

# **ALGORITMO *K*-MEANS PARA AGRUPAMIENTO DE SEGMENTOS EN SISMOGRAMAS MEDIANTE DESCRIPTORES: 2 CASOS DE ESTUDIO**

Juan Manuel Jiménez Vergara

Pregrado en Geología

Curso: Herramientas Computacionales 2022-2

## OBJETIVOS

1. **Segmentación de un sismograma mediante algoritmos de aprendizaje no supervisado (*unsupervised machine learning*), en este caso *K-MEANS*, y por descriptores estadísticos.**
2. Determinar posibles cambios en el régimen de un sistema por medio de la evolución temporal de las “familias”.
3. Determinar las características mas importantes de cada “familia”.

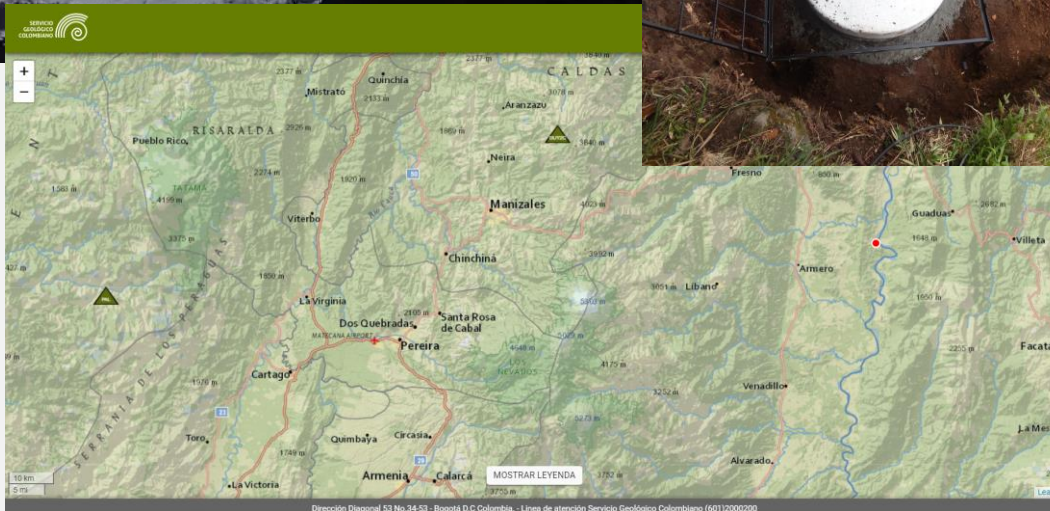
## MOTIVACIÓN

Encontrar patrones en señales sísmicas capaces de brindar un conocimiento mayor de cambios en el régimen de un sistema, con miras a posible aplicación en el estudio de riesgo, en este caso, riesgo por actividad volcánica y sismicidad inducida.



# CASOS DE ESTUDIO

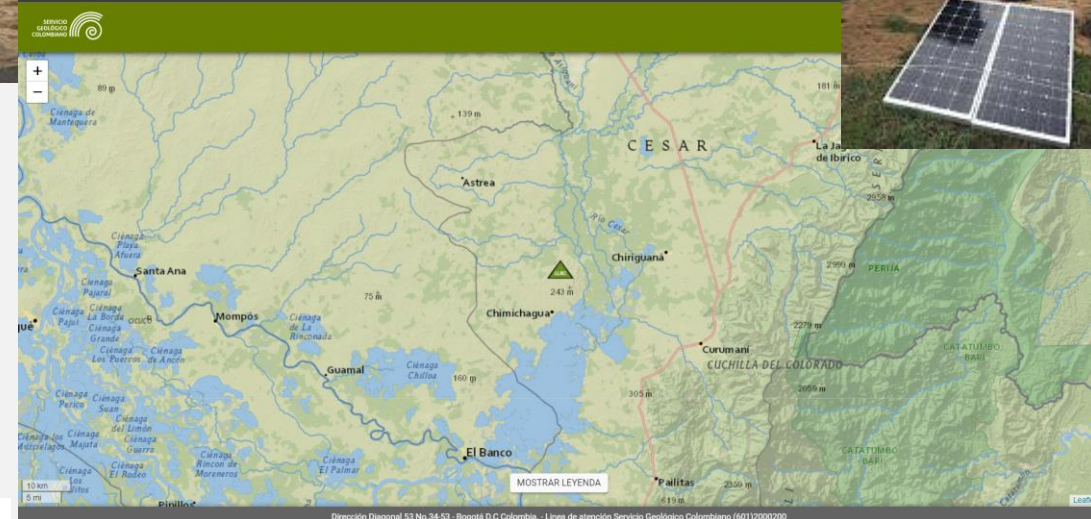
- Nevado del Ruiz (8-14 Nov 2022)



<https://siigeo.sgc.gov.co/stations/detail/374>

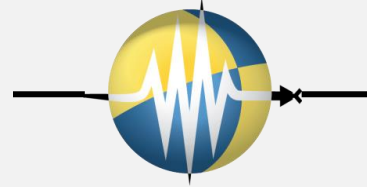
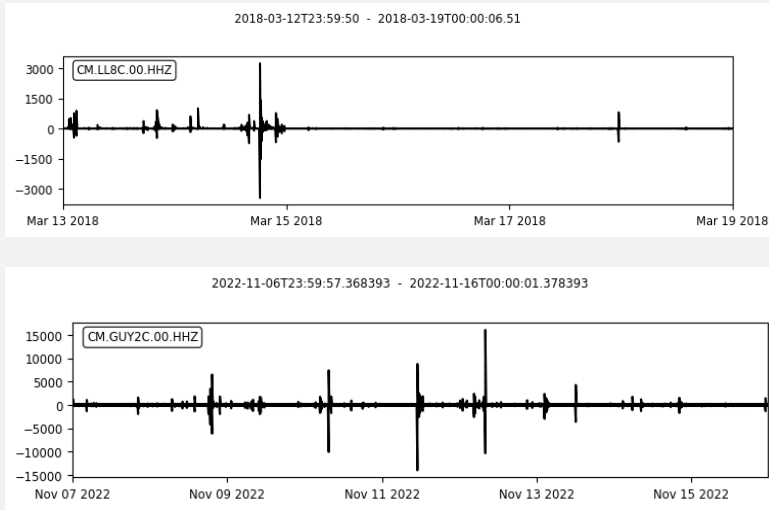


- La Loma, Cesar (14 Marzo 2018)



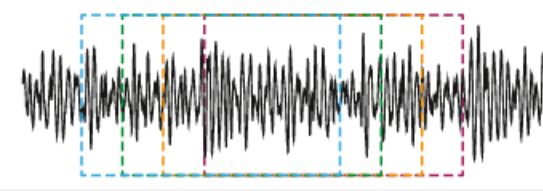
<https://siigeo.sgc.gov.co/stations/detail/386>

## Señal Cruda



1. Se escoge componente HHZ
2. Se filtra el ruido (highpass, 0,1 Hz)

## Señal Segmentada



**window:** Segmenta una serie temporal con un porcentaje de traslape (60,80)

T

Descriptores

1. Desviación estándar
2. Curtosis
3. Asimetría

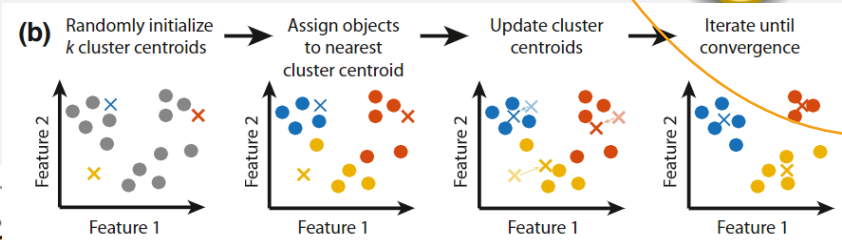
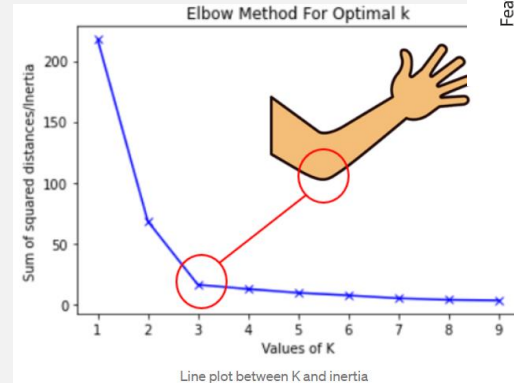
F

1. Pico
2. Asimetría
3. Percentil 50

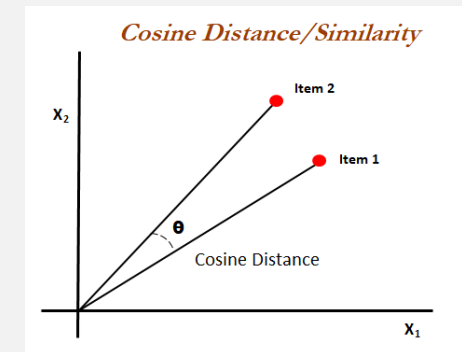


**descrp:** Da una lista de descriptores de la señal basado en : **Watson, L. M. (2020). Using unsupervised machine learning to identify changes in eruptive behavior at Mount Etna, Italy. Journal of Volcanology and Geothermal Research, 405, 107042.**

## Método del "codo"



## Análisis



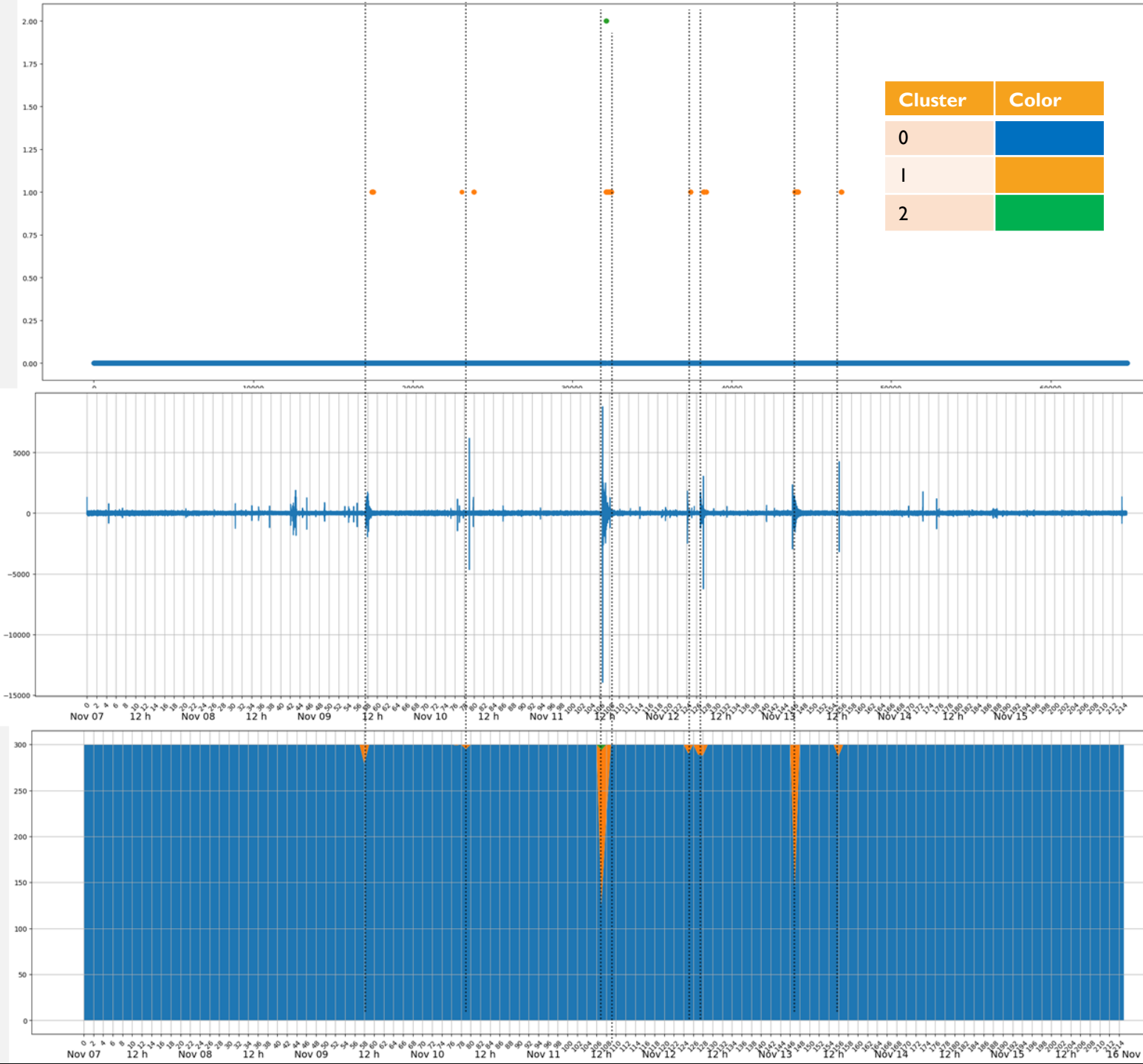
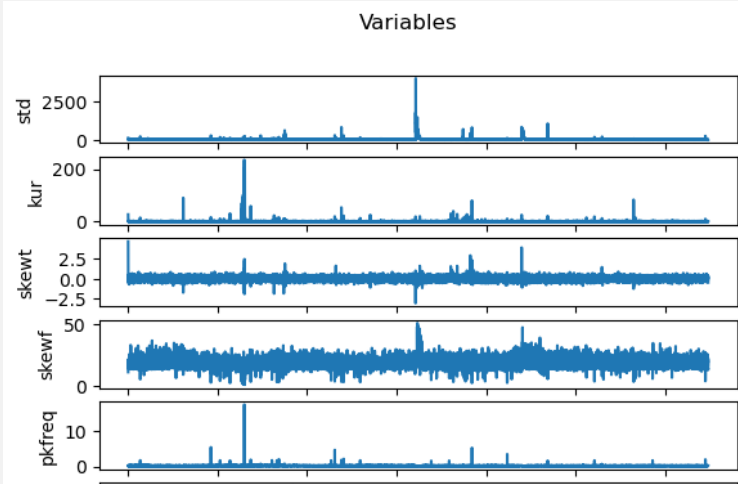
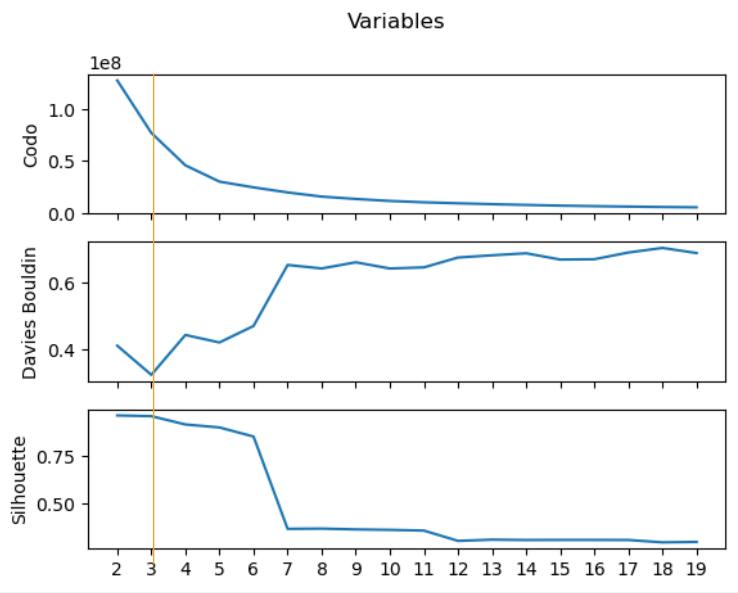
Señales Madre

## Dataframe con Descriptores de cada segmento

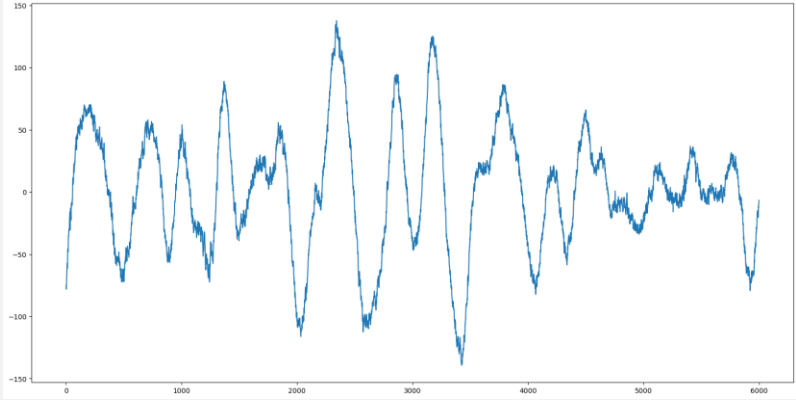
pd.DataFrame					



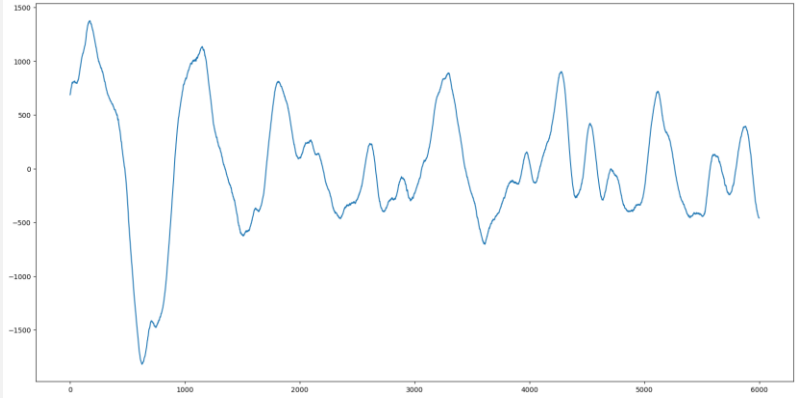
# Caso I: Nevado del Ruiz



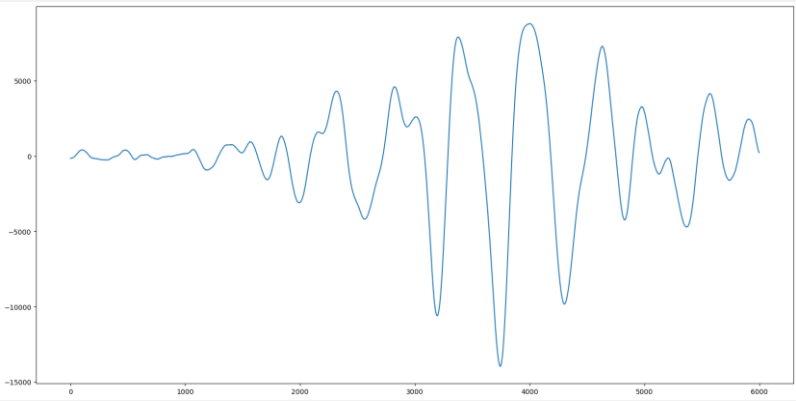
# Caso I: Nevado del Ruiz



Familia 0

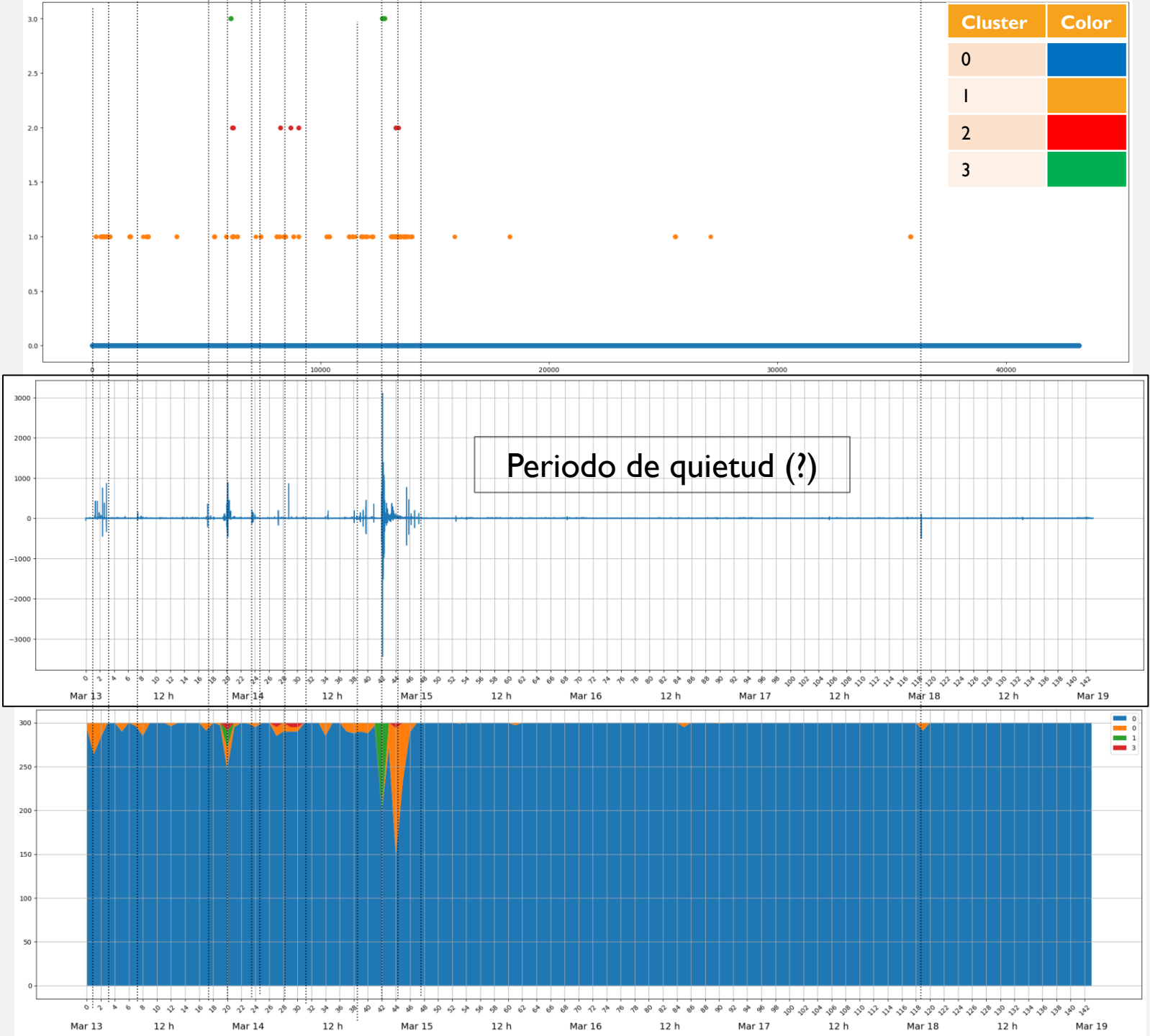
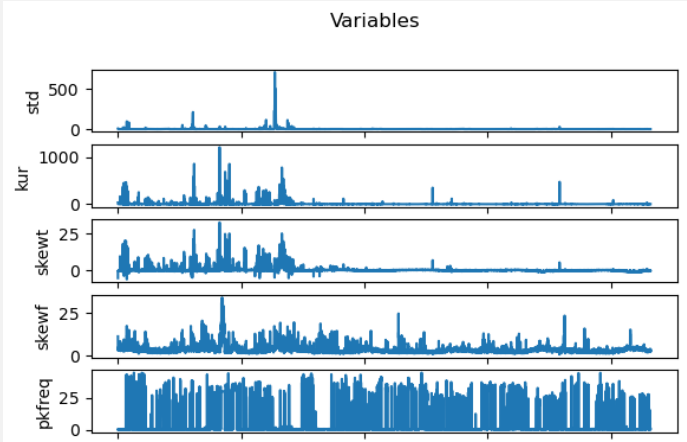
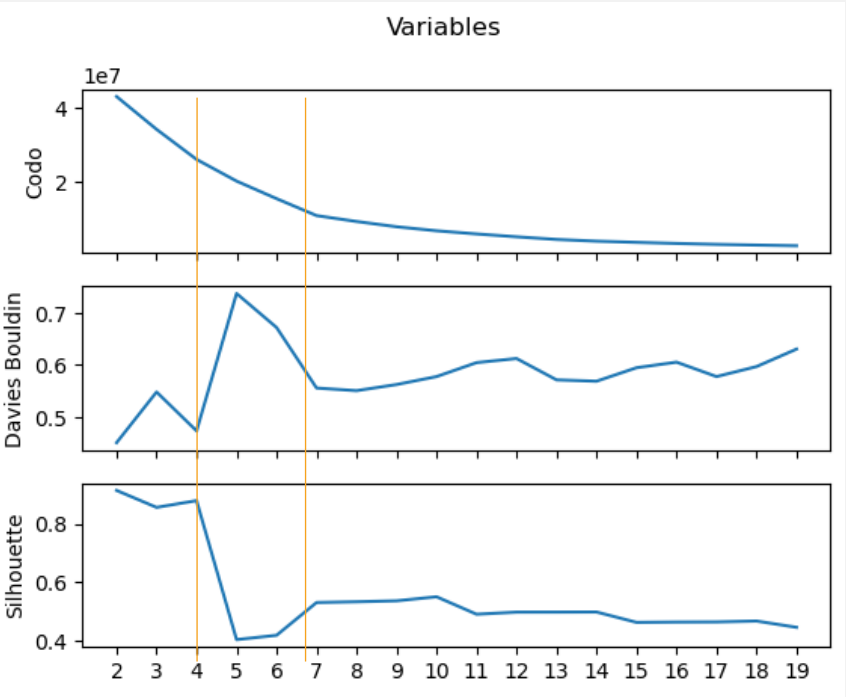


Familia 1

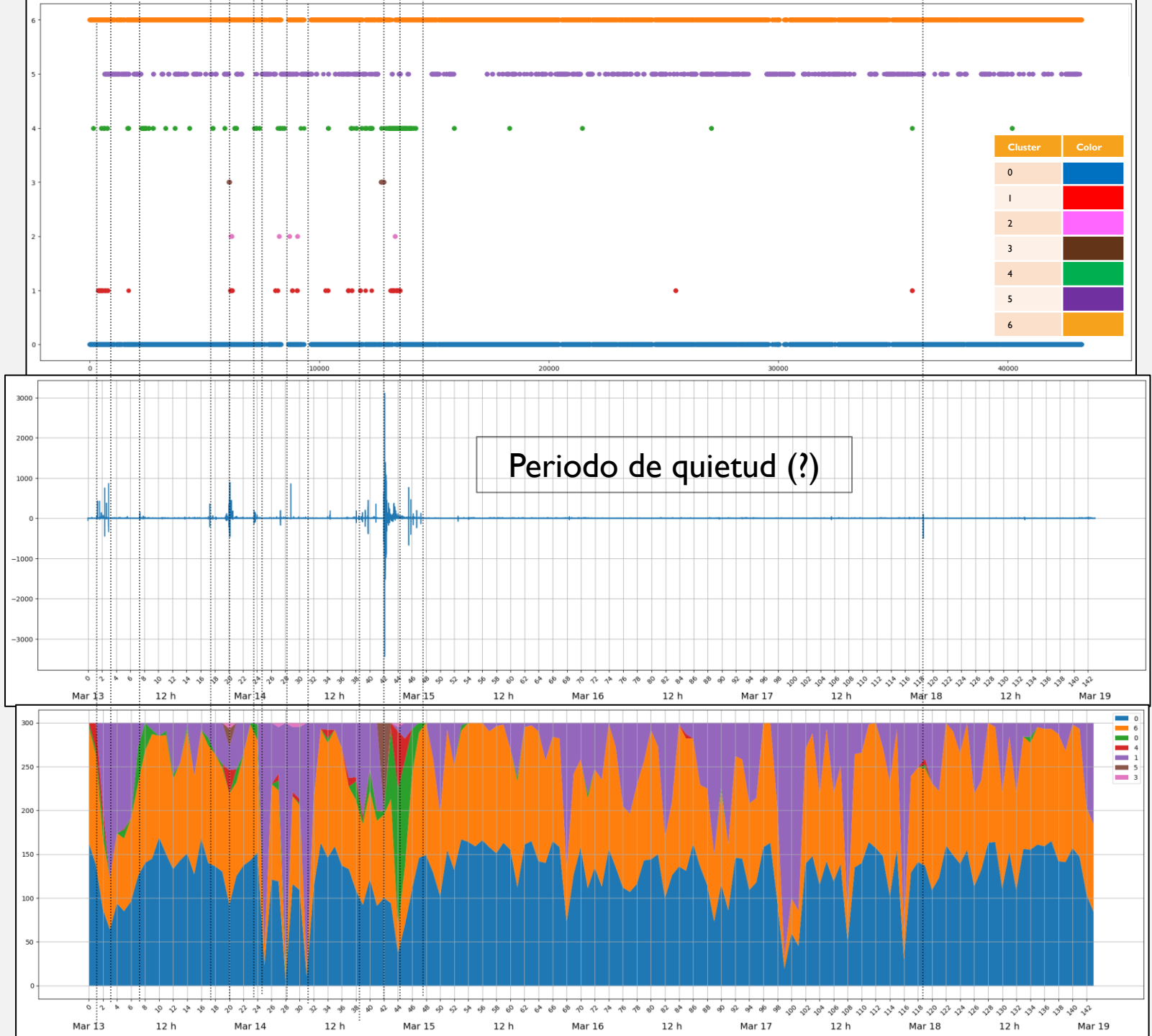
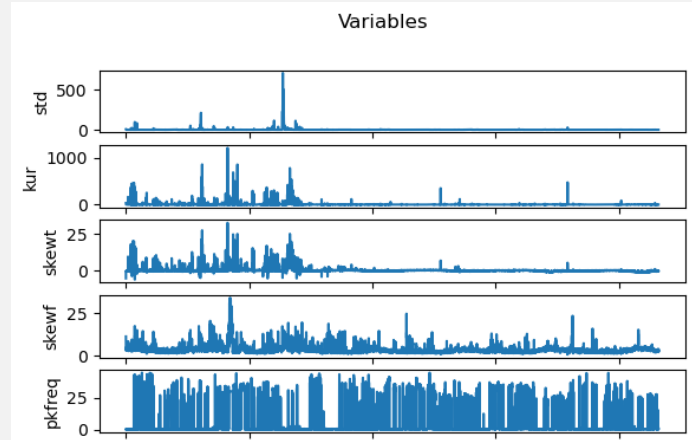
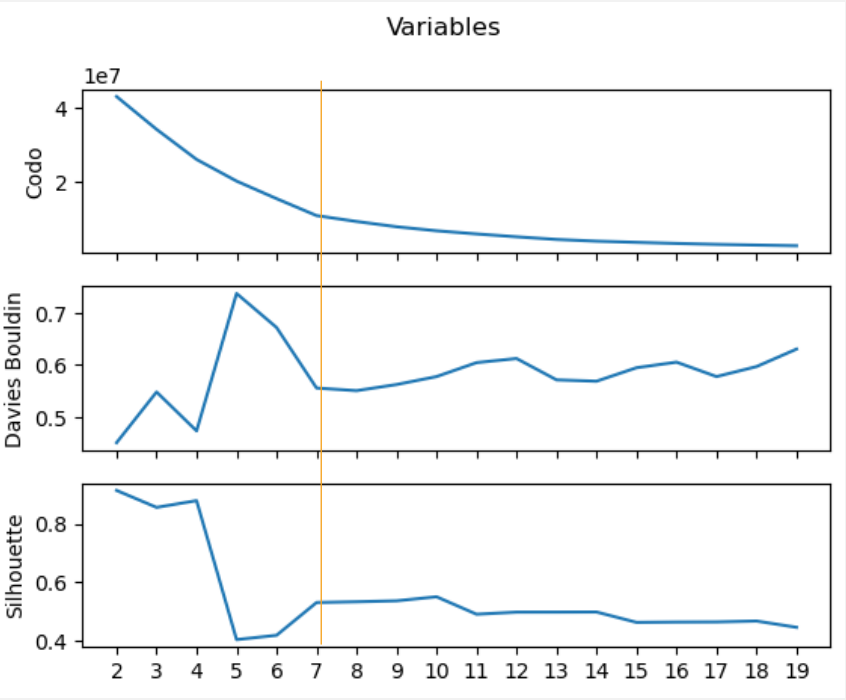


Familia 2

# Caso II(I): La Loma, Cesar

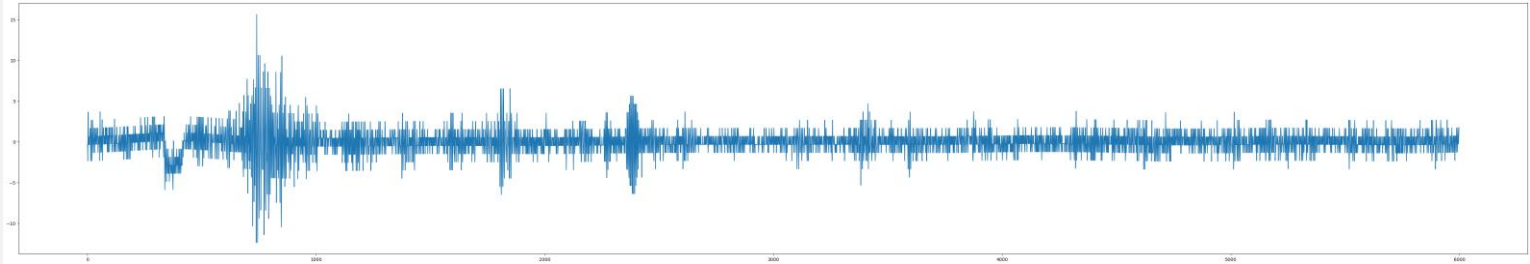


# Caso II(2): La Loma, Cesar

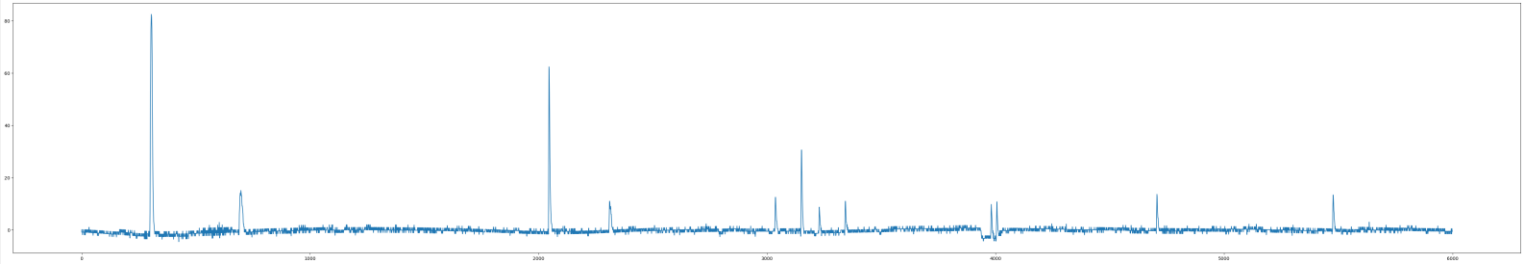




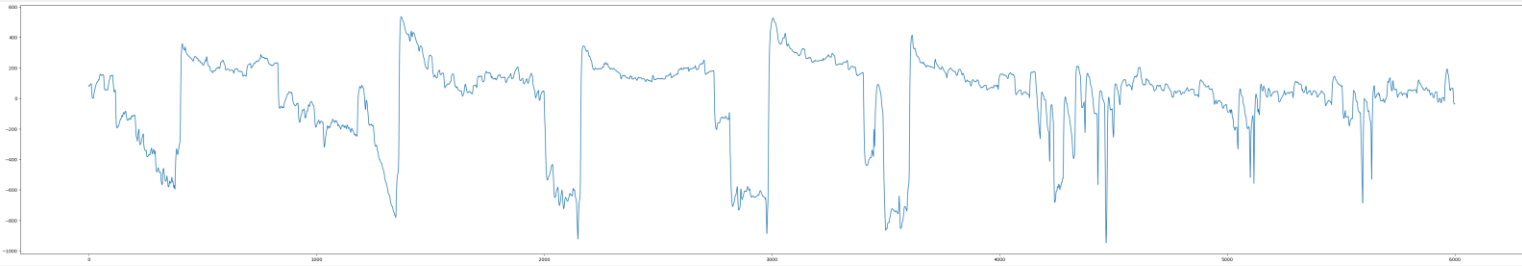
# Caso 11: La Loma, Cesar



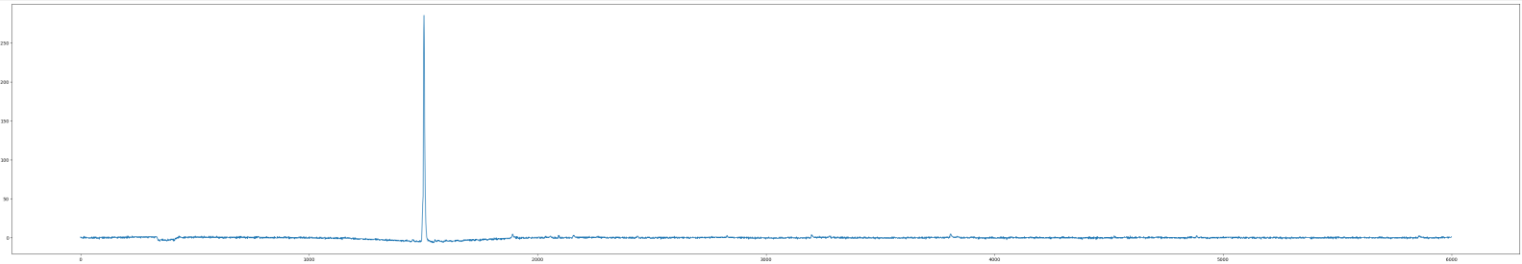
Familia 0



Familia 1



Familia 3



Familia 2

## SIGUIENTES PASOS

1. **Añadir otros descriptores** , por ejemplo, los contemplados en [Cadena Ibarra, O. \(2021\)](#)
2. **Posibilidad de una ventana de tiempo mas extensa, y una segmentación menor (mayor capacidad de computo).**
3. **Analizar de manera mas profunda las señales “madre” de cada familia**
4. **Correlacionar con eventos mas discretos (sismos, inyección de agua, flujos piroclásticos, etc).**
5. **Efectuar en otras zonas con calidad de datos mejor**
6. **Evaluar una ventana de frecuencias dependiendo el tipo de estudio**

## CONCLUSIONES

1. Se observan patrones los suficientemente diferentes para indicar anomalías a nivel sísmico.
2. Las señales cuentan con errores de carácter instrumental, los cuales pueden afectar la agrupación.
3. En algunos casos, hay discrepancia entre el método codo y los demás.

## REFERENCIAS

- <https://towardsdatascience.com/cheat-sheet-to-implementing-7-methods-for-selecting-optimal-number-of-clusters-in-python-898241e1d6ad>
- <http://sismo.sgc.gov.co:8080/fdsnws/dataselect/1/builder>
- Watson, L. M. (2020). Using unsupervised machine learning to identify changes in eruptive behavior at Mount Etna, Italy. *Journal of Volcanology and Geothermal Research*, 405, 107042.
- Cadena Ibarra, O. (2021). *Modelos de fuente de sismicidad LP para la actividad del volcán Galeras 2004-2010 (Colombia)*. Universidad Nacional de Colombia.
- <https://stackoverflow.com/questions/18424228/cosine-similarity-between-2-number-lists>