Spatial Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Spatial Statistics

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

Prediction of intensity and location of seismic events using deep learning

Orietta Nicolis [a,b,*], Francisco Plaza [c,d], Rodrigo Salas [d]

[a] Facultad de Ingenieria, Universidad Andres Bello, Chile
[b] National Research Center for Integrated National Disaster Management (CIGIDEN), Chile
[c] Instituto de Fomento Pesquero, Chile
[d] Universidad de Valparaíso, Chile

ARTICLE INFO

Article history:
Received 18 November 2019
Received in revised form 27 March 2020
Accepted 28 March 2020
Available online xxxx

Keywords:
Earthquake
Intensity function
Deep learning
LSTM
CNN

ABSTRACT

The object of this work is to predict the seismic rate in Chile by using two Deep Neural Network (DNN) architectures, Long Short Term Memory (LSTM) and Convolutional Neural Networks (CNN). For this, we propose a methodology based on a three-module approach: a pre-processing module, a spatial and temporal estimation module, and a prediction module. The first module considers the Epidemic-Type Aftershock Sequences (ETAS) model for estimating the intensity function, which will be used for estimating the seismic rate on a 1×1 degree grid providing a sequence of daily images covering all the seismic area of Chile. The spatial and temporal estimation module uses the LSTM and CNN for predicting the intensity and the location of earthquakes. The last module integrates the information provided by the DNNs for predicting future values of the maximum seismic rate and their location. In particular, the LSTM will be trained using the maximum intensity of the last 30 days as input for predicting the maximum intensity of the next day, and the CNN will be trained on the last 30 images provided by the application of the ETAS model for predicting the probability that the next day the maximum event will be in certain area of Chile. Some performance indexes (such as R^2 and accuracy) will be used for validating the proposed models.

* Corresponding author at: Facultad de Ingenieria, Universidad Andres Bello, Chile.
E-mail addresses: orietta.nicolis@unab.cl (O. Nicolis), francisco.plaza.vega@ifop.cl (F. Plaza), rodrigo.salas@uv.cl (R. Salas).

https://doi.org/10.1016/j.spasta.2020.100442
2211-6753/

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

1. Introduction

Earthquakes represent one of the most destructive yet unpredictable natural disasters along the world, with a massive physical, psychological and economical impact in the population worldwide. Chile is considered one of most seismic-active country in the world, having the world's largest instrumentally documented earthquake occurred in Valdivia (on May 22, 1960), as well as recently been affected by three major earthquakes with magnitudes > 8.0 (on Richter scale). Thus, to have a better approximation or additional information on where and when an event of that magnitude could occur, it would represent an invaluable tool for managing and designing public policies regarding natural disasters. However, earthquake prediction is a very challenging task due to the high complexity associated to the process itself, and also due to the fact that their occurrences depend on a multitude of variables, that in most cases could be unidentified (Sobolev, 2015; Cimellaro and Marasco, 2018; Joffe et al., 2018). Most of the proposed prediction models are focused on some form of seismic hazard estimation (Budnitz et al., 1997; Woessner et al., 2015; Petersen et al., 2018). This term is defined as the probability that an earthquake will occur in a given geographic area, within a given window of time, and with a magnitude exceeding a given threshold. In fact, according to Allen (1976) and Joffe et al. (2018), a valid approach for earthquake prediction should consider a spatio-temporal window, a magnitude estimation, a scientific sound validation process, and an appropriate visualization procedure.

Self-exciting point process models have become essential components in the assessment of seismic hazard. A particular class is given by the Epidemic Time Aftershock Sequence (ETAS) models which have proven to be extremely useful in the description and modeling of earthquake occurrence times and locations. In that sense, Ogata (1988, 1998), introduced ETAS models for temporal and spatio-temporal seismic hazard estimation, respectively. Those models use a given parametrization of the conditional intensity function associated with the occurrence rate of an earthquake and its triggering function at time t and within an (x, y) location. Aftershocks are then estimated following the seismic aftershock propagation law, or Omori's law (Utsu, 1961). Recently, many improvements and extensions have been proposed for incorporating local features of the seismic events (Lombardi et al., 2010; Ogata, 2011; Bansal and Ogata, 2013; Kumazawa and Ogata, 2014; Guo et al., 2015; Nicolis et al., 2015).

Although ETAS models have shown to be very useful for estimating the triggering earthquakes, they often fail when predicting future events. Nicolis et al. (2017) show that the ETAS normally underestimate the real number of seismic events, depending on the precise time that the main shock happened. For solving this problem, they introduced a correction factor that takes into account when the main earthquake happens just before the forecasting day. They also show the superiority of the temporal ETAS for predicting future values of the intensity respect the spatio-temporal ETAS model.

Joffe et al. (2018) stated that contemporary techniques are insufficiently sensitive to allow for precise modeling of future earthquake events. This raises the importance for new approaches that consider broader and bigger sources of information. In that sense, Deep Neural Networks (DNN) have state-of-art accuracy for most of the problems where statistical learning models are applied and where a precise mathematical formulation is hard to obtain. Moreover, DNN architectures, like Long Short Term Memory (LSTM) networks and Convolutional Neural Networks (CNN) have appeared in the last few years, with positive results in a variety of problems such as speech recognition, language modeling, translation, image classification and captioning, time series anomaly detection, stock market prediction, to name a few (Liu et al., 2017; Kumar et al., 2018).

Some of the first machine learning applications on earthquake analysis appeared in the 1990s, and used multilayer perceptrons or artificial neural networks for event detection and phase picking (Wang and Teng, 1995; Tiira, 1999; Zhao and Takano, 1999). In the next decades, along with the further development of new techniques, several new machine learning methods have been successfully applied, for instance, Asim et al. (2018a) used Support Vector Machine Regression and Hybrid Neural Networks to predict earthquake occurrences based on a combination of relevant seismic features such as Gutenberg–Richter law, seismic rate changes, foreshock frequency, seismic energy release and total recurrence time, and Reyes et al. (2013) showed a Neural Network approach for earthquake prediction in Chile from 1960 to 2011, taking into account seismic areas of Talca, Pichilemu, Santiago and Valparaíso.

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

Furthermore, the application of DNN to seismological problems is at its dawn (Kriegerowski et al., 2019). In that sense, recent works proposed the application of DNNs for seismic analysis, most of these works used CNN or LSTM networks. Kislov and Gravirov (2018) analyzed the potential of using Deep Neural Networks for the analysis of seismic data, where DNN has a higher level of abstraction which, consequently, improves the generalization ability of the model. Zhou et al. (2019) introduced a hybrid CNN-RNN to detect earthquake events from seismograms signals. Linville et al. (2019) applied CNN and RNN to discriminate between quarry blasts and tectonic sources from event catalogs and the spectrograms of the sensors, with a performance higher than 99%. Kriegerowski et al. (2019) have applied CNN to accurately predict de hypocenter locations from full-waveform records of multiple stations. Perol et al. (2018) proposed the ConvNetQuake model, a CNN for earthquake detection and location from waveforms. Vijayasankari and Indhuja (2018) used LSTM networks for spatio-temporal earthquake forecast. Geng et al. (2019) proposed a dilated causal temporal convolution network and a CNN-LSTM network to forecast seismic events. Huang et al. (2018) proposed to project the seismic events into a topographic map and generated a dataset of images where the earthquake with magnitude > 6 is marked with a label "1". The authors used a CNN to detect and predict if these large earthquake events will occur in the next 30 days. Li et al. (2019) proposed a method for seismic fault detection using a CNN, moreover, to augment the dataset they developed an encoder–decoder CNN to enrich very small training set. Wang et al. (2018) trained the ResNets for seismic data antialiasing interpolation, where the model is used to reconstruct dense data with halved trace intervals. The generated data can be used to improve the accuracy of subsequent algorithms. On the other hand, Oliveira et al. (2018) assessed the performance of a conditional generative adversarial network to interpolate and generate seismic data. Recently, Plaza et al. (2019) used a LSTM for predicting the intensity function of a temporal ETAS in Chile. LSTM were also applied by Reyes et al. (2013), Wang et al. (2017), Asim et al. (2018b) and Vardaan et al. (2019) for the temporal and spatial prediction of earthquakes.

In this work we explore a new approach based on a Long Short Term Memory (LSTM) network (Hochreiter and Schmidhuber, 1997) and a Convolutional Neural Network (CNN) (LeCun et al., 1989)) for predicting the intensity function and the probability that a seismic event with a given magnitude occurs in a certain area. In particular, first we use the seismic catalogue of Chile for estimating the ETAS model on a grid of 1×1 degree for the period 2000–2017 using three different threshold magnitudes (≥ 4.0, ≥ 5.0 and ≥ 6.0). Then, we take the spatial maximum value of the intensity and we use a LSTM for predicting the maximum rate of seismic event occurrence on the next day. A CNN is also applied to the same data for predicting the probability that the maximum value of intensity is in a certain area. For this goal we divide the data in 6 areas and a classification of the areas are provided. By crossing the outputs of the two neural networks we can predict the maximum intensity and the location of the next seismic event.

This work is structured as follows. In Section 2 we briefly show the methodological modular approach with a brief description of the ETAS model, and the LSTM and CNN neural networks. In Section 3 we apply the latter models for data preprocessing and intensity and location prediction of future seismic events. Some results are shown in Section 4. Conclusions and further developments are provided in Section 5.

2. Methodology

The general purpose for this work is to use a DNN approach with a Long Short Term Memory (LSTM) network and a Convolutional Neural Network (CNN) for the conditional intensity prediction and classification of seismic events. To achieve that goal, three modules are developed, the data preprocessing module, the spatio-temporal estimation module, and the output module. The data preprocessing module processes the original data and prepare the inputs for the DNN models (LSTM and CNN). The spatio-temporal module take those inputs and estimate magnitude and location with LSTM and CNN models, respectively. Finally, the output module, fuse the outputs of the previous model and evaluate their performance (Fig. 1). For this work we will consider events with seismic magnitudes greater than or equal to 4.0, 5.0 and 6.0 (on Richter scale).

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

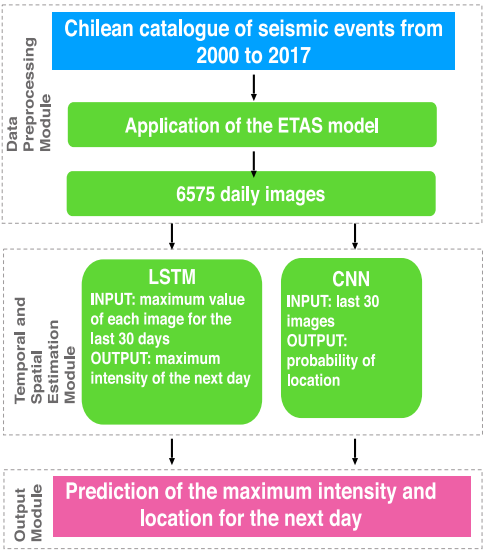


Fig. 1. Scheme for the proposed DNN approach.

2.1. Data

The database consists of 86,000 seismic event records for a total of 6575 days, obtained from the National Seismological Center (http://www.sismologia.cl). Each record is composed by a time location (Year, Month, Day, Hour, Minute and Second), a spatial location (Latitude and Longitude), depth (in kilometers), and magnitude (on Richter scale). Fig. 2 shows all the events with magnitude grater or equal to 6.0. From Fig. 2 (left) we note that most of the events are concentrated in the central northern part of Chile, as well as the three events with magnitude greater than 8.0 whose epicenters are located close to Concepción, Iquique, and Coquimbo, respectively. Fig. 2 (right) shows the magnitudes of the events as time series (in days), and shows the three aforementioned events occurred with a magnitude of 8.8, 8.2, and 8.4 on February, 27th, 2010 (time 3710), April, 1st, 2014 (time 5204), and September 16th, 2015 (time 5737), respectively.

2.2. Data preprocessing module

The data preprocessing module consists in estimating the conditional intensity function, by assuming that the number of seismic events follow a Poisson distribution. Point process models have become essential components in the assessment of seismic hazard. A particular class is given by the self-exciting spatio-temporal point process which model events whose rate at time t may depend on the history of events at times preceding t, allowing events to trigger new events (see Reinhart et al., 2018 and the references within). These models appeared for the first time in applications to population genetics and for this they are also known as Epidemic-Type Models. Ogata (1988) introduced the Epidemic-Type Aftershock Sequence (ETAS) models for the temporal modeling of seismic events.

Self exciting models can be defined in terms of the conditional intensity function given by

$$\lambda_g(t|\mathcal{H}_t) = \lim_{\Delta t \downarrow 0} \frac{E\left[N\left\{(t, t+\Delta t)|\mathcal{H}_t\right\}\right]}{\Delta t}, \tag{1}$$

where $N(A)$ the number of events occurring at time $t \in A$ and $\{\mathcal{H}_t : t \geq 0\}$ is the history of all events up to time t. By denoting $t_i \in [0, T)$ a simple point process with $t_i < t_{i+1}$ and Δt the time difference

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

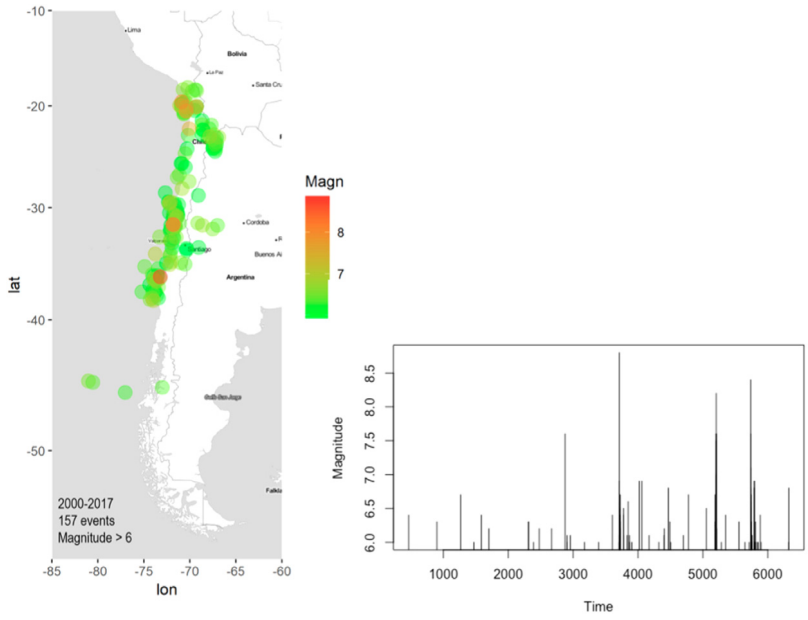


Fig. 2. (Left) Maps of Chile with seismic events with magnitude greater than or equal to 6.0 (Right) Magnitude vs time.

between t_i and t_{i+i}, the intensity function can be written as

$$\lambda_g(t|\mathcal{H}_t) = \mu + \sum_{i:t_i<t} c(m_i)g(t-t_i) \tag{2}$$

where the component μ is the background seismicity rate which prevents the process from dying out, m_i is the magnitude at the time t_i, and g is the triggering function which determines the form of the self-excitation (Ogata, 1988). This process with intensity function $\lambda_g(t|\mathcal{H}_t)$ is also known as marked self-exciting point process, where the mark is given by the magnitude associated to each event.

Different parametrizations have been proposed for the function m and f. Ogata (1988) proposed the use of $c(m) = e^{\alpha(m-M_t)}$ and $f(t) = \frac{K}{(t+c)^p}$, where α measures the effect of magnitude in the production of aftershocks and f is the modified Omori formula (Utsu, 1961), with t representing the time of occurrence of the shock, K a normalizing constant depending on the lower bound of the aftershocks, and c and p are characteristic parameters of the seismic activity of a given region.

The first extension of the temporal ETAS to the space component have been proposed by Ogata (1998) . The space–time ETAS models are characterized by a parametric intensity function which represents the occurrence rate of an earthquake at time t and the location (x, y) conditional on the past history of the occurrence.

Their conditional space time intensity $\lambda(x, y, t)$ can be written as :

$$\lambda(x,y,t) = \mu f(x,y) + \sum_{t_j<t} \frac{Ke^{(\alpha-\gamma)}(m_j-m_0)}{(t-t_j+c)^p} \left\{ \frac{(x-x_j)^2+(y-y_j)^2}{e^{\gamma(m_j-m_0)}} + d \right\}^{-q} \tag{3}$$

where $f(x, y)$ is estimated through a weighted kernel gaussian estimator. Weights are given by the estimated probabilities of being a background event; for the ith event this is given by $\rho_i = \frac{\mu f(x_i,y_i)}{\lambda(x_i,y_i,t_i)}$. In Eq. (3), μ measures the background general intensity (assumed temporally homogeneous); K is a scale parameter related to the importance of the induced seismicity; c and p

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

are the characteristic parameters of the seismic activity of the given region (c is a shift parameter while p is the exponent parameter of the modified Omori law for temporal decay rate of aftershocks) α and γ measure the efficiency of an event of given magnitude in generating aftershock sequences; d and q are two parameters related to the spatial influence of the mainshocks (see, Nicolis et al., 2015; Chiodi and Adelfio, 2017).

2.3. Spatial and temporal deep learning module

In this section, we present a brief description of the LSTM and the CNN models used for intensity function and location prediction, respectively. Both models use an iterative process in order to estimate the inner parameters for the models. In the case of the LSTM models, the parameters are associated with each LSTM cell (see, Section 2.3.1), and in the case of the CNN model, the parameters are associated with the convolution operation (see, Section 2.3.2). This iterative process starts by splitting the data in two subsets, the training and the testing, which correspond to the data used for the estimation and for the prediction, respectively. The training subset is then used for the learning (or estimation) process which compares the observed data with the estimated data, adjusting the parameters of the model with a gradient-descent correction process. When the network has reached a stopping criteria (e.g. performance measures from internal validation, a given number of iterations, among others), the network parameters (or weights) are saved and used in the testing subset for evaluating the performance of each model. Finally, the best model is chosen for the prediction. While in the LSTM model the training and testing subsets are composed by time series, in the CNN, the same subsets are composed by images.

2.3.1. Long Short Term Memory Recurrent Neural Networks (LSTM-RNN)

Recurrent Neural Networks (RNNs) (Medsker and Jain, 1999; Rumelhart et al., 1986) are a type of neural networks which have a primitive type of memory in the form of recurrent layers. A recurrent layer of a RNN takes two kinds of inputs: the output of the preceding layer and the output of the same recurrent layer from the last point it processed (Cady, 2017). Thus, the RNN can solve the purpose of sequence handling, being a more appropriate choice for dealing with time series data. However, these networks work well when it comes to short contexts, as they are not able to understand and remember the context behind longer sequences in order to build a story and remember it, just like a human brain could do (Srivastava, 2017). Thus, Long Short Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) are a type of RNN, precisely designed to deal with the long term dependency issue of recurrent networks.

Analog to multilayer perceptron networks that have neurons as their basic processing units, LSTM recurrent networks have cells as processing units that apply information loops inside each one of the cells (inner recurrence) and between the cells themselves (outer recurrence) (Goodfellow et al., 2016). Moreover, each cell $s_i^{(t)}$ is the basic unit of a LSTM-RNN, having a similar parameter structure than a recurrent neural network in addition to a structure of gating units located inside each cell, defined as $f_i^{(t)}$ (for each time step t and cell i), that manage the information stream inside and between cells, then a sigmoid unit is applied in order to scale the weights in $[0-1]$, as follows:

$$f_i^{(t)} = \sigma\left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)}\right), \tag{4}$$

where: $x^{(t)}$ is the input vector; $h^{(t)}$ is the hidden layer vector, containing the outputs of all the cells; b^f, U^f, W^f are biases, input weights, and recurrent weights for the forget gates, respectively, and $\sigma(\cdot)$ is a sigmoid function (also called sigmoid unit). The internal state of each cell is then updated with a conditional self-loop weight $f_i^{(t)}$ as follows:

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma\left(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)}\right), \tag{5}$$

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

Table 1
Parameter estimates of the intensity function using ETAS model for magnitudes greater than or equal to 4.0, 5.0, and 6.0 (on Richter scale).

	μ	K,	α	c	p
Intensity ($\geq$ 4.0)	0.001	2.3526	0.2336	0.0247	1.1205
Intensity ($\geq$ 5.0)	0.0340	0.0007	1.4980	0.0197	1.0340
Intensity ($\geq$ 6.0)	0.0041	0.0001	1.6452	0.0062	1.0080

where b, U and W denote the biases, input weights, and recurrent weights into each cell, respectively. The external input gate unit $g_i^{(t)}$ is similarly computed to the forget gate as shown below:

$$g_i^{(t)} = \sigma\left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)}\right), \tag{6}$$

The LSTM cell output, defined as $h_i^{(t)}$, can be controlled by two ways: via the hyperbolic tangent function, that scales the output values between 1 and -1, and through the output gate $q_i^{(t)}$ that uses the sigmoid function:

$$h_i^{(t)} = \tanh(s_i^{(t)}) q_i^{(t)}, \tag{7}$$

$$q_i^{(t)} = \sigma\left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)}\right), \tag{8}$$

with b^o, U^o and W^o its biases, weights and recurrent weights, respectively.

2.3.2. Convolutional Neural Networks (CNNs)

The CNN is a deep neural network which is most commonly used for image and data classification. A CNN is composed by layers with filters, which search for particular patterns or features in the data. Each filter has a kernel, that is represented by a matrix of numbers, which, resembles the pattern that the filter is trying to detect (Cady, 2017). A CNN then is conformed by the convolution of an input vector or matrix and a series of learned filters, with every neuron having groups of weights with inputs at different shifted frequencies (Abdel-Hamid et al., 2013; Murphy, 2012). A typical CNN layer consists of three phases: firstly, the layer performs several convolutions (i.e., the applications of kernel based filters) in parallel to produce a set of linear matrix transformations; secondly, each linear activation (i.e., the linear matrix operation that goes through all the pixels of the image) is run through a nonlinear activation function which could be a "relu", a "leaky-relu", "Sigmoid" etc.; finally, the third phase uses a pooling function to modify the output of the layer further (Goodfellow et al., 2016). The pooling layer is used to down sampling feature maps by summarizing the presence of these features. The CNN can be composed by different combinations of these layers, producing an output according to the nonlinear function used in the last one.

3. Application to the Chilean catalogue

As stated in Section 2.2, the preprocessing module based on the temporal ETAS model has been used for estimating the conditional intensity function in Chile in the period 2000–2017 on a 42° by 15° grid including the area between latitudes $(-57, -15)$ and longitudes $(-80, -65)$. Each pixel of the grid had a size of $1° \times 1°$ which represents an area of approximately 111 km$m \times 111$ km for each pixel. Firstly, the parameters of the ETAS model had been estimated on all area by considering seismic magnitudes greater than or equal to 4.0, 5.0 and 6.0 (see Table 1). Secondly, the estimated parameters were used for estimating the intensity on each pixel of the grid over all the period, assuming a rather small spatial variability of the parameters alongside the country. The package `PtProcess` (Harte, 2010) of R software (R Core Team, 2013) had been used for the estimation.

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

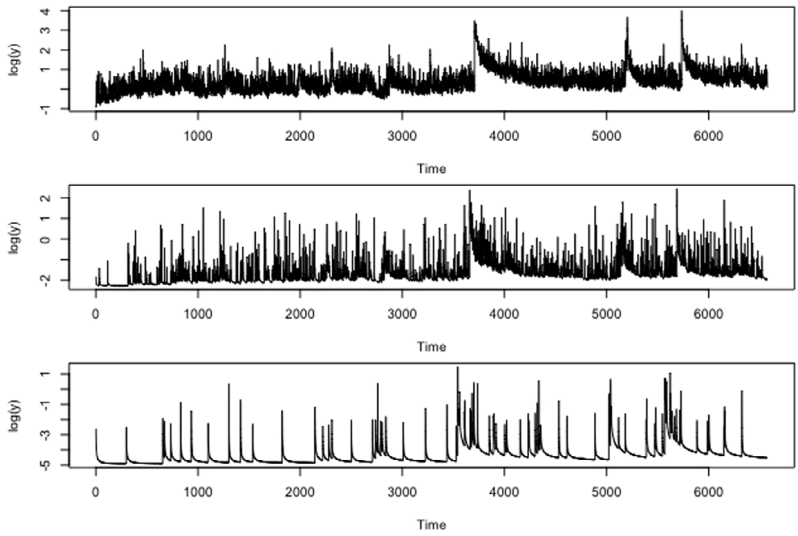


Fig. 3. Estimated Intensity function using ETAS model for magnitude greater than or equal to 4.0 (upper), 5.0 (middle), and 6.0 (bottom) .

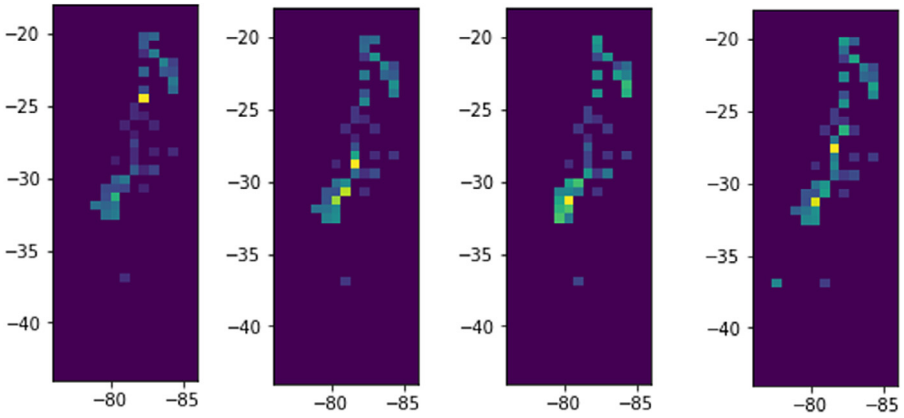


Fig. 4. Example of the spatial representation of the intensity function in four different times (in days).

The estimation pixel by pixel for all the period (from 2000 to 2017) provided 6575 daily images. Fig. 3 shows the estimated intensity functions for magnitudes greater than or equal to 4.0, 5.0, and 6.0.

Fig. 4 shows a spatial representation (or images) of the estimated intensities at four different times (days). An alternative way to obtain these images is to estimate the intensity function of a space time ETAS model. However, due to very large data set which make complex the estimation of the parameters and the lower temporal performance of this model (as explained in Nicolis et al., 2015) only the temporal ETAS had been considered for the purpose of this work.

As a third step, two deep learning neural networks were applied: a LSTM model for predicting the maximum intensity for the next day and a CNN model for predicting the probability that the maximum intensity would be in a certain location.

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

Table 2
Latitude macro-zone ranges for the CNN earthquake location prediction.

Zone	Latitude range
1	(−44.00; −39.67)
2	(−39.67; −35.34)
3	(−35.34; −31.00)
4	(−31.00; −26.67)
5	(−26.67; −22.34)
6	(−22.33; −18.00)

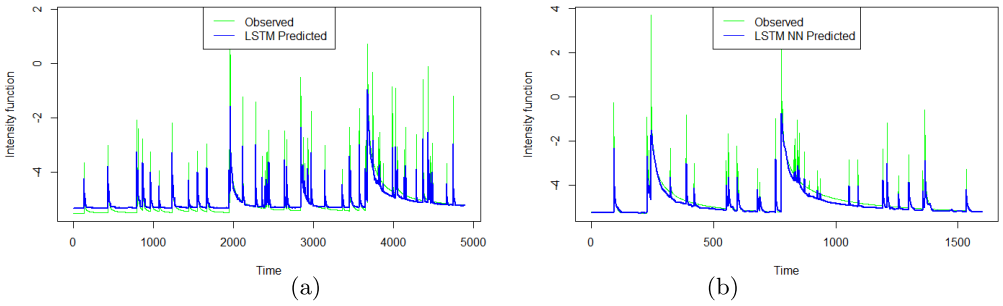


Fig. 5. Estimated maximum intensities by ETAS model and their predictions using the training set (a) and testing set (b).

In order to perform the temporal prediction of the maximum intensities, the maximum value of the intensity function for each day over all the study area was considered as database of the LSTM model. Then, the maximum intensities of the past 30 days was used as inputs for the LSTM to predict the maximum intensity of the next day (output). The proposed DNN was composed by 50 LSTM cells where the hyperbolic tangent was used as the activation function. A number of 100 epochs were used to train the model, which considered the 75% of the total data (4931 days) as the training set, and the remaining 25% (1644 days) as testing set. In particular, the training set was used to estimate the parameters (e.g. weights) of the LSTM model, while the testing set was used for predicting new values of the maximum intensities by considering the parameters estimated in the training process.

To predict the maximum intensity location, the maximum intensity occurrences were classified into one of six latitudinal zones defined by dividing the study area in six macro zones (provided in Table 2). Then, a CNN model was trained using the images of the last 30 days as input, and the macrozones 1 to 6 that represented maximum intensity location of the next day as the output for the CNN model. The structure of the CNN model was composed by three convolutional layers (with their respective pooling layers each), then 5 fully connected multilayer perceptron layers. Each layer considered a dropout rate of 25% (meaning one on 4 inputs were randomly excluded from each update cycle) in order to avoid the overfitting of the model due to the large number of parameters.

The DNN models (LSTM and CNN) were implemented using the module `keras` and `tensonflow` of R and Python softwares (Abadi et al., 2016; Chollet et al., 2017, 2015; Rossum, 1995).

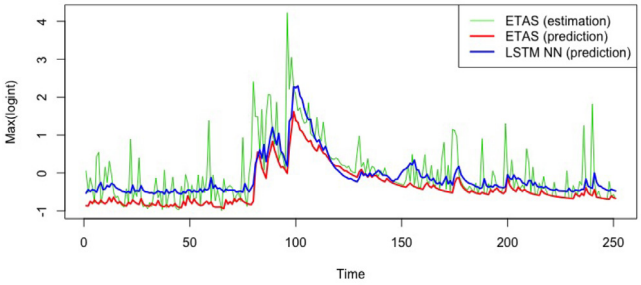
4. Results

Fig. 5 represents the maximum intensity LSTM prediction results for events with magnitude greater than or equal to 6.0 (on Richter scale). The figure shows that the LSTM is able to both identify patterns that characterize high-magnitude earthquakes, and to predict the maximum intensity with a certain approximation. The goodness of fit of the model was confirmed by the R^2 determinant coefficient which resulted 0.70 for the training set, and 0.66 for the testing set.

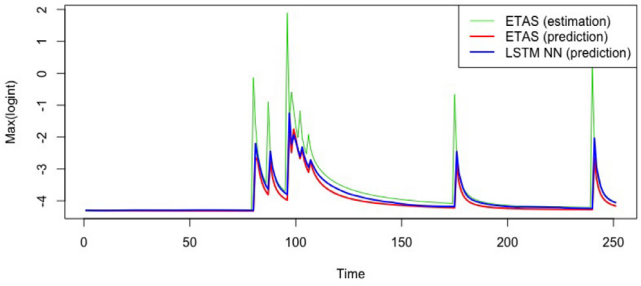
Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

Table 3
Prediction performance measures for the ETAS and LSTM models using magnitudes $\geq 4.0, 5.0$, and 6.0 (on the Richter scale).

	MSE	MAE	RMSE	R^2
LSTM ($M \geq 4.0$)	0.1853	0.2780	0.4424	0.4489
ETAS ($M \geq 4.0$)	0.3685	0.3807	0.6071	0.3844
LSTM ($M \geq 5.0$)	0.2893	0.3092	0.5380	0.3961
ETAS ($M \geq 5.0$)	0.2844	0.2918	0.5333	0.3249
LSTM ($M \geq 6.0$)	0.2804	0.2136	0.5296	0.6561
ETAS ($M \geq 6.0$)	0.2998	0.3484	0.5475	0.6241



(a)



(b)

Fig. 6. Predictions using the ETAS (red lines) and LSTM (blue lines) for a period of 250 days considering seismic events with magnitude ≥ 4.0 (a) and ≥ 6.0 (b). The green lines represents the estimated ETAS models.

In order to compare the performance of the LSTM with the classical temporal ETAS model an analysis of residuals for the two models was performed. The ETAS prediction of the maximum intensities had been implemented by using the past seismic events until the day t for predicting the conditional intensity at the day $t+1$ for each pixel of size $1^\circ \times 1^\circ$) of the studied area. Then, the maximum intensities were chosen for all the prediction period. Table 3 reports the results in terms of MAE (Mean Absolute Error), MSE (Mean Square Error), RMSE (Root Mean Square Error) and R^2. In all cases, the LSTM seemed outperforming the ETAS model. Fig. 6(a) and (b) show the behavior of the predictions with ETAS and LSTM (using data with magnitude greater than or equal to 4.0 and 6.0, respectively) for 250 days including the main shock (of magnitude 8.2) happened on April 1, 2014, corresponding to the maximum peak of the estimated ETAS (green line).

For predicting the location probability of the next day we used a CNN which was composed by three convolutional layers with pooling layers in each one of them. This structure considered

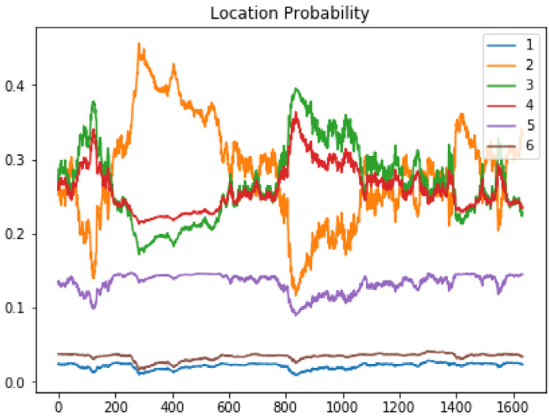


Fig. 7. Classification results for the location estimation process.

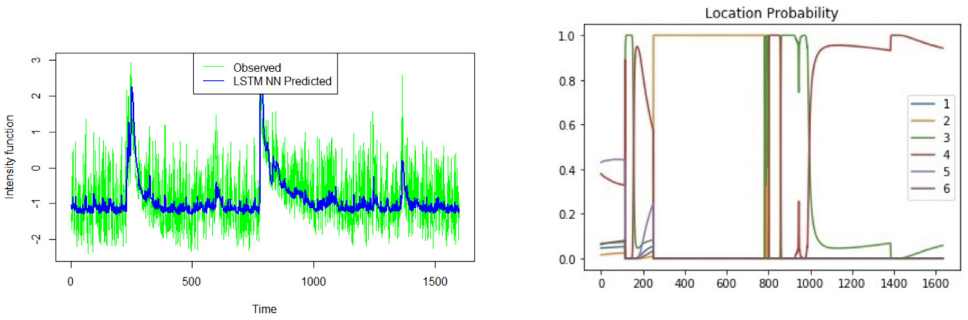
fully connected layers with "relu" and "leaky-relu" transfer functions. In the last layer of the CNN, a multilayer perceptron neural network was used, which has a non-linear activation function that assessed the probability of occurrence of each of the classes (form 1 to 6 in our case). The results for seismic events with magnitude greater than or equal to 6.0 are shown in Fig. 7 (for the testing set). In this case the accuracy resulted equal to 45%. Although this performance index was not so high, the model seemed to correctly capture the periods in which there were areas with high probabilities of earthquake occurrence (in terms of maximum intensities). In particular, these results showed that zones 2, 3 and 4 were rather more seismic than the other zones, with higher probabilities of maximum intensity location among all zones. This findings had been validated with real data which confirmed the good classification for high probability occurrences. For instance, the highest probability occurrence corresponded to the April, 1*st*, 2014 earthquake which had its epicenter in the city of Iquique (located in macro-zone zone 2). The second high probability corresponded to the 2015 earthquake that took place in the city of Coquimbo (zones 3 and 4). Finally, the last peaks in probability show a correspondence with earthquakes occurred in Atacama (zone 2) and a few months later, in the city of Valparaíso (zones 3 and 4).

For seismic events with magnitudes greater than or equal to 4.0, Fig. 8(a) represents the prediction for the testing set using the LSTM and Fig. 8(b) shows the location probability for the same period. Since the number of events increase significantly, the LSTM model provides worse predictions due to the difficulty to identify regular patterns. In this case the R^2 for the testing set was only 0.45. However, the accuracy of the CNN prediction improved significantly, reaching the 80%.

5. Discussion

In this work we propose a deep learning approach based on a Long Short Term Memory (LSTM) and a Convolutional Neural Network (CNN) for predicting the intensity and location of future seismic events in the Chilean catalogue with magnitudes greater than or equal to 4.0, 5.0, and 6.0. The results showed that LSTM can predict the maximum intensity of the seismic events with an R^2 of 0.66 on the testing set. This performance index considerably decreases when considering the LSTM prediction on events with lower magnitudes. Moreover, the CNN seemed to be able to predict the probability of the next seismic events with an accuracy of 80% and 45% for magnitudes greater than or equal to 4.0 and 6.0, respectively. The proposed methodology extends the work of Plaza et al. (2019) by including the prediction to the spatial component and improving the temporal prediction for seismic event with magnitude greater than or equal to 6.0. Other approaches have been proposed in the literature for predicting earthquakes using similar deep learning LSTM architectures. In particular, Vardaan et al. (2019) use the LSTM and FFNN approach for predicting some features of the next event (such

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.



(a) (b)

Fig. 8. LSTM and CNN predictions for seismic events with magnitudes greater than or equal a 4.0: (a) LSTM intensity function prediction for the testing set; (b) CNN location probability prediction.

as latitude, longitude, magnitude, and time). Although the LSTM outperformed the classical FFNN, in all cases the R^2 scores resulted negative, showing a poor prediction ability of the models. Wang et al. (2017) designed an LSTM network with two-dimensional input including information of the events and on the location. The model reached the accuracy of 86% for predicting if at the time $t + 1$ will be an event in a certain zone. Asim et al. (2018b) proposed a binary ensemble classifier with the aim to generate temporal prediction for earthquakes of magnitude 5.0 and greater, 15 days prior to an earthquake. The accuracy of the proposed model for the central zone of Chile was about 85.7% slightly better than that obtained by Reyes et al. (2013) of 83.5% for the same area, using a classical neural network. In summary, differently to our approach which is aimed to predict the seismic hazard using a preprocessing of data based on ETAS model, the majority of the work uses classifiers for detecting the presence or absence of an earthquake or apply the DL model directly on the catalogue with worse performance.

Regarding the CNN location prediction, the proposed model approach is similar to that presented by Huang et al. (2018), who consider 30 to 210 days in order to predict earthquakes occurrences with a 30-days time window, using a database consisting of approximately 470 examples for training and 95 for testing the model. The structure of the model is composed by three convolutional layers with pooling layers in each one of them. Our approach considers more fully connected layers with relu and leaky-relu transfer functions, due to the six categories (or macrozones) considered for the location prediction. The results obtained by our model are slightly better than the approach presented by Huang et al. (2018), although the results are not fully comparable due to the different considered zones, the data quantity, and the magnitudes considered for the prediction.

Differently to the classical parametric models (i.e., ETAS and its extension), in which the estimation of the parameters could be complex when the dataset is very large, the DL models can process big datasets including the worldwide seismic events. Big earthquake database such as the Quake Template Matching (QTM) catalog (Ross et al., 2019) and the STanford EArthquake Dataset (STEAD) (Mousavi et al., 2019) are now available to be processed with artificial intelligence and deep learning techniques. A review on the potential of artificial intelligence in seismology is provided by Jiao and Alavi (2019).

6. Concluding remarks and further developments

As the prediction of seismic events still constitutes an area that needs further development, this work constitutes a preliminary analysis on the joint use of Deep Learning methods, such as LSTM and CNN. The results can be easily extended for predicting the number of events or the maximum magnitude in a certain area. Also, this work establishes a baseline from which the proposed DL model could be further improved by incorporating some exogenous variables

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

such as the seismic depth, crustal displacement, the distance from the main fault and other geological variables, especially for the prediction of main shocks where normally ETAS models fail. Other improvements could be done by improving the resolution of the space–time hazard map using spatio-temporal ETAS model and considering the findings (van Lieshout and Stein, 2012) for avoiding edge effects and identify correct clusters of events. Finally, further developments could also include the exploration of other types and architectures of neural networks such as Generative Adversarial Networks (GAN, Goodfellow et al., 2014) and the spatio-temporal autoencoder neural network for better incorporating the spatial component (Li et al., 2019).

Acknowledgments

Orietta Nicolis was partially supported by the National Research Center for Integrated National Disaster Management (CIGIDEN)ANID/FONDAP/15110017, by the Andres Bello University grant DI-03-19/R, and by the national grant FONDECYT Regular ID1201478. Francisco Plaza was partially supported by the CONICYT-PFCHA/DOCTORADO-BECAS-CHILE/2018–21182037.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: A system for large-scale machine learning. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. Berkeley, CA, USA. pp. 265–283.

Abdel-Hamid, O., Deng, L., Yu, D., 2013. Exploring convolutional neural network structures and optimization techniques for speech recognition. In: Interspeech. pp. 3366–3370.

Allen, C.R., 1976. Responsibilities in earthquake prediction: to the seismological society of America, delivered in Edmonton, Alberta, May 12, 1976. Bull. Seismol. Soc. Am. 66 (6), 2069–2074.

Asim, K.M., Idris, A., Iqbal, T., Martí nez Álvarez, F., Li, X., 2018a. Earthquake prediction model using support vector regressor and hybrid neural networks. PLoS One 13.

Asim, K.M., Idris, A., Iqbal, T., Martínez-Álvarez, F., 2018b. Seismic indicators based earthquake predictor system using genetic programming and AdaBoost classification. Soil Dyn. Earthq. Eng. 111 (April), 1–7.

Bansal, A., Ogata, Y., 2013. A non-stationary epidemic type aftershock sequence model for seismicity prior to the December 26, 2004 M 9.1 Sumatra-Andaman Islands mega-earthquake. J. Geophys. Res. Solid Earth 118 (2), 616–629.

Budnitz, R., Apostolakis, G., Boore, D., 1997. Recommendations for probabilistic seismic hazard analysis: Guidance on uncertainty and use of experts. Technical Report, Nuclear Regulatory Commission, Washington, DC (United States), 1 (6372).

Cady, F., 2017. The Data Science Handbook. John Wiley & Sons, Incorporated.

Chiodi, M., Adelfio, G., 2017. Mixed non-parametric and parametric estimation techniques in R package etasFLP for earthquakes' description. J. Stat. Softw. 76 (3), 1–29.

Chollet, F., Allaire, J., et al., 2017. R interface to Keras. https://github.com/rstudio/keras.

Chollet, F., et al., 2015. Keras. https://github.com/fchollet/keras.

Cimellaro, G.P., Marasco, S., 2018. Earthquake prediction. In: Introduction to Dynamics of Structures and Earthquake Engineering. Springer, pp. 263–280.

Geng, Y., Su, L., Jia, Y., Han, C., 2019. Seismic events prediction using deep temporal convolution networks. J. Electr. Comput. Eng. 2019.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, http://www.deeplearningbook.org.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680.

Guo, Y., Zhuang, J., Zhou, S., 2015. An improved space-time ETAS model for inverting the rupture geometry from seismicity triggering. J. Geophys. Res. Solid Earth 120 (5), 3309–3323.

Harte, D., 2010. PtProcess: An R package for modelling marked point processes indexed by time. J. Stat. Softw. 35 (8), 1–32.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Huang, J., Wang, X., Zhao, Y., Zhao, C., Xiang, H., 2018. Large earthquake magnitude prediction in Taiwan based on deep learning neural network. Neural Netw. World 8.

Jiao, P., Alavi, A.H., 2019. Artificial intelligence in seismology: Advent, performance and future trends. Geosci. Front..

Joffe, H., Rossetto, T., Bradley, C., O'Connor, C., 2018. Stigma in science: the case of earthquake prediction. Disasters 42 (1), 81–100.

Kislov, K.V., Gravirov, V.V., 2018. Deep artificial neural networks as a tool for the analysis of seismic data. Seism. Instrum. 54.

Kriegerowski, M., Petersen, G.M., Vasyura-Bathke, H., Ohrnberger, M., 2019. A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms. Seismol. Res. Lett..

Kumar, J., Goomer, R., Singh, A.K., 2018. Long short term memory recurrent neural network (LSTM-RNN) based workload forecasting model for cloud datacenters. Procedia Comput. Sci. 125, 676–682.
Kumazawa, T., Ogata, Y., 2014. Nonstationary ETAS models for nonstandard earthquakes. Ann. Appl. Stat. 8 (3), 1825–1852.
LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1 (4), 541–551.
Li, S., Yang, C., Sun, H., Zhang, H., 2019. Seismic fault detection using an encoder–decoder convolutional neural network with a small training set. J. Geophys. Eng..
van Lieshout, M.N., Stein, A., 2012. Earthquake modelling at the country level using aggregated spatio-temporal point processes. Math. Geosci. 44 (3), 309–326.
Linville, L., Pankow, K., Draelos, T., 2019. Deep learning models augment analyst decisions for event discrimination. Geophys. Res. Lett..
Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E., 2017. A survey of deep neural network architectures and their applications. Neurocomputing 234, 11–26.
Lombardi, A.M., Cocco, M., Marzocchi, W., 2010. On the increase of background seismicity rate during the 1997–1998 Umbria-Marche, Central Italy, sequence: apparent variation or fluid-driven triggering? Bull. Seismol. Soc. Am. 100 (3), 1138–1152.
Medsker, L.R., Jain, L.C., 1999. Recurrent Neural Networks: Design and Applications. CRC Press.
Mousavi, S.M., Sheng, Y., Zhu, W., Beroza, G.C., 2019. Stanford earthquake dataset (STEAD): A global data set of seismic signals for AI. IEEE Access 7, 179464–179476.
Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning Series). The MIT Press.
Nicolis, O., Chiodi, M., Adelfio, G., 2015. Windowed ETAS models with application to the Chilean seismic catalogs. Spatial Stat. 14, 151–165.
Nicolis, O., Chiodi, M., Adelfio, G., 2017. Space-time forecasting of seismic events in Chile. In: Zouaghi, T. (Ed.), Earthquakes - Tectonics, Hazard and Risk Mitigation. IntechOpen, pp. 169–184.
Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes. J. Amer. Statist. Assoc. 83 (401), 9–27.
Ogata, Y., 1998. Space-time point-process models for earthquake occurrences. Ann. Inst. Statist. Math. 50 (2), 379–402.
Ogata, Y., 2011. Significant improvements of the space-time etas model for forecasting of accurate baseline seismicity. Earth Planets Space 63 (3), 217–229.
Oliveira, D.A.B., Ferreira, R.S., Silva, R., Brazil, E.V., 2018. Interpolating seismic data with conditional generative adversarial networks. IEEE Geosci. Remote Sens. Lett..
Perol, T., Gharbi, M., Denolle, M., 2018. Convolutional neural network for earthquake detection and location. Sci. Adv. 4.
Petersen, M.D., Mueller, C.S., Moschetti, M.P., Hoover, S.M., Rukstales, K.S., McNamara, D.E., Williams, R.A., Shumway, A.M., Powers, P.M., Earle, P.S., et al., 2018. 2018 one-year seismic hazard forecast for the central and eastern United States from induced and natural earthquakes. Seismol. Res. Lett. 89 (3), 1049–1061.
Plaza, F., Salas, R., Nicolis, O., 2019. Assessing seismic hazard in Chile using deep neural networks. In: Tiefenbacher, J.P. (Ed.), Natural Hazards. IntechOpen, Rijeka.
R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, p. 201, http://www.R-project.org/.
Reinhart, A., et al., 2018. A review of self-exciting spatio-temporal point processes and their applications. Statist. Sci. 33 (3), 299–318.
Reyes, J., Morales-Esteban, A., Martínez-Álvarez, F., 2013. Neural networks to predict earthquakes in Chile. Appl. Soft Comput. 13 (2), 1314–1328.
Ross, Z.E., Trugman, D.T., Hauksson, E., Shearer, P.M., 2019. Searching for hidden earthquakes in Southern California. Science.
Rossum, G., 1995. Python Reference Manual. Technical Report. CWI (Centre for Mathematics and Computer Science), Amsterdam, The Netherlands, The Netherlands.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (6088), 533.
Sobolev, G.A., 2015. Methodology, results, and problems of forecasting earthquakes. Her. Russ. Acad. Sci. 85 (2), 107–111.
Srivastava, P., 2017. Introductory guide to generative adversarial networks (GANs) and their promise! https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/. (Accessed: 2018-07-22).
Tiira, T., 1999. Detecting teleseismic events using artificial neural networks. Comput. Geosci. 25 (8), 929–938.
Utsu, T., 1961. A statistical study on the occurrence of aftershocks. Geophys. Mag. 30, 521–605.
Vardaan, K., Bhandarkar, T., Satish, N., Sridhar, S., Sivakumar, R., Ghosh, S., 2019. Earthquake trend prediction using long short-term memory RNN. Int. J. Electr. Comput. Eng. 9 (2), 1304–1312.
Vijayasankari, S., Indhuja, P., 2018. Earthquake prediction based on spatio-temporal data mining approach. Int. J. Sci. Eng. Res. 9 (4), 1573–1579.
Wang, Q., Guo, Y., Yu, L., Li, P., 2017. Earthquake prediction based on spatio-temporal data mining: an LSTM network approach. IEEE Trans. Emerg. Top. Comput..
Wang, J., Teng, T.-L., 1995. Artificial neural network-based seismic detector. Bull. Seismol. Soc. Am. 85 (1), 308–319.
Wang, B., Zhang, N., Lu, W., Wang, J., 2018. Deep learning based seismic data interpolation: a preliminary result. Geophysics.
Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., Demircioglu, M.B., et al., 2015. The 2013 European seismic hazard model: key components and results. Bull. Earthq. Eng. 13 (12), 3553–3596.

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.

Zhao, Y., Takano, K., 1999. An artificial neural network approach for broadband seismic phase picking. Bull. Seismol. Soc. Am. 89 (3), 670–680.

Zhou, Y., Yue, H., Kong, Q., Zhou, S., 2019. Hybrid event detection and phase-picking algorithm using convolutional and recurrent neural networks. Seismol. Res. Lett. 90.

Please cite this article as: O. Nicolis, F. Plaza and R. Salas, Prediction of intensity and location of seismic events using deep learning. Spatial Statistics (2020) 100442, https://doi.org/10.1016/j.spasta.2020.100442.