Seismotectonics Considered Artificial Neural Network Earthquake Prediction in Northeast Seismic Region of China

Jian Sheng¹, Dongmei Mu²*, Hongyan Zhang¹ and Han Lv¹

¹Earthquake Administration of Jilin Province, Changchun, Jilin, 130117, P.R. China
²School of Public Health, Jilin University, Changchun, Jilin, 130021, P.R. China

Abstract: It is well known that earthquakes are a regional event, strongly controlled by local geological structures and circumstances. Reducing the research area can reduce the influence of other irrelevant seismotectonics. A new sub-region-dividing scheme, considering the seismotectonics influence, was applied for the artificial neural network (ANN) earthquake prediction model in the northeast seismic region of China (NSRC). The improved set of input parameters and prediction time duration are also discussed in this work. The new dividing scheme improved the prediction accuracy for different prediction time frames. Three different research regions were analyzed as an earthquake data source for the ANN model under different prediction time duration frames. The results show: (1) dividing the research region into smaller sub-regions can improve the prediction accuracies in NSRC, (2) larger research regions need shorter prediction durations to obtain better performance, (3) different areas have different sets of input parameters in NSRC, and (4) the dividing scheme, considering the seismotectonics frame of the region, yields better results.

Keywords: Earthquake prediction, neural network, seismotectonics unit, northeast seismic region.

1. INTRODUCTION

Earthquakes are among the most terrifying natural disasters and have seized hundreds of thousands of lives as well as destroyed millions of buildings. Statistically, 15 earthquake events have respectively caused more than 20,000,000 people's deaths throughout history. Although other natural disasters can be well predicted such as hurricanes [1], typhoons [2], landslides [3], or volcanic eruptions [4], earthquake prediction is difficult, but some scientists are trying to find methods to do it. There have been numerous efforts using different applications during the past few decades [5]. The earliest earthquake prediction parameters date back to more than 70 years ago [6]. Unfortunately, the progress in this field is very slow and until recently there were still no successful methods to satisfy the requirements described by Allen (1982) [7], which included predictions of when, where, how big and how probable an impending earthquake would be.

The reason for this failure is that the occurrence of earthquakes is a very complex process influenced by a large number of factors for which the effects are still not exactly understood. Traditional statistical and mathematical methods are not easily able to analyze such complex processes and overcome the distorted control force due to time-delay [8], the artificial intelligence technique seems a feasible method to solve this issue, due to its powerful ability to process complex variable data and non-linear variable data. The artificial intelligence earthquake prediction model was developed and continues to arouse attention. Nevertheless, the ANN stands out from other artificial intelligent techniques due to its strong non-linear fitting capabilities that can be mapped to any complex non-linear relationship [9]. ANN also has simple usage instructions and easy access to computer implementation. The resilience, large memory capacity, strong non-linear mapping ability and self-learning capability enable the neural network to process a large amount of complex non-linear data.

The first ANN model for earthquake prediction was modified by a financial market forecasting model [10]. This model includes three inputs: time, intensity and location. These three inputs are known as the three basic elements of an earthquake. This model correctly predicted two earthquakes using major groups 1° in longitude in the Azores with a range of ±5-6 months [10]. Three different ANN models used a novel set of seismicity indicators which were applied to forecast the earthquake magnitude for southern California and San Francisco bay. These ANN models achieved acceptable results for the earthquake of a magnitude between 6.0 and 7.5 [11]. What is remarkable about their work is that they compared the prediction accuracies with a recurrent neural network, a radial basis function neural network and a Levenberg-Marquardt back-propagation neural network. The recurrent neural network yielded the best prediction result so they tried to predict the time and location of an earthquake in southern California by recurrent neural network [12]. Probabilistic neural network, also a type of ANN, was used to forecast the earthquake in southern California and presented good prediction accuracies for earthquakes of magnitudes between 4.5 and 6.0 [12]. The radial basis function neural network yielded more accurate and effective prediction re-
The accuracies of the earthquake predictions are improved by dividing the research region into rectangular sub-regions and tectonic units. The earthquake prediction accuracies caused by the two different fragment schemes are discussed. The result shows: (1) dividing the research region into sub-regions can improve the prediction accuracies in NSRC, (2) bigger research regions need shorter prediction duration to obtain better performance, (3) different areas have different improved sets of input parameters in NSRC, (4) the dividing scheme considering the seismotectonics frame of the region yields better results than others. This result indicates that the occurrences of earthquakes rely on the geological structure but the regional seismotectonics frame must also be considered during the procedure of the earthquake data processing.

2. BACKGROUND INFORMATION

This section serves to introduce some background information about this work, such as the research region (NSRC), the smaller research region divided by regular coordinates and by seismotectonics frame, and the data source of this research.

2.1. NSRC

The northeast seismic region of China, which consists of Heilongjiang Province, Jilin Province and some parts of Inner Mongolia Autonomous Region and the Liaoning Province. Considering tectonic unit boundary, earthquake positioning error and earthquake catalog integrity the NSRC also includes a little part of Russia, Korea and Mongolia, shown in Fig. (1). The history records no $M_\geq 7.0$ earthquake events occurred in this region and the seismic activity in this region following the method from Martinez-Álvarez et al. (2013) [16].
was relatively weak not only in intensity but also in quantity in the past. A few studies have been made regarding its activity and potential seismic risks in the future [22]. However, since 2013, 7 events of magnitude from M 5.0 to M 5.9 have occurred in this region. This indicates that the NSRC is entering its earthquake-active period that was influenced by the western Pacific plate subduction exacerbation, which was manifested by the Tohoku M 9.0 event, 2011 and Okhotsk Sea M8.3 event, 2013. The seismic time sequence diagram (Fig. (2)) and seismic energy release curve (Fig. (3)) of NSRC also show a significant enhancement in seismic activities in 2013. Statistics show the majority of earthquakes were M≥5.0 in NSRC with a magnitude of M5.0- M5.9, and only 4 events were more than M6.0, the biggest was about M6.5/4. Although earthquakes with magnitude of M5.0- M5.9 may not directly cause the deaths of civilians, it can cause suffering to people as well as huge financial loss. An example of financial loss is the Songyuan M 5.8 earthquake Sequence, 2013 affected 60,000 residents, seriously damaged 16,000 houses, 40,000 buildings somewhat damaged, 310 houses collapsed and caused the economic loss of 2 billion RMB (about 300 million US dollars). Furthermore, these earthquakes will generate social problems such as rumor or panic. An earthquake of this magnitude has a higher probability of occurrence than a M≥6.0 one. So we will focus on predicting the earthquakes with a magnitude of M 5.0- M 5.9 in the NSRC.

2.2. Research Region Dividing

Dividing the seismic region into smaller areas can promote the prediction accuracies of ANN’s earthquake prediction model [12]. In this work, the NSRC was also divided into rectangular research regions and tectonic units. The NSRC has relatively low seismicity intensity and frequency. In order to ensure sufficient earthquake data, regular rectangular research regions were divided into 2°×2° and four regions were randomly selected as research areas. These are named REC-1 (between 121°E and 123°E longitude and 47°N and 49°N latitude), REC-2 (between 129°E and 131°E longitude and 47°N and 49°N latitude), REC-3 (between 118°E and 120°E longitude and 44°N and 46°N latitude) and REC-4 (between 123°E and 125°E longitude and 44°N and 46°N latitude). Four tectonic units were also randomly selected as research areas. These are named as TEC-1 (Daxinganling uplift block), TEC-2 (central depression of Songliao basin), TEC-3 (Jiamusi uplift block) and TEC-4 (Southern uplift of Songliao basin) (Fig. (1)). The four rectangular regions and four tectonic units were studied and predicted as research areas and the prediction accuracies are discussed in Section 5.

2.3. Data Source

The earthquakes in NSRC can be divided into two categories based on focal depth: the first is the shallow earth-
Seismotectonics Considered Artificial Neural Network Earthquake Prediction

The Open Civil Engineering Journal, 2015, Volume 9

Table 1. The best set of input indicators of the research regions.

<table>
<thead>
<tr>
<th>Research Region</th>
<th>Input Parameters Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSRC</td>
<td>$x_1$, $x_2$, $x_3$, $T$, $M_{\text{mean}}$, $dE_{1/2}$, $c$, $\eta$</td>
</tr>
<tr>
<td>REC-1</td>
<td>$x_4$, $x_5$, $x_6$, $x_7$, $M_{\text{max}}$, $dE_{1/2}$, $c$</td>
</tr>
<tr>
<td>REC-2</td>
<td>$x_1$, $x_2$, $x_3$, $x_4$, $x_5$, $T$, $M_{\text{max}}$, $\mu$, $c$</td>
</tr>
<tr>
<td>REC-3</td>
<td>$x_6$, $x_7$, $T$, $M_{\text{min}}$, $\mu$, $c$, $\eta$</td>
</tr>
<tr>
<td>REC-4</td>
<td>$x_4$, $x_5$, $x_6$, $x_7$, $M_{\text{mean}}$, $dE_{1/2}$, $c$</td>
</tr>
<tr>
<td>TEC-1</td>
<td>$x_1$, $x_2$, $x_3$, $x_4$, $x_5$, $T$, $c$</td>
</tr>
<tr>
<td>TEC-2</td>
<td>$x_6$, $x_7$, $T$, $M_{\text{max}}$, $dE_{1/2}$, $c$, $\eta$</td>
</tr>
<tr>
<td>TEC-3</td>
<td>$x_6$, $x_7$, $T$, $M_{\text{min}}$, $dE_{1/2}$, $c$, $\eta$</td>
</tr>
<tr>
<td>TEC-4</td>
<td>$x_6$, $x_7$, $T$, $M_{\text{mean}}$, $\mu$, $c$, $\eta$</td>
</tr>
</tbody>
</table>

Earthquake (the focal depth $< 60\text{km}$), and the second is the deep earthquake (the focal depth $> 300\text{km}$). In this work, when it comes to earthquakes we just consider the shallow earthquakes. The main seismicity catalog used in this work was derived from 2 resources. The data of moderate to strong earthquakes (the magnitude is equal to or larger than M 4.7) was derived from the following: China Earthquake Catalogue (Monitoring and Forecasting Division of Earthquake Administration of China, 2010), Modern China Earthquake Catalogue (Earthquake Disaster Prevention Division of Earthquake Administration of China, 1999), Northeast Historical Earthquake Series (Seismological Press, 1992) and China Earthquake Details Catalogue (Earthquake Analysis and Prediction center of Earthquake Administration of China, 1970-2013). Then, the data of small earthquakes (the magnitude is smaller than M 4.7) was derived from instrumental records which are recorded in China Earthquake Details Catalogue (Earthquake Analysis and Prediction Center of Earthquake Administration of China, 1970-2013) and referred to the earthquake catalog of the Earthquake Administration of Heilongjiang Province, Jilin Province, Liaoning Province and Inner Mongolia Autonomous Region.

3. METHODOLOGY

In this section, the earthquake prediction model, the methods of determining the best set of input parameters and the calculation results assessment are briefly introduced.

3.1. Earthquake Prediction Model

Martínez-Álvarez et al. (2013) [16] had compared the prediction accuracy of the artificial neural network with Naive Bayes (NB), K-nearest neighbors (KNN) and support-vector machines (SVM). The comparison result showed ANN had better performance than any other classifier. This work builds the earthquake prediction model by ANN. Many researchers have discussed and described the concept of ANN in details [8, 23, 24]. We will not introduce the concept of BPN here, but focus on presenting how to build the ANN earthquake prediction model. The model is composed of three layers: the input layer, hidden layer and output layer. The input layer has 7 input indicators, the hidden layer has 15 neurons and the output layer has 2 output parameters. Firstly, there are 16 alternative parameters for input set, derived from two previous research papers [12,17]. Input=$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, T, M_{\text{max}}, dE_{1/2}, \beta, a, AM, \mu, c, \eta\}$. The first 7 parameters came from Panakkat&Adeli (2009) [12], and the next 9 parameters were obtained from Reyes et al. (2013) [17]. For detailed information about these parameters please refer to these research papers. The best input set of every research region, which includes 7 parameters was obtained by Weka software (see section 3.2). The output includes 2 parameters: one is the maximum magnitude $M_{\text{max}}$ observed in the next 30 days and the other is the time t (measured by days) recorded the temporal duration when the maximum earthquake occurred. The Weka software was used to build and calculate the previously mentioned model.

3.2. Better Set of Features

Martínez-Álvarez et al. (2013) [16] applied the Weka software to measure the information gain associated with each feature with respect to the class, and discussed the significance of choosing the rational input parameters. It is the first time the Weka software has been applied for earthquake prediction [16]. We adopted this method in our work to choose the best input parameters of every research region (Table 1), the details of this method are proposed in [16].

3.3. Evaluating Method

The assessment methods are modified from past research [11, 16, 17, 19]. First of all, several important parameters have to be introduced:

1. Double true (DT). The number of times that an impending earthquake was properly predicted in magnitude and time.
2. Magnitude true (MT). The number of times that the ANN model was properly predicted an earthquake only in magnitude.
3. Time true (TT). The number of times that the ANN model was properly predicted an earthquake only in time.
4. Double false (DF). The number of times that the prediction result was wrong in magnitude and time.
The rate of perfect prediction result (denoted by $P_{DT}$), the rate of proper magnitude prediction (denoted by $P_M$), the rate of proper time prediction (denoted by $P_T$), and the rate of entirely false prediction result (denoted by $P_{DF}$) are calculated by the equation below:

\[ P_{DT} = \frac{DT}{DT + MT + TT + DF} \]  
\[ P_M = \frac{DT + MT}{DT + MT + TT + DF} \]  
\[ P_T = \frac{DT + TT}{DT + MT + TT + DF} \]  
\[ P_{DF} = \frac{DF}{DT + MT + TT + DF} \]

Additionally, the chi-square statistic test was applied to check the prediction accuracy difference (denoted by $P_C$) between the different research region definition criterions. For the detailed concept of the chi-square statistic test can refer to Moor (1976) [25].

Finally, these parameters are comprehensively compared to find out which research region definition criterions can yield better prediction accuracy.

### 4. CALCULATION RESULTS

This section serves to present the modeling result of the ANN earthquake prediction model by different research region definition criterions: the entire region (presented by NSRC), the smaller rectangular regions (presented by REC-1 to REC-4), and the smaller seismotectonics unit regions (presented by TEC-1 to TEC-4). The prediction time frame was also considered in this work, it changed from one month (Table 2) to 15 days (Table 3) and 7 days (Table 4).

### 5. DISCUSSION

This section is to summarize and compare the data and tables from previous sections. It also serves to show the prediction performance of different research region definition criterions and prediction time frames.

Table 1 lists the best set of input indicators of the research regions analyzed. It shows the research regions have different best sets of input features. It indicates earthquake occurrence has regional characteristics, because it depends on local special and complex circumstance.

Tables 2-4 show the prediction performance for research regions with the prediction time frame from one month (Table 2) to 15 days (Table 3) and 7 days (Table 4). For NSRC, the 7 days prediction time frame yields better prediction accuracy than either 30 days or 15 days. The result is in line with the previous research [12] about how reducing the prediction time frame can improve the prediction accuracy to some extent. However, the smaller research regions are not able to meet this conclusion. The rectangular sub-regions produced better results when the prediction duration is 30 days, while the seismotectonics unit smaller sub-regions need 15 days of prediction duration. The reason is that the mean time between typical events is relatively longer than the prediction duration due to the fact that smaller sub-regions have more mean time between typical events. In other words, the 15 days and 7 days prediction time frames are considerably shorter than the duration of a typical earthquake of the rectangular sub-regions and the seismotectonics unit sub-regions, respectively.

#### Table 2. The prediction performance for research regions of 30 days prediction time frame.

<table>
<thead>
<tr>
<th>Region</th>
<th>DT</th>
<th>MT</th>
<th>TT</th>
<th>DF</th>
<th>$P_{DT}$ (%)</th>
<th>$P_M$ (%)</th>
<th>$P_T$ (%)</th>
<th>$P_{DF}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSRC</td>
<td>44</td>
<td>140</td>
<td>33</td>
<td>125</td>
<td>12.87</td>
<td>53.80</td>
<td>22.51</td>
<td>36.55</td>
</tr>
<tr>
<td>REC-1 to REC4</td>
<td>55</td>
<td>27</td>
<td>16</td>
<td>9</td>
<td>51.40</td>
<td>76.64</td>
<td>66.36</td>
<td>8.41</td>
</tr>
<tr>
<td>TEC-1 to TEC4</td>
<td>105</td>
<td>56</td>
<td>27</td>
<td>9</td>
<td>53.30</td>
<td>81.73</td>
<td>67.01</td>
<td>4.57</td>
</tr>
</tbody>
</table>

#### Table 3. The prediction performance for research regions of 15 days prediction time frame.

<table>
<thead>
<tr>
<th>Region</th>
<th>DT</th>
<th>MT</th>
<th>TT</th>
<th>DF</th>
<th>$P_{DT}$ (%)</th>
<th>$P_M$ (%)</th>
<th>$P_T$ (%)</th>
<th>$P_{DF}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSRC</td>
<td>70</td>
<td>138</td>
<td>43</td>
<td>91</td>
<td>20.47</td>
<td>60.82</td>
<td>33.04</td>
<td>26.61</td>
</tr>
<tr>
<td>REC-1 to REC4</td>
<td>54</td>
<td>15</td>
<td>12</td>
<td>26</td>
<td>50.47</td>
<td>64.49</td>
<td>61.68</td>
<td>24.30</td>
</tr>
<tr>
<td>TEC-1 to TEC4</td>
<td>106</td>
<td>61</td>
<td>28</td>
<td>2</td>
<td>53.81</td>
<td>84.77</td>
<td>68.02</td>
<td>2.06</td>
</tr>
</tbody>
</table>

#### Table 4. The prediction performance for research regions of 7 days prediction time frame.

<table>
<thead>
<tr>
<th>Region</th>
<th>DT</th>
<th>MT</th>
<th>TT</th>
<th>DF</th>
<th>$P_{DT}$ (%)</th>
<th>$P_M$ (%)</th>
<th>$P_T$ (%)</th>
<th>$P_{DF}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSRC</td>
<td>90</td>
<td>122</td>
<td>51</td>
<td>80</td>
<td>26.23</td>
<td>80.61</td>
<td>53.61</td>
<td>23.32</td>
</tr>
<tr>
<td>REC-1 to REC4</td>
<td>54</td>
<td>13</td>
<td>11</td>
<td>29</td>
<td>50.47</td>
<td>62.62</td>
<td>60.75</td>
<td>27.10</td>
</tr>
<tr>
<td>TEC-1 to TEC4</td>
<td>106</td>
<td>27</td>
<td>26</td>
<td>38</td>
<td>53.81</td>
<td>67.51</td>
<td>67.01</td>
<td>19.29</td>
</tr>
</tbody>
</table>
Seismotectonics Considered Artificial Neural Network Earthquake Prediction

The result of this work is in accordance with the research of Panakkat & Adeli (2009) [12], which suggested that dividing the seismic region into smaller areas could yield better prediction accuracy. Both the regular smaller sub-regions and the seismotectonics unit smaller sub-regions yield obviously better results than the large whole region, according to $P_{\chi^2} < 0.05$ by chi-square statistic test. The Table 2 to Table 4 also show that under different prediction time frames the smaller research regions produced better results than the NSRC and the seismotectonics unit smaller sub-regions have higher accuracy than the regular smaller sub-regions. Although, the prediction accuracy differences between the rectangular sub-regions and the seismotectonics unit sub-regions are not significant, a small increase can still be observed. The seismotectonics unit sub-regions have higher accuracy than the regular smaller sub-regions, in spite of the seismotectonics unit sub-regions have larger area. This effect can be explained by the fact that the earthquake is a geological event controlled by local geo-structure and has an inner-relationship with the tectonic unit [21]. It is well known that the distribution of earthquakes is uneven and controlled by local geo-structure so the regional seismotectonics frame must be considered during the procedure of the earthquake data processing.

The geological structures have analogous characteristics and an activity rhythm in the same seismotectonics unit. The different seismotectonics units have different seismicity rhythms controlled by their respective geological structures characteristics and activity rhythm. The large research region is composed of several seismotectonics units; the different seismicity rhythms of these seismotectonics units interfere and influence each other, which lead to a relatively inaccurate prediction result. Dividing the research region into smaller areas can reduce above mentioned interference and influence by reducing the involved number of seismotectonics. However, dividing the research regions into rectangular smaller areas cannot reduce this interference and influence due to seismotectonics units not being rectangular in shape. The rectangular smaller sub-regions consist of less seismotectonics units than the large region but they normally involve two or more seismotectonics units. Dividing the large region into seismotectonics units can reduce these impacts to a minimum by ensuring a uniform seismicity rhythm in the smaller areas and generates higher accuracy.

CONCLUSION

Earthquake prediction, a well-accepted and difficult issue, is deserving of our efforts to explore due to the fact there is still a lot of space for improvement. The authors follow the previous researches [11, 12, 16, 17, 19, 26] to promote the earthquake prediction accuracy of the ANN model, and to find out the seismotectonics model considered to generate better results because earthquakes are a regional geological event and strongly depend on local geological circumstance. In general, the following conclusions can be drawn from this work:

1. Under the premise of adequate seismicity data, dividing the research region into smaller sub-regions can improve the prediction accuracies in NSRC.

2. Shortening the prediction duration can improve the prediction results but the prediction duration must be longer than the mean time between typical events. That is to say larger research regions need shorter prediction duration to obtain better performance, because larger regions have more earthquake data and relatively shorter mean time between typical events.

3. Earthquakes are regional geological events and have the regional characteristics of the region. This leads to the regional seismicity characteristics. Different areas have different improved sets of input parameters due to regional seismicity characteristics.

4. Dividing the large region into seismotectonics units can reduce the interference and influence from other seismotectonics units and promote better prediction accuracy. So the dividing scheme considering the seismotectonics frame of the region yields better results than the others of options.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work is supported by the Internal Contract Science Research Project of Earthquake Administration of Jilin Province (201512) (201507).

REFERENCES


