Analysis of El Centro's Seismic Signals Using Time-Varying Auto-regressive Models

Arjun Verma
Department of Electronics and Communications Engineering
Indian Institute of Technology, Guwahati
Email: arjun.verma@iitg.ac.in

***Abstract*—The report discusses the analysis of the non-stationary seismic signal from El Centro's Earthquake using a 3rd-order Time-Varying Auto-regressive (TVAR) model. The study involves estimating the Auto-Regression (AR) coefficients over time through Auto-correlation method and Linear Regression side by side and then visualizing them through a coefficient cluster map, and interpreting the patterns to understand the signal's behavior.**

I. Introduction

A non-stationary signal is one whose statistical properties, such as mean and variance, vary over time. The seismic data from the El Centro earthquake is an excellent example of a non-stationary signal. In a Time-Varying Auto-regressive (TVAR) model, these non-stationary signal at any time t can be represented as a linear combination of its past values. Our goal is to estimate the coefficients associated with the different lagged or past values.

Apart from the conventional method of calculating the Auto-Regression (AR) coefficients using the autocorrelation method by solving Yule-Walker equations, as discussed in previous studies [1], this problem can also be approached by fitting a linear model to a set of past data points to predict the current value. The weights associated with the input features—namely, the past values—can then be interpreted as the AR coefficients. The AR coefficients obtained from these two methods can further be compared with the output of prebuilt Python functions to analyze the accuracy and validity of the methods.

In the El Centro seismic data, ground motion was recorded by an accelerometer, providing a signal of acceleration versus time. This signal exhibits periods of high and low amplitude, corresponding to various phases of the earthquake, such as initial tremors and aftershocks. The seismic data obtained from the accelerometer includes ground motion in three directions: North-South, East-West, and Up-Down. These three signals are analyzed individually to obtain the AR coefficients using the methods mentioned above.

II. Methodology

To begin, let $\mathcal{X}$ represent the three signals corresponding to the seismic motion components:

$$\mathcal{X} = \{x_{\text{NS}}, x_{\text{EW}}, x_{\text{UP}}\},$$

where x_{NS}, x_{EW}, and x_{UP} represent the north-south, east-west, and up-down components of the seismic signals, respectively. These signals represent acceleration a recorded at discrete but equidistant time intervals separated by 0.02 seconds. Let the signal components be denoted as $x_{\text{NS}}(t)$, $x_{\text{EW}}(t)$, and $x_{\text{UP}}(t)$ for the north-south, east-west, and up-down components, respectively, where t represents the discrete time index. The third-order TVAR model for each component of the seismic signal can be written as:

$$x_{\text{NS}}(t) = \sum_{i=1}^{3} \phi_{i,\text{NS}}(t) x_{\text{NS}}(t-i) + \epsilon_{\text{NS}}(t),$$

$$x_{\text{EW}}(t) = \sum_{i=1}^{3} \phi_{i,\text{EW}}(t) x_{\text{EW}}(t-i) + \epsilon_{\text{EW}}(t),$$

$$x_{\text{UP}}(t) = \sum_{i=1}^{3} \phi_{i,\text{UP}}(t) x_{\text{UP}}(t-i) + \epsilon_{\text{UP}}(t),$$

where:

- $x_{\text{NS}}(t)$, $x_{\text{EW}}(t)$, and $x_{\text{UP}}(t)$ are the seismic signal components at time t,
- $\phi_{i,\text{NS}}(t)$, $\phi_{i,\text{EW}}(t)$, and $\phi_{i,\text{UP}}(t)$ are the time-varying auto-regressive coefficients for each component at time t (for $i = 1, 2, 3$),
- $\epsilon_{\text{NS}}(t)$, $\epsilon_{\text{EW}}(t)$, and $\epsilon_{\text{UP}}(t)$ are the error terms (noise) for each component at time t.

A. Sliding-Window Extraction

To estimate the time-varying AR coefficients, we partition each seismic signal into overlapping windows. Let N denote the window size and S the step size. The i-th window extracted from a signal x is given by: $w_i = x[i : i + N], \quad \text{where } i = 0, S, 2S, \ldots,$ such that $i + N \leq \text{len}(x)$. For each component of $\mathcal{X}$, the corresponding set of windows is:

$$\mathcal{W}_{\text{NS}} = \{w_0^{\text{NS}}, w_S^{\text{NS}}, w_{2S}^{\text{NS}}, \ldots\},$$
$$\mathcal{W}_{\text{EW}} = \{w_0^{\text{EW}}, w_S^{\text{EW}}, w_{2S}^{\text{EW}}, \ldots\},$$
$$\mathcal{W}_{\text{UP}} = \{w_0^{\text{UP}}, w_S^{\text{UP}}, w_{2S}^{\text{UP}}, \ldots\}.$$

The number of windows M in each set is determined by:

$$M = \left\lfloor \frac{\text{len}(x) - N}{S} \right\rfloor + 1,$$

where $\text{Shape}(\mathcal{W}_{\text{NS}}) = (M, N)$. This approach preserves the signal's time-dependent features while allowing detailed analysis within each window.

B. Auto-regressive Coefficients Estimation

We then estimate the AR coefficients for each signal component using the ‘AutoReg‘ method from the StatsModels library. For a given window w and AR model order, the coefficients are computed as:

$$\mathbf{a} = \text{AutoReg}(w, \text{order}),$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_{\text{order}}]$ represents the set of AR coefficients.

Applying this method to all windows of the three signal components yields:

$$\mathbf{a}_{\text{NS}} = [\text{AutoReg}(w_{\text{ns}}, \text{order}) \mid w_{\text{ns}} \in \mathcal{W}_{\text{NS}}],$$

$$\mathbf{a}_{\text{EW}} = [\text{AutoReg}(w_{\text{ew}}, \text{order}) \mid w_{\text{ew}} \in \mathcal{W}_{\text{EW}}],$$

$$\mathbf{a}_{\text{UP}} = [\text{AutoReg}(w_{\text{up}}, \text{order}) \mid w_{\text{up}} \in \mathcal{W}_{\text{UP}}].$$

These coefficients serve as inputs for subsequent comparison with other modeling techniques, such as autocorrelation and linear regression.

C. Auto-regressive Modeling via Yule-Walker Equations

To estimate the AR coefficients we used the classical Yule-Walker equations. These equations establish a relationship between the autocorrelation function of a signal and its AR coefficients.

For a signal $x(t)$, the Yule-Walker equations for an AR(3) model are expressed as:

$$r(k) = \sum_{i=1}^{3} \phi_i r(k-i), \quad \text{for } k = 1, 2, 3,$$

where:

- $r(k)$ is the autocorrelation of the signal at lag k,
- ϕ_i are the AR coefficients to be estimated for $i = 1, 2, 3$.

These equations can be represented in matrix form as:

$$\begin{bmatrix} r(0) & r(1) & r(2) \\ r(1) & r(0) & r(1) \\ r(2) & r(1) & r(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ r(3) \end{bmatrix}.$$

This linear system is solved for ϕ_1, ϕ_2, ϕ_3 using the Toeplitz structure of the autocorrelation matrix:

$$\boldsymbol{\phi} = \mathbf{R}^{-1}\mathbf{r},$$

where $\mathbf{R}$ is the autocorrelation matrix, and $\mathbf{r}$ is the autocorrelation vector. The AR coefficients for each window are estimated using by solving this matrix equation for each of the signal components which yields:

$$\boldsymbol{\xi}_{\text{NS}} = [\text{SolveToeplitz}(r_{\text{ns}}, \text{order}) \mid w_{\text{ns}} \in \mathcal{W}_{\text{NS}}],$$

$$\boldsymbol{\xi}_{\text{EW}} = [\text{SolveToeplitz}(r_{\text{ew}}, \text{order}) \mid w_{\text{ew}} \in \mathcal{W}_{\text{EW}}],$$

$$\boldsymbol{\xi}_{\text{UP}} = [\text{SolveToeplitz}(r_{\text{up}}, \text{order}) \mid w_{\text{up}} \in \mathcal{W}_{\text{UP}}].$$

D. Linear Regression for AR Coefficients Estimation

To estimate the AR coefficients, we also apply the Linear Regression (LR) method. The goal is to predict the future value of the signal based on its past values to obtain the weights associated with it. For a given window w of length N, the signal value at time t, denoted as $w[t]$, is predicted using a linear combination of the past p values, where $p = 3$ in this case. The AR model can be expressed as:

$$w[t] = \sum_{i=1}^{3} \phi_i w[t-i] + \epsilon(t),$$

where:

- ϕ_i are the AR coefficients for lag i,
- $\epsilon(t)$ is the residual error,
- $t \in \{3, 4, \ldots, N-1\}$.

This equation can be rewritten in matrix form as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\phi} + \boldsymbol{\epsilon},$$

where:

- $\mathbf{y} = \begin{bmatrix} w[3] \\ w[4] \\ \vdots \\ w[N-1] \end{bmatrix}$ is the target vector, containing the future signal values,
- $\mathbf{X} = \begin{bmatrix} w[2] & w[1] & w[0] \\ w[3] & w[2] & w[1] \\ \vdots & w[N-2] & w[N-3] \end{bmatrix}$ is the design matrix formed by the lagged signal values,
- $\boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix}$ is the vector of AR coefficients.

To estimate the AR coefficients, we solve the linear system by minimizing the residual sum of squares. The coefficients $\boldsymbol{\phi}$ are obtained as:

$$\boldsymbol{\phi} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}.$$

For each signal window, the AR coefficients are estimated using the Linear Regression method, with no intercept term (i.e., the regression is fit with fit_intercept = False). The AR coefficients for the seismic signal components are then computed as:

$$\boldsymbol{\eta}_{\text{NS}} = [\text{Linear}(w_{\text{ns}}, 3) \mid w_{\text{ns}} \in \mathcal{W}_{\text{NS}}],$$

$$\boldsymbol{\eta}_{\text{EW}} = [\text{Linear}(w_{\text{ew}}, 3) \mid w_{\text{ew}} \in \mathcal{W}_{\text{EW}}],$$

$$\boldsymbol{\eta}_{\text{UP}} = [\text{Linear}(w_{\text{up}}, 3) \mid w_{\text{up}} \in \mathcal{W}_{\text{UP}}].$$

Having obtained the three sets of coefficients $\{\mathbf{a}_{\text{NS}}, \mathbf{a}_{\text{EW}}, \mathbf{a}_{\text{UP}}\}$, $\{\boldsymbol{\xi}_{\text{NS}}, \boldsymbol{\xi}_{\text{EW}}, \boldsymbol{\xi}_{\text{UP}}\}$ and $\{\boldsymbol{\eta}_{\text{NS}}, \boldsymbol{\eta}_{\text{EW}}, \boldsymbol{\eta}_{\text{UP}}\}$ derived from the three methods for each component of the seismic signal, we will proceed to compare the results and analyze the signal's behavior based on the estimated AR coefficients.

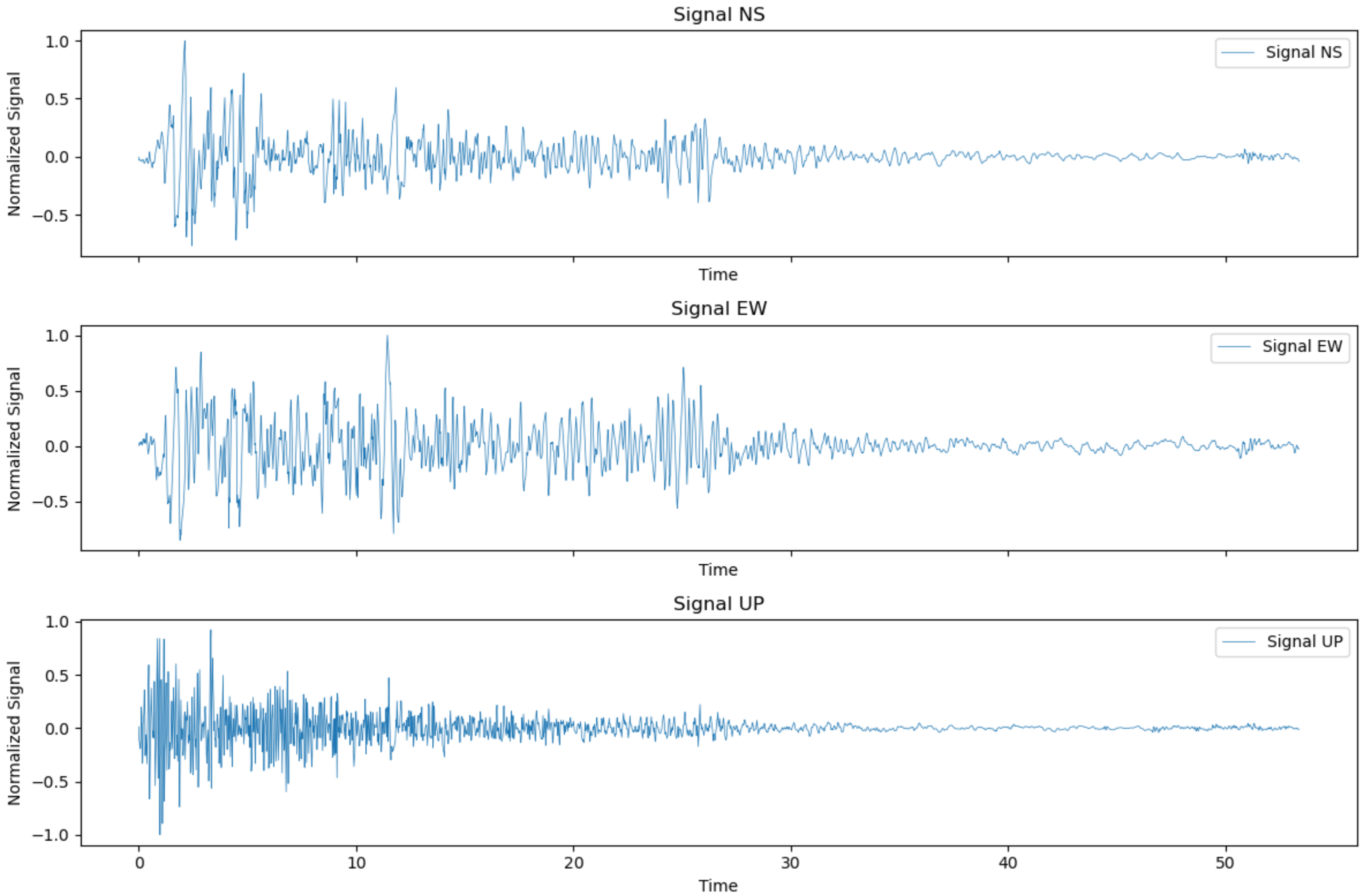

Fig. 1. The three components of the seismic signal, visualized to highlight the earthquake shocks and the associated ground movements.

E. Experiment and Results

The three seismic signals of acceleration versus time, showing the movement of the ground in the three directions, were plotted for the three components $x_{\mathrm{NS}}(t)$, $x_{\mathrm{EW}}(t)$, and $x_{\mathrm{UP}}(t)$, as shown in Figure 1. These plots clearly illustrate the time duration during which the earthquake tremors were most intense. Aftershocks can also be observed, occurring a few time steps after the initial tremors, somewhere in the middle of the observation period. These aftershocks are less intense than the initial tremors of the earthquake.

The AR coefficients cluster map, obtained from the three methods $\{\mathbf{a}_{\mathrm{NS}}, \mathbf{a}_{\mathrm{EW}}, \mathbf{a}_{\mathrm{UP}}\}$, $\{\boldsymbol{\xi}_{\mathrm{NS}}, \boldsymbol{\xi}_{\mathrm{EW}}, \boldsymbol{\xi}_{\mathrm{UP}}\}$, and $\{\boldsymbol{\eta}_{\mathrm{NS}}, \boldsymbol{\eta}_{\mathrm{EW}}, \boldsymbol{\eta}_{\mathrm{UP}}\}$, is shown in Figure 2 for the NS component, Figure 3 for the EW component, and Figure 4 for the UP component. The cluster map for the NS component shows the following observations:

- The AR coefficients change a lot over time, showing that the seismic signal is non-stationary. For example, in all three cluster maps, values between 0.5 and 1.5 (yellow areas) are more common in some time windows, while values between -0.5 and -1.0 (dark purple areas) dominate in others.
- ϕ_1 (Coefficient 0) is the most stable across time compared to ϕ_2 and ϕ_3. This means it represents the baseline energy of the seismic signal, which is mainly driven by $x(t-1)$, the previous value of the signal. ϕ_1 has higher values (yellow regions) during the first 100 time windows, showing stronger energy due to the arrival of primary seismic waves (like P-waves and S-waves). After this, the energy decreases (darker colors) as the waves lose strength and transition to aftershocks or surface waves.
- ϕ_2 (Coefficient 1) changes more suddenly, switching between positive (green/yellow) and negative (dark purple) values. These abrupt changes are most visible during:
  1) Time windows 40–80: A sharp shift from positive to negative values, likely caused by initial wave arrivals (like surface or reflected waves) or interactions between seismic waves.
  2) Time windows 150–200: Another big shift, possibly due to aftershocks or damping of earlier seismic waves.

This coefficient represents the oscillations or ups and downs in the signal. It captures fast changes in the strength and direction of the seismic waves, which might happen due to wave interference or sudden bursts of energy.

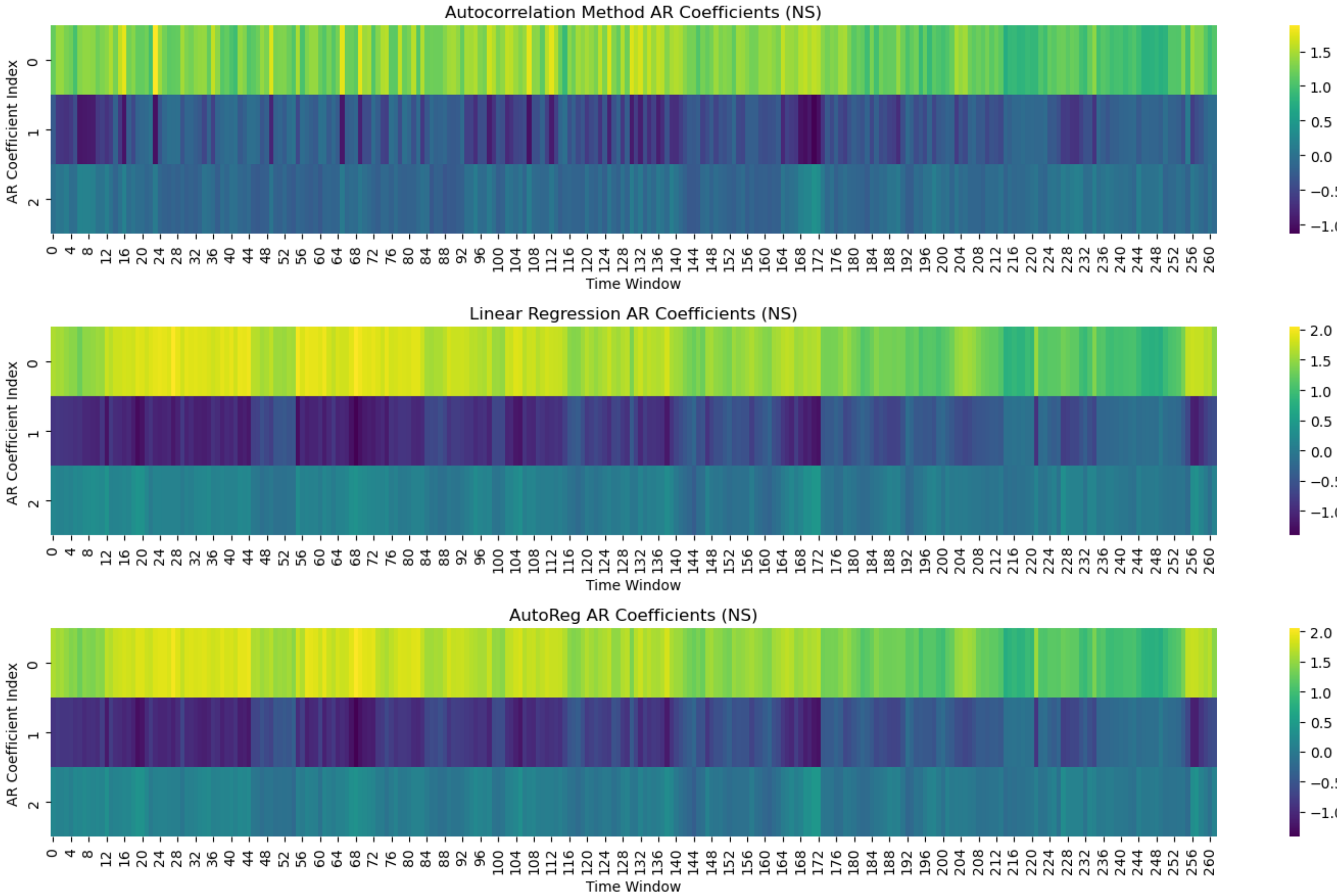

Fig. 2. AR Coefficients cluster map for the North-South component of the seismic signal, obtained using the three discussed methods. The time windows are on the X-axis and the AR coefficients on the Y-Axis.

- ϕ_3 **(Coefficient 2) represents slower changes in the signal**. While it still changes over time, its shifts are not as dramatic as ϕ_2. It captures smaller, smoother oscillations in the signal, like longer-term trends or less intense wave movements.
- Sudden changes in color show high-energy events like the arrival of strong seismic waves. These spikes are more visible in the Linear Regression and AutoReg maps for ϕ_2 and ϕ_3, showing their sensitivity to energy bursts. In quieter periods, a gradual change in color (dark to light) is seen across all indices, indicating steady energy transitions. During these intervals, the coefficient values are between -0.5 and 0.5, representing lower seismic activity.

The movement in the NS, EW, and UP components, along with similar trends in the AR coefficients, suggests that the earthquake affected all directions equally. ϕ_1 remains stable, showing the base energy from the primary seismic waves (P-waves). The sudden changes in ϕ_2 and ϕ_3 point to the arrival of secondary waves (S-waves and surface waves) and aftershocks and the oscillatory behavior of the signal. The similar patterns in all directions indicate that the earthquake was spread evenly and had a significant impact in all directions.

The Autocorrelation Method AR coefficients demonstrate smoother transitions with well-defined patterns. The Coefficients show stability across time with subtle variations in trends. The Linear Regression method, while showing similar overall patterns, exhibits sharper transitions and higher sensitivity to abrupt changes, closely resembling the behavior captured by AutoReg.

F. Links

The seismic data used for this analysis is obtained from here: https://www.vibrationdata.com/elcentro.htm. The complete code for implementation is available on the GitHub repository: https://github.com/who-else-but-arjun/TVAR_ElCentro

References

[1] G. Eshel, "The Yule Walker Equations for the AR Coefficients," stat.wharton.upenn.edu. Archived (PDF) from the original on July 13, 2018. [Online]. Available: https://stat.wharton.upenn.edu. [Accessed: Jan. 27, 2019].

[2] Q. Zhang and L. Wei, "Time-varying autoregressive model for real-time earthquake prediction based on linear regression," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 543085, 2012.

[3] A. Sağlam and A. H. Özdemir, "Application of time-varying autoregressive models for predicting seismic data," *International Journal of Earthquake Engineering*, vol. 49, no. 3, pp. 311-323, 2015.

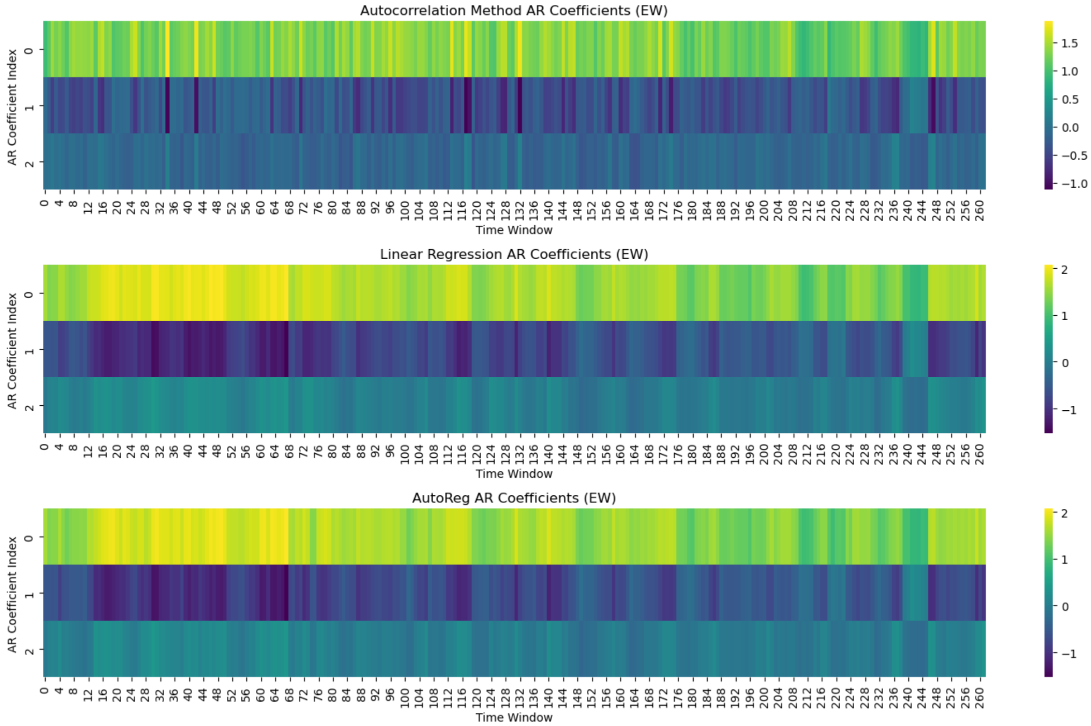

Fig. 3. AR Coefficients cluster map for the East-West component of the seismic signal, obtained using the three discussed methods. The time windows are on the X-axis and the AR coefficients on the Y-Axis.

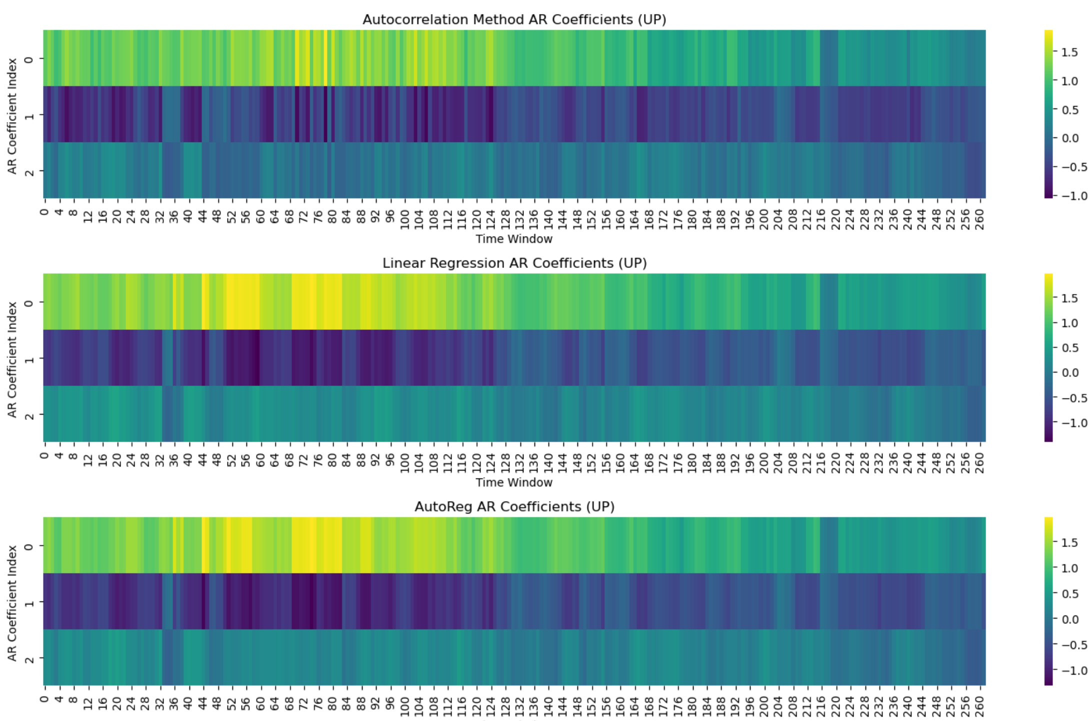

Fig. 4. AR Coefficients cluster map for the Up-Down component of the seismic signal, obtained using the three discussed methods. The time windows are on the X-axis and the AR coefficients on the Y-Axis.