Extended Abstract

Forecasting Tsunami Hazard Using Ocean Bottom Sensor Data and Classification Predictive Modeling †

Yao Li *

Department of Statistical and Actuarial Sciences, Western University, London, ON N6A 3K7, Canada; yli3285@uwo.ca

* Correspondence: yli3285@uwo.ca

† The 4th International Electronic Conference on Geosciences, online, 01-15 December 2022.

Abstract: This study compares the results of analyzing tsunami simulations that are based on two approaches of characterizing earthquake slips, i.e., uniform (simplistic) and heterogeneous (complex) distributions. The aim of this study is to compare how heterogeneous and uniform distributed data affect the classification of tsunami maximum near-shore tsunami amplitudes. Due to the lack of historical earthquake and tsunami data to train the forecasting model, 4000 stochastic tsunami simulations are employed. The focused location is Iwanuma, Japan, where ocean bottom sensors (OBS) S-net network has been deployed. Multiple linear regression combined with Akaike Information Criterion (AIC) is applied to the simulated off-shore wave amplitude data to fit the model. The estimated tsunami amplitude is classified into four levels of warning classes. The performance of the models is quantified by the accuracy of the confusion matrices and is compared with the base model that only uses earthquake information. The forecasting accuracy can be improved by 30% when the wave amplitude data are used as additional information. The heterogeneous slip-based model reaches a higher accuracy than the uniform-slip based model. The result of this study is particularly valuable for setting up an OBS-based system for monitoring the physical phenomena of tsunamis and choosing heterogeneous as a preferable slip distribution when tsunami events are simulated.

Keywords: Tsunami forecast classification; ocean bottom sensor; stochastic tsunami simulation

1. Introduction

Tsunamis are rare events, but they can be destructive. In the last two decades, there are two most notable tsunami events (2004 Indian Ocean tsunami and 2011 Tohoku tsunami) that were triggered by mega earthquakes whose magnitude (M) were 9.0 or greater. Because of the increasing vulnerability of coastal regions on account of growing population and building asset development [1], research on tsunami early warning has been conducted by utilizing various techniques, such as using seismic source (earthquake moment) data [2] and using real-time wave data [3,4].

A tsunami early warning system can foretell the information about the amplitude of the tsunami wave and the tsunami arrival time. This study focuses on generating early warning at Iwanuma in Japan based on the maximum near-shore tsunami amplitude using a classification model by multiple linear regression (MLR). To fine-tune the model so that the model can capture the uncertainty from a variety of earthquakes that may happened in the future, synthetic earthquake source data and tsunami wave data are used because of lack of historical data. The synthetic slip distribution generated by a stochastic earthquake source model has an important effect on tsunami wave amplitude [5]. Thus, this study considered both uniform and heterogeneous slip distribution for data simulation and model comparison.

This study builds on the result by Li and Goda [6] and develops a tsunami classification model utilizing both uniform and heterogeneous event data. Two types of classification models, i.e. one model trained with earthquake magnitude and location only, denoting base model, and the other model with additional parameters from OBS, denoting sensor model, are developed under each distribution of data. The model performance of the sensor models is compared with the base model, and also compared against models trained with data with different distributions. The aim of this study is to use statistical regression-based methods to develop a tsunami amplitude classification model for early warning announcement and investigate how uniformly and heterogeneously distributed slip distributions affect the prediction of tsunami.

2. Tsunami data and classification model

2.1. Study area and S-net sensors

Iwanuma is in the Tohoku region of Japan, bordered by the Pacific Ocean to the east (Figure 1). The epicenter of the 2011 Tohoku earthquake was about 80 km east of Iwanuma and Iwanuma faced the triggered tsunami directly. There are a total of 99 off-shore sensors that are employed in this study (Figure 1a) and all of them are used to fit the MLR model firstly. Then, a variable selection technique using AIC is applied to select six most informative sensors (Figure 1b), and an efficient simpler model is built [6]. The base model using earthquake magnitude (x_{mag}), earthquake epicenter latitude (x_{lat}), and longitude (x_{lon}) and the sensor model with additional wave amplitudes from six sensors ($x_i, i = 1, 2, ..., 6$), and they can be presented as follows:

$$Base\ model = \beta_0 + x_{mag}\beta_{mag} + x_{lat}\beta_{lat} + x_{lon}\beta_{lon} \tag{1}$$

$$Sensor\ model = \beta_0 + x_{mag}\beta_{mag} + x_{lat}\beta_{lat} + x_{lon}\beta_{lon} + \sum_{i=1}^{6} x_i\beta_i \tag{2}$$

There are two sensor models that fitted by the different data (i.e. uniform versus heterogeneous slip distributions). The number of explanatory variables (3 parameters from earthquake information plus 6 parameters from OBS) is the same for the two sensor models but the coefficients of parameters (β) are different since one is fitted by heterogeneous slip data and the other is fitted by the uniform slip data.
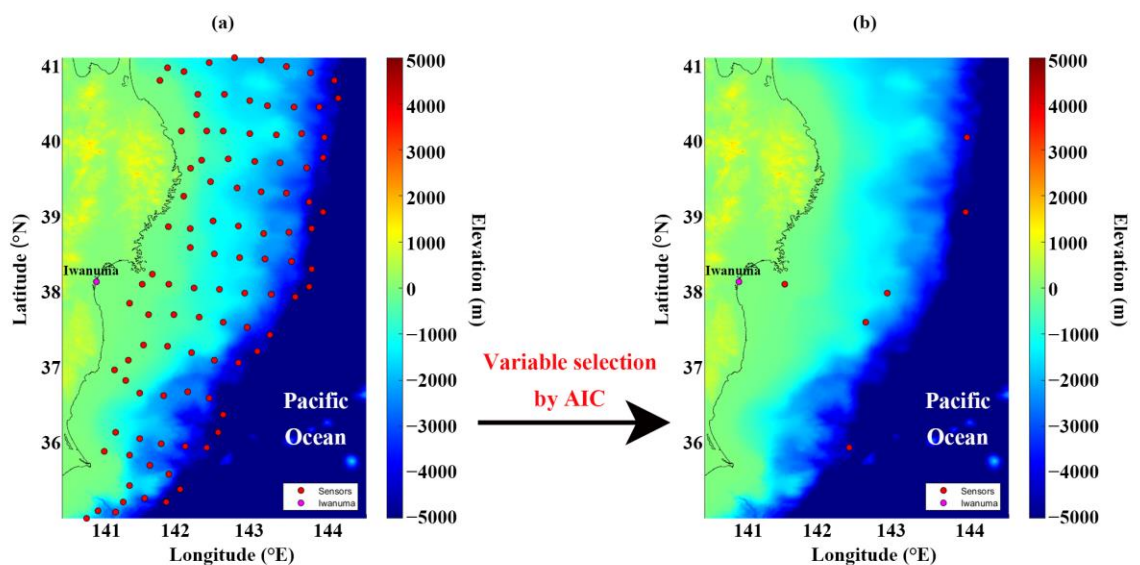

Figure 1. Map of Iwanuma and off-shore sensors: (**a**) 99 sensors before the AIC variable selection, (**b**) 6 sensors after the AIC variable selection.

2.2. Data and statistical methods

This study generates 4000 stochastic earthquake and tsunami events for both uniform and heterogeneous distributions [7]. This study considers the range of earthquake magnitude M7.5-9.1 with a magnitude interval of 0.2. The wave amplitude is calculated based on physical governing equations [8] and the wave amplitude simulation starts when an earthquake occurs and ends after 120 minutes. The maximum wave amplitude during the first 3 minutes that collected by the off-shore sensors is used as explanatory variable to fit the model since enough information is contained in the first 3 minutes data collected by the off-shore sensors after an earthquake [6].

Figure 2 shows the plot of the maximum near-shore tsunami amplitude (response variable) of the heterogeneous slip distribution against the uniform slip distribution. At the same earthquake event, the heterogeneous distributed maximum near-shore tsunami amplitude tends to be greater than the uniform one (see Figure 2). The maximum near-shore tsunami amplitude becomes greater as the earthquake magnitude increases, with average amplitude for the heterogeneous slip distribution increases from 0.58 m to 8.10 m and the average amplitude for the uniform slip distribution increases from 0.35 m to 3.6 m.
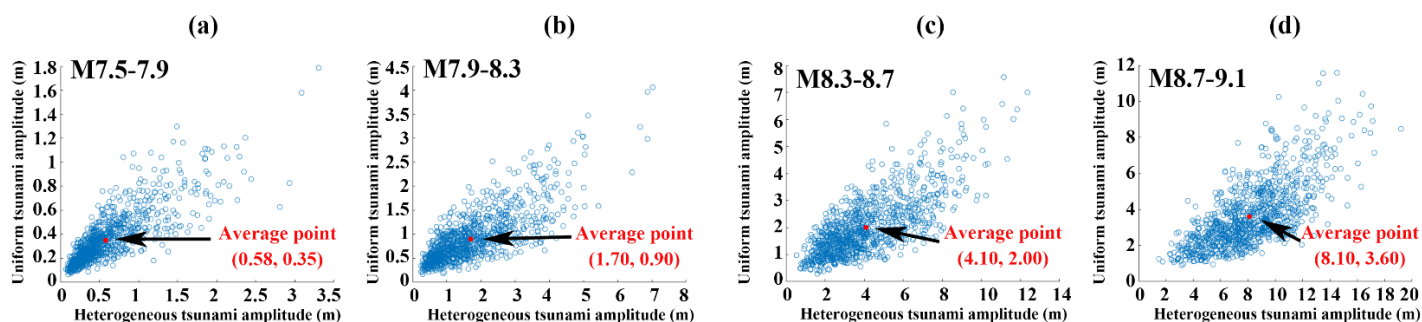

Figure 2. Comparison plots of simulated maximum near-shore tsunami amplitude: (**a**) the amplitude plot under earthquake magnitude 7.5-7.9, (**b**) the amplitude plot under earthquake magnitude 7.9-8.3, (**c**) the amplitude plot under earthquake magnitude 8.3-8.7, (**d**) the amplitude plot under earthquake magnitude 8.7-9.1.

The response variable is the maximum near-shore tsunami amplitude, which is a continuous variable originally. Because the assumption of data distribution is normal in MLR, logarithmic transformation (based on Box-Cox transformation) is applied to the response variable to hold the normality. After the model predicts the value, the value will transform back to original unit. Then, it is transformed to a category variable (four classes: 1st class is 0 – 2 m, denoting 'low'; 2nd class is 2 – 5 m, denoting 'moderate'; 3rd class is 5 – 10 m, denoting 'high' and 4th class is over 10 m, denoting 'critical') under a criterion for tsunami hazard assessment at water level height scale [9]. The classification tsunami forecasting is more commonly for early waring announcement purpose, and it can be more easily accepted by the public. The accuracy of a confusion matrix (CM) is used to describe the model performance for classification problems.

3. Results

The tsunami classification model result is presented in Figure 3. Figure 3a shows the accuracy of the base model (39%) has a much lower accuracy than the sensor models. It is insufficient that only earthquake magnitude and location data are used for regression to forecast a tsunami amplitude. When the wave amplitude data collected by OBS are added, the accuracy of the heterogeneous slip-based model (78%) is doubled (Figure 3a,b), and the accuracy of the uniform slip-based model increases by 30% (Figure 3a,c). The heterogeneous slip-based model has approximately 10% higher accuracy than the uniform

slip-based model (Figure 3b,c). The latter is because the coefficients of the uniform slip-based model are calibrated with simulated tsunami data with unrealistic earthquake slip distributions.
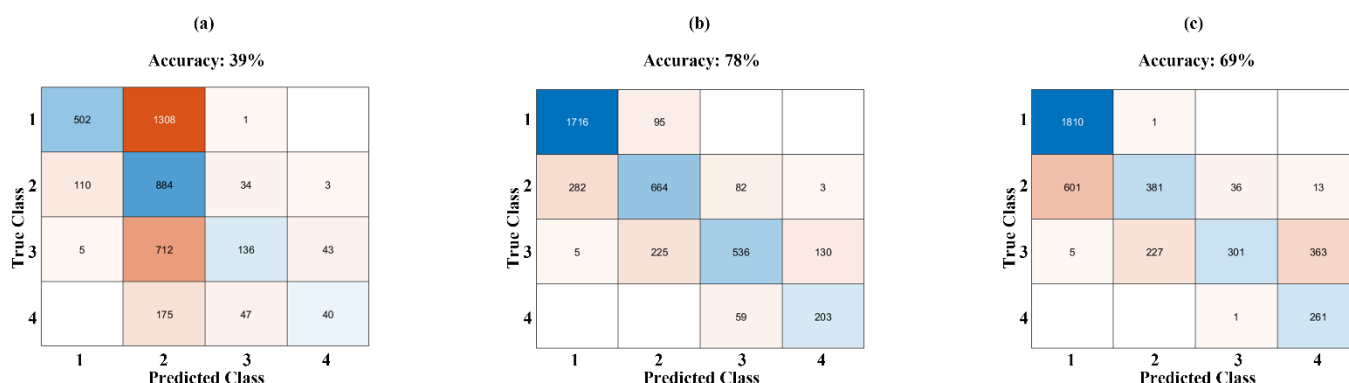
(a)
Accuracy: 39%

(b)
Accuracy: 78%

(c)
Accuracy: 69%

Figure 3. Confusion matrixes with accuracy of three models: (**a**) CM of base model, (**b**) CM of heterogeneous slip-based sensor model, (**c**) CM of uniform slip-based sensor model.

4. Discussion

This study investigated that using wave amplitude data collected by OBS improves the tsunami classification by its wave amplitude. It is difficult to classify a tsunami amplitude using the model that is only trained with earthquake information in this study. The heterogeneous slip-based model achieves a higher accuracy than the uniform slip-based model. For hypothetical tsunami simulations, a heterogeneous earthquake slip distribution is preferable, even though the process of generating heterogeneous distributed data is more complex than generating uniform distributed earthquake slip data.

For studies in the future, several fields can be improved. Firstly, applying the tsunami early warning algorithm (model) to different regions with similar geophysical characteristics as Iwanuma. Besides, a different approach to reduce the number of explanatory variables by applying Lasso and Ridge regression.

References

1. Imamura, F.; Boret, S. P.; Suppasri, A.; Muhari, A. Recent Occurrences of Serious Tsunami Damage and the Future Challenges of Tsunami Disaster Risk Reduction. *Progress in Disaster Science* **2019**, *1*, 100009.
2. An, C; Liu, H.; Ren, Z.; Yuan, Y. Prediction of Tsunami Waves by Uniform Slip Models. *Journal of Geophysical Research: Oceans* **2018**, *123(11)*, 8366-8382.
3. Maeda, T.; Obara, K.; Shinohara, M.; Kanazawa, T.; Uehira, K. Successive Estimation of a Tsunami Wavefield without Earthquake Source Data: A Data Assimilation Approach toward Real-time Tsunami Forecasting. *Geophysical Research Letters* **2015**, *42(19)*, 7923-7932.
4. Wang, Y.; Su, H.Y.; Ren, Z.; Ma, Y. Source Properties and Resonance Characteristics of the Tsunami Generated by the 2021 M8.2 Alaska Earthquake. *Journal of Geophysical Research: Oceans* **2022**, *123(3)*, e2021JC018308.
5. Geist, E. L. Complex Earthquake Rupture and Local Tsunami. *Journal of Geophysical Research: Solid Earth* **2002**, *107(B5)*, ESE-2.
6. Li, Y.; Goda, K. Hazard and Risk-Based Tsunami Early Warning Algorithms for Ocean Bottom Sensor S-Net System in Tohoku, Japan, Using Sequential Multiple Linear Regression. *Geosciences* **2022**, *12(9)*, 350.
7. Goda, K. Multi-hazard Parametric Catastrophe Bond Trigger Design for Subdection Earthquakes and Tsunamis. *Earthquake Spectra* **2021**, *37(3)*, 1827-1848.
8. Goto, C.; Ogawa, Y.; Shuto, N.; Imamura, F. Numerical Method of Tsunami Simulation with the Leap-frog Scheme. *IOC Manuals and Guides* **1997**, *35*, 130.
9. Santos, A.; Fernandes, J.; Mileu, N. Tsunami Hazard Assessment at Oeiras Municipality, Portugal. *Journal of Marine Science and Engineering* **2022**, *10(8)*, 1120.