σ_2 in order to obtain the best effectiveness, depending on the zone and prediction method.

4.4.4. Results for study #4: on the sensitivity of parameter ϕ in μ and c

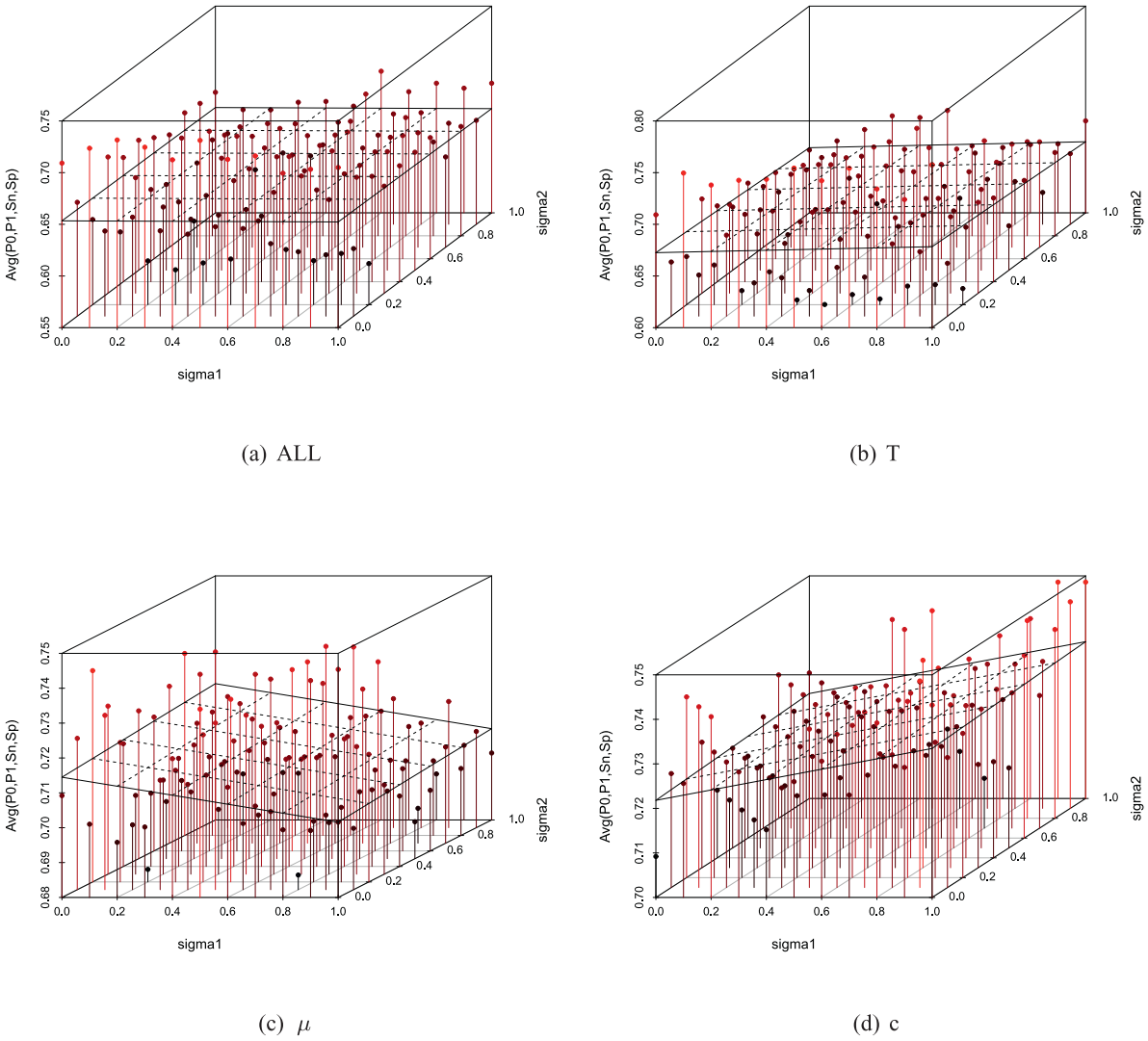
In this study values of parameter ϕ have been varied from 0 to 1, by increments of 0.05. From the observation of Fig. 11, it can be concluded that there has not been a remarkable difference in the effectiveness in function of ϕ for Pichilemu, Santiago and Valparaíso. Nonetheless, the effectiveness oscillates between 0.42 and 0.58 for Talca. Therefore, it has been found that for some seismic zones, such as Talca, there are ϕ values that produce significantly better predictions. Consequently, it could be promising to optimize ϕ with regard to the seismic zone and the method selected to predict.

4.4.5. Results for study #5: on the sensitivity of the number of previous events considered

In this study the parameters n_M and n_A of the proposed model have been analyzed in terms of their impact on prediction effectiveness. The parameters n_M and n_A affect to the assessment of attributes b_M and b_A. Values of n_M and n_A have been simultaneously changed ($n_M = n_A = n$) and they belong to the interval [4, 60] in this study. It is important to highlight that both parameters n_M and n_A refer to the same conceptual parameter (n, number of previous events) in the b-value assessment. The only difference between n_M and n_A is the application to a set of attributes of another. Since the goal of this study is to check the sensitivity on the number of previous events using both set of attributes ($x_1, \ldots, x_7$ and b, a, η, ΔM, T, μ, c, $dE^{1/2}$, M_{mean}), the parameters n_M and n_A must be equal. The results of this study are illustrated in Fig. 12.

Excepting Santiago, where there has not been a noticeable variation of the effectiveness in function of n, in the other three cities, the effectiveness variation has been remarkable. Thus, n could be automatically calculated to maximize the effectiveness of an established training set.

Increasing trends in effectiveness when the n value increases have been found for Pichilemu and Valparaíso. By contrast, decreasing trends have been found for Talca. A smaller value of n allows using minor training sets and, hence, smaller catalogues. This could be especially relevant for zones where the catalogue is short or when the minimum magnitude threshold for prediction is high, $M > 6$, for example. It should be noted that the effectiveness values range between 0.5 and 0.7 for all the n values. Thus, the goal could be finding the minimum value of n that maximizes the effectiveness in a training set for a zone and for a specific prediction method.



(a) ALL (b) T

(c) μ (d) c

Fig. 10. Study #3: on the sensitivity of parameters σ_1 and σ_2 in T, μ and c for Valparaíso.

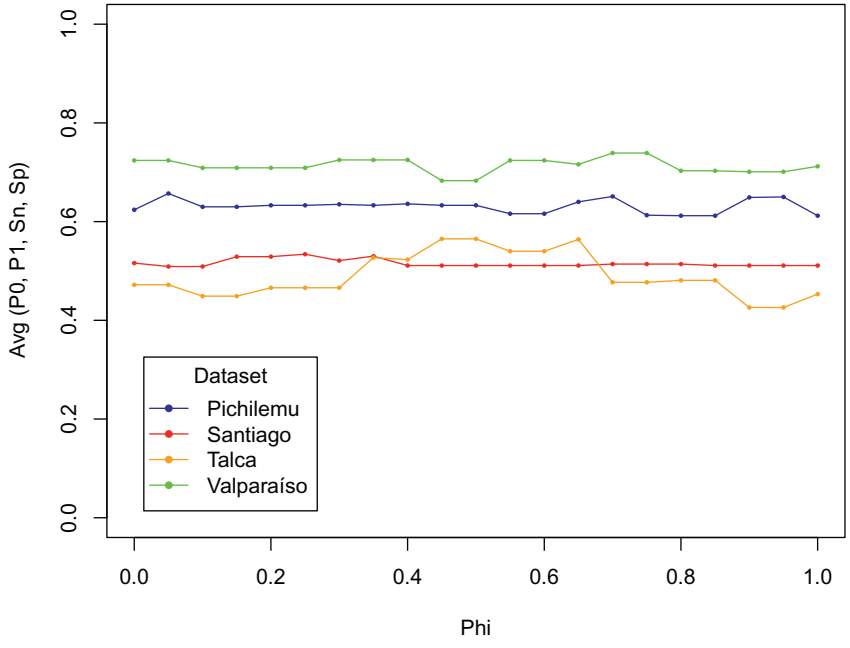


Fig. 11. Results for study #4: on the sensitivity of ϕ in μ and c.

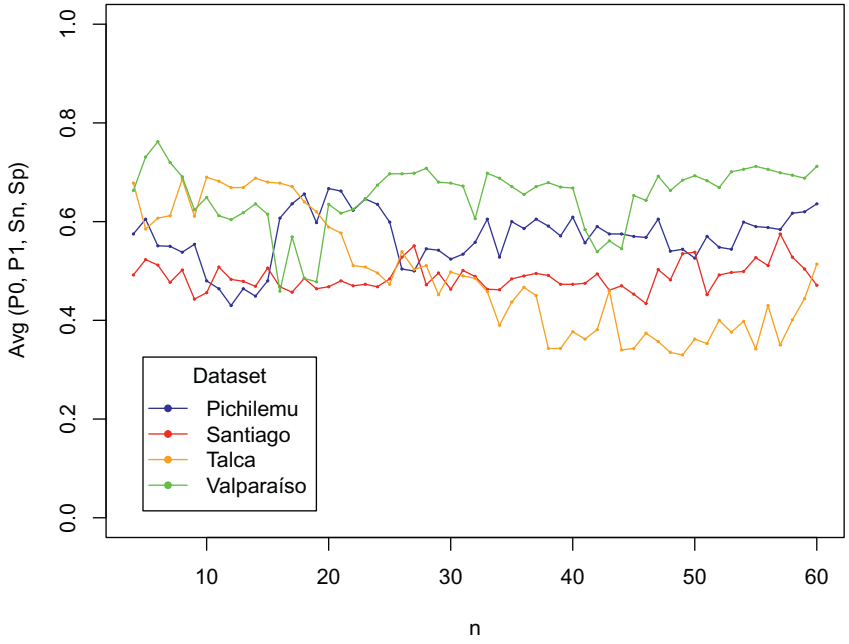


Fig. 12. Results for study #5: on the sensitivity of the number of previous events considered, n_M and n_A.

5. Conclusions

This work has thoroughly examined how an adequate use of certain seismicity indicators may lead to the obtention of better accuracy when the earthquake prediction problem has been tackled. In particular, these indicators have been used as inputs in several well-known machine learning classifiers. A set of indicators already proposed in the literature has been analyzed showing that if they would have been adequately tuned in the original works, the prediction accuracy could have been improved. In order to make robust predictions and to generalize the tuning of inputs, a new methodology that exhaustively explores how certain parameters must be set up has been proposed. Although its performance has been assessed in four Chilean cities, it has been designed as a general purpose method. The steps described can be followed regardless of the zone under study. The results achieved show that by properly adjusting the inputs, the prediction ability for certain classifiers has been clearly outperformed. These results pose new challenges in earthquake prediction such as optimizing and developing *ad hoc* systems making use of all the available information, as well as discovering new indicators in order to provide the system with more information.

The optimization of the parameters that have produced significant fluctuations of the effectiveness in this work (r, s, t, u, b_A/b_M, σ_1, σ_2 and n; but not ϕ) should consider the following procedure to find the best value of the parameters that maximize the effectiveness. First, a zone and a prediction algorithm must be selected. Next, the catalogue of earthquake events is divided into three consecutive parts: the training set, the validation test set and the blind test set. Then, an optimization algorithm (for example, a metaheuristic-based algorithm) should be applied by varying the value of the parameters. For each iteration, a specific set of values for parameters are used to train the prediction algorithm using the training set. Next, the validation test set is predicted and the error (or the desired goodness measure) is assessed. At the end of such process, the set of parameter values that minimizes the error is selected. Finally, these parameter values are used to predict the blind test set.

Acknowledgements

This work has been supported by the Spanish Government through research project P12-TIC-1728, and by NT2 TGT through Grants no. 2210 and 2211.

References

[1] K.F. Tiampo, R. Shcherbakov, Seismicity-based earthquake forecasting techniques: Ten years of progress, Tectonophysics 522-523 (2012) 89–121.
[2] H. Adeli, A. Panakkat, A probabilistic neural network for earthquake magnitude prediction, Neural Netw. 22 (2009) 1018–1024.
[3] A. Alexandridis, E. Chondrodima, E. Efthimiou, G. Papadakis, Large earthquake occurrence estimation based on radial basis function neural networks, IEEE Trans. Geosci. Remote Sens. 52 (9) (2014) 5443–5453.
[4] F. Martínez-Álvarez, A. Troncoso, A. Morales-Esteban, J.C. Riquelme, Computational intelligence techniques for predicting earthquakes, Lect. Notes Artif. Intell. 6679 (2) (2011) 287–294.
[5] A. Zamani, M.R. Sorbi, A.A. Safavi, Application of neural network and ANFIS model for earthquake occurrence in Iran, Earth Sci. Inf. 6 (2) (2013) 71–85.
[6] A. Morales-Esteban, F. Martínez-Álvarez, J. Reyes, Earthquake prediction in seismogenic areas of the Iberian Peninsula based on computational intelligence, Tectonophysics 593 (2013) 121–134.
[7] A. Panakkat, H. Adeli, Neural network models for earthquake magnitude prediction using multiple seismicity indicators, Int. J. Neural Syst. 17 (1) (2007) 13–33.
[8] J. Reyes, A. Morales-Esteban, F. Martínez-Álvarez, Neural networks to predict earthquakes in Chile, Appl. Soft Comput. 13 (2) (2013) 1314–1328.
[9] F. Martínez-Álvarez, J. Reyes, A. Morales-Esteban, C. Rubio-Escudero, Determining the best set of seismicity indicators to predict earthquakes. Two case studies: Chile and the Iberian Peninsula, Knowl. Based Syst. 50 (2013) 198–210.
[10] A. Morales-Esteban, F. Martínez-Álvarez, A. Troncoso, J.L. de Justo, C. Rubio-Escudero, Pattern recognition to forecast seismic time series, Expert Syst. Appl. 37 (12) (2010) 8333–8342.
[11] J. Reyes, V. Cárdenas, A Chilean seismic regionalization through a Kohonen neural network, Neural Comput. Appl. 19 (2010) 1081–1087.
[12] R.J. Geller, Earthquake prediction: A critical review, Geophys. J. Int. 131 (3) (1997) 425–450.
[13] T.H. Jordan, Earthquake predictability, brick by brick, Seismol. Res. Lett. 77 (2006) 3–6.
[14] T. Matsuzawa, T. Igarashi, A. A. Hasegawa, Characteristic small-earthquake sequence off Sanriku, northeastern Honshu, Japan, Geophys. Res. Lett. 29 (11) (2002).1543–147
[15] M. Wyss, Cannot earthquakes be predicted? Science 278 (1997) 487–490.
[16] C.R. Allen, Responsibilities in earthquake prediction, Bull. Seismol. Soc. Am. 66 (1982) 2069–2074.
[17] D. Schorlemmer, M.C. Gerstenberger, RELM testing center, Seismol. Res. Lett. 78 (1) (2007) 30–36.

[18] Y. Ben-Zion, V. Lyakhovsky, Accelerated seismic release and related aspects of seismicitypatterns on earthquake faults, Pure Appl. Geophys. 159 (2002) 2385–2412.
[19] T.H. Jordan, L.M. Jones, Operational earthquake forecasting: Some thoughts on why and how, Seismol. Res. Lett. 81 (4) (2010) 571–574.
[20] S. Wiemer, D. Schorlemmer, ALM: An Asperity-based likelihood model for California, Seismol. Res. Lett. 78 (1) (2007) 134–143.
[21] V.G. Kossobokov, L.L. Romashkova, V.I. Keilis-Borok, Testing earthquake prediction algorithms:statistically significant advance prediction of the largest earthquakes in the Circum-Pacific, 1992–1997, Phys. Earth Planet. Inter. 111 (1999) 187–196.
[22] A. Peresan, V. Kossobokov, L. Romashkova, G.F. Panza, Intermediate-termmiddle-range earthquake predictions in Italy: A review, Earth Sci. Rev. 69 (2005) 97–132.
[23] R.D. Giovambattista, Y.S. Tyupkin, Spatial and temporal distribution of the seismicitybefore the Umbria–Marche September 26, 1997 earthquakes, J. Seismol. 4 (2000) 589–598.
[24] G.A. Sobolev, On applicability of the RTL prognostic algorithms and energy estimation to Sakhalin seismicity, J. Volcanol. Seismol. 1 (2007) 198–211.
[25] C. Yin, P. Mora, Stress reorientation and LURR: Implication for earthquake prediction using LURR, Pure Appl. Geophys. 163 (2006) 2363–2373.
[26] C. Yin, H.L. Xing, P. Mora, H.H. Xu, Earthquake trend around Sumatra indicated by a new implementation of LURR method, Pure Appl. Geophys. 165 (2008) 723–736.
[27] F.F. Evison, D.A. Rhoades, The precursory earthquake swarm in New Zealand: Hypothesis tests II, J. Geol. Geophys. 40 (1997) 537–547.
[28] Q. Huang, Search for reliable precursors: A case study of the seismic quiescence of the 2000 western Tottori prefecture earthquake, J. Geophys. Res. 11 (2006) B04301.
[29] D. Vere-Jones, The development of statistical seismology: A personal experience, Tectonophysics 413 (2006) 5–12.
[30] Y. Ogata, Statistical models for earthquake occurrences and residual analysis for point processes, J. Am. Stat. Assoc. 83 (401) (1988) 8–27.
[31] J.D. Zechar, T.H. Jordan, Simple smoothed seismicity earthquake forecasts for italy, Ann. Geophys. 53 (3) (2010) 99–105.
[32] A. Morales-Esteban, F. Martínez-Álvarez, S. Scitovski, R. Scitovski, A fast partitioning algorithm using adaptive mahalanobis clustering with application to seismic zoning, Comput. Geosci. 73 (2014) 132–141.
[33] A.S.N. Alarifi, N.S.N. Alarifi, S. Al-Humidan, Earthquakes magnitude predication using artificial neural network in northern Red Sea area, J. King Saud Univ. Sci. 24 (2012) 301–313.
[34] A. Ikram, U. Qamar, Developing an expert system based on association rules and predicate logic for earthquake prediction, Knowl. Based Syst. 75 (2015) 87–103.
[35] A. Ikram, U. Qamar, A rule-based expert system for earthquake prediction, J. Intell. Inf. Syst. 43 (2) (2014) 205–230.
[36] B. Gutenberg, C.F. Richter, Frequency of earthquakes in California, Bull. Seismol. Soc. Am. 34 (1944) 185–188.