

べき乗回帰モデル及び多層パーセプトロンを用いた津波浸水深予測

課題責任者

馬場 俊孝 徳島大学大学院社会産業理工学研究部

著者

上谷 政人*¹*¹ 徳島大学大学院創成科学研究科

本研究では、南海トラフ地震を想定した 3480 ケースの波源断層モデルから構築された津波浸水データベースを用いて、クラスタ解析で予測点を削減したうえで、べき乗回帰モデル及び多層パーセプトロンを用いて津波浸水予測を実施した。浸水深の予測精度は共役勾配法よりも多層パーセプトロンが優れていた。多層パーセプトロンは重回帰式の説明変数に複数の中間層を加えることができる。これにより、複雑な事象に対応することができたと考えられる。内閣府 11 ケースモデルを検証データとして予測してみたところ、あるクラスタ域内の津波浸水深の平均値の予測で RMSE=0.3~1.1m だった。クラスタ数の増加や学習データの追加による浸水予測についても解析を行ったが、予測精度の向上は見られなかった。更なる予測精度の向上のためには、教師データの充実と過学習への対策、予測モデルの改良が必要である。

キーワード：津波浸水予測，回帰，べき乗則，多層パーセプトロン

1. はじめに

日本海溝のプレート境界では 2011 年にマグニチュード 9 の東北地方太平洋沖地震が発生した。地震を検知してから約 3 分後に津波警報が発令された。震災後の詳細な調査によれば、津波の高さは最大で 40.1m であったが、その地域の津波予報第一報は 3m 未満であった。第一報から 30 分後の第二報では予測が大幅に切り上げられたものの、地震による停電の影響で第二報以降の津波警報を知ることができなかった地域もあった。東北沖地震の被害は広範囲に生じ、死者・行方不明者は約 23000 人と報告されている。

大規模災害時には電源消失などの事態が生じるため、実用を考える上では予測システムの堅牢さにも気を配る必要がある。実際、気象庁の津波警報システムは東京と大阪で重複化されている。しかしながら、万が一の事態を考えて、リアルタイム高速・大規模計算を必要としない別の方法も研究開発されるべきであろう。それには沖合と沿岸の津波の相関を利用する回帰モデル型^[1,2]が最も適している。回帰モデル型は学習データの構築に計算資源が必要であるものの、一旦、回帰式を構築してしまえば津波観測の特徴量のみで予測できる。予測時には高速・大規模計算を必要としない軽いシステムである。つまり、小システムによる分散型処理が可能なロバストな方法である。先行研究^[1]ではガウス過程を用いて海底ケーブルシステムで観測できる津波高から沿岸の津波高を予測する方法を提案された。ガウス過程は内挿に強いが外挿には弱い欠点があり、それを補うために Yoshikawa et al.^[2]は沖合津波と沿岸津波の相関の傾向を調べて、べき乗則を用いる手法を提案した。それは、内挿区間ではガウス過程モデルとほぼ同等の性能で、外挿区間ではガウス過程モデルよりもよい性能を示した。

しかし、先行研究の回帰モデルは海岸の任意の 1 点のみの高さを予測するだけで、最大浸水深分布のような面的な分布を求めるに至っていない。津波災害発災後の応急対応などを考えた場合、沿岸津波高だけでなく、浸水深分布も予測できることが望ましい。既存の回帰モデルで浸水深分布を求めるには、単純には空間上のすべての点

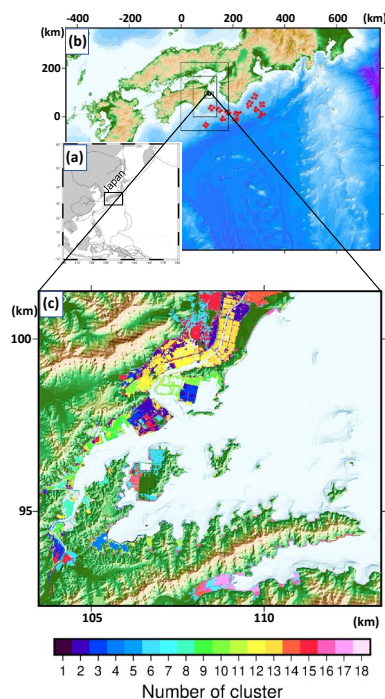


図 1 (a) 研究対象地域。(b) 津波計算領域。黒い矩形領域はネスティングレイヤーを示す。赤の三角形は DONET 観測所の海底圧力計の位置、座標系は日本の平面直角座標系 IV である。(c) 津波予測領域。色は k-means 法による津波の浸水深のクラスタ化された領域を意味する。

について予測を実施すればよいわけだが、予測点が膨大で処理時間が長くなるという問題があり、ここでは回帰モデルの利点が失われる。解決策として、本研究では、津波による浸水深が常に類似しているエリアを予めグループ化して予測点を減らす手法を提案する。手順は次のとおりである。まず、回帰モデル構築のための学習データ（津波データベース）を準備する。そのデータベースの浸水深分布データに対してクラスタ解析を実施して、どの津波イベントに対しても浸水深が似ている区域を判別し、クラスタ分けする。クラスタ内の浸水深の最大値を特徴量として、3480 ケースを用いて回帰モデルを構築する。このようにすれば、少ない回帰モデルで浸水深の空間分布を予測できる。最終的に、2012 年に日本政府が発表したマグニチュード9の地震シナリオに対して、回帰モデルを用いて予測を実施し、予測精度を評価する。なお、本研究では回帰モデルの構築に共役勾配法、多層パーセプトロンの2つ方法の性能を調査した。

2. 解析手法

藤原ら^[3]は南海トラフ地震を想定した 3480 ケースの断層モデルを提案した。本研究では、3480 ケースの断層モデルによるすべての津波を地球シミュレータで計算して、津波浸水データベースを構築した。研究対象地域を図1に示す。各断層モデルについて、海底地殻変動を半無限均質弾性体モデルを用いて計算し、海底斜面の水平変位による津波励起の効果と、線形ポテンシャル理論に基づく水理フィルタを考慮した上で津波の初期水位を計算した。また、初期水位のライズタイムは 60 秒と設定した。津波の伝搬路上計算には次の非線形長波理論を、スタaggerド格子のリーブフロッグ差分法を用いて解いた。津波計算に用いた地形データは、徳島県の津波浸水想定で使ったデータを利用した。津波襲来時の潮位は平均潮位 (T.P. 0m) とした。図1cに示した予測対象地域に向かって5層にネスティング(図1b)を行った。格子サイズは大きいものから 810m, 270m, 90m, 30m, 10m と設定した。

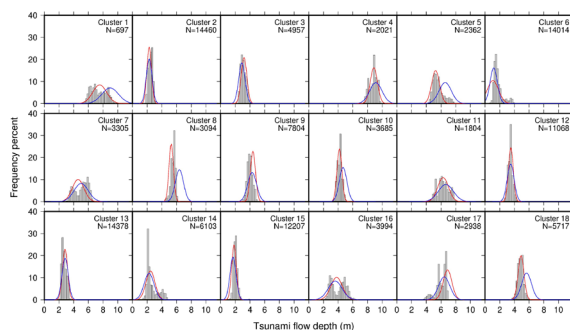


図2 M9 シナリオ No. 8 (テストデータ) の各クラスタにおける津波浸水深の頻度を表したヒストグラム。灰色のヒストグラムは、フォワード計算によって得られたものである。青色はCG法、赤色はMLP法により海底圧力データから予測されたもの。

津波伝搬を評価地域のすべてに最大波が含まれるように計算時間を6時間とした。計算ステップ幅については、計算の安定条件を満たすように 0.1 秒と設定した。津波防波堤等の海岸構造物を線状の構造物として準備したデータを計算に利用しており、海岸構造物は越流した場合に構造物が倒壊したとみなして、それ以降は線状構造物を削除して計算を継続した。南海トラフでは津波被害想定のための地震のシナリオ (以降, M9 シナリオ) が 11 個想定されている。本研究ではこれらの津波を上述の手法で計算しテストデータとして利用する。

ここではどの断層モデルにおいても津波浸水深が同一として扱えるエリアをクラスタ化する。トレーニングデータのうち津波が有意に浸水している 14 の断層モデルをランダムに選び出して、その浸水深分布に k-means 法を適用した。利用した断層モデルは藤原ら^[3]の断層モデル番号で 101, 315, 884, 1562, 1596, 1816, 1838, 2125, 2512, 2645, 2668, 2725, 2842, 2850 とした。k-means 法では研究対象領域内のクラスタ数は解析者が事前に指定しなければならない。そこでクラスタ数を変化させた解析を複数回行い、クラスタ分けされたエリア内の浸水深データのばらつきを参考にしてクラスタ数を決めた。本研究では、できる限り少ない津波浸水の予測点 (小さいクラスタ数) で浸水深を予測することを目的としているので、便宜上、クラスタ数 18 を採用した。この場合、1 クラスタ内の浸水深データの標準偏差は平均 0.46 m であった。テストデータセットの M9 シナリオの 8 番の各クラスタ内浸水深データの頻度分布図 (図 2) を例として示す。クラスタ解析で得られた結果を図 1c に示す。

本研究で利用した津波予測式は次のように表現される。

$$y_j = \sum_{i=1}^n a_i x_{i,j}^{b_i} \quad (1)$$

ここで、 i は観測点番号、 j は津波シナリオ番号で、 $n = 51$ は沖合観測点数 (図 1)、 a, b は推定する回帰係数である。 x は津波襲来中の海底水圧の最大偏差の絶対値であり、 y は目的変数でクラスタ内の津波浸水深の平均値、標準偏

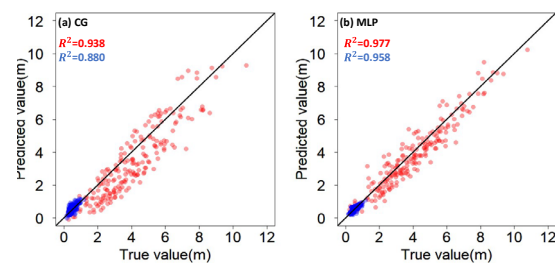


図3 (a)CG法、(b)MLP法による真値 (フォワード計算結果) と予測値との散布図。赤丸はクラスタ化した領域の浸水深の平均値、青丸は標準偏差を示す。

差の2つの場合を試した。つまり、図2のクラスタ内の浸水深の頻度分布の形を推定したことになる。回帰係数 (a, b) を、3480 ケースのすべてのデータを使って共役勾配法（以降、CG）により推定した。CG法は式(1)を直接解いた。解析には統計処理ソフトRのoptim関数を利用した。なお、optimのCG法では切片(c)も入っている。なお、反復計算を行う際の初期条件としてaとcに0を与え、bに1を与えた。

本研究ではCG法により構築したべき乗回帰式に加えて、yを出力層、xを入力層として複数の中間層を持つ多層パーセプトロン（以降MLPと呼ぶ）を用いた浸水深の予測式を構築した。浸水予測式の構築には、3480 ケースのすべてを利用した。機械学習ライブラリであるTensorflowを利用した。本研究では浸水予測式を構築する際の条件として、多層パーセプトロンの中間層を9層、ノード数をはじめの層を8ノードとして、9つの中間層が2のべき乗となる様に設定した。中間層にはL2正則化項を考慮した、損失関数には平均二乗誤差を適用して、活性化関数にReLU活性化関数を用いた。オプティマイザーにはAdamを用いた。予備調査として、正則化項及びAdamのハイパーパラメータの値を変更してRMSEが最小となる値を検証した結果、ハイパーパラメータの値には0.001を採用した。

3. テストデータの予測結果

3480 ケースとは別の南海トラフ地震断層モデルである11個の内閣府M9シナリオ^[4]により、本研究で構築した津波浸水深の予測式の性能を評価した。津波データベースの構築と同じ手法でM9シナリオの津波を計算し、説明変数である51か所の海底水圧偏差の絶対値を求めて、先に構築したべき乗回帰モデルと多層パーセプトロンの予測式に代入し、当該地域の浸水深分布を予測した。それと断層モデルから直接フォワード計算した真値を比較した。2つの手法による11個のM9シナリオの18クラスタにおける予測と真値の比較の散布図を図3に示す。

浸水深の平均値に予測については、CG法と比べてMLP法が優れていた。標準偏差は手法間でさほど違いがなかった。より定量的に予測誤差を示したものが図4で、シナ

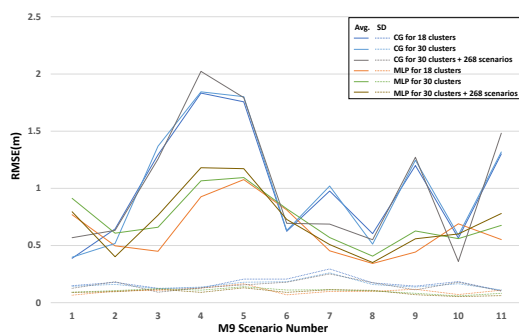


図4 M9シナリオのNo.1～No.11の真値（フォワード計算結果）と予測値のRMSEを折れ線グラフ。実線と破線はそれぞれ平均値と標準偏差を示す。

リオ毎に平均と標準偏差の真値との中の残差二乗和平均平方根（RMSE）を示す。シナリオ1, 6, 10はCG法の方がよかったが、その他はMLP法の予測精度がよかった。どの手法でも、シナリオ4, 5は他のシナリオと比較してやや予測精度が低かった。シナリオ4, 5について、MLP法を用いることでCG法と比較してRMSEが49%, 39%改善された。図5はM9シナリオ8のフォワード計算結果とMLP法によるべき乗則の予測結果（平均値）の比較である。図2では、平均値、標準偏差の予測結果を用いた正規分布曲線を赤線（MLP）と青線（CG）で示した。フォワード計算による津波浸水深分布を、概ね再現できていることがわかる。

4. 考察

CG法とMLP法の予測精度に差が生じた理由として、予測式の基底関数の仮定の有無が挙げられる。CG法は基底関数としてべき乗回帰モデルを予め設定しているが、MLP法では基底関数を決める必要がない。言い換えれば、MLPはデータに適合する関数形を自動で判別することにより、より複雑な事象に対応できる。また、中間層の数の違いが予測精度の与える影響を評価したところ、中間層の数が多いほど予測精度は良かった。

図2～5では、双方の予測手法において、研究対象地域に特に大きな津波を襲来させるシナリオ（シナリオ4, 5）の予測精度が相対的に悪かった。この原因として、一般に機械学習では膨大なデータの学習が必要であり、教師データ数が不足しているのではないかと考えられた。そこで、クラスタ数を18から30に増やすことによりデータ数を疑似的に増やして同じ解析を行ってみたが、残念ながらCG法、MLP法の予測精度の向上は見られなかった（図4）。

M9シナリオ4, 5の津波規模は、学習させた3480ケースのうち最大のもと同程度である。一般に大津波ほど非線形性も強くなり、複雑な挙動を示し、予測が難しくなる。極大シナリオの学習データが足りない可能性があり、そこで学習データにより大きな津波を含めて再解析してみた。3480ケースの地震シナリオのうち大きい方から268ケースを取り出して、そのすべり量を1.5倍にした極大

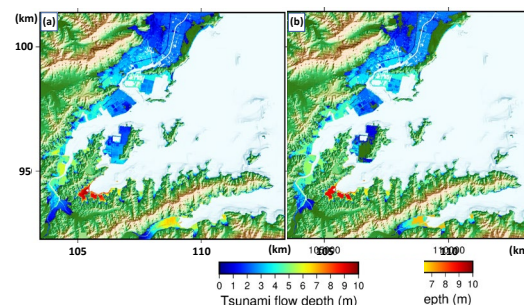


図5 (a)M9シナリオNo.8からフォワード計算された津波浸水深分布。(b)MLP法により予測された津波浸水深の深度分布。

津波シナリオを作成し、津波を計算し、学習データに含んだ。これを用いて再度回帰モデルを構築し、シナリオ 4, 5 の津波を予測したが、この場合も残念ながら精度が向上しなかった。クラスタ数を増やしても、学習データを増やしても CG 法及び MLP 法の精度向上は見られなかった。

クラスタ数を増やしても学習データ数を増やしても精度が向上しないのならば、さらにどのようにして予測精度を向上していったらよいだろうか。ひとつは性質の異なる説明変数を利用することが挙げられる。具体的には津波の大きさだけでなく、津波の到達時刻を利用することである。津波の到達時刻は、言い換えれば津波の襲来方向（波源の位置）の情報を持っている。到達時刻を説明変数に利用することにより、疑似的に波源位置の情報を取り込めるようになり、予測精度が向上する可能性がある。また、これをさらに発展させて、水圧の時系列波形すべてを学習させるスキームも利用可能であろう。水圧の時系列波形を並べた画像を作成し、それを学習させて予測するのである。これらの調査については今後の課題とした。

5. まとめ

本研究ではべき乗則を用いた回帰式^[1,2]を拡張させて津波の浸水深分布を予測するスキームを新たに提案した。津波浸水データベースから無作為に選んだ浸水深分布について、浸水深が常に似ている領域を k-means 法を用いてクラスタ化した。津波浸水データベースを学習データとして、各クラスタ領域の浸水深（平均値と標準偏差）を目的変数、DONET 観測点の水圧変動の最大値を説明変数とした予測モデルを構築した。予測モデルの構築には、べき乗則による回帰モデルと近年発展が目覚ましい機械学習の一手法である MLP を利用した。べき乗則による回帰モデルの構築においては CG 法を試した。浸水深の平均値に予測については、CG 法と比べて MLP 法が優れていた。内閣府 11 ケースモデルを検証データとして予測してみたところ、CG 法のクラスタ域内の津波浸水深の平均値の予測値と真値の差は RMSE=0.4~1.8m, MLP 法は RMSE=0.3~1.1m だった。

本手法をベースにすれば、スパコンによる高速計算や大規模データベースの必要ない軽いかつロバストな予測システムの構築が可能になる。計算資源は学習データの作成には必要だが実行時には高速計算機もデータベースも必要ない。大規模災害時には不測の事態が想定されるため、現在主流になりつつある大規模システムによる津波予測に加えて、このようなスタンドアロン処理が可能な予測システムもあったほうがよいだろう。また、津波災害は世界共通の課題であるが、高速リアルタイム計算機が利用できる国は限られており、本手法はそのような地域での活用も期待される。

謝辞

本研究は科学研究費補助金 (19H02409) の支援を受けています。津波計算には地球シミュレータを利用させていただきました。記して、感謝いたします。

文献

- [1] Igarashi Y, Hori T, Murata S, Baba T, Okada M. Maximum tsunami height prediction using pressure gauge data by a Gaussian process at Owase in the Kii Peninsula, Japan. *Mar. Geophys. Res.* **37**, 361-370, 2016. <https://doi.org/10.1007/s11001-016-9286-z>
- [2] Yoshikawa M, Igarashi Y, Murata M, Baba T, Hori T, Okada M. A nonlinear parametric model based on a power law relationship for predicting the coastal tsunami height, *Mar. Geophys. Res.* **40**, 467-477, 2019. <https://doi.org/10.1007/s11001-019-09388-4>
- [3] 藤原広行ほか, 南海トラフ沿いの地震に対する確率論的津波ハザード評価 -第一部 本編-, 防災科学技術研究所資料, 第 439 号 (第一部) 71pp, 2020. https://dil-opac.bosai.go.jp/publication/nied_tech_note/pdf/n439_01m_1.pdf,
- [4] 内閣府, 2012 年南海トラフの巨大地震モデル検討会, <http://www.bousai.go.jp/jishin/nankai/model/index.html>, 最終アクセス日時 2022 年 1 月 25 日

Tsunami Inundation Predictions Applying Machine Learning Techniques to Calculated Database

Project Representative

Toshitaka Baba, Graduate School for Social and Industrial Science and Technology,
Tokushima University

Author

Masato Kamiya*¹

*¹Graduate School for Sciences and Technology for Innovation, Tokushima University

This study proposes a new tsunami prediction method using tsunami database calculated from 3480 fault models in the Nankai Trough. We applied machine learning techniques to the database for predictions in tsunami inundation. We investigated two machine learning models of multiple regressions using power law and multilayer perceptron. To speed up the analysis, we reduced tsunami inundation prediction points by clustering before the predictions. The predictions were made for the Cabinet Office 11-case earthquake scenarios as test data. The multilayer perceptron was superior to the power law regression in terms of predicting tsunami inundation depth. It estimated RMSE from 0.3 to 1.1 m for the average tsunami inundation depth within clustered area.

Keywords : Tsunami prediction, regression, power law, multilayer perceptron

1. Introduction

Yoshikawa et al.^[1] predicted the tsunami height at a point on the coast by a regression model using power law, but did not obtain the spatial distribution of the maximum tsunami flow depth. Emergency response after a tsunami disaster requires information on the flow depth distribution in the damaged area, in addition to the tsunami height along the coast. To obtain it using the regression models, we predict all the points in the inundated area, but the number of predicted points is enormous, and the processing time will be long. Hence, this study proposes a method to reduce the number of prediction points by pre-grouping the areas where the flow depths are always similar. The analysis procedure was as follows. First, we calculated the tsunamis of 3480 cases in the Nankai Trough^[2] for training data to construct a regression model. Then, we applied cluster analysis to the flow depths of the training data to identify areas with similar flow depths for all tsunami events in the training data. Regression relationships were estimated using the conjugate gradient (CG), and the multilayer perceptron (MLP) methods, in which objective variables are the average flow depths in the clustered areas, and the explanatory variables are the maximum ocean bottom pressure deviations during a tsunami at DONET stations. Finally, we used the constructed regression models to predict the tsunami flow depths calculated from hypothetical earthquake scenarios released by the Japanese government^[3] to evaluate the prediction accuracy.

2. Analysis methods

Fujiwara et al.^[2] proposed fault models with 3480 cases of interplate earthquakes in the study area. We constructed the training data by calculating all the tsunamis generated by 3480 fault models. Fault motion in each model caused seafloor

displacement assuming a semi-infinite homogeneous elastic body. We estimated the initial tsunami water level using the vertical component of the seafloor displacement, the effect of tsunami excitation by horizontal displacement of the seafloor slope, and the filter of the linear potential theory. Numerical tsunami simulations used the rise time of the initial water level at 60s. The nonlinear long-wave equations solved by the staggered grid leapfrog difference method estimated the tsunami propagation from the initial water level and tsunami run-up on land. The topographic data used in the tsunami calculations were obtained from the local government in the study area, i.e., Tokushima Prefecture. Topographic nesting consisted of five layers. The grid intervals in the layers were 810m, 270m, 90m, 30m, and 10m from the coarsest layer to the study area. The tide level when the tsunami attack occurred was assumed to be the mean tide level (0m in Tokyo Peil). Coastal tsunami defense structures smaller than the grid intervals, such as breakwaters, were modeled as line structures in the calculations. When a tsunami overtops the coastal structures, we considered the collapsed structures and continued the calculation by excluding the line structures. The integral time was 6h, so that the maximum tsunami waves arrived in all the evaluation areas. To satisfy the stability condition of the computation, the computational time step width was set to 0.1s.

The cabinet office of Japan^[3] re-investigated geological and geophysical features and historical inter-plate earthquakes in the Nankai Trough. They released 11 earthquake scenarios (fault slip distributions) with a magnitude of 9.1 (hereafter referred to as M9 scenarios) which may occur in the Nankai Trough. M9 scenarios are realistic, which include the heterogeneity of the slip distribution on the fault plane. This study calculated tsunamis from the M9 scenarios using the above procedure and used them

as test data.

We created clustered areas where the tsunami flow depth was similar to the training data to reduce the prediction points. Fourteen fault models that cause significant tsunami inundation in the study area were randomly selected from the training data, and the k-means method was applied to their flow depth distributions. Of the 3480 cases, the selected fault model id numbers were 101, 315, 884, 1562, 1596, 1816, 1838, 2125, 2512, 2645, 2668, 2725, 2842, and 2850 in Fujiwara et al.^[2]

In the k-means method, an analyst should specify the number of clustered areas in advance. This study determined the number of clustered areas based on the variance in flow depth in the clustered area. Additionally, a small number of clustered areas is preferable because the purpose of this study is to predict tsunami flow depths with low computational cost. We found that 18 clustered areas were optimal in this study, in which the standard deviation of the flow depth data in one cluster was 0.46 m on average.

The tsunami prediction equation used in this study is as follows:

$$y_j = \sum_{i=1}^n a_i x_{i,j}^{b_i} \quad (1)$$

where, $n = 51$ is the number of offshore observation points, i is the number of observation points, j is the number of tsunami cases, \mathbf{a} , \mathbf{b} are the regression coefficients to be estimated. \mathbf{x} is the absolute value of the maximum deviation of seafloor water pressure. \mathbf{y} is the objective variable, which is the mean and standard deviation of the tsunami flow depths in the clustered area.

The CG estimated the regression coefficients (\mathbf{a} , \mathbf{b}) using all the training data from 3480 cases. We used the function *optim* of the statistical processing software *R*. The CG method included the intercept (c), and gave 0 to \mathbf{a} and c, and 1 to \mathbf{b} as the initial conditions for iteration.

In addition to the power law regression in equation (1) solved by the CG method, this study predicted the tsunami flow depth for the clustered area using the MLP method with \mathbf{y} as the output

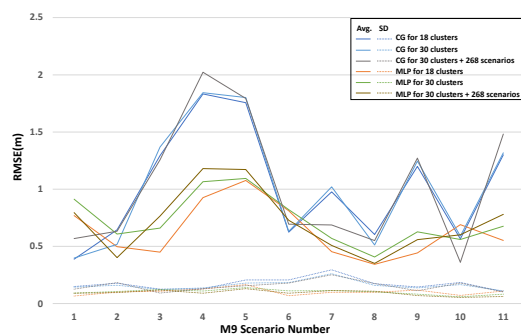


Fig.1 Line graph of the RMSE between true (forward calculations) and predicted values for the M9 scenario from No.1 to No. 11. Solid and dashed lines indicate mean and standard deviation, respectively.

layer and \mathbf{x} as the input layer. We used Tensorflow libraries with the following analysis settings. The number of the intermediate layers was 9. The first intermediate layer had eight nodes, and the number of nodes in the other intermediate layers was set to be a power of 2. We applied the ReLU activation function and the L2 regularization in the intermediate layers. Adam algorithm was used to minimize the loss function of the mean square error. We also performed cross-validation tests by changing hyperparameter values to avoid overfitting.

3. Results

Here, we show the performance of the tsunami flow depth prediction using the test data of the 11 M9 scenarios^[3], which are different Nankai Trough earthquake models from the training of 3480 cases. The absolute values of the seafloor water pressure deviations at 51 DONET stations from the M9 scenarios are explanatory variables that were substituted into the prediction equations from the CG and MLP methods. The MLP method is more accurate than the CG method in predicting the average flow depth. The prediction of the standard deviation of the clustered flow depth data was similar between the two methods.

Figure 1 shows more quantitative comparisons in the prediction error by the root mean square of residual errors (RMSE) between the predicted and the true value for each scenario. The M9 scenarios No. 1, 2, 8, and 10 were better predicted by the CG method, while the MLP method had better prediction accuracy for the other M9 scenarios. The M9 scenarios No. 4 and 5 had lower prediction accuracy than the other scenarios.

Acknowledgement

This study was supported by JSPS KAKENHI, 19H02409. The tsunami calculations were conducted on the Earth Simulator. We deeply thank staff who served on the operations of the Earth Simulator.

References

- [1] Yoshikawa M, Igarashi Y, Murata M, Baba T, Hori T, Okada M. A nonlinear parametric model based on a power law relationship for predicting the coastal tsunami height, *Mar. Geophys. Res.* **40**, 467-477, 2019. <https://doi.org/10.1007/s11001-019-09388-4>
- [2] Fujiwara H, Hirata K, Nakamura H et al., Probabilistic Tsunami Hazard Assessment for Earthquakes Occurring along the Nankai Trough –Volume 1 Part I–. Technical Note of the National Research Institute for Earth Science and Disaster Resilience, 439, 71pp. 2020.
- [3] Cabinet Office, Government of Japan (2012) Massive Earthquake Model Review Meeting of the Nankai Trough. <http://www.bousai.go.jp/jishin/nankai/model/index.html> in Japanese, last accessed on 25 January 2022.