IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 September 2022, accepted 23 September 2022, date of publication 28 September 2022,
date of current version 13 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3210554

RESEARCH ARTICLE

Spatio-Temporal Seismicity Prediction in Chile Using a Multi-Column ConvLSTM

ALEX GONZÁLEZ FUENTES[1], ORIETTA NICOLIS[1,2], (Member, IEEE), BILLY PERALTA[1,3], (Member, IEEE), AND MARCELLO CHIODI[4]

[1]Facultad de Ingeniería, Universidad Andres Bello, Viña del Mar 2520000, Chile
[2]Research Center for Integrated Disaster Risk Management (CIGIDEN), Santiago 7810000, Chile
[3]Universidad Andres Bello, Santiago 7500000, Chile
[4]Dipartimento di Scienze Economiche, Aziendali e Statistiche, Universitá degli Studi di Palermo, 90128 Palermo, Italy

Corresponding author: Orietta Nicolis (orietta.nicolis@unab.cl)

The work of Orietta Nicolis was supported in part by the Fondecyt Project Grant 1201478; and in part by the CIGIDEN Project Fondap/15110017.

ABSTRACT One way to characterize the seismicity in a given zone is through the study of the conditional intensity function of the ETAS model (Epidemic Type Aftershock Sequence) which represents the average number of seismic events greater than given magnitude. Being Chile one of the most seismic country in the world, it is very important to predict where the seismic events will happen with more frequency. In this work we propose a parallel neural network based on the Convolutional Network (CNN) and the Long Short Term Memory (LSTM) network, called Multi-Culumn ConvLSTM, using the accumulated crustal velocity and the intensity data as input for predicting the daily mean number of seismic events in Chile with magnitude greater than a given value. For the application, the central zone of Chile between the regions of Coquimbo and Araucanía, in the period from 2010 to 2017 was considered. At the spatial level, each region was partitioned considering a 20×20 dimension grid, while at the temporal level, input data from the last 20 days were used to predict the mean number of seismic events for the following day. The experiments showed that the Multi-column ConvLSTM network obtained the best results in the test set with an average coefficient of determination of 0.81.

INDEX TERMS Deep learning, ETAS model, prediction, seismic events.

I. INTRODUCTION

Chile is one of the most seismic country in the world due to its proximity to the Nazca plate which beneath South American continental tectonic plate with an average converging rate of 6.5 mm per year, one of the fastest rates on Earth [3]. Depending on the depth, subduction earthquakes can be classified into two main categories: the crustal and interplate events which occur at a plate boundary at a depth less than 70 km, approximately; and the deep and intermediate intraplate events which occur at greater depths (major than 70km) in the interior of the tectonic plate (see, [4], [34], and [53]). Most of earthquakes which are concentrated on interplate faults have caused major disasters in Chile with many dead and/or economic losses. In the last century, numerous earthquakes have occurred on the Chilean subduction zone with magnitude larger than 8.0, including the 1960 Valdivia earthquake, the largest instrumentally recorded earthquake in the world, which it is supposed to have caused an increased stress that leaded to other earthquakes [25]. Successively, about 230 km north of the 1960 quake, another major earthquake in Chile was registered on February 27th, 2010, that affected almost 80% of the Chilean population in a certain degree [39]. Recently, two main earthquakes with magnitude major than 8 happened on Abril 1, 2014 in the North of Chile (Iquique, Tarapaca region) and on September 16, 2015 in the central part on Chile (Illapel, Coquimbo region), respectively. The last zone is also characterized by outer rise stress changes related to the subduction of the Juan Fernandez Ridge [16].

Hence, any information which could improve prediction of future seismic events could be very useful for preventing

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

disastrous consequences. Unlike other natural phenomena such as the rains or hurricanes where their occurrences can be predicted with a certain precision, for earthquake events there is not a consolidated mechanism that allows to predict when and where the next big earthquake will occur. However, there are different perspectives on earthquake prediction: while some studies concluded that earthquakes cannot be predicted [64], others proposed different methods for predicting the occurrence of future seismic events (see, for example, [7], [9], and [28]). Considering the hypothesis that a seismic event is predictable, seismic prediction models are grouped into different categories that consider the use of statistical and spatial tools, analysis of precursor signals, classic machine learning algorithms and deep learning algorithms [60]. One way to represent the seismicity in a region is through the ground intensity function which represent the expected number (or occurrence rate) of earthquake events with magnitude greater than a particular degree. This work proposes to predict the average number of daily seismic events with magnitudes greater than 2.8 on the Richter scale in the central zone of Chile (between the regions of Coquimbo and Araucanía) during the years 2010-2017 using modern machine learning techniques, especially deep neural networks. Specifically, we consider neural network models for spatio-temporal data such as the Convolutional network Long Short Term Memory (ConvLSTM) and the Multicolumn Convolutional Long Short Trem Memory (Multicolumn ConvLSTM) networks.

The data used in these models come from the the public catalog of the National Seismological Center of Chile and from the e Global Position System (GPS) instruments. In particular, the data are pre-processed in the following way: from the the seismic events of the Chilean catalog we estimate the intensity function using the spatio-temporal ETAS model, and from the GPS data, we evaluate the accumulated velocity associated to the main seismic event. Both data sets are build on a 20×20 spatial grid covering the area under study. This spatial distribution facilitates the application of convolution operators of the neural networks used. Sequences of these matrices are then used as inputs of the networks for predicting the average future number of seismic events in the following days on a 20×20 grid. Since we assume that the use of velocity time series could help to improve the prediction of the number of seismic events, we propose a network designed for multiple data sources, the Multi-column ConvLSTM. The proposed model is then compared to some traditional neural network models such as feed-forward neural network (FFNN) Long Short Term Memory (LSTM), and the Convolutional LSTM, in order to assess the goodness of the results.

II. RELATED WORK

In the last years various approaches have been proposed for the prediction of seismic events by applying deep neural networks, most of them using CNN or LSTM networks

Gravirov [31] analyzed the potential of using deep neural networks (DNN) for seismic data analysis and they concluded that DNNs are promising tools, although the major limitation of these data is the presence of few events of large magnitude. Zhou et al. [68] introduced a hybrid CNN-RNN model to detect seismic events from seismic waves captured by seismograms, outperforming traditional algorithms in terms of selection stability (low false detection) and high selection accuracy. Linville et al. [38] explored the use of convolutional and recurrent neural networks to achieve explosive and tectonic source discrimination for local distances, using a 5-year event catalog generated by seismograph stations, which performed better than 99%. Kriegerowski et al. [33] used the full waveform recording of three components from multiple seismic stations while a convolutional network (CNN) was used as a neural model. CNN successfully located events in validation. The CNN successfully located 908 events in the validation. Perol et al. [48] presented ConvNetQuake, a highly scalable convolutional neural network for earthquake detection and localization from waveform. This technique was applied to study induced seismicity in Oklahoma (USA), detecting 20 times more earthquakes than previously cataloged by the Oklahoma Geological Survey. Vijayasankari and Indhuja [60] used LSTM networks for spatio-temporal earthquake forecasting. Simulation results showed that the developed two-dimensional input LSTM network was capable of discovering and exploiting spatio-temporal correlations between earthquakes. Geng et al. [20] addressed the problem of long-term historical dependence of seismic time series prediction by proposing deep temporal convolution neural networks (DCTCNN and CNN-LSTM), demonstrating that they are superior to the other five presented algorithms, successfully completing the seismic prediction task. Huang et al. [27] proposed to project seismic events on a topographic map and generated an image dataset where earthquake with magnitude > 6 is labeled "1". The authors used a CNN to detect and predict whether these large earthquakes will occur in the next 30 days. Li et al. [37] proposed a method for earthquake fault detection using a CNN that requires only a small training set, treated the fault detection process as a semantic segmentation task, and trained a CNN encoder-decoder, to perform pixel-by-pixel prediction to determine whether each pixel is a fault or not. Wang et al. [61] trained ResNets for seismic data interpolation, where the model is used to reconstruct dense data with halved trace intervals. The generated data can provide reasonable interpolation results and can be used to improve the accuracy of subsequent algorithms. Due to physical or financial constraints, seismic data sets can often be undersampled, occasionally these data sets can also present bad data or dead traces that the geoscientist must deal with. Even neural generative models have been applied to this problem. Oliveira et al. [46] evaluated the performance of a conditional generative adversarial network for the interpolation problem on post-stack seismic datasets.

LSTM and CNN neural networks were also applied by [1], [19], [29], [30], [36], [40], [43], [49], [63] for temporal and spatial earthquake prediction. Wang et al. [63] employed a long short-term memory (LSTM) network to learn the spatio-temporal correlations among earthquakes in different locations and make predictions. Fabregas et al. [19] developed a system based on a Rule Based Algorithm and a Long Short-Term Memory (LSTM) Network which is able to forecast the following variables: frequency, maximum magnitude, and average depth of earthquake events in a specific region in a given year. Li et al. [36] proposed a Deep Learning model, called DLEP, for Earthquake Prediction called DLEP by using a CNN on eight precursory pattern-based indications and explicit features. Kavianpour et al. [30] and [1] propose a novel prediction method based on attention mechanism (AM), convolution neural network (CNN), and bi-directional long short term memory (BiLSTM) models for the earthquake prediction in China [30] and in Bangladesh [1], respectively. Kail et al. [29] proposed a deep neural network classification model based on a LSTM and a convolutional network to predict if an earthquake with the magnitude above a threshold takes place at a given area of Japan for the next 30−180 days.

Few works have dealt deep neural network model for predicting seismic events in Chile. A novel approach combining ETAS (Epidemic type aftershock sequence) models (see, [41] and [42]) with neural networks have been recently proposed by [49] and [43] to forecast seismic events for the next days in Chile. In particular, in [43], a CNN is used for predicting the region where the the seismic event with maximum magnitude will occur, and a LSTM for predicting the average number of seismic events for the next days.

While these works are mostly based on applications of CNN or LSTM for analyzing and predicting seismic events using historical data, we propose a Multi-column Convolutional LSTM (MConvLSTM) network which takes into account the past values of the earth crustal velocity as covariate besides to the ETAS intensity function values. The introduction of this new covariate was also considered by [14] where they empirically show that this component can improve the description of triggered seismicity in the ETAS model.

III. BACKGROUND

A. ETAS (EPIDEMIC TYPE AFTERSHOCK SEQUENCE)

Space-time point processes can be uniquely characterized by their conditional intensity function, which represents the mean number of events in a region and time, conditional on the past [17]. These models first appeared in applications to population genetics and are therefore also known as Epidemic Type Models. Ogata [44] and [45] introduced Epidemic-Type Aftershock Sequence (ETAS) models for modeling seismic events in the temporal and spatio-temporal domains, respectively. The space time conditional intensity function can be represented as follows:

$$\lambda(t, s|H_t) = \lim_{L(\Delta t), L(\Delta s)\to 0} \frac{E[N([t, t+\Delta t]\times[s, s+\Delta s])|H_t]}{L(\Delta t)L(\Delta s)} \quad (1)$$

where $L(\cdot)$ is the Lebesgue measure, Δt and Δs are increments in time and space, $s = (x, y)$. The intensity function $\lambda(\cdot)$ depends on a set of parameters to be estimated.

Seismic events are generally collected in a seismic catalogue, which includes the time when the event occurred, the coordinates of the hypocenter, and the magnitude for n observed events. The ETAS model is normally used for describing the spatio-temporal seismicity and the clustering features of aftershocks. The intensity function of the ETAS model is based on the history of event occurrences H_t and depends on a set of parameters, θ, to be estimated. According to [13], [45] and [14], the spatio-temporal conditional intensity can be written as:

$$\lambda_\theta(t, s|H_t) = \mu f(s) + \sum_{t_j<t} g(t - t_j|m_j)\ell(s - s_j|m_j) \quad (2)$$

where $\mu f(s)$ represents the background seismicity, stationary in time, with $f(s)$ space density. The second part of Eq. describes the triggering component given by the product the temporal density of aftershocks describing the occurrence rate of aftershocks at time t, after the earthquake at time t_j with magnitude m_j, and the density of aftershocks in space. The rate of earthquake occurrence at time t is described by the parametric model [13]:

$$g(t - t_j|m_j) = \frac{Ke^{(\alpha-\gamma)(m_j-m_0)}}{(t - t_j + c)^p} \text{ with } t > t_j \quad (3)$$

where K is an aftershock productivity constant, c and p are parameters of the modified Omori's law proposed by [59]; p is used to characterize the seismicity pattern, indicating the rate of decrease of aftershocks over time. The parameter α and γ measure the influence on the relative weight of each sequence, and m_0 is the magnitude threshold (lower limit for which earthquakes with higher magnitude values are certainly recorded in the catalogue).

For the spatial distribution conditional on the magnitude of the event, the following distribution is usually used:

$$\ell(s - s_j|m_j) = \left\{\frac{(s - s_j)^2}{e^{\gamma\ (m_j-m_0)}} + d\right\}^{-q} \quad (4)$$

where d and q are two parameters related to the spatial influence of the main earthquake. Finally, the equation for the conditional intensity can be written as follows:

$$\lambda_\theta(t, s|H_t) = \mu f(s) + \sum_{t_j<t} \frac{Ke^{(\alpha-\gamma)(m_j-m_0)}}{(t - t_j + c)^p}\left\{\frac{(s - s_j)^2}{e^{\gamma(m_j-m_0)}} + d\right\}^{-q}. \quad (5)$$

B. NEURAL NETWORKS

Artificial neural networks (ANNs) are important tools in the field of artificial intelligence. They are based on the behavior of the brain, referring to neurons and their connections. ANN models are normally designed to allow the resolution of complex problems where the relationship between the the desired output and the input variables is non-linear. Next, the neural network models used in this work will be described below.

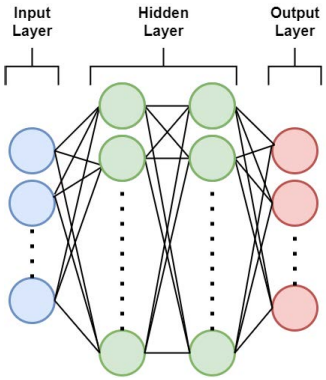


FIGURE 1. Diagram of a FFNN.

1) FEED-FORWARD NEURONAL NETWORK

One of the most popular neural network paradigms is the feed-forward neural network (FFNN) [57]. In a FFNN neural network, neurons are generally organized in layers. This network is denoted as $X \times H \times O$, where X, H, and O represent the number of input units, the number of hidden layers, and the number of output units, respectively. Figure 1 shows a fully connected feed-forward network.

The neurons of the hidden layer and the output layer are processing units. Each neuron has an activation function that is chosen, this can be a sigmoid function, ReLU, etc. The network input to neuron j is given by:

$$\text{net}_j = f(\sum_i w_{ij} x_i + \theta_j) \tag{6}$$

where x_j are the outputs of the previous layer, w_{ij} are the link weights connecting neuron i to neuron j, θ_j is the bias and f is the activation function. This model is the most basic of those available, which is why it is considered as base method to compare the proposals.

2) LONG SHORT TERM MEMORY

Long Short Term Memory (LSTM) network was proposed by Hochreiter and Schmidhuber [26], and can be considered an evolution of the Recurrent Neuronal Network (RNN). LSTM networks are models designed for data in the form of sequences, capable of learning long-term dependencies and remembering information over long periods of time. Figure 2 represents the structure of the LSTM neural network:

This network is composed of memory blocks called *cells*. These transfer two states to the next *cells*, the *cell state* and the *hiddent state*. The *cell state* is the main data flow chain, which allows data to go forward without major changes. Data can be added to or removed from the *cell state* through the sigmoid layer. The sigmoid function takes values from the output of the last LSTM unit h^{t-1} and the current input x^t at time step t. The sigmoid function will determine which part of the previous output should be removed. This gate is called the *forget gate* which will be denoted by f^t [23]; f^t is a vector with values ranging from 0 to 1, corresponding

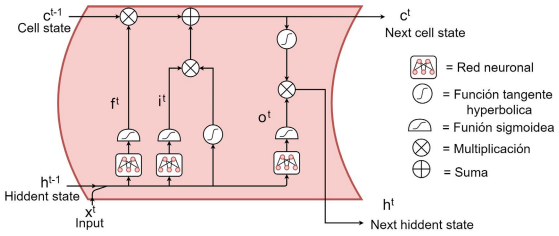


FIGURE 2. Structure of the neural network long short Trem Memory (LSTM).

to each number in the *cell state* C^{t-1}.

$$f^t = \sigma(b^f + \sum_j U_j^f x_j^t + \sum_j W_j^f h_j^{t-1}) \tag{7}$$

x^t is the input vector, σ is the sigmoid function, b^f, U^f and W^f are the bias, input weights and recurrent weights of the gate respectively.

The next step is to store the new information in the new input x^t by updating the cell state. This step contains two parts, the sigmoid layer and the hyperbolic tangent function. The sigmoid layer decides whether the new information should be updated or ignored with values between 0 or 1, and the hyperbolic tangent function gives weight to the values passed, through an importance level between -1 and 1. These two values are multiplied to update the new x^t *cell state*. This new memory is added to the old memory in C^{t-1} resulting in C^t.

$$i^t = \sigma(b^i + \sum_j U_j^i x_j^t + \sum_j W_j^i h_j^{t-1}) \tag{8}$$

$$C^t = C^{t-1} f^t + i^t \tanh(b^c + \sum_j U_j^c x_j^t + \sum_j W_j^c h_j^{t-1}) \tag{9}$$

C^{t-1} and C^t represent the cell states at time steps t and $t-1$, while b, U and W are the bias, input weights and recurrent weights of each *cell state* respectively.

For the final step, the output values h^t are based on the output *cell state* O^t. First, a sigmoid layer decides which parts of the *cell state* reach the output. Next, the output of this is multiplied by the new values created by the hyperbolic tangent function of the *cell state* C^t, with a value varying between -1 and 1.

$$O^t = \sigma(b^o + \sum_j U_j^o x_j^t + \sum_j W_j^o h_j^{t-1}) \tag{10}$$

$$h^t = O^t tanh(C^t) \tag{11}$$

b^o, U^o y W^o are the bias, the input weights, and the recursive weights of the *output gate*, respectively. Although it does not consider the spatial disposition of the events, this model is adequate to the available temporal data, for which it is considered as a base method in the experiments.

3) CONVOLUTIONAL NEURONAL NETWORK

The Convolutional neural network (CNN) is a specialized type of neural network for processing gridded data [23].

Convolutional networks are neural networks that use convolution instead of general matrix multiplication. This operation refers to a mathematical combination of two functions to produce a third function. In the case of CNNs, the convolution is performed between the input data and a kernel that performs the filter function, so as to obtain a *feature map*, this operation is usually indicated by an asterisk (see, Figure 3).

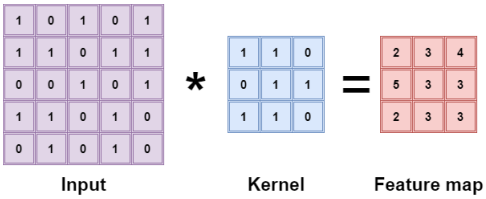


FIGURE 3. Convolution example.

In machine learning applications, the input is usually an array of data and the kernel is usually an array of parameters to be learned. Often convolutional ones are used on more than one dimension at a time, which is why if a two-dimensional image I is used as input, a two-dimensional kernel K should probably also be used:

$$s[i,j] = (I * K)[i,j] = \sum_m \sum_n I[m,n]K[i-m, j-n]. \tag{12}$$

This model is a component of the ConvLSTM and Multi-column ConvLSTM models, which are the ones proposed in this work.

4) ConvLSTM

LSTM networks have been widely used to process sequential data however it can have difficulties modeling grid data, due to the large number of weights and large matrix multiplications required to parameterize each cell [10]. A more suitable alternative in this type of data is the LSTM convolutional network (ConvLSTM), which differs from LSTM in that it replaces the matrix product operations with convolution operations. This model has been applied to spatio-temporal data such as rainfall [56] or videos [62].

Figure 4 shows the structure of a ConvLSTM cell. The architecture of this network allows large matrices to be operated using convolutions sequentially, which drastically reduces the number of parameters needed to train. Simultaneously, the convolution operation allows to model the relations between the neighboring cells of the input. The mathematical formulation of the ConvLSTM model is:

$$\begin{aligned} i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \\ f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \\ C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\ o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \\ H_t &= o_t \circ \tanh(C_t) \end{aligned} \tag{13}$$

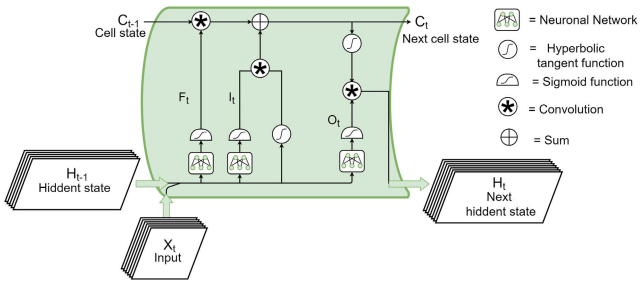


FIGURE 4. Convolutional long short Trem memory (ConvLSTM) neural network structure.

where σ is the sigmoidal function, $W_{x\sim}$ and $W_{h\sim}$ are 2D convolutional kernels, X_t is defined as the input, C_t the cell output, H_t the hidden state and the gates i_t, f_t, o_t. In general, all variables correspond to 3D tensors. The symbol "$*$" denotes the convolution operator, while "$\circ$" denotes the Hadamard product. Our proposal uses this architecture because the input spatio-temporal data considers a grid shape in the spatial representation of the geographic regions under study.

5) MULTI-COLUMN CONVOLUTIONAL NEURAL NETWORK

The Multi-column Convolutional Neural Network (MCNN) is a neural network architecture proposed by [15] in the context of visual object recognition. This neural network differs from a typical convolutional neural network in that it contains multiple branches which are connected at the output layer, thus handling multiple inputs simultaneously. This model has been applied to counting multiple people [67] or emotion recognition using EEG [66]

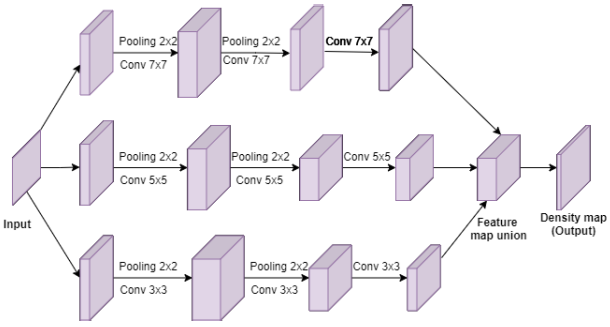


FIGURE 5. Multi-column convolutional neuronal network architecture [67].

Figure 5 shows the architecture used in [67]. This model shows the use of different filters in each component CNN which allow parallel processing of patterns at different scales within the input image. Subsequently, the outputs of these networks are concatenated to form the output density map. Our proposal consists in the use of a variant of this architecture to process different input data through multiple components.

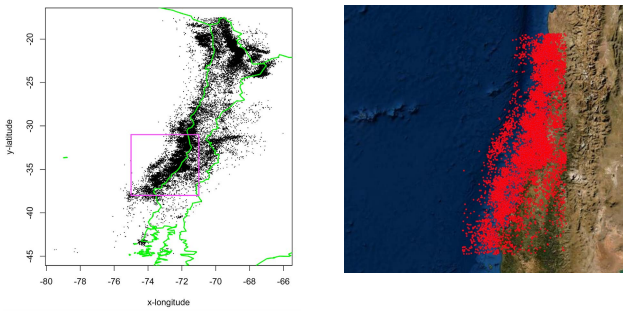


FIGURE 6. Seismic events in Chile from 2010 to 2017 with the studied area in the square (left panel); Google map of the study area with seismic events (right panel).

IV. PROPOSED METHODOLOGY

In this work, data from the earthquake Chilean catalogue, covering the Coquimbo and Araucanía regions, located between latitudes $[-38.00, -31.00]$ and longitudes $[-75.00, -71.00]$ are considered (see, Figure 6). The selection of this area is mainly related to its sismicity. In particular, it is characterized by an high number of interplate events (with depth less than 70 km) occurred mainly along the cost of Chile. Also, this area includes the epicenters of two big earthquakes: the Constitución earthquake occurred on the south of the area on February 27, 2010 and the Illapel earthquake occurred on September 16, 2015 located in the north of the study area. Two types of variables are considered for the earthquake prediction: the daily accumulated velocity and the values of the intensity function from the estimation of the ETAS model. In summary, the proposed methodology mainly consists of (i) data pre-processing and (ii) the application of neural models. First, data pre-processing is described to facilitate the convolution operation used in neural models. Then the specific architectures of the proposed networks based on ConvLSTM are described. These two data sets are described below.

A. VELOCITY BASED ON GPS SENSORS

Several studies show that GPS data provide useful information for the earthquake prediction (see, for example, [24] and [21]). In this work, we use the accumulated displacement velocity, obtained by the GPS measurements, as a predictor of the average number of daily seismic events. To obtain the daily accumulated velocities associated to each seismic event in the study area, we downloaded the GPS measurements from the web page *http://geodesy.unr.edu/NGLStationPages/GlobalStationList* and we use the seismic catalogue provided by the National Seismological Center of the University of Chile (https://www.sismologia.cl/). While GPS stations capture daily coordinates in a defined time interval and spatial point, the seismic catalogue includes some main features associates to each seismic event, such as longitude, latitude, magnitude, date of occurrence, and depth.

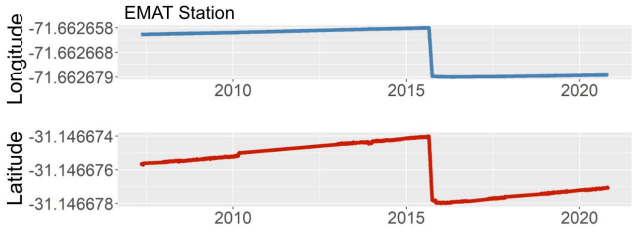


FIGURE 7. Longitude y latitude time series for the EMAT station.

Figure 7 shows an example of time series of the coordinates (longitude y latitude) recorded by the EMAT GPS station located in the Coquimbo region, in the period from 2007 to 2021. The velocity field considered in this work has been calculated by estimating the regression line at the GPS stations as mentioned in [18]. In this case an interval of 5 days prior to the event was considered for the adjustment of the line. The slopes representing the vertical and horizontal velocity were obtained to finally perform the Pythagorean theorem and thus obtain the movement in mm/5days. The imputTS library of R software has been used for interpolating missing data. The preparation of the input data for the neural networks in 20×20 grid format required the following steps:

1) Estimation of missing values of the time series, for which two steps were required:
   - Missing value estimation through the R library imputTS
   - Denoising process of time series through R wmtsa library
2) Assignment of the velocity to each seismic event through the following algorithm:

Algorithm 1 Assignment of Velocity to Seismic Catalogue

input : C,G,n
[C= Catalog of seismic events
G= Time series of GPS stations
n= Number of seismic events
j= Number of columns of C]
for $i = 0$ **to** n **do**
  $G'[i] \leftarrow$ Find nearby station$(C[i,], G)$;
  $D[i] \leftarrow$ Calculate velocity $(G'[i])$;
  $C[i, j+1] \leftarrow D[i]$
output: C

3) Generation of the 20×20 grid in the geographic space.

When creating this data set representing the displacement velocity, a 20×20 grid per day was generated in the defined geographic space. Each cell was associated to a velocity value measured by the nearest GPS station. In the case of the presence of multiple GPS stations within a cell area, the cell was assigned to the velocity with the largest value. Finally, in

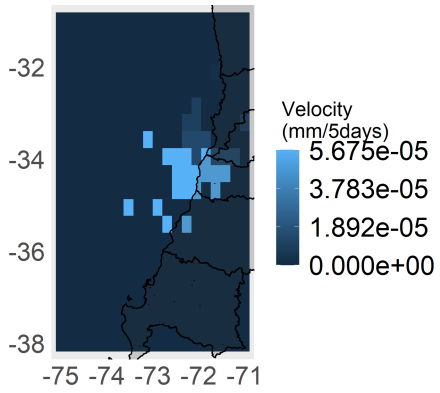


FIGURE 8. Displacement velocity for the day 02/03/2010.

case of absence of velocity in a cell, it was assigned the value of zero.

Figure 8 shows the velocity calculated in the defined zone corresponding to March 2, 2010, that is, three days after the big Constitucion earthquake of magnitude 8.8.

B. INTENSITIES OF THE ETAS MODEL

The Chilean seismicity is here represented through the conditional intensity function of the ETAS model. In particular, the conditional intensity describes the average number of seismic events greater than a predetermined threshold in a spatial region. The input data of the ETAS model are the seismic events (in terms of longitude, latitude, magnitude, and time) recorded by the National Seismological Center of Chile. For the spatial estimation, we consider a 20×20 grid over the geographic area of interest. This zone has been defined between latitudes $[-38.00,-31.00]$ and longitudes $[-75.00, -71.00]$ in the years 2010-2017 corresponding to the Coquimbo and La Araucania regions of Chile. In the estimation of the ETAS intensity function, we considered a magnitude threshold of 2.8 on the Richter scale. This value has been selected by analyzing the log-cumulate distribution of the magnitudes of the data in the studied area as suggested by [14]. For each cell, the intensity value is calculated using Eq. 5. The estimation of the parameters of this function has been implemented by using the *EtasFLP* library [13] of the R software.

Table 1 shows the estimated parameters according to the estimation of the intensity function. According to [14], we fix the parameter p to 1 1and γ to 0.

Figure 9 shows the values of the intensity function for the day 03/02/2010 in Chile between the regions of Coquimbo and Araucanía. Finally, the intensity functions are processed by logarithmic transformation as proposed by [43], in order to reduce the high asymmetry in the data.

V. THE PROPOSED MULTI-COLUMN LSTM NEURAL NETWORK

The neural network model proposed in this work, called Multi-column ConvLSTM, is based on ConvLSTM neural networks arranged in two parallel branches. This model can be seen as a variant of the Multi-column CNN network seen in section III-B5. The proposed network is then compared with a ConvLSTM network where both input data are processed in a single branch as shown by Figure 10.

In Figure 11, we present the structure of the Multi-column ConvLSTM network with two parallel inputs proposed to treat independently the velocity and intensity. The two models are then combined in the last layer to perform the prediction.

The structure of the proposed Multi-column ConvLSTM network is composed of two ConvLSTM layers for each column, with Dropout and Batch Normalization components between layers. Finally, a third ConvLSTM layer is added where the two columns that process intensity and velocity are connected.

TABLE 1. Parameter estimation of the ETAS model.

Parameters	Estimates	Std Error
μ	0,6796	0,0182
k	0,2440	0,0449
c	0,0135	0,0007
p	1,0000	0,0000
d	45,6977	2,9065
q	1,7382	0,0271

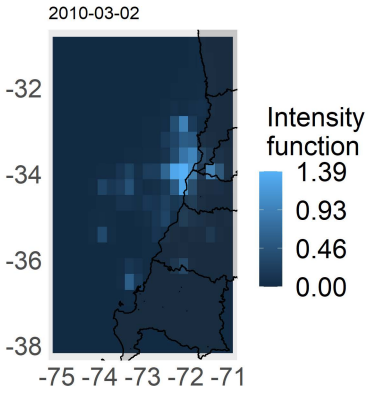


FIGURE 9. ETAS model estimation for the day 03/02/2010.

VI. EXPERIMENTAL PROCEDURE

This section describes the preparation of the data and the experimental results for the validation of the proposed models.

A. DATA PREPARATION

Experiments have been performed following the block-based cross-validation scheme [11]. This method is based on dividing the data sequentially into blocks. In this case the blocks are chosen considering the different seismicity in each period.

The database, covering the period from 01/31/2009 to 12/31/2017, is sequentially partitioned into three blocks considering the years 2015, 2016 and 2017 within the test data set. The validation set corresponds to the previous year of the testing year, while the training set corresponds to all

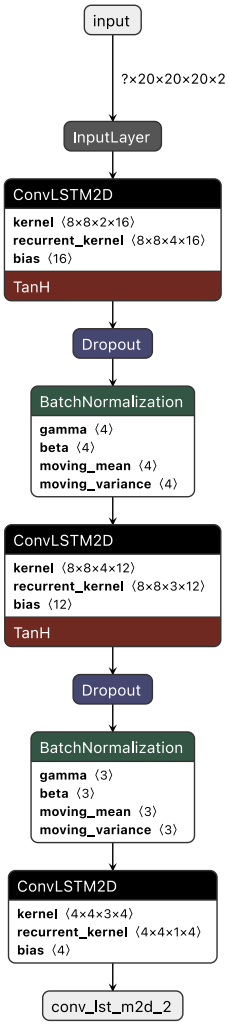


FIGURE 10. Structure of the tested ConvLSTM model.

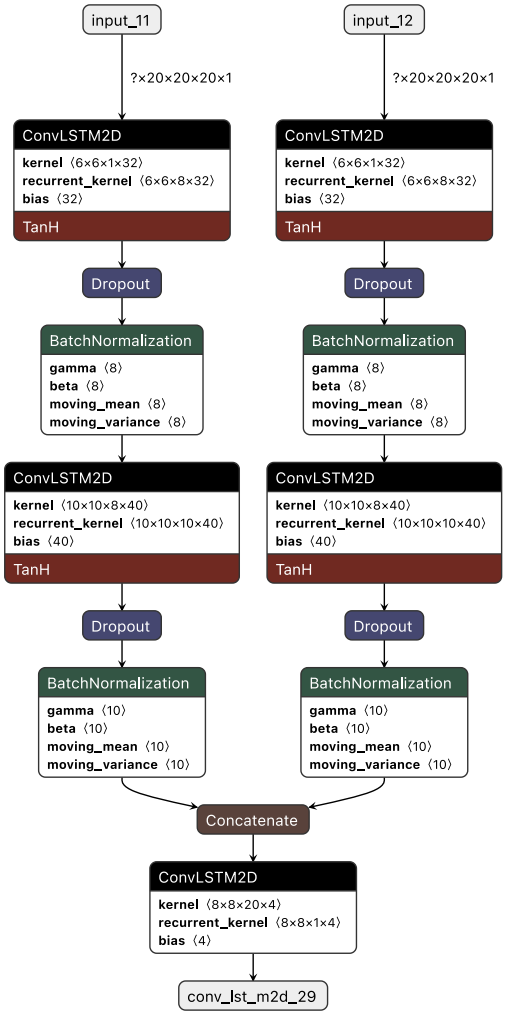


FIGURE 11. Structure of the proposed multi-column ConvLSTM model.

previous years. All blocks are initialized following Xavier's [22] method. For the prediction, 20 previous days are used to make the forecast for the next day.

In the Table 2 the sequential blocks of data are detailed with the corresponding dates. For the choice of hyperparameters of the neural networks, the Keras-Tuner tool proposed by [47] was used. This optimization is done considering only the training and validation sets. Regarding the parameters of this optimization, we considered 5 iterations for the hyperparameter optimization algorithm based on hyperbands as well as 25 iterations per model [47]. After finding the hyperparameters of the models, the training of each model was carried out considering 240 epochs and a batch-size of size 20.

TABLE 2. Partition of the data in training, validation and testing sets for three different periods.

Period	Dates	Partition	Time interval
1	31/12/09 - 31/12/15	Train	(31/12/09 - 31/12/13)
		Val	(01/01/14 - 31/12/14)
		Test	(01/01/15 - 31/12/15)
2	31/12/09 - 31/12/16	Train	(31/12/09 - 31/12/14)
		Val	(01/01/15 - 31/12/15)
		Test	(01/01/16 - 31/12/16)
3	31/12/09 - 31/12/17	Train	(31/12/09 - 31/12/15)
		Val	(01/01/16 - 31/12/16)
		Test	(01/01/17 - 31/12/17)

B. EXPERIMENTAL RESULTS

The experiments were performed at a quantitative and qualitative level. In the quantitative experiments, the neural models are tested and the metrics are reported to evaluate the quality of each model. Qualitative experiments are complementary. In this part, some predictions at the spatial and temporal level are visually shown.

1) QUANTITATIVE EXPERIMENTS

The quantitative experiments were performed considering the proposed neural networks, ConvLSTM and Multi-column ConvLSTM. Base methods such as multidimensional FFN and LSTM were added for comparison. The FFN network

consisted of 3 hidden layers, while the LSTM network has 2 hidden layers. Each neural network was hyperoptimized considering the number of neurons per layer. The metrics considered to evaluate the quality of the models are the mean squared errors (MSE), the mean absolute error (MAE) and the coefficient of determination (R^2). These metrics are generally used to evaluate time series forecast models. On the other hand, since it is not obvious how to comparatively evaluate metrics with non-predefined scales such as MSE or MAE, we have chosen to additionally compare them using the relative improvement criterion, which is applied in the machine learning community ([12], [65]). For example, if two techniques obtain an MSE of 0.128 and 0.064, then the relative improvement of the second over the first is $(1-0.064/0.128)*100\% = 50\%$, which expresses the error reduction to the half. For uniformity, this criterion was also adapted for the R^2. In the following, we will focus on the results for the test set although we will also show the results for the training and validation sets.

a: ConvLSTM

This neural network considers historical conditional intensity and velocity data as input. The architecture is specified in Figure 10. Table 3 shows the results of the ConvLSTM neural network for each block.

The results show that this network obtained on average a MSE equal to 0.0051, a MAE equal to 0.0039 and an R^2 equal to 0.761 on the test set. In all periods, a determination coefficient greater than 0.74 was obtained. When analyzing the results by blocks, it is observed that the best metrics are in the second block with MSE, MAE and R2 with values 0.0038, 0.033 and 0.78, respectively.

TABLE 3. Resulting metrics with their standard deviations (in parenthesis) in the training and testing sets for different periods using ConvLSTM network.

Period	Train			Val			Test		
	MSE	MAE	R2	MSE	MAE	R2	MSE	MAE	R2
1	0.0050	0.0434	0.841	0.0043	0.0398	0.713	0.0066	0.0475	0.747
	(0.1E-4)	(0.2E-4)	(3.0E-4)	(0.2E-4)	(0.4E-4)	(0.001)	(0.3E-4)	(0.6E-4)	(0.001)
2	0.0044	0.0361	0.847	0.0054	0.0388	0.785	0.0038	0.0336	0.780
	(0.1E-4)	(0.2E-4)	(3.0E-4)	(0.2E-4)	(0.6E-4)	(0.001)	(0.2E-4)	(0.5E-4)	(0.001)
3	0.0043	0.0345	0.846	0.0036	0.0317	0.787	0.0049	0.0364	0.756
	(0.1E-4)	(0.2E-4)	(3.0E-4)	(0.2E-4)	(0.5E-4)	(0.001)	(0.2E-4)	(0.5E-4)	(0.001)
	Train			Val			Test		
	MSE	MAE	R2	MSE	MAE	R2	MSE	MAE	R2
Mean	0.0046	0.0380	0.845	0.00446	0.0368	0.761	0.0051	0.0392	0.761
Sd	(0.1E-4)	(0.2E-4)	(0.3E-4)	(0.2E-4)	(0.5E-4)	(0.001)	(0.2E-4)	(0.5E-4)	(0.001)

b: MULTI-COLUMN ConvLSTM (MConvLSTM)

This neural network considers both historical velocity and seismic intensity data as inputs. Each component column of this network processes each input. Table 4 presents the results of this neural network by block.

The results show that this network obtained on average an MSE of 0.0039, MAE of 0.031 and R^2 of 0.812. About MSE and MAE, the best result was given in the second block with values of 0.034 and 0.0294; while with respect to R^2 it reached 0.827. We observe that in the first block the best value of R^2 was reached, and also the highest values of MSE and MAE, which indicates that the metrics are complementary. When comparing these results with those of the ConvLSTM, the MConvLSTM manages to significantly outperform them in all metrics. In MSE, MAE and R^2 it achieves on average a relative improvement of 23%, 20% and 6%, which indicates that MConvLSTM clearly outperforms ConvLSTM. We propose that this improvement is due to the use of multicolumns, which allow the data to be initially processed independently, while ConvLSTM mixes them from the beginning, which can make it difficult to find the best hypotheses of the predictive model.

c: BASE METHODS

For comparison, other classic neural network architectures are applied: Feed Forward Neural Network (FFNN) and the multidimensional Long Short Trem Memory (LSTM). The FFNN network maintains the data in the form of sequences, however, it does not consider local connections between nodes, losing local temporal and spatial connectivity. On the other hand, the LSTM network does consider temporal memory, however, it does not consider spatial connectivity, unlike the ConvLSTM and Multi-column ConvLSTM networks.

TABLE 4. Resulting metrics with their standard deviations (in parenthesis) for the Multi-column ConvLSTM network processed with intensities and logarithm of velocities as inputs.

Period	Train			Val			Test		
	MSE	MAE	R2	MSE	MAE	R2	MSE	MAE	R2
1	0.0038	0.0304	0.880	0.0031	0.0269	0.789	0.0044	0.0330	0.827
	(0.1E-4)	(0.2E-4)	(3.0E-4)	(0.2E-4)	(0.5E-4)	(0.001)	(0.3E-4)	(0.5E-4)	(0.001)
2	0.0038	0.0302	0.869	0.0045	0.0321	0.822	0.0034	0.0294	0.805
	(0.1E-4)	(0.3E-4)	(0.001)	(0.3E-4)	(0.5E-4)	(0.001)	(0.2E-4)	(0.4E-4)	(0.001)
3	0.0036	0.0299	0.873	0.0030	0.0273	0.825	0.0040	0.0313	0.804
	(0.1E-4)	(0.2E-4)	(0.001)	(0.2E-4)	(0.6E-4)	(0.001)	(0.2E-4)	(0.5E-4)	(0.001)
	Train			Val			Test		
	MSE	MAE	R2	MSE	MAE	R2	MSE	MAE	R2
Mean	0.0037	0.0302	0.874	0.0035	0.0288	0.812	0.0039	0.0312	0.812
Sd	(0.1E-4)	(0.2E-4)	(3.0E-4)	(0.2E-4)	(0.5E-4)	(0.001)	(0.2E-4)	(0.5E-4)	(0.001)

Table 5 shows the metrics of the LSTM and FFNN models considering the intensity and velocity as inputs. The LSTM metric values are better than those given by FFN. In relation to MSE, MAE and R^2, LSTM obtains an average of 0.012, 0.075 and 0.408, while FFN has 0.0168, 0.091 and 0.222, respectively. These results occur because LSTM considers the sequential nature of the data leading to a clear advantage over FFN. However, both networks have much lower performance than the ConvLSTM and Multi-column ConvLSTM network models, which indicates that spatial information is very relevant in this problem.

d: ConvLSTM (USING ONLY SEISMIC INTENSITY)

In this way we complement the ConvLSTM neural network using only the seismic intensity data as input. The architecture was similar to the previous experiment using ConvLSTM where only the input data was changed.

Although this work studies the use of intensity as velocity, we have observed that the latter variable is comparatively highly dispersed due to the fact that there are few GPS sensors compared to seismic sensors. This may imply that not necessarily both data sources are similarly reliable.

Table 6 shows the behavior of the ConvLSTM network without the velocity variable as input. For this network,

TABLE 5. Resulting metrics with their standard deviations (in parenthesis) for the LSTM and FFNN models processed with intensities and logarithm of velocities as inputs.

Model	Period	Train MSE	Train MAE	Train R2	Val MSE	Val MAE	Val R2	Test MSE	Test MAE	Test R2
LSTM	1	0.0035 (0.1E-4)	0.0342 (0.2E-4)	0.889 (2.0E-4)	0.0058 (0.2E-4)	0.0504 (0.5E-4)	0.609 (0.001)	0.0138 (0.5E-4)	0.0735 (0.8E-4)	0.470 (0.001)
	2	0.0033 (0.1E-4)	0.0319 (0.2E-4)	0.885 (2.0E-4)	0.0152 (0.5E-4)	0.0687 (0.9E-4)	0.406 (0.001)	0.0117 (0.3E-4)	0.0788 (0.7E-4)	0.324 (0.001)
	3	0.0032 (0.1E-4)	0.0324 (0.2E-4)	0.886 (2.0E-4)	0.0063 (0.2E-4)	0.0561 (0.5E-4)	0.633 (0.001)	0.0115 (0.3E-4)	0.0719 (0.7E-4)	0.431 (0.001)
FFNN	1	0.0170 (0.3E-4)	0.0870 (0.4E-4)	0.460 (0.001)	0.0126 (0.2E-4)	0.0936 (0.5E-4)	0.159 (0.001)	0.0256 (0.6E-4)	0.1190 (0.8E-4)	0.011 (0.002)
	2	0.0148 (0.2E-4)	0.0774 (0.4E-4)	0.489 (4.0E-4)	0.0209 (0.7E-4)	0.0880 (0.1E-4)	0.186 (0.002)	0.0141 (0.3E-4)	0.0823 (0.8E-4)	0.185 (0.001)
	3	0.0080 (0.1E-4)	0.0588 (0.2E-4)	0.717 (3.0E-4)	0.0077 (0.2E-4)	0.0647 (0.5E-4)	0.552 (0.001)	0.0108 (0.2E-4)	0.0703 (0.6E-4)	0.468 (0.001)

Model	Metric	Train MSE	Train MAE	Train R2	Val MSE	Val MAE	Val R2	Test MSE	Test MAE	Test R2
LSTM	Mean	0.0033	0.0328	0.887	0.0091	0.0584	0.550	0.0123	0.0747	0.408
	Sd	(0.1E-4)	(0.2E-4)	(2.0E-4)	(0.3E-4)	(0.6E-4)	(0.001)	(0.4E-4)	(0.4E-4)	(0.001)
FFNN	Mean	0.0133	0.0744	0.556	0.0137	0.0821	0.299	0.0168	0.0905	0.222
	Sd	(0.2E-4)	(0.3E-4)	(6.0E-4)	(0.4E-4)	(0.4E-4)	(0.001)	(0.4E-4)	(0.7E-4)	(0.001)

the average values of MSE, MAE and R^2 resulted 0.0042, 0.0339 and 0.802, respectively. It is notable that the ConvLSTM network outperforms the ConvLSTM network with both intensity and velocity data sources. Nonetheless, it still has a lower result (on average) than Multi-column ConvLSTM which considers both data sources. Note that Multi-column only makes sense when considering multiple data sources. When analyzing each metric, it is observed that, on average, Multi-colum ConvLSTM obtains a relative improvement compared to ConvLSTM (with only intensity) of 7%, 8% and 1.2% in the MSE, MAE and R^2, respectively.

TABLE 6. Resulting metrics with their standard deviations (in parenthesis) for the ConvLSTM network with only intensity as input.

Period	Train MSE	Train MAE	Train R2	Val MSE	Val MAE	Val R2	Test MSE	Test MAE	Test R2
1	0.0043 (0.1E-4)	0.0387 (0.2E-4)	0.864 (3.0E-4)	0.0037 (0.2E-4)	0.035 (0.4E-4)	0.754 (0.001)	0.0051 (0.3E-4)	0.0401 (0.5E-4)	0.805 (0.001)
2	0.0040 (0.1E-4)	0.0315 (0.2E-4)	0.861 (4.0E-4)	0.0046 (0.2E-4)	0.0321 (0.5E-4)	0.817 (0.001)	0.0034 (0.2E-4)	0.0292 (0.5E-4)	0.802 (0.001)
3	0.0038 (0.1E-4)	0.0327 (0.2E-4)	0.867 (2.0E-4)	0.0031 (0.2E-4)	0.0292 (0.4E-4)	0.819 (0.001)	0.0041 (0.2E-4)	0.0321 (0.5E-4)	0.799 (0.001)

	Train MSE	Train MAE	Train R2	Val MSE	Val MAE	Val R2	Test MSE	Test MAE	Test R2
Mean	0.0040	0.0343	0.864	0.0038	0.0321	0.797	0.0042	0.0339	0.802
Sd	(0.1E-4)	(0.2E-4)	(3.0E-4)	(0.2E-4)	(0.4E-4)	(0.001)	(0.2E-4)	(0.5E-4)	(0.001)

2) QUALITATIVE EXPERIMENTS

Qualitative experiments are complementary to quantitative experiments due to the difficulty of visually analyzing the results of space-time predictions. This subsection visually shows a set of predictions at the spatial level considering a specific day and, at the temporal level, considering the cumulative number of events in the study area. In particular, in this analysis, we consider the period including the Illapel earthquake of magnitude 8.4 on the Richter scale that occurred on September 16, 2015

a: SPATIAL PREDICTION

Figure 12 shows the spatial behavior of the predictions considering all the models that used the intensity and velocity data. For comparing the performances of the models we considered the fifth day after the great Illapel seismic event of September 16, 2015, as shown in Figure 12, since the prediction of the main event is still a challenge open problem. It is evident that the proposed Multi-column ConvLSTM

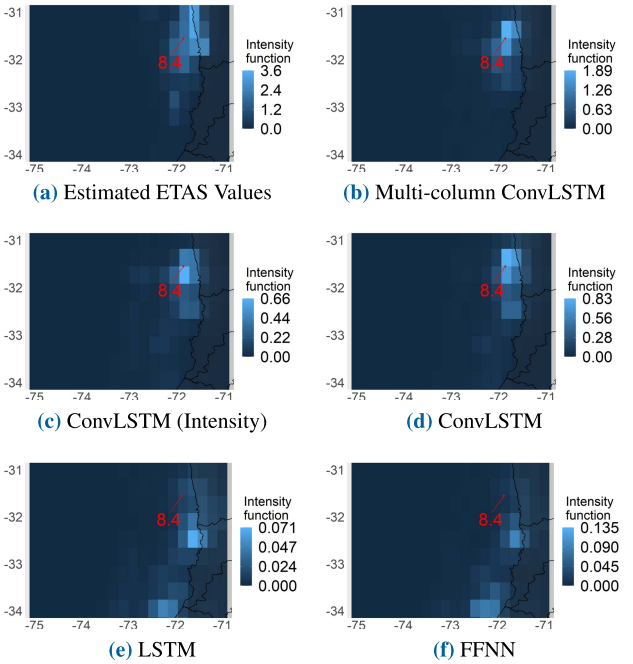


FIGURE 12. Spatio-temporal predictions obtained by different models considering intensity and velocity. This graph refers to the fifth day after the main Illapel earthquake occurred on September 16, 2015. The text in red colour indicates the location and the magnitude of the main event.

network (Figure 12 b) obtains much better predictions than other models. The LSTM and FFNN models (Figure 12 d and e) present poor performances in terms of the number of events in this day.

b: TEMPORAL PREDICTION

Figure 13 shows the behavior of the intensity function at the location where the Illapel 2015 earthquake occurs through neural models considering data intensity and velocity. It is observed that the both ConvLSTMs and Multi-column ConvLSTM models fit the real intensity values better than the FFNN and LSTM models. Moreover, both models have close levels of prediction, which is why the analysis of metrics is required to be able to compare both models considering all the blocks of the data set. In this aspect, when we analyze the 15 consecutive days to the fifth day after the main event main event, given in Figure 14, it is observed that the Multi-column ConvLSTM model obtains a great advantage over the ConvLSTM model. This interestingly suggests that velocity information is more relevant on days close to a major event, in this case in the first 7 days later, after which the models tend to have similar performance.

However, these models can be used to predict the number of triggering events but none of them can predict the main event.

Figure 13 shows the behavior of the estimates of the logarithm of the intensity function for the different networks in the period 1 test, as well as the ETAS model estimates and the observed values. Among the estimates of the different

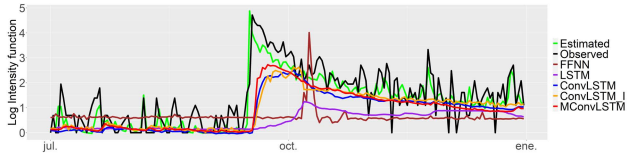


FIGURE 13. Time series of the logarithm of the intensity function where the main event occurs, between latitude [−33.5,−31] and longitude [−73,−71], for the different models.

models, it is observed that the M-ConvLSTM network shows a better performance. For the time series of the observed events, a linear interpolation treatment was performed when no seismic event was recorded in the area, in order to avoid problems with the logarithmic transformation.

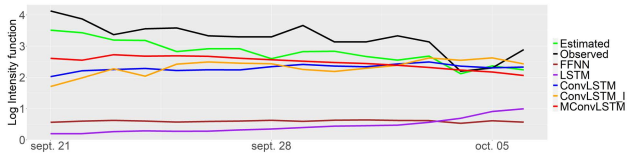


FIGURE 14. Logarithm of the intensity time series between latitudes [−33.5,−31] and longitudes [−73,−71] considering 15 consecutive days up to the fifth day after the main event on September 16, 2015.

3) ANALYSIS OF RESULTS RESPECT THE OBSERVED DATA

Additionally, we measured the performance of the models respect the observed data. Although the models were optimized considering the logarithm of the intensity function according to the ETAS model as the output variable, we also assessed models with reference to the observed seismic events. For this, we considered the period with the greatest telluric movement, that is, the first period where the year 2015 was tested.

Figure 15a) and b) show the predicted intensity function by the M-ConvLSTM which the seismic events overlapped for five and ten days after the main event, respectively. It can be seen that the highest predicted intensity coincides with the area where the most earthquakes occurred. This indicates that the Multicolum ConvLSTM network is able to capture the spatio-temporal behaviour of the aftershock events. To quantify this result, we calculated the metric values by comparing the observed data and the model results considering the logarithm of the total number of daily seismic events as well as the number of daily predicted events during the year 2015.

By observing Table 7, it is evident that MconvLSTM is again the model that best predicts the logarithm of the observed number of seismic events leaving both ConvLSTM models in second place according to the evaluated metrics. Therefore, despite the fact that the model focused on values of the logarithm of the number of events, it is still able to recover the patterns of the number of daily events.

Note that in this work, we considered the prediction of the logarithm transformation of the intensity function due to the high asymmetry of the number of seismic events [5]. In particular, the presence of high magnitude earthquakes

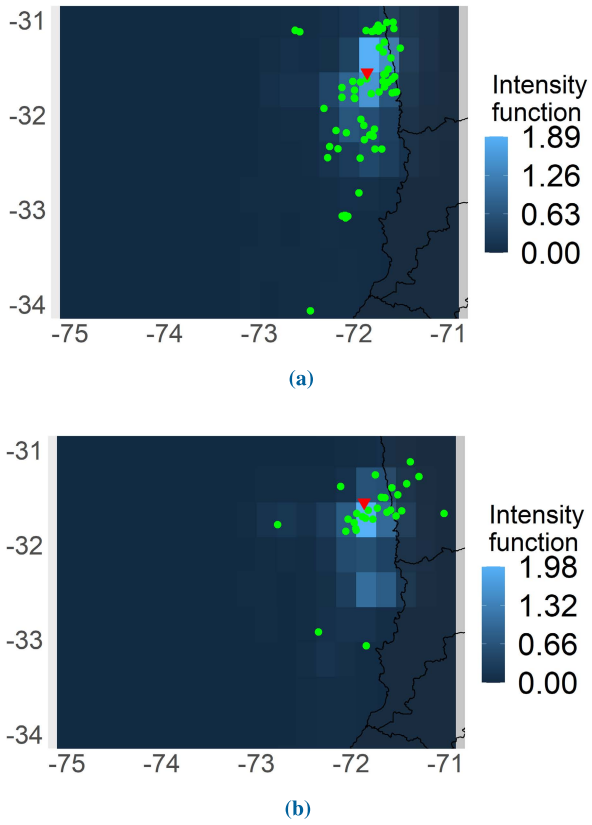


FIGURE 15. Spatio-temporal predictions using the M-ConvLSTM network with the observed seismic events overlapped (green dots) on the fifth (a) and tenth (b) days after the main Illapel earthquake occurred on September 16, 2015. The red triangle indicates the main event.

TABLE 7. Resulting metrics with their standard deviations (in parenthesis) respect the observed number of events.

Method	MSE	MAE	R2
MConvLSTM	**0.486** (0.015)	**0.501** (0.006)	**0.448**(0.012)
ConvLSTM(Only intensity)	0.510 (0.017)	0.522 (0.006)	0.421 (0.011)
ConvLSTM	0.531 (0.017)	0.517 (0.006)	0.396 (0.012)
LSTM	0.819 (0.023)	0.623 (0.008)	0.070 (0.009)
FFN	0.977 (0.025)	0.757 (0.008)	-0.110 (0.015)
Etas model	0.235 (0.007)	0.377 (0.004)	0.733 (0.008)

(greater than 7.0) originate a very high number of aftershocks, which decreases rapidly in the following days. This behavior caused some difficulties in the models to produce relatively good predictions, so that logarithm transformation was considered. In subsequent works we hope to refine the neural models to predict the number of events directly.

VII. DISCUSSION AND RESULTS

When visualizing the resulting metrics of all the experiments, it can be concluded that the FFNN approach is the weakest, failing to capture the continuous changes of the intensity function in the space-time of the three periods. LSTM networks are able to capture some spatio-temporal change, however their determination coefficients are still low. On the contrary, the ConvLSTM and Multi-column ConvLSTM networks better capture the continuous changes of the

intensity function. When comparing the best models, that is, the ConvLSTM network without the use of velocity data and the Multicolum ConvLSTM network with both intensity and velocity data, it is observed that the Multicolum-ConvLSTM network outperforms the ConvLSTM, on average. On the other hand, when comparing the results of the ConvLSTM neural network using intensity and velocity data with respect to the same network using only intensity data, the last one obtains better results on average. This leads us to suggest that architectures with multiple columns are the best option when using multiple data sources. In general, the experiment that obtained the best results was the Multi-column ConvLSTM network, where the aggregation of the velocity information contributes slightly to the improvement of the performance. One aspect of interest is that the greatest improvement compared to ConvLSTM in the MSE, MAE and R^2 metrics occurs in the first block corresponding to the 2015 test, when the great Illapel earthquake occurred (relative improvements of 13.7%, 17.7% and 2.6%, respectively). On the other hand, the smallest improvements occur in the second block (being similar in MSE and MAE with an improvement in R^2 of 0.4%) corresponding to the 2016 test.

We can note that in 2016, no major earthquake occurred within the study area. A similar pattern happens in the third block where the year 2017 is tested. Therefore, this leads us to suggest that the advantage of using velocity information with a Multi-column ConvLSTM network is more exploitable when there are periods with larger earthquakes, which are precisely the most relevant cases from the perspective of disastrous effects, which highlights the importance of this study. On the other hand, the experiments show that if the prediction is performed in periods where there are no major earthquakes, the advantage may not be so high, although it should be noted that these periods are the ones of least interest.

Finally, we can note that the results given in this work improved those observed in [43] where the maximum values of the intensity functions were processed in an LSTM network for different areas of Chile, where determination coefficients were obtained that varied between 0.32 to 0.65 for different magnitudes of cut.

VIII. CONCLUSION AND FUTURE WORK

Spatio-temporal prediction of the conditional intensity values is a highly challenging task. In this work, the ConvLSTM and Multi-column ConvLSTM neural network architectures were proposed for the prediction of the rate of occurrence of seismic events in Chile, using as input the values of the intensity function of the ETAS model and the previous velocity of the displacement associated to a seismic catalog. The results showed that the Multi-column Convolutional LSTM network, using both intensity and velocity, turns out to be the best model among those tested, since it obtained the best average values of MSE, MAE and determination coefficient in the test set corresponding to 0.004, 0.031 and 0.81, respectively. Interestingly, it is observed that the use of velocity in a Multi-column ConvLSTM network generates a large relative advantage over the use of only intensity in a ConvLSTM network when a period containing a large earthquake within the study region is evaluated, as in the case of the great Illapel earthquake of 2015.

Despite the relatively good results obtained in this work, further questions need to be addressed in the future. One of these is to further refine the pre-processing of the velocity data using spatial interpolations in the points where the GPS measurement is missing. Also, the real number of events (and a pre-processing of them) should be considered instead of the ETAS intensity values for making predictions more interpretable and realistic. Finally, new advanced neural network should be experimented, such as the neural network with attention.

REFERENCES

[1] M. H. A. Banna, T. Ghosh, M. J. A. Nahian, K. A. Taher, M. S. Kaiser, M. Mahmud, M. S. Hossain, and K. Andersson, "Attention-based bi-directional long-short term memory network for earthquake prediction," *IEEE Access*, vol. 9, pp. 56589–56603, 2021.

[2] K. M. Asim, A. Idris, T. Iqbal, and F. Martínez-Álvarez, "Seismic indicators based earthquake predictor system using genetic programming and AdaBoost classification," *Soil Dyn. Earthq. Eng.*, vol. 111, pp. 1–7, Aug. 2018.

[3] S. Barrientos, "The seismic network of Chile," *Seismol. Res. Lett.*, vol. 89, no. 2A, pp. 467–474, Mar. 2018.

[4] S. E. Barrientos, *Earthquakes in Chile*. London, U.K.: Geological Society Special Publication, Jan. 2007, pp. 263–287.

[5] K. Benoit, "Linear regression models with logarithmic transformations," London School Econ., London, U.K., Tech. Rep., 2011, pp. 23–36, vol. 22, no. 1.

[6] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," *Inf. Sci.*, vol. 191, pp. 192–213, May 2012.

[7] A. Berhich, F.-Z. Belouadha, and M. I. Kabbaj, "LSTM-based models for earthquake prediction," *Int. J. High Perform. Syst. Archit.*, vol. 10, no. 1, pages 1–7, Mar. 2020.

[8] T. Bhandarkar, N. Satish, S. Sridhar, R. Sivakumar, and S. Ghosh, "Earthquake trend prediction using long short-term memory RNN," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 9, no. 2, p. 1304, Apr. 2019.

[9] J. D. Brehm and W. L. Braile, "Intermediate-term earthquake prediction using precursory events in the New Madrid Seismic Zone," *Bull. Seismol. Soc. Amer.*, vol. 88, no. 2, pp. 564–580, 1998.

[10] J. Burge, M. Bonanni, M. Ihme, and L. Hu, "Convolutional LSTM neural networks for modeling wildland fire dynamics," 2020, *arXiv:2012.06679*.

[11] V. Cerqueira, L. Torgo, and I. Mozetič, "Evaluating time series forecasting models: An empirical study on performance estimation methods," *Mach. Learn.*, vol. 109, no. 11, pp. 1997–2028, Nov. 2020.

[12] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[13] M. Chiodi and G. Adelfio, "Mixed non-parametric and parametric estimation techniques in R package etasFLP for Earthquakes' description," *J. Stat. Softw.*, vol. 76, no. 3, 2017.

[14] M. Chiodi, O. Nicolis, G. Adelfio, N. D'Angelo, and A. Gonzàlez, "ETAS space–time modeling of Chile triggered seismicity using covariates: Some preliminary results," *Appl. Sci.*, vol. 11, no. 19, p. 9143, Oct. 2021.

[15] C. Dan Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," 2012, *arXiv:1202.2745*.

[16] V. Clouard, J. Campos, A. Lemoine, A. Perez, and E. Kausel, "Outer rise stress changes related to the subduction of the Juan Fernandez Ridge, central Chile," *J. Geophys. Res.*, vol. 112, no. B5, pp. 1–15, May 2007.

[17] D. J. Aley and D. Vere-Jones, *An Introduction to Theory Point Processes*. New York, NY, USA: Springer-Verlag, 2007.

[18] L. E. Sjöberg, P. Ming, and A. Erick, "An analysis of the Äespöe crustal motion-monitoring network observed by GPS in 2000, 2001 and 2002," Svensk Kärnbränslehantering AB, Stockholm, Sweden, Tech. Rep. R-02-33, 2002.

[19] A. C. Fabregas, P. B. V. Arellano, and A. N. D. Pinili, "Long-short term memory (LSTM) networks with time series and spatio-temporal approaches applied in forecasting earthquakes in the Philippines," in *Proc. 4th Int. Conf. Natural Lang. Process. Inf. Retr.*, New York, NY, USA, Dec. 2020, pp. 188–193.

[20] Y. Geng, L. Su, Y. Jia, and C. Han, "Seismic events prediction using deep temporal convolution networks," *J. Electr. Comput. Eng.*, vol. 2019, pp. 1–14, Apr. 2019.

[21] V. G. Gitis, A. B. Derendyaev, and K. N. Petrov, "Earthquake prediction based on combined seismic and GPS monitoring data," in *Computational Science and Its Applications—ICCSA* (Lecture Notes in Computer Science), O. Gervasi, B. Murgante, S. Misra, C. Garau, I. Blečić, D. Taniar, B. O. Apduhan, A. Maria, A. C. Rocha, E. Tarantino, and C. M. Torre, Eds. Cham, Switzerland: Springer, 2021, pp. 601–612.

[22] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[23] J. Ian Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[24] M. Hao, Y. Li, and W. Zhuang, "Crustal movement and strain distribution in East Asia revealed by GPS observations," *Sci. Rep.*, vol. 9, no. 1, p. 16797, Dec. 2019.

[25] J. L. Hardebeck and T. Okada, "Temporal stress changes caused by earthquakes: A review," *J. Geophys. Res., Solid Earth*, vol. 123, no. 2, pp. 1350–1365, Feb. 2018.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[27] J. Huang, X. Wang, Y. Zhao, C. Xin, and H. Xiang, "Large earthquake magnitude prediction in Taiwan based on deep learning neural network," *Neural Netw. World*, vol. 28, no. 2, pp. 149–160, 2018.

[28] Y. Y. Kagan and L. Knopoff, "Statistical short-term earthquake prediction," *Science*, vol. 236, no. 4808, pp. 1563–1567, Jun. 1987.

[29] R. Kail, E. Burnaev, and A. Zaytsev, "Recurrent convolutional neural networks help to predict location of earthquakes," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[30] P. Kavianpour, M. Kavianpour, E. Jahani, and A. Ramezani, "A CNN-BiLSTM model with attention mechanism for earthquake prediction," 2021, *arXiv:2112.13444*.

[31] K. V. Kislov and V. V. Gravirov, "Deep artificial neural networks as a tool for the analysis of seismic data," *Seismic Instrum.*, vol. 54, no. 1, pp. 8–16, Jan. 2018.

[32] A. Konstantaras, N. S. Petrakis, T. Frantzeskakis, E. Markoulakis, K. Kabassi, I. O. Vardiambasis, T. Kapetanakis, A. Moshou, and E. Maravelakis, "Deep learning neural network seismic big-data analysis of earthquake correlations in distinct seismic regions," *Int. J. Adv. Technol. Eng. Explor.*, vol. 8, no. 84, pp. 1410–1423, Nov. 2021.

[33] M. Kriegerowski, G. M. Petersen, H. Vasyura-Bathke, and M. Ohrnberger, "A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms," *Seismol. Res. Lett.*, vol. 90, no. 2A, pp. 510–516, Mar. 2019.

[34] F. Leyton, J. Ruiz, J. Campos, and E. Kausel, "Intraplate and interplate earthquakes in Chilean subduction zone: A theoretical and observational comparison," *Phys. Earth Planet. Interiors*, vol. 175, nos. 1–2, pp. 37–46, Jun. 2009.

[35] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," *J. Mach. Learn. Res.*, vol. 18, no. 185, pp. 1–52, 2018.

[36] R. Li, X. Lu, S. Li, H. Yang, J. Qiu, and L. Zhang, "DLEP: A deep learning model for earthquake prediction," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[37] S. Li, C. Yang, H. Sun, and H. Zhang, "Seismic fault detection using an encoder–decoder convolutional neural network with a small training set," *J. Geophys. Eng.*, vol. 16, no. 1, pp. 175–189, Feb. 2019.

[38] L. Linville, K. Pankow, and T. Draelos, "Deep learning models augment analyst decisions for event discrimination," *Geophys. Res. Lett.*, vol. 46, no. 7, pp. 3643–3651, Apr. 2019.

[39] S. Lorito, F. Romano, S. Atzori, X. Tong, A. Avallone, J. McCloskey, M. Cocco, E. Boschi, and A. Piatanesi, "Limited overlap between the seismic gap and coseismic slip of the great 2010 Chile earthquake," *Nature Geosci.*, vol. 4, no. 3, pp. 173–177, Mar. 2011.

[40] I. M. Murwantara, P. Yugopuspito, and R. Hermawan, "Comparison of machine learning performance for earthquake prediction in Indonesia using 30 years historical data," *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 18, no. 3, p. 1331, Jun. 2020.

[41] O. Nicolis, M. Chiodi, and G. Adelfio, "Windowed ETAS models with application to the Chilean seismic catalogs," *Spatial Statist.*, vol. 14, pp. 151–165, Nov. 2015.

[42] O. Nicolis, M. Chiodi, and G. Adelfio, "Space-time forecasting of seismic events in chile," in *Earthquakes—Tectonics, Hazard and Risk Mitigation*. IntechOpen, 2017, doi: 10.5772/66339.

[43] O. Nicolis, F. Plaza, and R. Salas, "Prediction of intensity and location of seismic events using deep learning," *Spatial Statist.*, vol. 42, Apr. 2020, Art. no. 100442.

[44] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," *J. Amer. Statist. Assoc.*, vol. 83, no. 401, pp. 9–27, 1988.

[45] Y. Ogata, "Space-time point-process models for earthquake occurrences," *Ann. Inst. Stat. Math.*, vol. 50, no. 2, pp. 379–402, 1998.

[46] D. A. B. Oliveira, R. S. Ferreira, R. Silva, and E. V. Brazil, "Interpolating seismic data with conditional generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1952–1956, Dec. 2018.

[47] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi. (2019). *Keras Tuner*. [Online]. Available: https://github.com/keras-team/keras-tuner

[48] T. Perol, M. Gharbi, and M. Denolle, "Convolutional neural network for earthquake detection and location," *Sci. Adv.*, vol. 4, no. 2, Feb. 2018, Art. no. e1700578.

[49] F. Plaza, R. Salas, and O. Nicolis, "Assessing seismic hazard in Chile using deep neural networks," *Natural Hazards*, J. P. Tiefenbacher, Ed. Rijeka, Croatia: IntechOpen, 2019, ch. 1.

[50] J. Reyes, A. F. Morales-Esteban, and F. Martínez-Álvarezc, "Neural networks to predict earthquakes in Chile," *Appl. Soft Comput.*, vol. 13, no. 2, pp. 1314–1328, 2013.

[51] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.

[52] S. Ruiz and R. Madariaga, "Historical and recent large megathrust earthquakes in Chile," *Tectonophysics*, vol. 733, pp. 37–56, May 2018.

[53] I. Santibanez, J. C. Perasso, G. González, F. Aron, and G. Yanez, "Crustal faults in the Chilean Andes: Tectonic significance and implications for geologic hazard," in *Proc. AGU Fall Meeting Abstr.*, Dec. 2013.

[54] M. Schnaubelt, "A comparison of machine learning model validation schemes for non-stationary time series data," Inst. Econ., Friedrich-Alexander-Univ. Erlangen-Nürnberg, FAU Discuss. Papers, 2019.

[55] F. P. Schoenberg, A. Chu, and A. Veen, "On the relationship between lower magnitude thresholds and bias in epidemic-type aftershock sequence parameter estimates," *J. Geophys. Res.*, vol. 115, no. B4, pp. 1–15, 2010.

[56] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," 2015, *arXiv:1506.04214*.

[57] Z. Tang and P. A. Fishwick, "Feedforward neural nets as models for time series forecasting," *ORSA J. Comput.*, vol. 5, no. 4, pp. 374–385, Nov. 1993.

[58] C. Tassara, S. Cesca, M. Miller, J. Á. López-Comino, C. Sippl, J. Cortés-Aranda, and B. Schurr, "Seismic source analysis of two anomalous earthquakes in northern Chile," *J. South Amer. Earth Sci.*, vol. 119, Nov. 2022, Art. no. 103948.

[59] T. Utsu, "A statistical study on the occurrence of aftershocks," *Geophys. Mag.*, vol. 30, pp. 521–605, Dec. 1961.

[60] S. Vijayasankari and P. Indhuja, "Earthquake prediction based on spatio-temporal data mining approach," *Int. J. Sci. Eng. Res.*, vol. 9, no. 4, pp. 1573–1579, 2018.

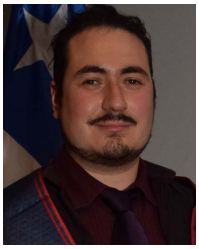
[61] B. Wang, N. Zhang, W. Lu, and J. Wang, "Deep-learning-based seismic data interpolation: A preliminary result," *Geophysics*, vol. 84, no. 1, pp. V11–V20, Jan. 2018.

[62] D. Wang, Y. Yang, and S. Ning, "DeepSTCL: A deep spatiotemporal ConvLSTM for travel demand prediction," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[63] Q. Wang, Y. Guo, L. Yu, and P. Li, "Earthquake prediction based on spatio-temporal data mining: An LSTM network approach," *IEEE Trans. Emerg. Topics Comput.*, vol. 8, no. 1, pp. 148–158, Jan. 2020.

[64] M. Wyss, R. Aceves, S. Park, R. Geller, D. Jackson, Y. Kagan, and F. Mulargia, "Cannot earthquakes be predicted?" *Sci.*, vol. 278, no. 5337, pp. 487–490, 1997.

[65] J. Xiao, H. Ye, X. He, H. Zhang, F. Wu, and T.-S. Chua, "Attentional factorization machines: Learning the weight of feature interactions via attention networks," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 3119–3125.
[66] H. Yang, J. Han, and K. Min, "A multi-column CNN model for emotion recognition from EEG signals," *Sensors*, vol. 19, no. 21, p. 4736, Oct. 2019.
[67] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 589–597.
[68] Y. Zhou, H. Yue, Q. Kong, and S. Zhou, "Hybrid event detection and phase-picking algorithm using convolutional and recurrent neural networks," *Seismol. Res. Lett.*, vol. 90, no. 3, pp. 1079–1087, May 2019.



ALEX GONZÁLEZ FUENTES was born in Valparaíso, Chile, in 1993. He received the Graduate degree in statistics from the University of Valparaíso, Chile, in 2018, and the master's degree in computer science from Universidad Andrés Bello, in 2021. He is currently teaches at the Faculty of Engineering, Universidad Andrés Bello. His research interests include deep learning, artificial intelligence, and geostatistics. He received a Scholarship from Universidad Andrés Bello for master's degree.

ORIETTA NICOLIS (Member, IEEE) received the degree in economics from the University of Verona, Italy, in 1995, and the Ph.D. degree in statistics from the University of Padua, Italy, in 1999. She did a postdoctoral fellowship in statistics at the University of Brescia, Italy, in the following two years. From 2002 to 2012, she worked as a Researcher and an Aggregate Professor in statistics with the University of Bergamo, Italy. From 2012 to 2018, she was worked at the University of Valparaiso, where she was the Director of the Ph.D. Program in statistics. Since August 2018, she has been a Full Professor at the Engineering Faculty, University Andres Bello en Vinã del Mar, Chile, where she is responsible of national projects on artificial intelligence and statistical models. She is the author of more than 60 international publications and a reviewer of several international scientific journals. Her research interests include the study of spatio–temporal models, machine learning methods, deep learning, big data, computer science, wavelet-transforms, fractional, and multifractal processes.

BILLY PERALTA (Member, IEEE) was born in Lima, Peru, in 1981. He received the B.S. degree in system engineering from the Universidad Nacional de Trujillo, Peru, in 2001, and the M.S. and Ph.D. degrees in computer science from Pontificia Universidad Católica, Chile, in 2008 and 2013, respectively. From 2004 to 2018, he was an Assistant Professor with the Universidad Católica de Temuco, Chile. Since 2018, he has been an Assistant Professor with the Universidad Andres Bello, Chile. He is the author of 22 articles. His research interests include the area of machine learning and computer vision. In recent years, he has focused on applications of deep learning applications. He has collaborated actively with researchers in several other scientific disciplines. He is a member of IEEE Society and Chilean Society of Computer Science.

MARCELLO CHIODI has been a Full Professor in statistics at the Department of Economics, Business and Statistics, University of Palermo, Italy, since 2000. Previously, he was a Researcher in statistics, in 1983, and an Associate Professor, in 1992. He has been the Coordinator of Ph.D. programs in statistical and economic areas, from 2008 to 2015, the Head of the Department, from 2015 to 2018, and the President of the Italian National Commission for the Scientific Qualification to the role of a Professor in statistics (2019–2022). He has been visiting at the University of Stuttgart, Germany, and Valparaiso, Chile, and elected to the National Board of the SIS (Italian Statistical Society), from 2004 to 2008. His scientific interests include the area of Statistics in seismology, space-time processes, parametric and non parametric dependence models, multivariate statistics, cluster analysis, simulation techniques, and approximations to sampling distributions. Presently, he has a course on linear models and a course on nonparametric statistical models and is an Academic Senator elected to represent a Full Professor at the University of Palermo.