В работе использованы данные подводных землетрясений — расстояния от гипоцентра землетрясения до глубоководного желоба, глубины землетрясения — для трех регионов: Японии, Чили и Восточных Алеутских островов.

Применяемые в работе методы заключались в стандартизации данных, создании модели робастной регрессии с помощью метода Хубера, поиске лучших гиперпараметров и кусочно-линейной аппроксимации полученных результатов. Для сравнения была также создана модель неробастной (линейной) регрессии и модель робастной регрессии на очищенных данных — без глубин 33 и 35 км, которые соответствуют землетрясениям с плохо определенными и недостоверными глубинами. Все применяемые в работе методы использованы отдельно для каждого из трех регионов.

Обработка данных выполнялась с помощью библиотеки Pandas. Стандартизация данных проводилась с помощью метода StandardScaler из библиотеки Scikit-learn. Стандартизация заключалась в том, чтобы каждая переменная имела нулевое среднее и единичное стандартное отклонение, что улучшает производительность моделей машинного обучения. Дополнительно было выполнено сохранение средних значений и масштабов для обратного преобразования.

Перед созданием модели выполнено преобразование исходных признаков в полиномиальные для улучшения возможностей модели по захвату нелинейных зависимостей в данных. Для этих целей применён метод PolynomialFeatures из библиотеки Scikit-learn. Одновременное использование полиномиальных признаков и робастной регрессии позволяет эффективно моделировать сложные зависимости даже в условиях присутствия выбросов. Такое сочетание подходов обеспечивает гибкость и устойчивость модели.

Для увеличения точности модели применён поиск лучших гиперпараметров с использованием метода RandomizedSearchCV из библиотеки Scikit-learn на основе кросс-валидации, выполнено обучение модели с найденными оптимальными параметрами.

Для интерпретируемости полученных результатов выполнено обратное преобразование предсказанных значений модели, коэффициентов, свободного члена в исходных масштаб.

В целях определения входных параметров при определении геометрии очага землетрясения использована кусочно-линейная аппроксимация полученных результатов.

Методика кусочно-линейной аппроксимации состоит из нескольких ключевых этапов:

1. Определение узлов (контрольных точек).

Узлы представляют собой ключевые точки на графике, через которые проходят линейные сегменты аппроксимации. В работе использовалась длина сегмента в качестве фиксированного параметра.

2. Расчёт расстояния между точками.

Для определения точек, в которых необходимо создать новые узлы, использовалась функция для расчёта евклидова расстояния между текущей точкой и следующей. Новый узел создавался, когда сумма указанного расстояния превышала установленную длину сегмента.

3. Вычисление углов наклона

После определения узлов и построения отрезков между ними, вычислялись углы наклона каждого отрезка относительно оси абсцисс.

Созданная в работе функция approximation() основана на библиотеке Pwlf и выполняет последовательно следующие действия:

- начальную инициализацию узловых точек;

- поочерёдный перебор всех точек набора данных;

- расчёт расстояния до следующей точки и проверку превышения суммарного расстояния заданной длины сегмента;

- добавление нового узла при необходимости;

- после обработки всех точек дополнение набора данных последней точкой;

- вычисление углов наклона между узловыми точками.

Визуализация всех данных и результатов работы выполнена с помощью библиотеки Matplotlib.pyplot. Для выполнения математических и статистических операций в коде также были использованы библиотеки Numpy, Scipy.stats.