Application of a New Machine Learning Model to Improve Earthquake Ground Motion Predictions

Authors: Anushka Joshi, Balasubramanian Raman, C. Krishna Mohan, Linga Reddy Cenkeramaddi

Contents:

Introduction

- Importance of Peak Ground Acceleration
  - PGA represents the maximum ground acceleration that occurs during earthquake shaking.
- Challenges with Current Ground Motion Prediction Equations
  - Geographical Limitations
  - Data Dependency
  - Simplification
- Introducing SeisEML: A new Approach
  - Ensemble Method
  - Cross Region Validation
- Objectives of SeisEML Application
  - Overcome the limitations of traditional GMPEs
  - Provide better predictions
  - Scalability and Adaptability

Model Overview

+ Hybrid
+ Sanitized Gray Wolf Optimizer (SGWO): Inspired by the social hierarchy and hunting behavior of gray wolves.
+ Optimization Function:
    - Formula:

$$F_f = (1 - M_{cvAcc}) * 100 \qquad M_{cvAcc} = \frac{c_1 + c_2 + c_3}{3}$$

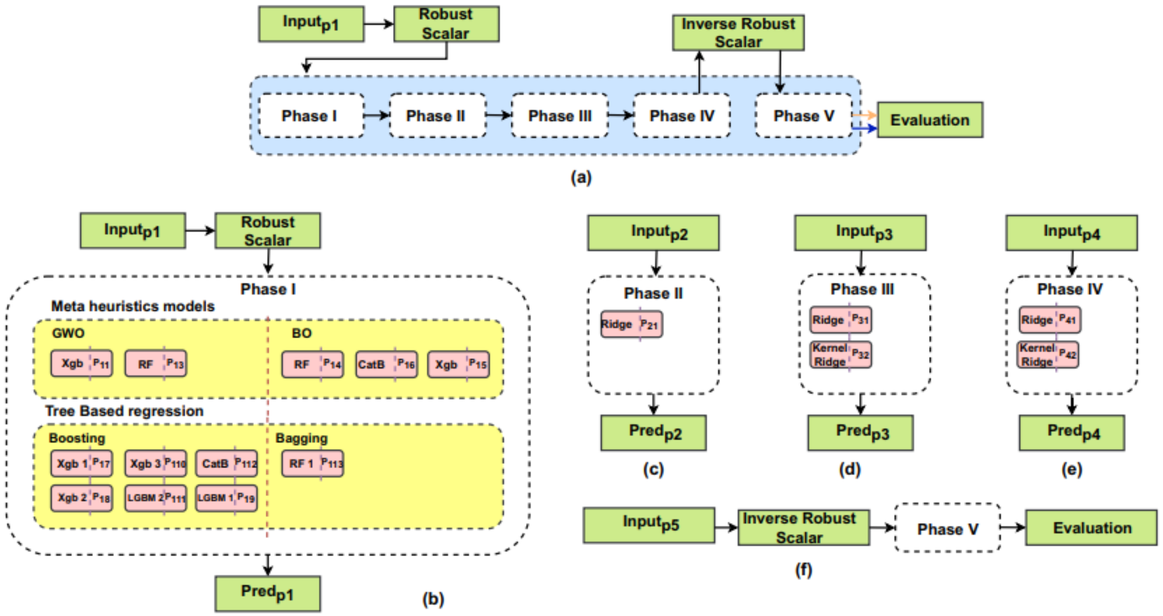
    - Explanation:
        + F_f: Fitness function aims to minimize the error rate.
        + M_{cvAcc}: Mean cross-validation accuracy from threefold scores c1,c2,c3.
    - Usage: Optimizes parameters for XGB and RF models.
    - Advantages: Balances exploration and exploitation, avoiding local optima.
+ Bayesian Optimizer (BO):
    - Key Feature: Uses probabilistic models to find the global optimum efficiently.
    - Application:
        + Enhances model parameters for XGB, RF, and CatB models.
        + Benefit: Reduces the number of training cycles required compared to other methods.

+ Kernel Ridge Regression
  - Polynomial Kernel Formula:

$$k(x, y) = (\gamma x^T y + c_0)^d$$

  - Parameters:
    + d: Degree of the polynomial, set to 4 in this study.
    + c_0: Coefficient, default value of 1.

Architecture of SeisEML:



- Phase I:
  - Meta-Learners: XGB-1, XGB-2, XGB-3, RF-1, RF-2, SGWO-XGB, SGWO-RF, BO-XGB, BO-RF, BO-CatB, LGBM-1, LGBM-2, CatB.
  - Inputs: Re, Rh, di, st, VS30, M.
- Phase II:
  - Ridge Layer: Uses predictions from Phase I as inputs.
  - Predictions: Denoted by Predp2.
- Phase III:
  - Input Set: Combines Phase I predictions and Phase II predictions.
  - Outputs: Predp3.
- Phase IV:
  - Input Set: Uses combined predictions from Phases I, II, and III.
  - Outputs: Predp4.
- Phase V:
  - Final Layer: Combines predictions from previous phases using averaging.
  - Outputs: Final predicted PGA values.

- Feature Scaling:
  - RobustScaler: Applied to handle outliers and standardize features.
- Key Features Used:
  - Re: Epicentral distance (km).
  - Rh: Hypocentral distance (km).
  - di: Focal depth.
  - st: Site characteristics.
  - VS30: Shear wave velocity at 30m depth.
  - M: Magnitude of the earthquake.
- Inverse Scaling: Applied to final predictions to match the original scale.

$$value = \frac{(X_i - X_{med})}{(p_{75} - p_{25})}$$

About the Data

+ The Dataset Source:
  - Kyoshin Network (K-NET), Japan:
    + High-Quality Near-Source Recordings: Collected from 1996 to 2022.
    + Total Records: 29,789 records from various earthquakes.
      (20852, 2681, and 6256 records from 412, 336, and 377 earthquakes)

Table 1 Range of various parameters used in the training, testing and validation of the model

Dataset	V_{S30}		Epicenter Distance(km)		Hypocenter Distance(km)		Magni-tude(M_{JMA})		Focal Depth(km)	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Train	39	1579	0.02	985.23	5.57	985.53	3	9	0	256
Test	39	1579	1.08	955.20	7.62	955.50	3	9	0	256
Validation	39	1579	3.02	946.25	7.79	946.32	3	9	0	256

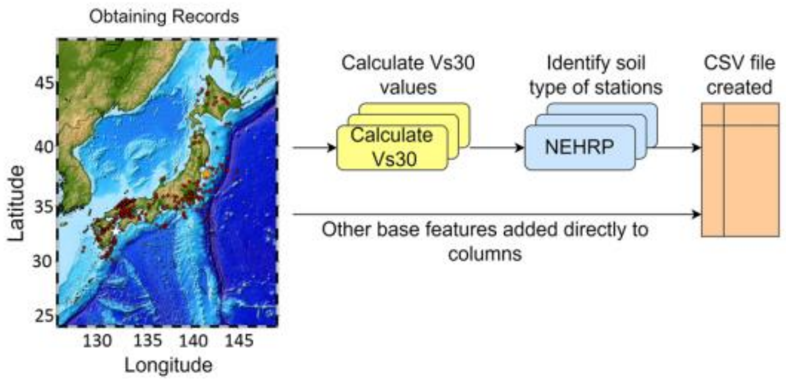
+ In the data V_{S30} is one of the major inputs that was calculated using shear wave velocity profile:

$$V_{S30} = \frac{30m}{\sum_i \Delta t_i}$$

+ In the above equation, the parameter ti represents the wave travel time of a shear wave in a particular layer, which is calculated by using the following equation:

$$\sum_i \Delta t_i = \frac{\Delta d_i}{v_{si}}$$

+ In the above equation, d_i represents the thickness of the intervening i th layer between the surface and 30 m depth, and v_{si} is the shear wave velocity in the respective layer.



Performance Testing

Models	Test		Validation	
	R^2	*MAE*	R^2	*MAE*
GPR	0.68	0.145	0.68	0.147
ExTR	0.69	0.1407	0.69	0.1419
RSubCatB	0.69	0.1477	0.70	0.147
AdaBoost	0.41	0.203	0.37	0.207
DTree	0.66	0.150	0.66	0.151
RR	0.41	0.204	0.41	0.204
SeisEML	**0.77**	**0.123**	**0.76**	**0.130**

SNo.	Feature set	R^2	MAE	$RMSE$
1	$[R_h, R_e, M, S_t, d_i, V_{S30}]$	0.74	0.49	0.62
2	$[R_e, M, S_t, d_i, V_{S30}]$	0.72	0.51	0.64
3	$[M, S_t, d_i, V_{S30}]$	0.16	0.89	1.12
4	$[S_t, d_i, V_{S30}]$	0.09	0.93	1.16
5	$[d_i, V_{S30}]$	0.09	0.93	1.16
6	$[V_{S30}]$	0.04	0.96	1.2

Degree (d)	*MAE*	*RMSE*
1	0.13	0.16
2	0.13	0.16
3	0.13	0.16
4	0.12	0.15
5	0.12	0.16

Discussion

- PGA compared with 2 different
  - Conversion Formula: Used to convert MJMA to Mw and Ms for GMPE calculations.

$$M_w = 0.58 M_{JMA} + 2.25, 3.0 \leq M_{JMA} \leq 5.5$$

$$M_w = 0.97 M_{JMA} + 0.04, 5.6 \leq M_{JMA} \leq 8.2$$

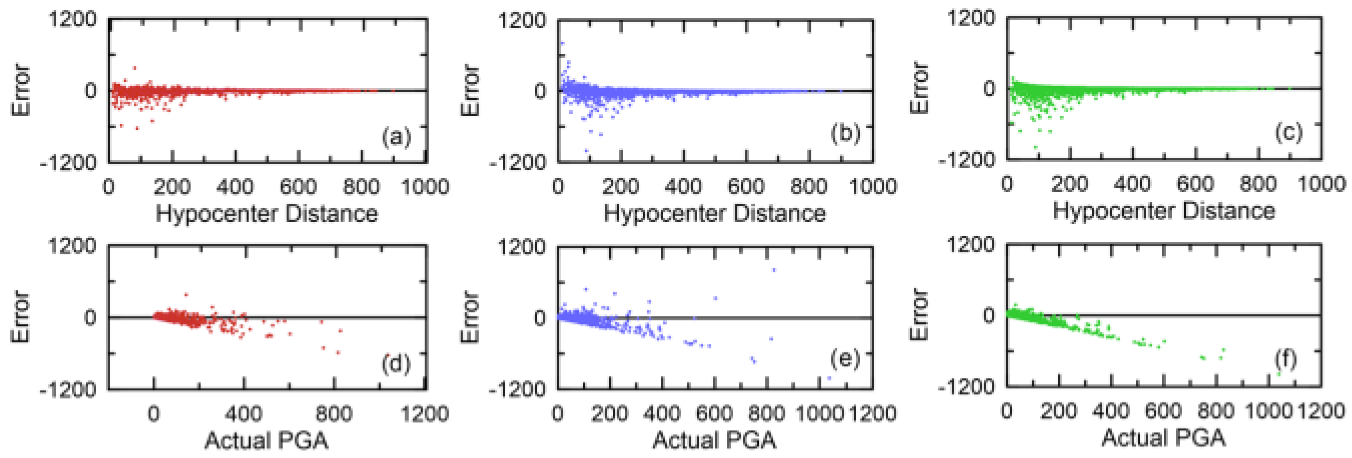
- GMPE formula calculation of each paper

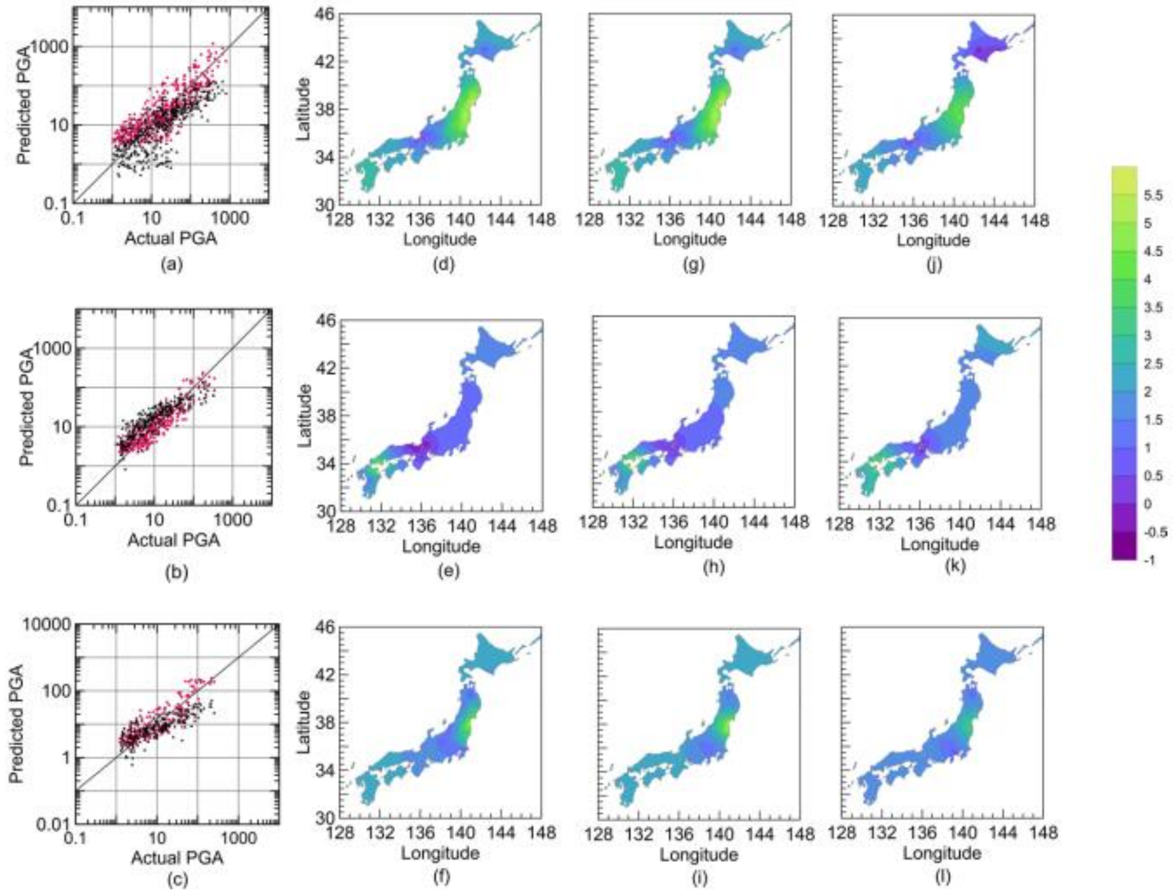
$$\log_{10} PGA_{pred} = aM_w + bR - \log_{10}(R + c10^{aM_w}) + d + s_T$$

$$\log_{10} PGA_{pred} = \alpha + \beta M_s - \bar{C}\log_{10}[r + e^{h_2 M_s}] + F\phi + Ebr$$

- The performance was evaluated by seeing the difference between actual and predicted PGA values.

$$ERR = PGA_{pred} - PGA_{actual}$$


1200
0
-1200
Error
(a)
0 200 400 600 800 1000
Hypocenter Distance
(b)
(c)
(d)
0 400 800 1200
Actual PGA
(e)
0 200 400 600 800 1000 1200
(f)



Metric	SeisEML	Ls-SVR RBF kernel	ANN/SA	GP/OLS	MEP
R^2	0.770	0.76	0.73	0.66	0.70
MAE	0.01	0.0316	0.13	0.488	0.697

Conclusion

- Key Findings:
  - Superior Performance: SeisEML outperforms traditional ground motion prediction equations in terms of R^2, MAE, and RMSE.
  - Reliable Predictions: Produces accurate PGA predictions for both regional and cross-region earthquakes.
  - Robust Against Outliers: Use of RobustScaler helps in handling outliers effectively.
- Applications:
  - Seismic Hazard Mapping: Enhances the reliability of PGA predictions, making it useful for seismic hazard assessments.
  - Early Warning Systems: Provides accurate real-time predictions for earthquake response.