We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Assessing Seismic Hazard in Chile Using Deep Neural Networks

Francisco Plaza, Rodrigo Salas and Orietta Nicolis

Abstract

Earthquakes represent one of the most destructive yet unpredictable natural disasters around the world, with a massive physical, psychological, and economical impact in the population. Earthquake events are, in some cases, explained by some empirical laws such as Omori's law, Bath's law, and Gutenberg-Richter's law. However, there is much to be studied yet; due to the high complexity associated with the process, nonlinear correlations among earthquake occurrences and also their occurrence depend on a multitude of variables that in most cases are yet unidentified. Therefore, having a better understanding on occurrence of each seismic event, and estimating the seismic hazard risk, would represent an invaluable tool for improving earthquake prediction. In that sense, this work consists in the implementation of a machine learning approach for assessing the earthquake risk in Chile, using information from 2012 to 2018. The results show a good performance of the deep neural network models for predicting future earthquake events.

Keywords: deep neural networks, conditional intensity function, DFANN, RNN-LSTM, seismic hazard prediction

1. Introduction

Chile is a one of the most seismic countries in the world, with an average of a major earthquake (> 8 in Richter scale) every 10 years. The last major earthquake in Chile was registered on February 27, 2010, that affected almost 80% of the Chilean population, registering 525 deaths and several wounded. Therefore, having a better approximation or additional information on where, when an event of that magnitude could occur would represent an invaluable tool for managing and designing public policies regarding natural disasters [1, 2]. However, earthquake prediction is a very challenging task, due to its highly complex, chaotic, or nonlinear nature, and also, their occurrence depend on a multitude of variables that in most cases are yet unidentified [3, 4].

Ogata [5] introduced epidemic-type aftershock sequence (ETAS) models for seismic hazard estimation; those models and their multiple extensions [6–11] are statistical models that use a given parametrization of the expected number of events in a given region conditional on the past events, also known as the conditional ground intensity function (GIF). The GIF is 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 [12]. Also, it is widely used for earthquake forecast applications [11, 13, 14]. Although the ETAS models are very good for estimating the intensity function and forecasting triggering events, they normally fail to predict the risk of main events

due to their limitations in identifying foreshock events. Then, their performance could also be affected by the use of very large datasets.

Joffe et al. [15] stated that current techniques are insufficiently sensitive to allow for precise modeling of future earthquake occurrences. The above raises the importance for new approaches that consider broader and bigger sources of information. In that sense, deep learning (DL) models 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, DL methods, like deep feedforward artificial neural networks (DFANNs) and recurrent neural networks with long short-term memory (RNN-LSTM), have appeared in the last few years, with incredible success to a variety of problems: speech recognition, language modeling, translation, time series anomaly detection, and stock market prediction, to name a few [16]. This paper presents a temporal deep learning approach for ground intensity function estimation in Chile, using historical information from seismic event catalogs.

2. Methods

The general purpose for this work is to use a deep learning (DL) approach with deep feedforward artificial neural networks (DFANNs) and a recurrent neural networks with long short-term memory (RNN-LSTM) for ground intensity function estimation. First, the data are preprocessed to estimate the daily ground intensity function; then the output is used as input for the DL networks (DFANN and RNN-LSTM). Finally, both DL approaches are compared to find the best model. A description of the proposed procedure is shown in **Figure 1**.

2.1 Data

The database consisted of 86,000 seismic event records occurred in Chile, from 2000 to 2017, obtained from the National Seismological Center (http://www.sismologia.cl); each record consists of a time location (year, month, day, hour, minute, and second), a spatial location (latitude and longitude), depth (in kilometers), and magnitude (on Richter scale). **Figure 2** shows the spatial distribution of seismic events with magnitude superior to 6 (in Richter scale).

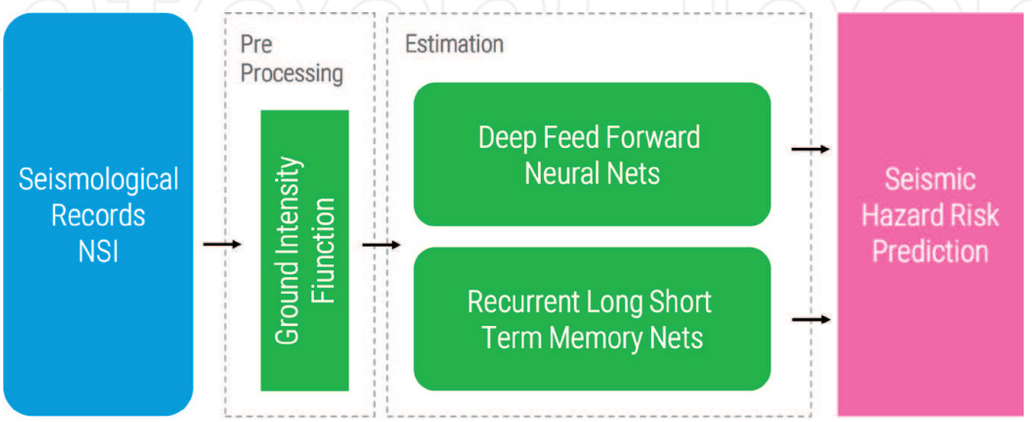

Figure 1.
Scheme for the two modular DL neural network framework: data preprocessing and estimation modules. In the data preprocessing module, all data are analyzed and prepared as inputs for the following modules; this considers estimating the daily ground intensity function. The estimation module will receive inputs from the previous model and use DFANN and RNN-LSTM DL to estimate and predict the ground intensity function.

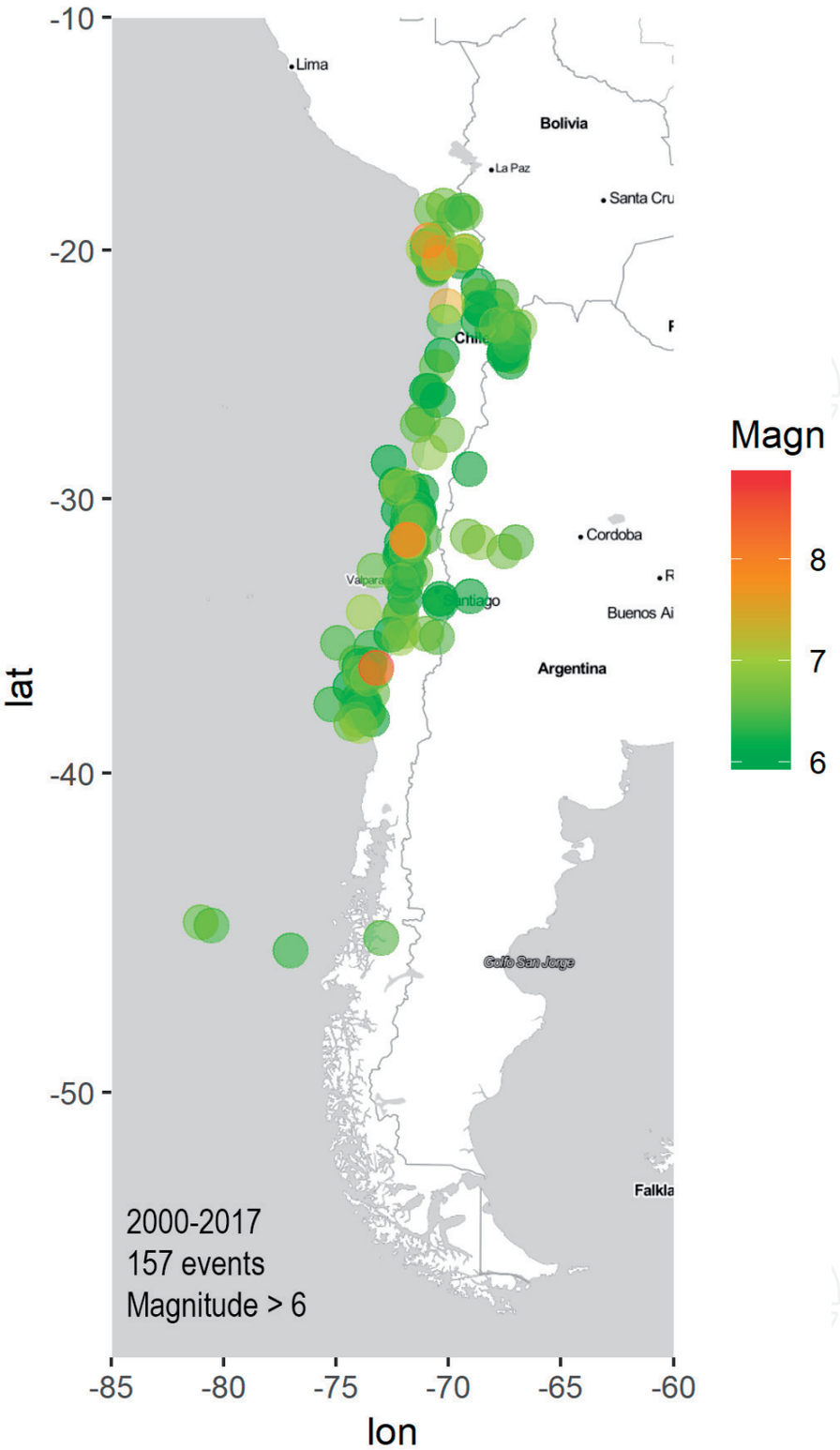

Figure 2.
Spatial distribution of seismic events (magnitude >6 Richter) for the period 2000–2017 in Chile.

2.2 Data preprocessing module

The data preprocessing module consists of estimating the conditional intensity function that represents a way of specifying how the present depends on the past in an evolutionary point process [17]. Point process models have become essential components in the assessment of seismic hazard. A particular class is given by the

self-exciting temporal point process which models events whose rate at time t may depend on the history of events at times preceding t, allowing events to trigger new events (see [18, 19] 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 [5, 20] introduced the epidemic-type aftershock sequence (ETAS) models for modeling seismic events. These models are characterized by a parametric intensity function which represents the occurrence rate of an earthquake at time t conditional on the past history of the occurrence.

ETAS models and its successive extensions have proven to be extremely useful in the description and modeling of earthquake occurrence times and locations. Self-exciting point process models [5, 19] were initially introduced in time and successively extended to the space [19]. The temporal self-exciting point processes can be defined in terms of the conditional ground intensity function (GIF):

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

where $N(A)$ is 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 $t_i < t_{i+1}$,the GIF 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 can be considered the base rate that prevents the process to die 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 [5]. 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. For example, the magnitude of an earthquake also influences how many aftershocks there will be.

Different parameterizations have been proposed for the functions m and f. Ogata [5] proposed the use of $c(m) = e^{\beta(m-M_t)}$ and $f(t) = \frac{K}{(t+c)^p}$, where the parameter β measures the effect of magnitude in the production of aftershocks and f is the modified Omori formula [12], 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 ground intensity function estimation can be estimated using the PtProcess library available in R [21].

3. Estimation module

Once the GIF databases are obtained for each magnitude (>3, >4, >5 and >6), they are structured for estimation with the DL models. The database is separated in two groups, training and test (67 and 33% of the data, respectively). A lookback of 3 is used, meaning that the output in time t will be estimated considering a window of t_{-1}, t_{-2}, t_{-3} inputs. Also both models were trained with 100 epochs.

3.1 Deep feedforward neural networks (DFANNs)

Deep feedforward artificial neural network (DFANN), also called feedforward neural networks or multilayer perceptron, is the most popular and widely known artificial neural network. In this network, the information is propagated in a

forward direction, from the input nodes through the hidden nodes (if any) and to the output nodes. As stated by [22, 23], DFANNs are universal approximators, and the universal approximation theorem states that "every bounded continuous function with bounded support can be approximated arbitrarily closely by a multilayer perceptron by selecting enough but a finite number of hidden neurons with appropriate transfer function" [22, 24].

The goal of a DFANN is to approximate some function f^* by mapping $\hat{y} = f(x;\theta)$ and learn the value of the parameters θ that result in the best function approximation for f^* [25].

The DFANN model consists a set of elementary processing elements called neurons. These units are organized in an architecture with three types of layers: the input or sensory layer, the hidden, and the output layers. The neurons corresponding to one layer are linked to the neurons of the subsequent layer without any type of bridge, lateral, or feedback connections. The connections symbolize the flux of information between neurons. **Figure 3** illustrates the architecture of this artificial neural network with r hidden layers.

DFANN operates as follows. The input signal is received by the neurons of the input layer; these neurons are just in charge of propagating the signal to the first hidden layer, and they do not make any processing. The first hidden layer processes the signal (applying a nonlinear transformation or transfer function) and transfers it to the subsequent layer; the second hidden layer propagates the signal to the third and so on. The number of hidden layers gives the depth of the model, hence the term "deep." When the signal is received and processed by the output layer, it generates the response.

The knowledge of the DFANN is registered, by the learning algorithms, in the connections between the neurons of each layer $\theta = \{\theta_1, \theta_2, \ldots, \theta_r\}$, called weights. Several learning algorithms have been created to estimate the weights, where the most

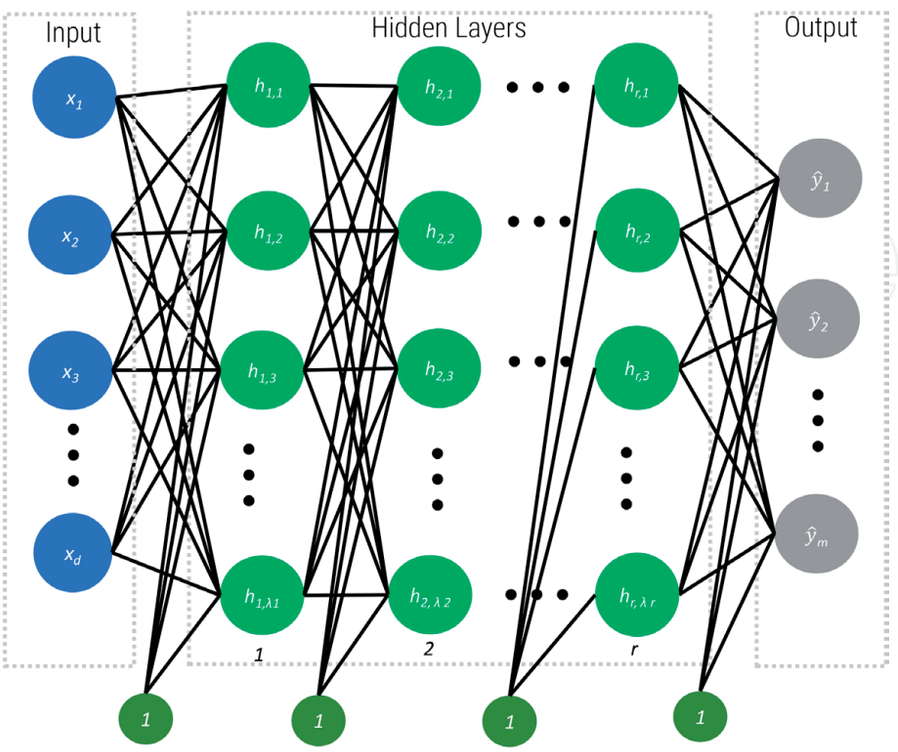

Figure 3.
Deep feedforward artificial neural network (DFANN).

popular and the first being the backpropagation, also known as generalized delta rule, popularized by [26]. The backpropagation learning algorithm is a supervised learning method and is an implementation of the Delta rule. It requires the desired output for any given input to be able to compute the output error. The main idea of the algorithm is to have a backward propagation of the errors from the output nodes to the inner nodes. For the construction of the backpropagation learning algorithm, we need to compute the gradient of the error of the network with respect to the network's modifiable weights. A DFANN network with 4 hidden layers and 12 neurons in each layer was implemented for this work.

3.2 Recurrent neural networks with long short-term memory (RNN-LSTM)

As firstly proposed by Rumelhart [26], recurrent neural networks have a primitive type of memory, in the form of recurrent layers that can operate in time [27]. Each recurrent layer takes both the output of the previous layer and an internal output of the current layer as inputs. Thus, RNNs are ideal for dealing with time series data [27]. RNNs can solve the purpose of sequence handling to a great extent but not entirely; they are great when it comes to short contexts, but to be able to build a story and remember it, the models need to be able to understand and remember the context behind longer sequences, just like a human brain. This is not possible with a simple RNN. Long short-term memory (LSTM) networks [28] are a type of RNN precisely designed to escape the long-term dependency issue of recurrent networks. LSTM recurrent networks (RNN-LSTM) have memory cells that have an internal recurrence (a self-loop), in addition to the outer recurrence of the RNN. The latter adds a nonlinear transformation to the inputs [28]. These memory cells, A, are controlled mainly by the memory door, the forgetting door ($\boldsymbol{h_t}$), and the output door. The memory door activates the entry of information to the memory cell, and the forgetting door selectively erases certain information in the memory cell and activates the storage to the next entry [29]. Finally, the output door decides what information the memory cell will emit [30]. The LSTM network structure is illustrated in **Figure 4**. Each cell has three gate activation functions σ and two output activation functions defined by tanh as a nonlinear transfer function.

In addition, they classify and predict based on time series data, since there may be delays of unknown duration between important events in a series of time. It allows clearly remembering events selected from far away in the past, which contrasts with basic NRs, for which the memory of an event decays over time [27].

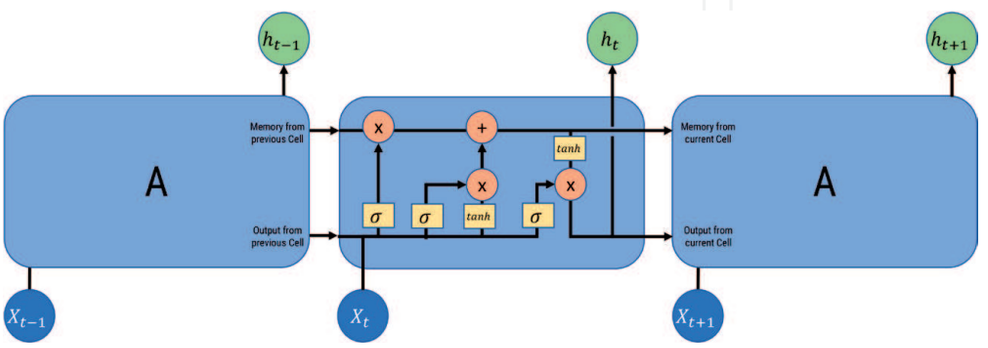

Figure 4.
LSTM cells structure, based on the work by [31].

A 1-layer RNN-LSTM with 12 cells was implemented for this work. Both DL models were implemented using Keras, with TensorFlow as backend, in Python.

4. Results

Figure 5 shows GIF estimation for the data preprocessing module, estimated for magnitudes >3, >4, >5, and >6, respectively. Note that with higher magnitudes, the GIF time series become thinner, due to the decrease of seismic events that fit in the category.

The structure implemented for both DFANN and RNN-LSTM models is shown in **Figure 6**.

The DFANN model performs slightly better than the RNN-LSTM models, in particular for lesser magnitudes (>3). **Table 1** shows the training and test performance measures (root mean square error, RMSE) for each magnitude group and DL model. Both models show better performances with magnitude >3, that is, when more information are available.

Also, a representation of the training and test results for the best model are shown in **Figure 7**. The model captures the trend very well; however, it does not perform accordingly in terms of the magnitude of the intensity function.

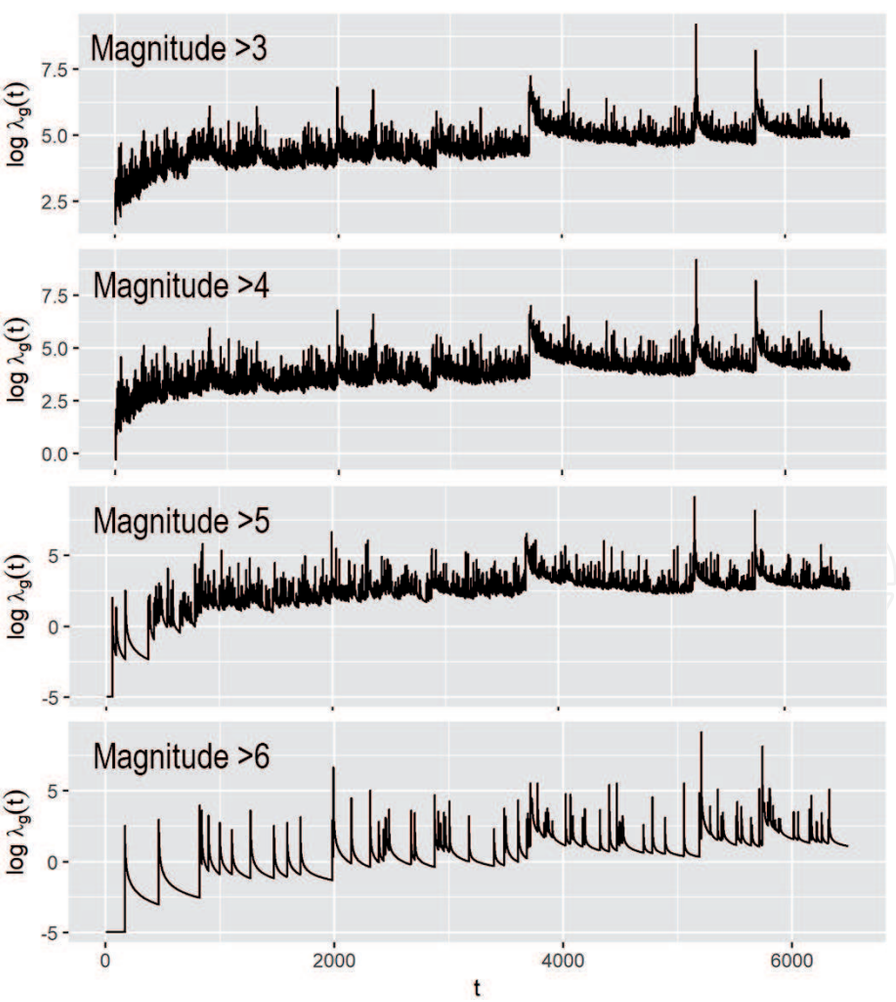

Figure 5.
Ground intensity function (GIF) estimation.

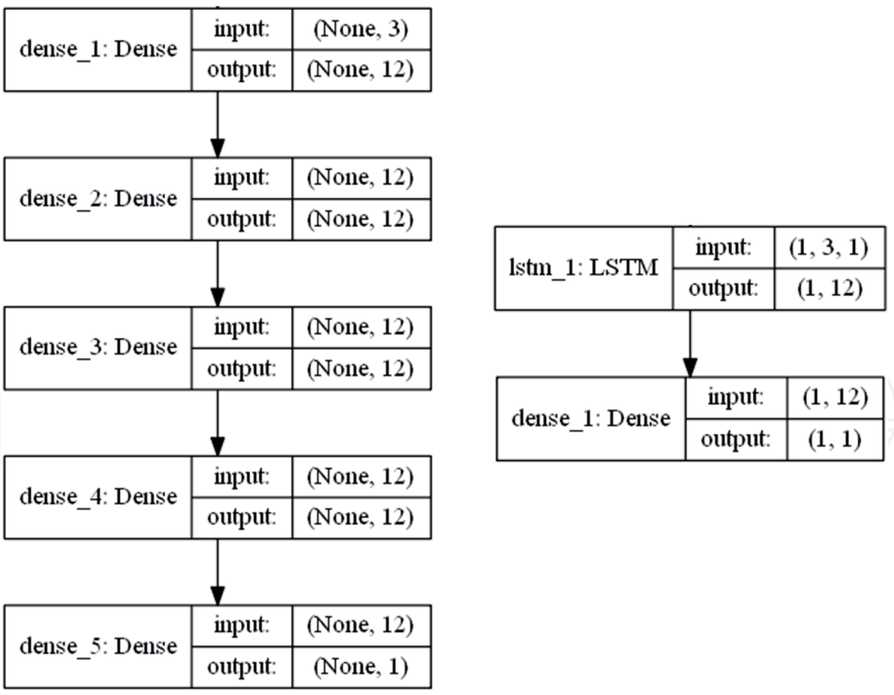

Figure 6.
Structure for the DL models, for both DFANN (on the left) and RNN-LSTM (on the right).

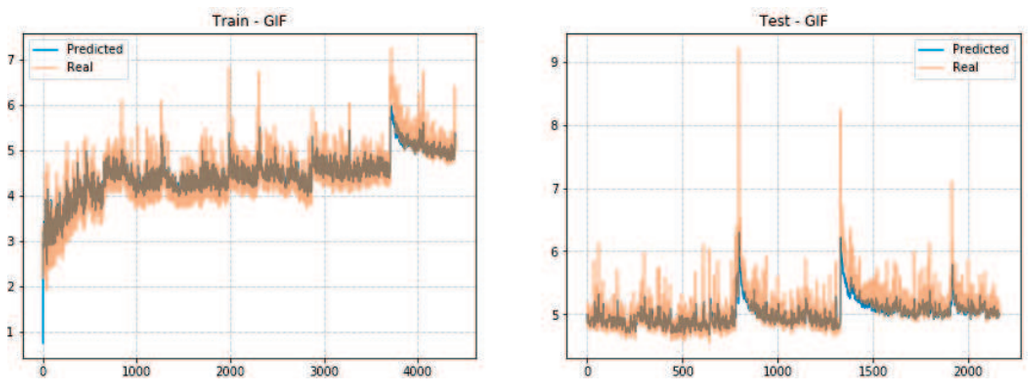

Figure 7.
Training and test groups for the best model (DFANN, Mag > 3).

RMSE training/test		
Mag	**DFANN**	**RNN-LSTM**
>3	**0.3478/0.2603**	0.5651/0.5167
>4	0.4624/0.3440	0.6698/0.4732
>5	0.5894/0.4457	0.7572/0.4449
>6	0.4226/0.4654	0.7941/0.4741

In bold the best model.

Table 1.
Root mean square error (RMSE) of the training and test groups for each DFANN and RNN-LSTM deep learning models.

5. Discussion

This work introduces a novel approach to predict the temporal ETAS-GIF alternative to the statistical approach proposed by [14]. The deep learning method has recently been used for predicting locations of aftershock events [31] especially based on ground motion data. The first use of a feedforward neural network for the prediction of seismic hazard was introduced by [32] in the spatial domain.

Possible extensions of the deep learning approach could be to include the ground motion together to other variables [30, 31] as inputs of the model and to incorporate the spatial dimension for a spatiotemporal prediction [33–35]. Some statistical techniques could be used for identifying possible patterns and inputs [36–37].

Also, since seismic events could be characterized by different features depending of the different locations of the principal events, we think that DL neural network models could be used for characterizing earthquakes in some specific seismic areas such as the local ETAS models [7, 11].

Different neural networks models could be used for comparing earthquake predictions [38]. For example, Bayesian DL neural networks could be used for a new prediction scenario considering the uncertainty of major earthquake occurrences and the probability of recurrence in a similar way to the Bayesian approach proposed by [32]. Additionally, other DL and machine learning approaches as convolutional neural networks (CNN), generative networks (GN), and random forest regression (RFR) could be implemented by incorporating the spatial component and allowing to "generate" new prediction seismic risk maps.

However, the main limitation of neural networks is that they are considered "black boxes" since it is difficult to quantify the correlation between the involved variables and their uncertainty.

6. Conclusion

This chapter deals with the estimation of seismic risk given by the temporal ETAS conditional intensity function. To achieve this goal, two deep learning models were implemented: a deep feedforward artificial neural network and a recurrent long short-term memory network. The results show a good estimation, in particular with the DFANN model. However, it should be pointed out that both implemented models could be improved by adding more hidden layers or stacking more LSTM layers in the DFANN and RNN-LSTM models, respectively. Also, exogenous variables (such as ground motion among others) could be considered for improving the predictions. Since the proposed model only considers a temporal model, extensions to the prediction of earthquake locations will be considered in future works. We think that deep learning algorithms could be useful tools for many earthquake prediction approaches.

Acknowledgements

The authors thank the National Research Center for Integrated National Disaster Management (CIGIDEN), CONICYT/FONDAP/15110017 (Chile) and CONICYT PFCHA/DOCTORADO BECAS CHILE/2018 – 21182037 for financing this work.

Author details

Francisco Plaza[1,2], Rodrigo Salas[1,3] and Orietta Nicolis[4,5]*

1 University of Valparaíso, Valparaíso, Chile

2 Institute for Fisheries Development (IFOP), Valparaíso, Chile

3 Center for Research and Development for Health Engineering, CINGS-UV, Chile

4 University of Andres Bello, Viña del Mar, Chile

5 National Research Center for Integrated National Disaster Management (CIGIDEN), Chile

*Address all correspondence to: orietta.nicolis@unab.cl

References

[1] Lomnitz C. Major earthquakes of Chile: A historical survey, 1535-1960. Seismological Research Letters. 2004;**75**(3):368. Available from: http://dx.doi.org/10.1785/gssrl.75.3.368

[2] Norio O, Ye T, Kajitani Y, Shi P, Tatano H. The 2011 eastern Japan great earthquake disaster: Overview and comments. International Journal of Disaster Risk Science. 2011;**2**(1):34-42

[3] Sobolev GA. Methodology, results, and problems of forecasting earthquakes. Herald of the Russian Academy of Sciences. 2015;**85**(2):107-111

[4] Cimellaro GP, Marasco S. Earthquake prediction. In: Introduction to Dynamics of Structures and Earthquake Engineering. Switzerland: Springer International Publishing AG a part of Springer Nature; 2018. pp. 263-280

[5] Ogata Y. Statistical models for earthquake occurrences and residual analysis for point processes. Journal of the American Statistical Association. 1988;**83**(401):9-27

[6] Lombardi AM, Cocco M, Marzocchi W. On the increase of background seismicity rate during the 1997-1998 Umbria-Marche, Central Italy, sequence: Apparent variation or fluid-driven triggering? Bulletin of the Seismological Society of America. 2010;**100**(3):1138-1152

[7] Ogata Y. Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. Earth, Planets and Space. 2011;**63**:217-229

[8] Bansal A, Ogata Y. A non-stationary epidemic type aftershock sequence model for seismicity prior to the December 26, 2004 m 9.1 Sumatra-Andaman islands mega-earthquake. Journal of Geophysical Research - Solid Earth. 2013;**118**(2013):616-629

[9] Kumazawa T, Ogata Y, et al. Nonstationary ETAS models for nonstandard earthquakes. Annals of Applied Statistics. 2014;**8**(3):1825-1852

[10] Guo Y, Zhuang J, Zhou S. An improved space-time ETAS model for inverting the rupture geometry from seismicity triggering. Journal of Geophysical Research - Solid Earth. 2015;**120**(5):3309-3323

[11] Nicolis O, Chiodi M, Adelfio G. Windowed ETAS models with application to the Chilean seismic catalogs. Spatial Statistics. 2015;**14**:151-165

[12] Utsu T. A statistical study on the occurrence of aftershocks. Geophysical Magazine. 1961;**30**:521-605

[13] Daley DJ, Vere-Jones D. An Introduction to the Theory of Point Processes. New York, USA: Springer; 2003. 469p

[14] Nicolis O, Chiodi M, Adelfio G. Space-time forecasting of seismic events in Chile. In: Earthquakes-Tectonics, Hazard and Risk Mitigation. Rijeka: InTech; 2017

[15] Joffe H, Rossetto T, Bradley C, O'Connor C. Stigma in science: The case of earthquake prediction. Disasters. 2018;**42**(1):81-100

[16] Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE. A survey of deep neural network architectures and their applications. Neurocomputing. 2017;**234**:11-26

[17] Rasmussen JG. Temporal point processes: The conditional intensity function. Lecture Notes; 2011

[18] Reinhart A et al. A review of self-exciting spatio-temporal point processes and their applications. Statistical Science. 2018;**33**(3):299-318

[19] Hawkes AG. Spectra of some self-exciting and mutually exciting point processes. Biometrika. 1971;**58**(1):83-90

[20] Ogata Y. Space-time point-process models for earthquake occurrences. Annals of the Institute of Statistical Mathematics. 1998;**50**(2):379-402

[21] Harte D et al. PtProcess: An R package for modelling marked point processes indexed by time. Journal of Statistical Software. 2010;**35**(8):1-32

[22] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks. 1989;**2**(5):359-366

[23] White H. Artificial Neural Networks: Approximation and Learning Theory. Cambridge, MA, USA: Blackwell Publishers, Inc; 1992

[24] Cybenko G. Approximation by superpositions of a sigmoidal function. Mathematics of Control, Signals, and Systems. 1989;**2**(4):303-314

[25] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep Learning. Vol. 1. Cambridge: MIT press; 2016

[26] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;**323**(6088):533-536

[27] Cady F. The Data Science Handbook. Hoboken, USA: John Wiley & Sons, Inc; 2017

[28] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation. 1997;**9**(8):1735-1780

[29] Gers FA, Schmidhuber J, Cummins F. Learning to Forget: Continual Prediction with LSTM. 9th International Conference on Artificial Neural Networks: ICANN '99, Edinburgh, UK; 1999. p. 850-855

[30] Gers FA, Schraudolph NN, Schmidhuber J. Learning precise timing with LSTM recurrent networks. Journal of Machine Learning Research. 2002;**3**(Aug):115-143

[31] Olah C. Understanding LSTM Networks [Internet]. 2015 [cited 2018 Nov 30]. Available from: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[32] Nomura S, Ogata Y, Komaki F, Toda S. Bayesian forecasting of recurrent earthquakes and predictive performance for a small sample size. Journal of Geophysical Research - Solid Earth. 2011;**116**(B4):1-18

[33] Harichandran RS, Vanmarcke EH. Stochastic variation of earthquake ground motion in space and time. Journal of Engineering Mechanics. 1986;**112**(2):154-174

[34] Atkinson GM, Boore DM. Earthquake ground-motion prediction equations for eastern North America. Bulletin of the Seismological Society of America. 2006;**96**(6):2181-2205

[35] Rezaeian S, Der Kiureghian A. Simulation of orthogonal horizontal ground motion components for specified earthquake and site characteristics. Earthquake Engineering & Structural Dynamics. 2012;**41**(2):335-353

[36] Plaza F, Salas R, Yáñez E. Identifying ecosystem patterns from time series of anchovy (Engraulis ringens) and sardine (Sardinops sagax) landings in northern Chile. Journal of Statistical Computation and Simulation. 2018;**88**(10):1863-1881

[37] Shekhar S, Evans MR, Kang JM, Mohan P. Identifying patterns in spatial

information: A survey of methods. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2011;**1**(3):193-214

[38] Ogata Y. A prospect of earthquake prediction research. Statistical Science. 2013;521-541