Journal of Information Systems Engineering and Management
2025, 10(12s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Modeling Seismic Signals of an Earthquake through Symbolic Dynamical System: The Case of Surigao Region, Philippines

Teotima Evangelista Gorres[1], Liane Vina G. Ocampo[2], Rachelle Santos[3], Reynold Campo[4]
[1]teotimaabato@gmail.com
Agusan del sur State College of Agriculture and Technology, Philippines
[2]lianeocampo@cvsu.edu.ph
Cavite State University, Philippines
[3]ramsantos@aims.edu.ph
Asian Institute of Maritime Studies (AIMS)
[4]reynold.campo@pcu.edu.ph
[4]Philippine Christian University, Philippines

ARTICLE INFO

Received: 22 Nov 2024

Revised: 30 Dec 2024

Accepted: 20 Jan 2025

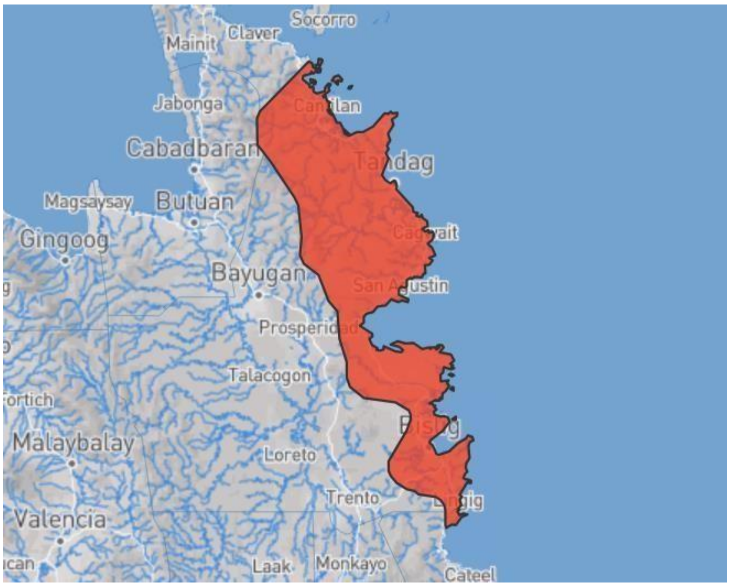
ABSTRACT

Earthquakes pose threats globally, most especially in developing hotspots such as the Philippines, given that it falls under the Pacific Ring of Fire. One of the most seismically active regions within the country is the Surigao region located in northeast Mindanao, where two large earthquakes hit in December 2023 with 7.4 and 6.8 magnitudes. The outcome of this study underscores the need for accurate earthquake forecasting models in such high-risk regions. Conventional methods for seismic signal modeling usually lose the complexity and nonlinear of seismic datasets. This paper uses symbolic regression with genetic programming as a means to come up with a more accurate and explainable model in estimating earthquake magnitudes in the Surigao area. A technique in machine learning, symbolic regression generates mathematical expressions which best fit empirical data, and it doesn't require a predefined model configuration. The developed model identifies critical parameters relevant to seismic phenomena and accurately estimates the magnitudes of earthquakes within the considered region. The model, such as the magnitude of 3.0, as happening in September 2024. The results obtained in this study contribute to the accuracy of forecasting for earthquakes by providing SR as an efficient tool for studying seismic records. These findings have the potential to further improve this effort in disaster preparedness and risk reduction that saves lives and reduces economic losses in earthquake-prone areas. Collaboration is encouraged with a continued comprehensive revision toward research for further exploration of broader applications in seismic hazard analysis.

Keywords: earthquake, magnitude, seismic signal, symbolic regression, genetic programming.

Introduction

Earthquake pose a significant threats to communities around the world, causing loss of life, property damage, and economic disruption. The Philippines, located in the Pacific Ring of Fire, is particularly vulnerable to seismic activity due to the convergence of multiple tectonic plate (Roque et al, 2023). The modeling seismic signals is very crucial particularly for regions prone to frequent seismic activity like Philippines.



Source; www.ThinkHazard.org

The Surigao region, situated in the northeastern part of Mindanao, is one of the most seismically active areas in the county that experience regular earthquakes. Last December 2023, two damaging earthquakes struck the offshore of the Province of Surigao del sur with a recorded 7.4 and 6.8 magnitude on the 2^{nd} and 4^{th} of the month (PhilVocs, 2024). It is also classified as medium level earthquake hazard according to the Global Facility for Disaster Reduction and Recovery. This implies that there is a 10 percent probability of experiencing earthquake tremors with the potential for causing damage in the vicinity for the next 50 years (GRFDRR, 2017). This necessitate the development of models to predict and explain this natural phenomenon, and accurate forecasting of earthquake magnitude in this region is crucial for disaster preparedness and mitigation efforts.

Traditional seismic signal modelling methods often rely on predefined models that not capture the complex, non-linear characteristics of seismic data. To address this challenge, the researchers' utilize the symbolic regression with genetic programming to develop a more accurate and interpretable seismic signal model for Surigao Earthquakes. Symbolic regression (SR) genetic programming (GP) is a stochastic (Poli, et. al, 2008) iteration techniques that propagate all possible symbolic models as valid (Vladislavleva, 2008) mathematical expression in a given set of input variables, basic functions and constants and searches (Lima Neto, 2009) a model (set of models) optimizing fitness objective such as prediction accuracy of the training set (observed data set).

Symbolic regression (SR), a type of machine learning, offers a promising alternative by evolving mathematical expressions that best fit the observed data without assuming predefined model structure (Schimidt & Lipson, 2009). SR combined with genetic programming (GP) is a well-established technique for generating the mathematical expression that illustrate relationship within dataset (Takaki & Tomoyuki, 2022).

This research aims to explore the application of symbolic regression for forecasting earthquake magnitude in the Surigao Region. This will contribute to the field of earthquake forecasting by introducing symbolic regression as a powerful tool for modeling seismic data. The insight gained from the evolved mathematical expression can enhance understanding of earthquake dynamics in the Surigao region and improve the accuracy of magnitude forecasts. Ultimately, this work seeks to aid in disaster preparedness and risk reduction efforts, potentially saving lives and reducing economic losses in earthquake-prone areas.

This paper uses symbolic regression (SR) with a genetic programming (GP) model to predict future values of earthquake magnitudes in Surigao, based on the concept of dynamical systems. Genetic programming is an evolutionary computation technique that is evolved using an evolutionary algorithm, which is particularly suited for big data. Genetic programs are developed based on genetic algorithms, which are inspired by the tools and procedures of evolutionary biology. During the evolutionary process, the computer program mutates or iterates to produce a potentially optimal model. In evolutionary biology, life evolves from simple to complex forms through natural selection, where individuals with higher "fitness qualities" are more likely to survive. These individuals are retained and paired with other highfitness individuals. Offspring are then generated with the aim of achieving higher fitness values than their parents. The algorithm stops when an offspring is found that meets a pre-determined fitness score (Gorres, T., 2021).

The development of Symbolic Regression Program

The development of Symbolic Regression (SR) programs has a rich history rooted in symbolic dynamical systems (SDS). Hadamard first applied SDS in 1898 to complex systems, and in the 1950s, Von Neumann described SDS as a powerful framework for artificial systems. Genetic programming models, utilizing machine learning algorithms, were developed to extend this idea, including the symbolic regression programs by Koza (1994), Schmidt and Lipson (2009), and Cohen, Beykal, and Bollas (2023). System dynamics theory, particularly coding theory or symbolic regression programs for symbolic dynamics, involves building a state space (function or model), identifying statistical properties (Lind & Marcus, 2021), and modeling sequences of infinite or bi-infinite symbolic shift spaces. These are measured in discrete time intervals with shifts or symbolic invariant mappings between spaces, forming pattern modeling-based codes in a symbolic system (Marcus & Rosenthal, 2013).

Investigating dynamic and complex systems—whether physical, biological, or ecological—requires significant effort to achieve meaningful results. As systems are inherently complex, researchers have developed system identification techniques such as symbolic regression (SR) through artificial intelligence and modern dynamical systems (genetic programming). In modeling and investigating real-world problems, symbolic regression algorithms based on genetic programming (GP) are used to identify iterated curves or functions. SR represents an abstract model space in topological systems for machine learning techniques (Schmidt & Lipson, 2009; Märtens, Kuipers, & Mieghem, 2017), which can be explained by physical systems through experimental data.

Computer systems are employed to elucidate system dynamics and complexity (chaoticity) through probability theory, measure theory, and ergodic processes. Symbolic regression, used for prediction models in real-world problems, fits the complexity or dynamics of these systems.

The development of modern dynamical systems and symbolic regression programs has been an active research area, with significant contributions from Koza (1992), Vladislavleva (2008), and Schmidt and Lipson (2009). They emphasize that genetic programming in symbolic regression does not assume any predefined model for the input-output data during the evolutionary process.

Symbolic regression involves inferring a non-linear and linear mathematical model in a symbolic analytical function $f\colon R^n \rightarrow R$, which fits two random variables X, and such that $y_n =$

$f\ (x_1, \dots \dots ..\ x_n)$ given $X_i = (x_1, \dots \dots ..\ x_n)$ and $Y_i = (y_1, \dots .\ y_n)\ i = 1, \dots .$ input-output data

(Tenachi, , Ibata, and Diakogiannis. (2023)

As explained by Poli, Langdon, and McPhee (2008), SR is a computer program or model used to fit numerical data within a specific interval. De Franca and Aldeia (2021) clarified that SR is used to model input-output data, identify systems, make predictions about unobserved samples (state space), and determine a system's statistical properties by simultaneously estimating coefficients or parameters and function space.

Conversely, symbolic regression, introduced by Vladislavleva (2008) and Gorres-Abato and Tarepe (2014), can be expressed as:

$$f^{\hat{}} = (X\ F\ W) \qquad (1)$$

This represents a measure-theoretic space over time. De Franca and Aldeia (2021) consider the equation $(X\ F\ W)$as the Interaction-Transformation for Symbolic Regression, where X is a state (function) space in an input-output pair (X, Y) F is a set of functions (function operators and analytic functions), and Crepresents the weights or coefficients. For example, equation (1) can be represented by a simple function resulting from mutation (iteration) using an evolutionary algorithm, a genetic programming (GP) approach (Poli, Langdon, & McPhee, 2008), to optimize the search space for an interpretable expression such as:

$$G(x_i) = 2 + 4x_1 - 4\sin x_2{}^2 x_1 + 7cos\ {}^{x}{}_{x}{}^{3}4^4 - 3\log x_5$$

and can be represented as in (X, F, W)

$X = [[1,0], [1, 2], [4, -1], [1,0]]$

$F = [[sin, cos, log]]$

$C = [2, 4, -4, 7, -3]$

Where M is a symbolic model representing a valid mathematical expression of the input variables X_1, $X_2 \ldots X_d$, basic functions operators F (+, -, sin, cos, etc.) and constant W. The set of constants can be finite and infinite, random, constant, drawn from a specific interval [a, b] (Gorres-Abato, T., and Tarepe, D., (2015), and Vladislavleva, (2008)). The general function will be transformed into symbolic form and filtered, resulting only a target functions (Rad, H. Feng,J.,and Iba, H., (2018)).

Aside from Koza and Poli (2005), Kadierdan, K., Nathan, K., and Brunton, S., (2020) and various other researchers have identified that SR programs in modern dynamical systems use simultaneous parameter and function space estimation (Kay & Mowbray et al., 2023; Austel & Dash et al., 2017) within symbolic dynamical systems through measure theory, random processes, and ergodic processes, optimizing iterations of function space (model) (De Franca, 2018; 2022; Koza, 1992; 1994).

Methods and Design

A symbolic regression program (genetic programming) is a stochastic process optimizing the desired prediction model, and the parameters. In a dynamical system, the Symbolic regression is a function that describes the relation of the input - output data and can be written as;

$$f = (X, F, \mu)$$

Where any input variables, F basic functions operators (+, -, sin, cos, etc.) and μ set of measures

$$(x_1, x_2, x_3, \ldots, x_n) = E \qquad \rightarrow (input\ data)$$

The task of symbolic regression with Genetic programming is to identify the target expression

(function) or model in symbolic form as in equation (1); which estimates the values of a specified target variable based on the values of a set of input variables, that is;

$$E\hat{} = f\hat{} = (X\ F\ W) \rightarrow output\ data$$

The study made use of the descriptive methods of research using evolutionary algorithms particularly the genetic programming based symbolic regression. Data were obtained from the Phivolcs, PAGASA Caraga and Accuweather. The maximization can be quantitatively measured by:

$$Minimize: MSE = \frac{\sum ni=1(Ei - E\hat{}i)2}{n-1}$$

Where: $E\hat{}_i$= predicted value of E_i using four (4) parameters.

The parameters are:

x_1= local temperature

x_2= rainfall

x_3= air pressure

x_4= sea pressure

$E = earthquake\ magnitude$

$$E\hat{}_i = \ = {}_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 \qquad (2)$$

Where: w_1, w_2, w_3, and w_4 are weights to be determined by using Symbolic Regression via genetic programming (GP). The program can generate the GP search process and establish proposed model. The software used is the license version of DATAROBOT.

RESULTS AND DISCUSSION

The analytic solution to the model using the Symbolic regression with Genetic programming mathematically expressed by:

$$\hat{E} = M(X\ F\ W)$$

The resulting model:

$$\hat{E} = 2.59 + 0.06 * co(sqrt(x_2)) - 0.04 * cos(2.69 + x_3 - x_2 * co(sqrt(x_2))) \qquad (3)$$

and can be represented as in $(X\ F\ W)$

$X = \{[0,1], [1,1]\}$

$F = [cos, sqrt]$

$W = [2.59, 0.06, 0.04, 2.69]$

Equations (3) is the model of an earthquake in Surigao Region. From the four (4) parameters used in the study only two (2) parameters shown in the output model as the result of the run. In the same way, using the model as in equation (3) above, the prediction of an earthquake in some part of Surigao is 3.0 for the month of September 2024.

Conclusion

Based on the result, it identified drivers of earthquake magnitude activities to be rainfall, and air pressure.

The result suggests that, within this study, the complexity of the magnitude of an earthquake in the Philippines relate more to changes in rainfall, and air pressure, rather than local temperature or sea pressure. Excluding the last sets of variables could lead to an inference that they may not directly and significantly affect the dynamics of an earthquake magnitude within the time period and conditions given in this research.

Recommendation

Further modification and cross-validation method should also be applied to check the model's efficiency in prediction. The research could be extended easily to include more predictors. Any collaboration for creating a software for this particular study is very much welcome.

REFERENCES

[1] Global Facility for Disaster Reduction and Recovery (2017). Surigao Del Sur. ThinkHazard. World Bank Group Philippine Institute of Volcanology and Seismology (2023). The December 2023 Magnitude 7.4 and 6.8 offshore Surigao del Sur earthquakes. Retrieved from: https://bit.ly/SurigaoDelSur-earthquakes

[2] Roque, Pamela & Violanda, Renante & Bernido, Christopher & Soria, Janneli Lea. (2023). Earthquake Occurrences in the Pacific Ring of Fire Exhibit a Collective Stochastic Memory for Magnitudes, Depths, and Relative Distances of Events. 10.2139/ssrn.4450059.

[3] Schmidt, M., & Lipson, H. (2009). Distilling Free-Form Natural Laws from Experimental Data. Science, 324(5923), 81-85.

[4] Takaki, Katsushi, Miyao, Tomoyuki [2022]. Symbolic regression for the interpretation of quantitative structure-property relationships, Artificial Intelligence in the Life Sciences,Volume 2, ISSN 2667-3185,https://doi.org/10.1016/j.ailsci.2022.100046.

[5] Ramdhane. F.B., (2023). Symbolic dynamical systems in topological spaces defined via edit distances. Mathematics [math]. Aix Marseille Université (AMU), Marseille, FRA.; University of Sfax, Tunisia, 2023. English. ffNNT : 2023AIXM0186ff. fftel-04146898f

[6] Koza, J.R., (1992). Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA, USA, 1992. [Koza, 1994] John R. Koza. Genetic Programming II: Automatic Discovery of Reusable Programs. MIT Press, Cambridge Massachusetts, May 1994.
[7] Koza, J.R., (1994). Genetic Programming II: Automatic Discovery of Reusable Programs. MIT Press, Cambridge Massachusetts, May 1994.
[8] Schmidt, M., and Lipson, H., (2009). Solving iterated functions using genetic programming. [PDF] from psu.edu. Book Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers
[9] Cohen, B., Beykal, B., and Bollas, G., (2023)."Dynamic System Identification from Scarce and Noisy Data Using Symbolic Regression," 2023 62nd IEEE Conference on Decision and Control (CDC), Singapore, Singapore, 2023, pp. 3670-3675, doi: 10.1109/CDC49753.2023.10383906. keywords:
[10] Lind, D., and Marcus, B., (2021). An Introduction to Symbolic Dynamics and Coding. 2nd Edition.
[11] Part of Cambridge Mathematical Library.
[12] https://www.cambridge.org/ph/universitypress/subjects/mathematics/differential-and-integralequations-dynamical-systems-and-co/introduction-symbolic-dynamics-and-coding-2ndedition?format=PB&isbn=9781108820288
[13] Marcus, B., and Rosenthal, J., (2013). Codes, Systems, and Graphical Models - The IMA Volumes in Mathematics and its Applications 123. https://www.waterstones.com/book/codes-systems-and-graphical-models/brian-marcus/joachimrosenthal/9781461265344
[14] Smith, R., (2013). Introduction to Supercritical Fluids. Richard Smith, ... Cor Peters, in Supercritical Fluid Science and Technology, 2013. https://www.sciencedirect.com/topics/chemistry/statisticalmechanics#:~:text=Statistical%20mechanics%20uses%20probability%20theory,the%20macroscopic% 20(bulk)%20properties
[15] Märtens, M., Kuipers, F., & Van Mieghem, P. (2017). Symbolic Regression on Network Properties. In J. McDermott, M. Castelli, L. Sekanina, I. Wolf, & P. García-Sánchez (Eds.), Genetic Programming: Proceedings - 20th European Conference, EuroGP 2017 (pp. 131-146). (Lecture Notes in Computer Science; Vol. 10196). Springer. https://doi.org/10.1007/978-3-319-55696-3_9
[16] de Franca, F. O. (2022). Transformation-Interaction-Rational Representation for Symbolic Regression.
[17] arXiv:2205.06807v1 [cs.NE] 25 Apr 2022
[18] de Franca, F.O. (2018). A greedy search tree heuristic for symbolic regression, Information Sciences, Volumes 442â€“443, 2018, Pages 18-32, ISSN 0020-0255,
https://doi.org/10.1016/j.ins.2018.02.040. ttps://doi.org/10.1016/j.ins.2018.02.040 0020-0255/© 2018 Elsevier Inc. All rights reserved
[19] Vladislavleva, E. (2008). *Model-based problem solving through symbolic regression via pareto genetic* programming. [Doctoral Thesis, Tilburg University]. CentER, Center for Economic Research.
[20] Schmidt, M., Lipson, H. (2009). Symbolic Regression of Implicit Equations. In: Riolo, R., O'Reilly, UM., McConaghy, T. (eds) Genetic Programming Theory and Practice VII. Genetic and Evolutionary Computation. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1626-6_5
[21] Tenachi, W., Ibata, R., and Diakogiannis. F., (2023). Symbolic regression driven by dimensional analysis for the automated discovery of physical laws and constants of nature. Journées 2023 de la Société Française d'Astronomie & d'Astrophysique (SF2A)., Jun 2023, Strasbourg, France. pp.107- 108. ffhal-04325284f
[22] Rad, H., Feng, J., and Iba, H., (2018). GP-RVM: Genetic Programing-based Symbolic. Regression Using Relevance Vector Machine. arXiv:1806.02502v3 [cs.NE] 26 Aug 2018 Poli, Langdon, and McPhee, (2008). A Field Guide toGenetic Programming.
[23] https://www.researchgate.net/publication/216301261_A_Field_Guide_to_Genetic_Programming
[24] Koza, J.R., Poli, R. (2005). Genetic Programming. In: Burke, E.K., Kendall, G. (eds) Search Methodologies. Springer, Boston, MA. https://doi.org/10.1007/0-387-28356-0_5
[25] Kadierdan, K., Nathan, K., and Brunton, S., (2020). 2020SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamicsProc. R. Soc.
[26] A.4762020027920200279http://doi.org/10.1098/rspa.2020.0279
[27] Kay, H., and Kay, S., and Mowbray, M., et al., (2023). Symbolic regression based interpretable datadriven soft-sensor for process quality control.
[28] https://www.sciencedirect.com/science/article/pii/B9780443152740502432

[29] Austel, V., Dash, S., Gunluk, O., Horesh, L., Liberti, L., Nannicini, G., and Schieber, B., (2017).
[30] Globally Optimal Symbolic Regression. ArXiv:1710.10720 [stat.ML] (or arXiv:1710.10720v3 [stat.ML] for this version) https://doi.org/10.48550/arXiv.1710.10720
[31] Gorres, T., (2021). Symbolic Regression via Genetic Programming; Philippines Population Prediction:
[32] 2010-2020 (May 2021). International Journal of Science and Research (IJSR) 2021, Available at
[33] SSRN: https://ssrn.com/abstract=3869624
[34] Gorres-Abato, T., and Tarepe, D., (2014). Unpublished master's degree thesis.

30 September 2024 - 02:42 AM (Surigao Del Sur)	08.67 126.71 031	3.0	048 km S 72° E of Marihatag
23 September 2024 - 08:36 PM Del Norte)	10.00 126.16 020	3.0	009 km S 75° E of Burgos (Surigao
11 September 2024 - 03:02 AM Del Sur)	08.12 127.02 018	3.0	068 km N 82° E of Lingig (Surigao
08 September 2024 - 06:02 PM Del Sur)	08.87 126.51 026	3.0	024 km S 76° E of Cagwait (Surigao
08 September 2024 - 09:35 AM Del Sur)	08.13 127.08 005	3.0	074 km N 82° E of Lingig (Surigao
06 September 2024 - 09:26 AM (Surigao Del Sur)	08.38 126.56 033	3.0	025 km N 87° E of Hinatuan