



Universidad Nacional de Asunción

# Arroyo Mburicão

Análisis de datos de precipitación y nivel para modelos  
de machine learning



Mathias Aguilar



Héctor Velázquez



Dr. Diego Stalder

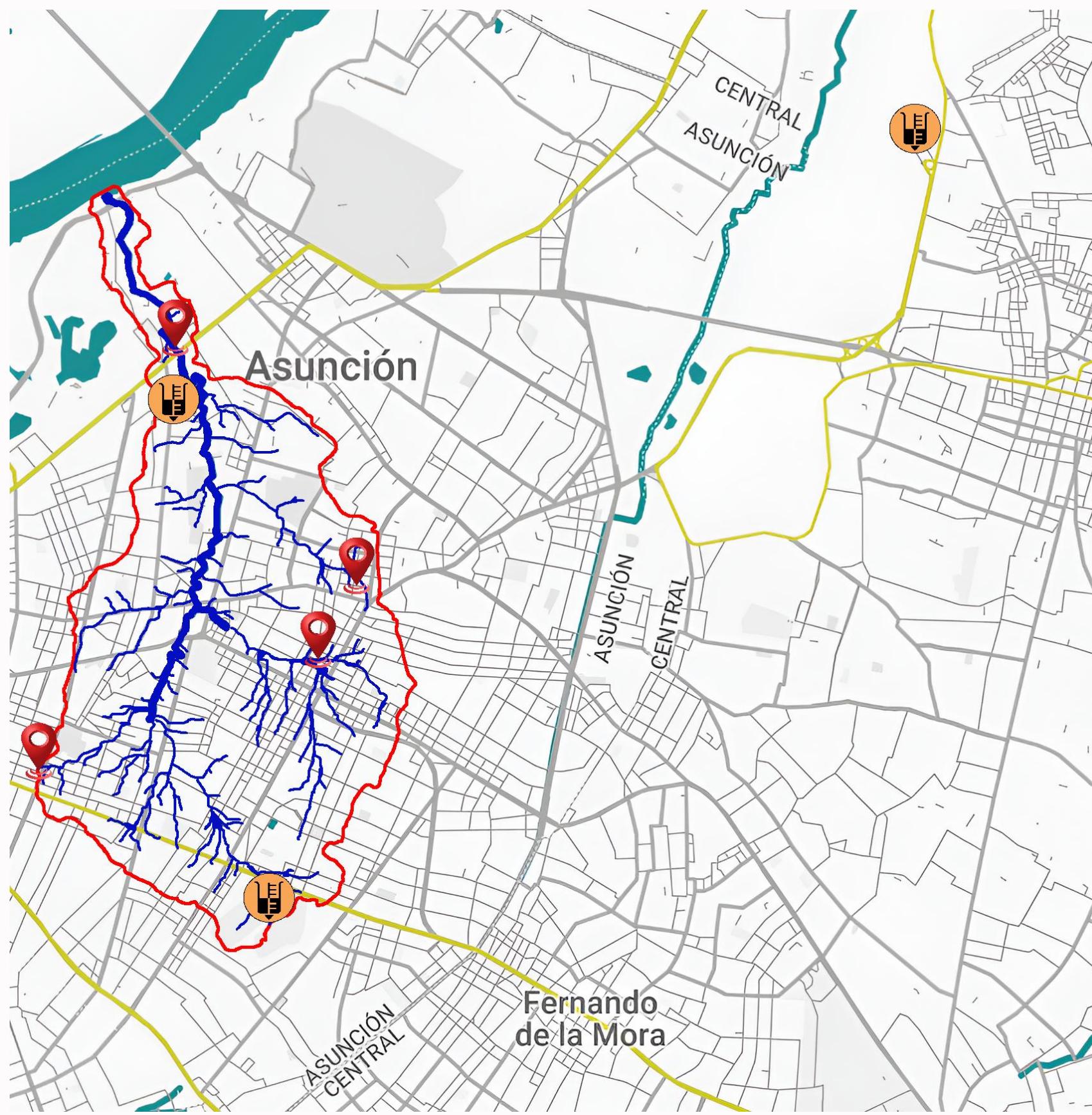


Dr. Andrés Wehrle



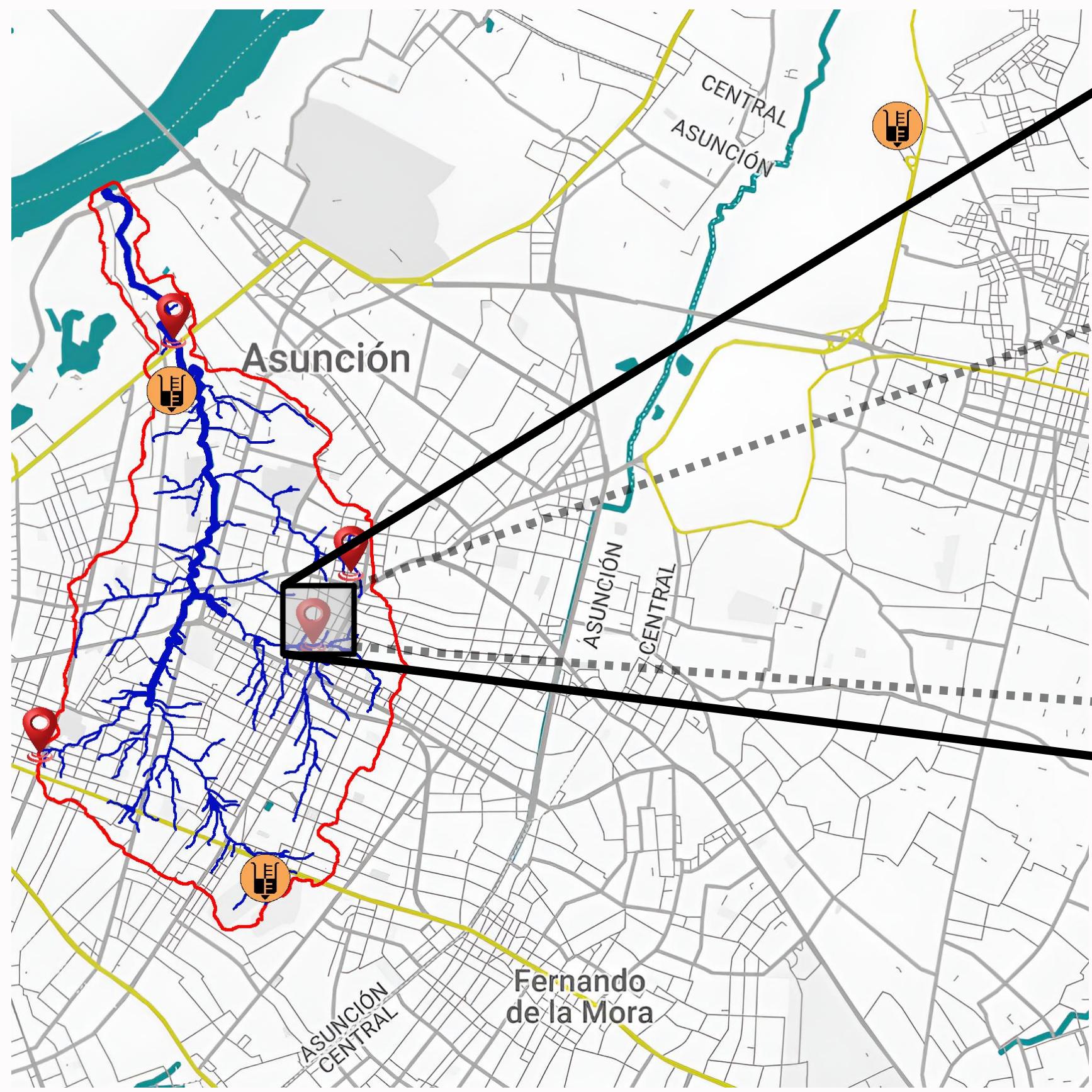
# Importancia

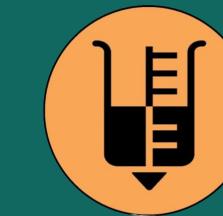
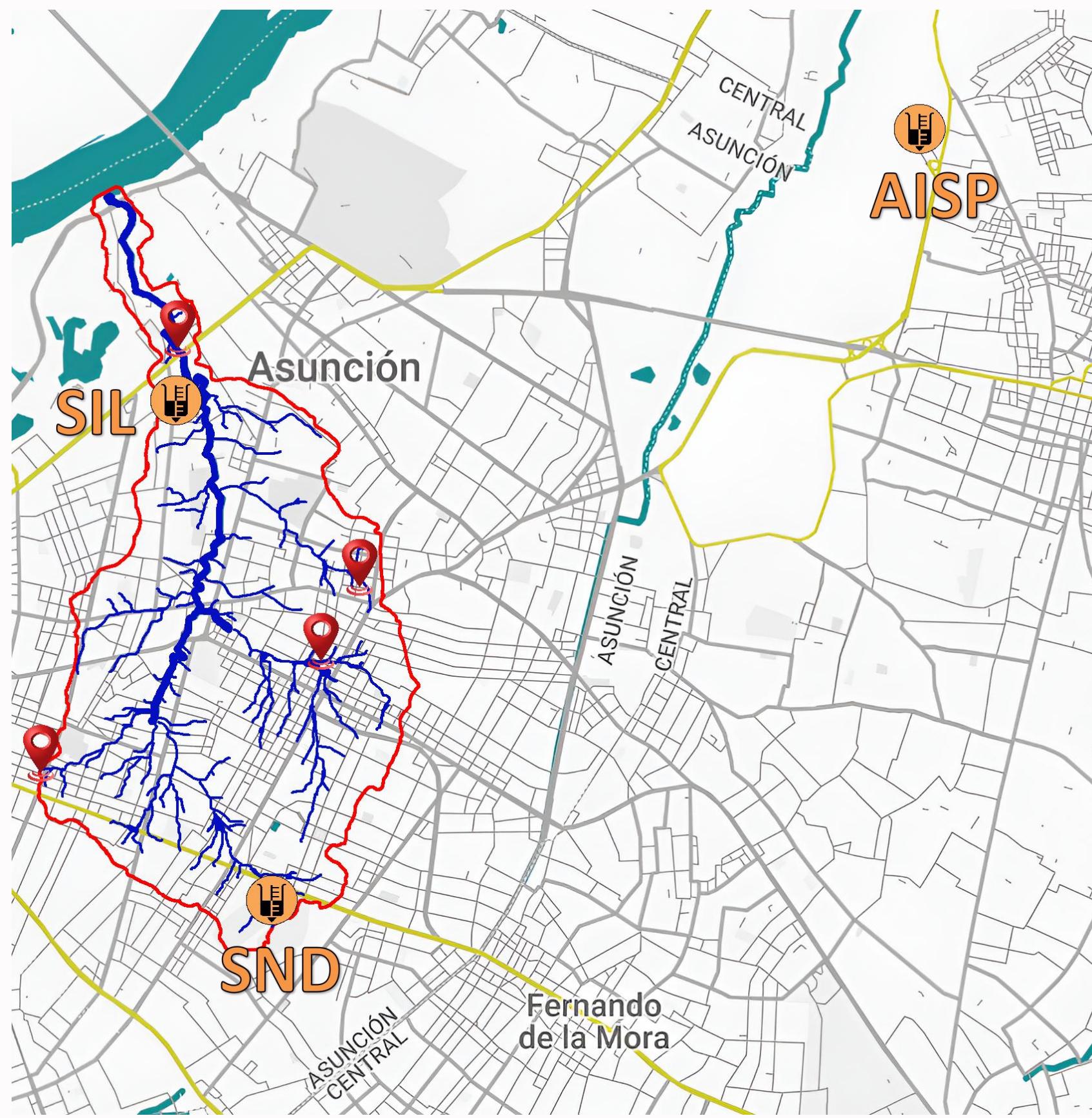
El arroyo Mburicao tiene un recorrido de 16.25 kilómetros, este cruza 7 barrios de Asunción (la capital).



# Puntos críticos

- Gral. Santos y San Antonio
- Perú y Artigas
- Montevideo entre 12 y 13  
proyectada
- 21 Proyectadas y Antequera
- Virgen del Carmen t Sargento  
Marecos
- Rca. Argentina y Caaguazú
- Eusebio Ayala y Boggiani
- Artigas y Mburicao
- Virgen del Carmen y Sargento  
Marecos
- Santa Ana y Roque Miranda
- Fdo. de la Mora y la Victoria





# Estaciones

- San Inacio de Loyola (**SIL**)
- Secretaría Nacional de Deporte (**SND**)
- Aeropuerto International Silvio Pettirossi (**AISP**)

# Etapas del proyecto





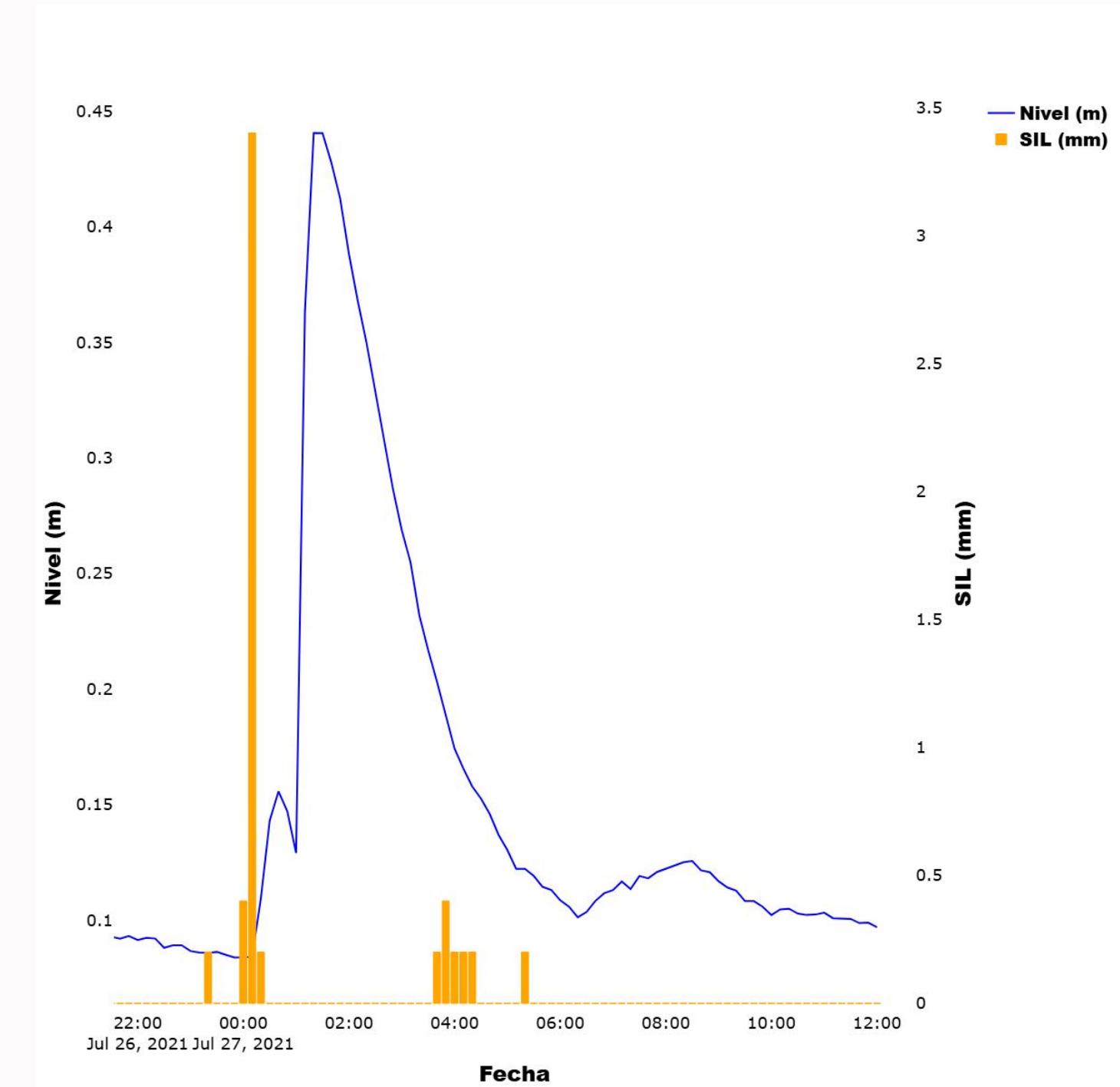
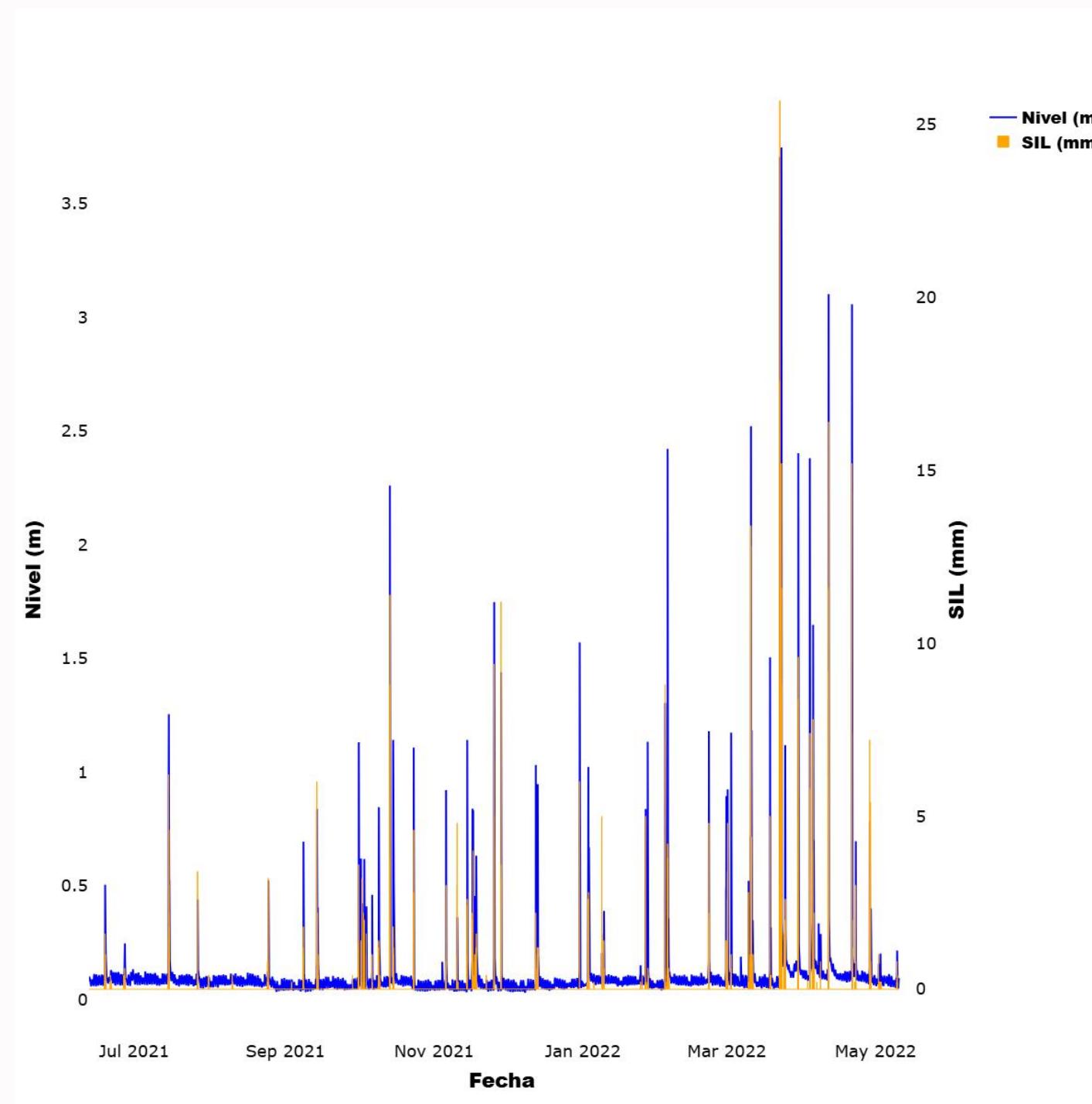
- Visualización de los Datos
- Relevancias de los Datos



*Recolección y  
Análisis de los  
Datos*



# SIL con Nivel

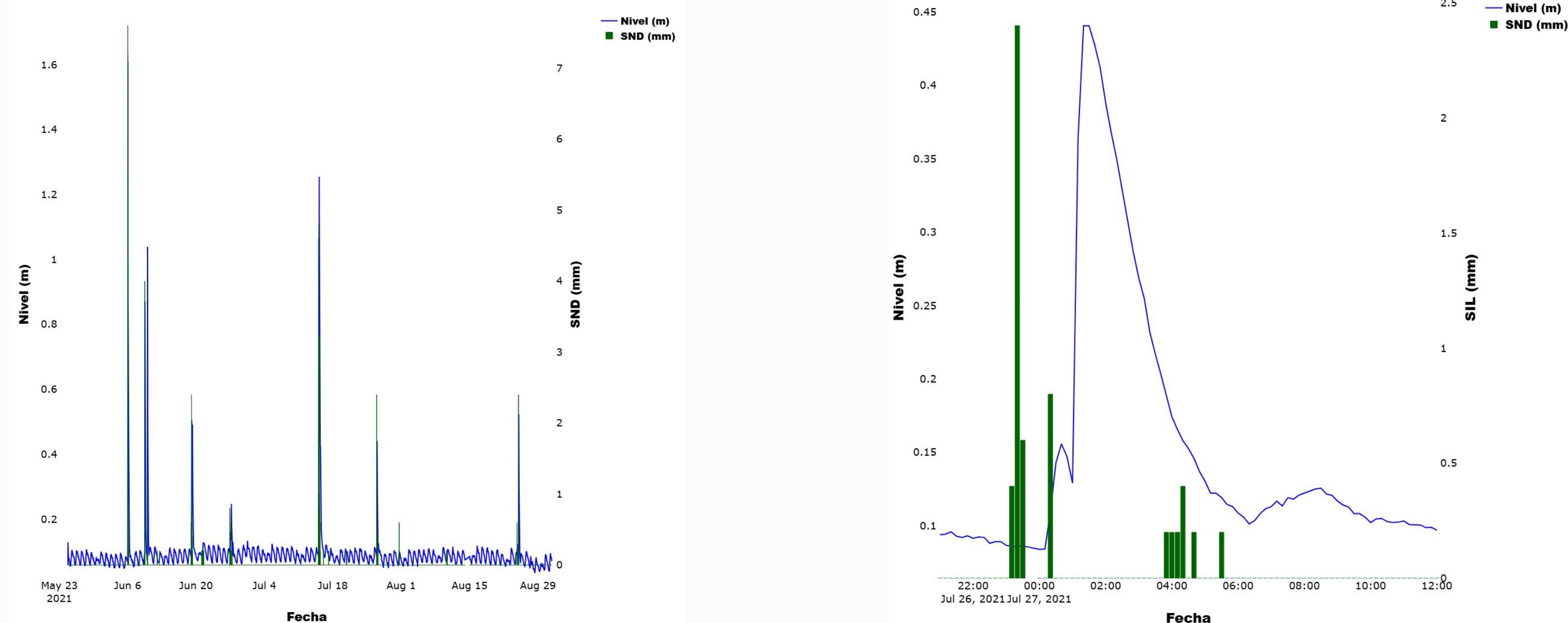


- SIL en milímetros
- Frecuencia de 10 minutos

- Zoom



# SND con Nivel

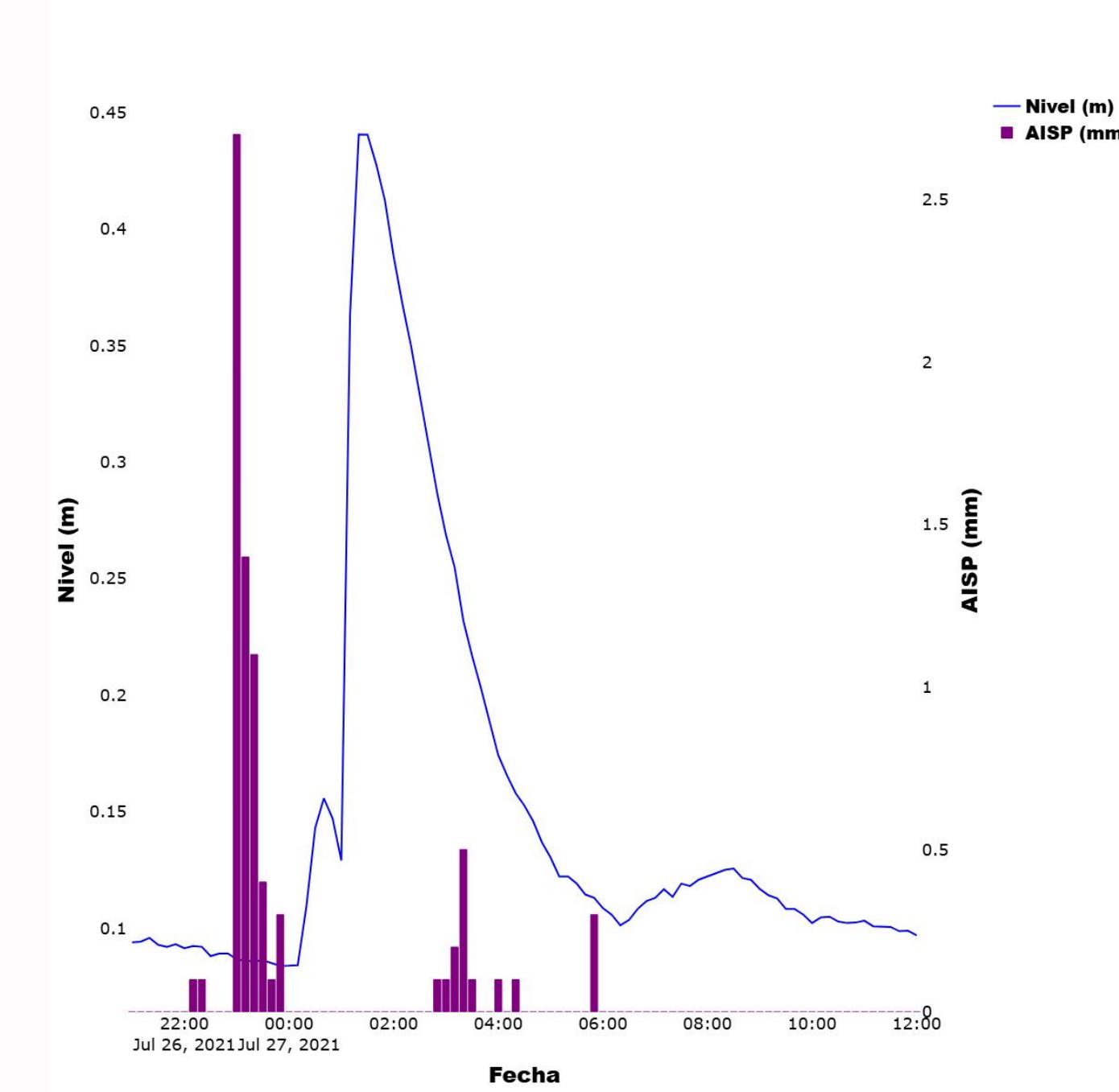
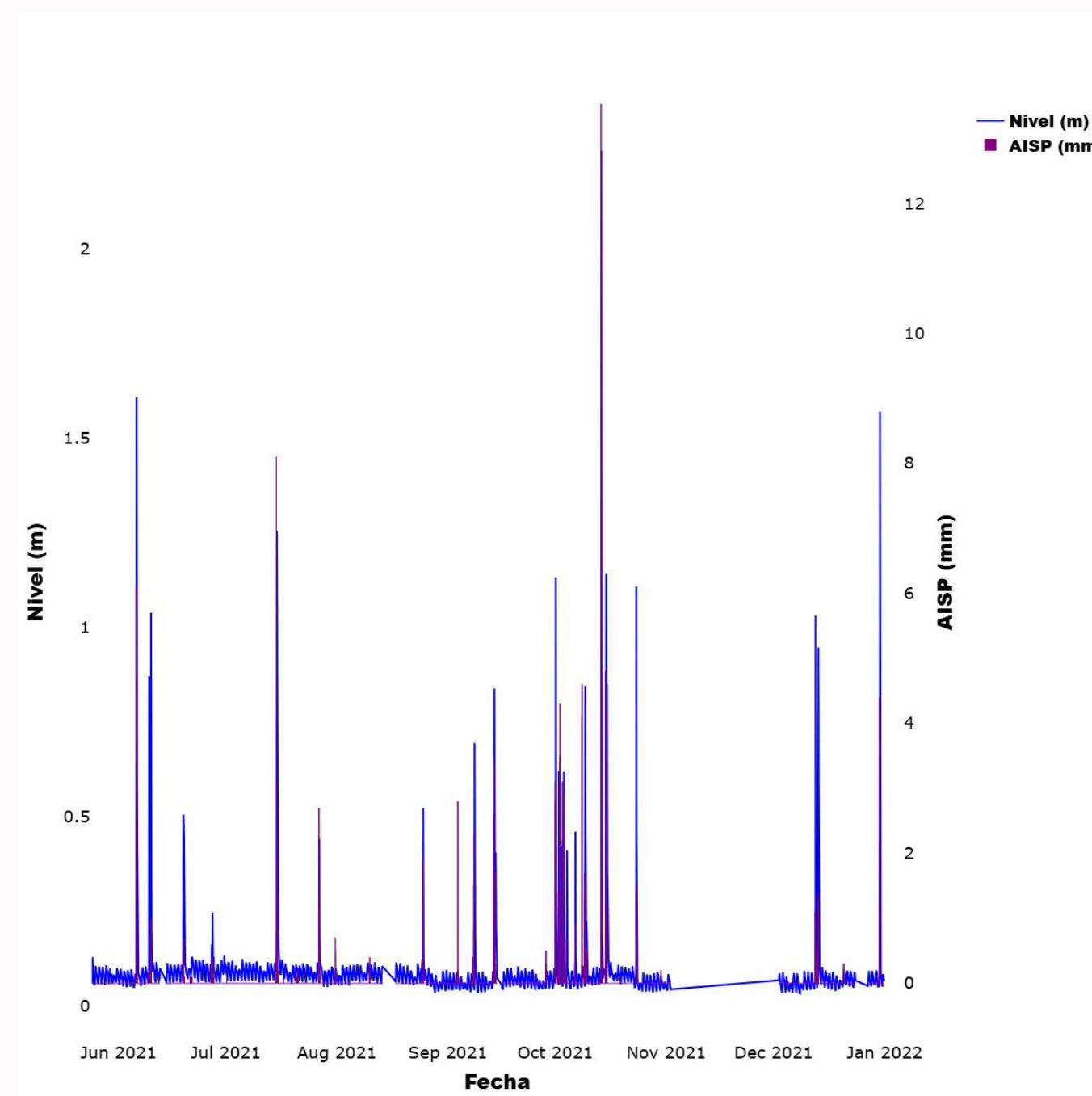


- SND en milímetros
- Frecuencia de 10 minutos

- Zoom



# AISP con Nivel

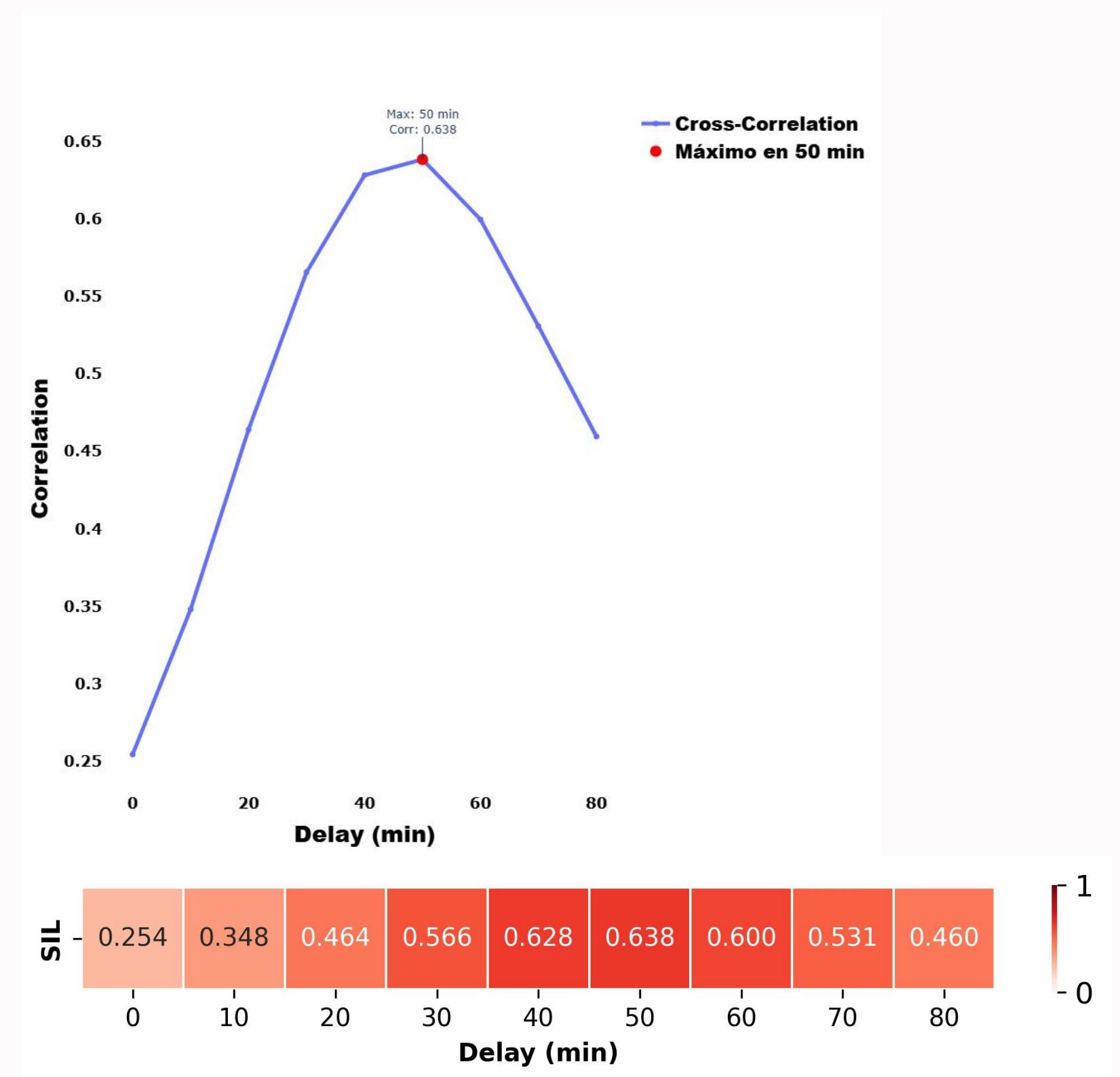


- AISP en milímetros
- Frecuencia de 10 minutos

- Zoom



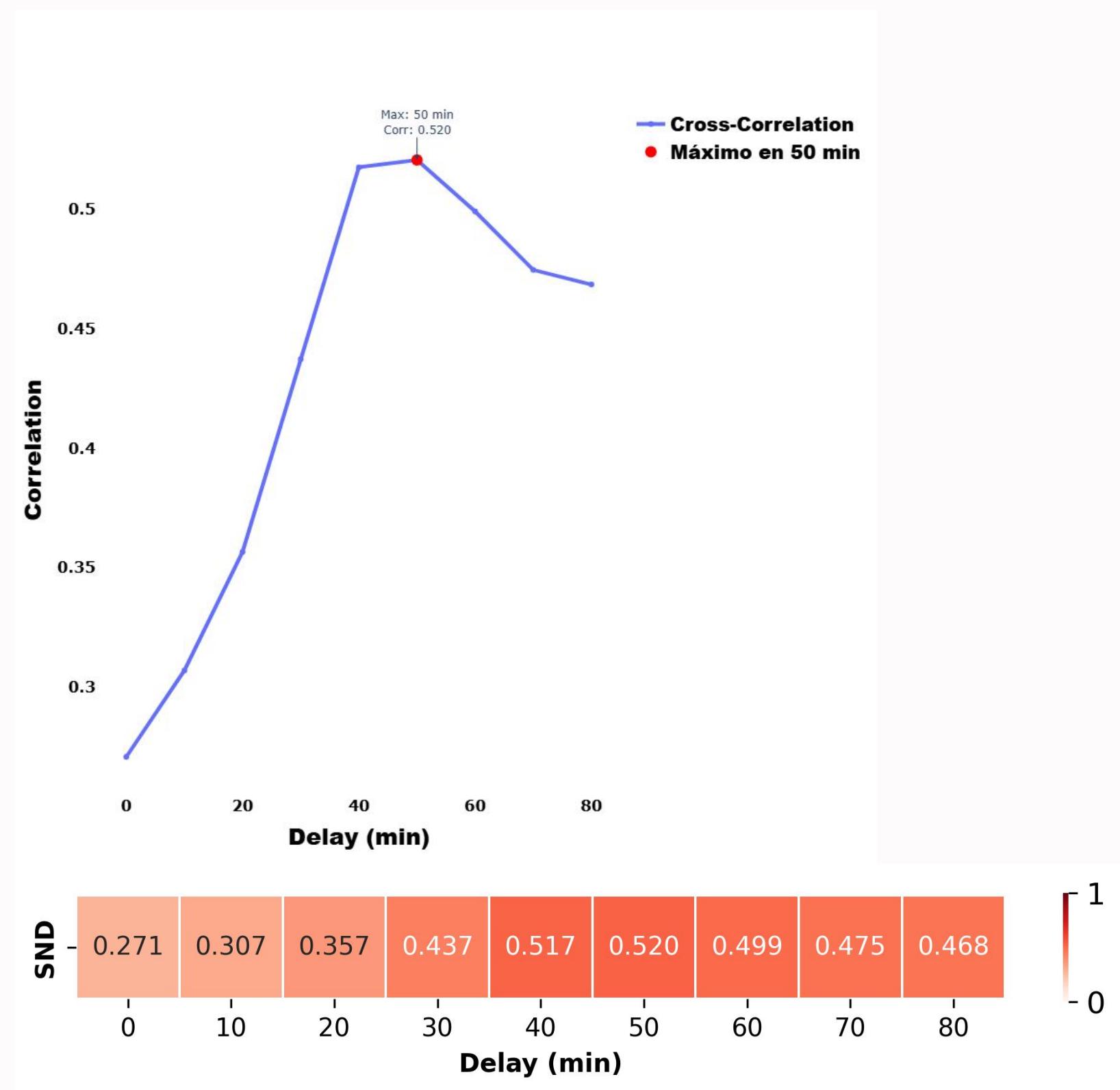
# Correlación Cruzada



Delay (min)	SIL	SND	AISP
0	0.2544	0.2709	0.1441
10	0.3480	0.3070	0.1471
20	0.4641	0.3566	0.1520
30	0.5656	0.4372	0.1590
40	0.6282	0.5174	0.1661
<b>50</b>	<b>0.6384</b>	0.5204	0.1751
60	0.5996	0.4989	0.1925
70	0.5308	0.4745	0.2231
80	0.4696	0.4683	0.2532



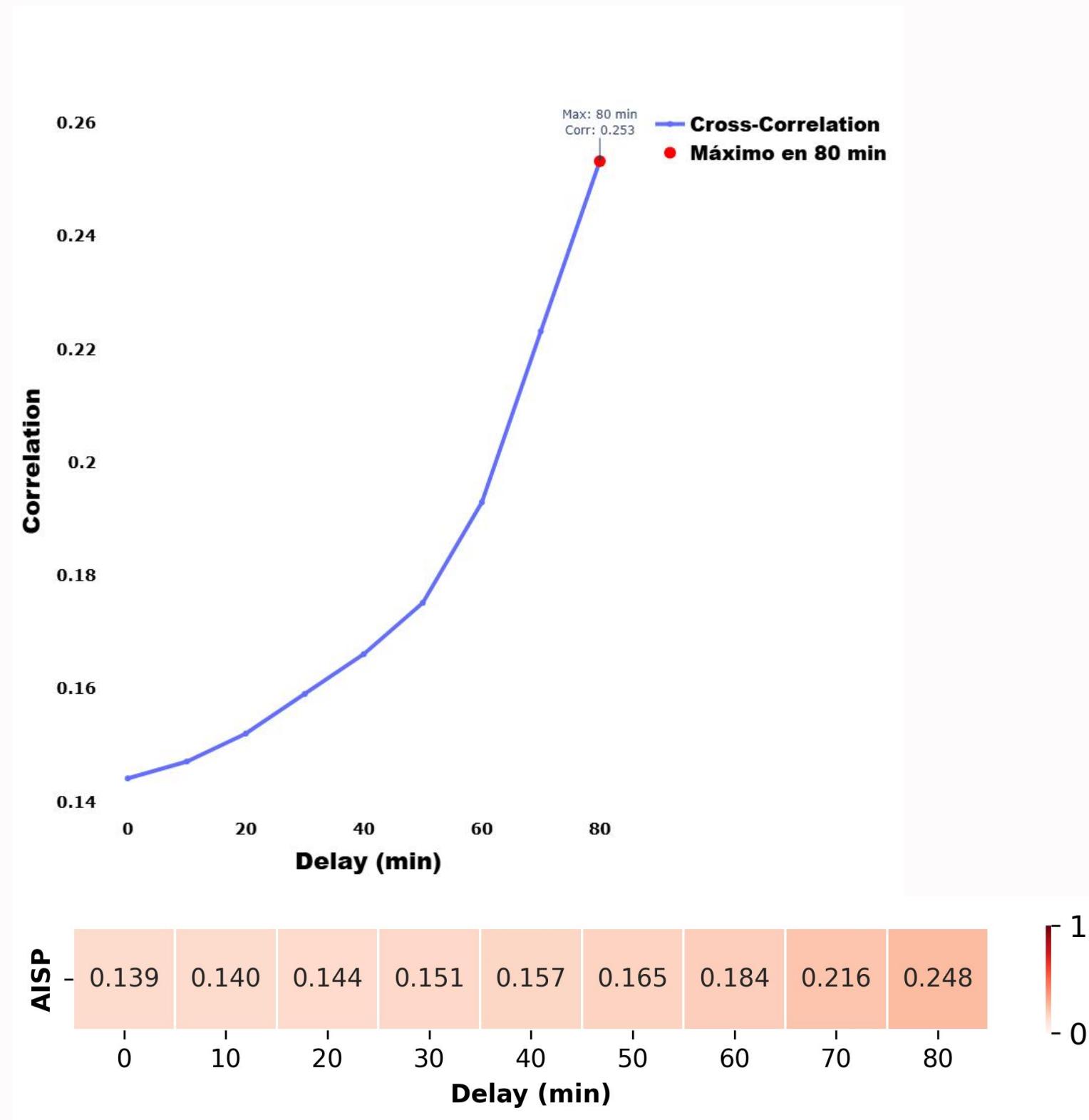
# Correlación Cruzada



Delay (min)	SIL	SND	AISP
0	0.2544	0.2709	0.1441
10	0.3480	0.3070	0.1471
20	0.4641	0.3566	0.1520
30	0.5656	0.4372	0.1590
40	0.6282	0.5174	0.1661
50	0.6384	0.5204	0.1751
60	0.5996	0.4989	0.1925
70	0.5308	0.4745	0.2231
80	0.4696	0.4683	0.2532



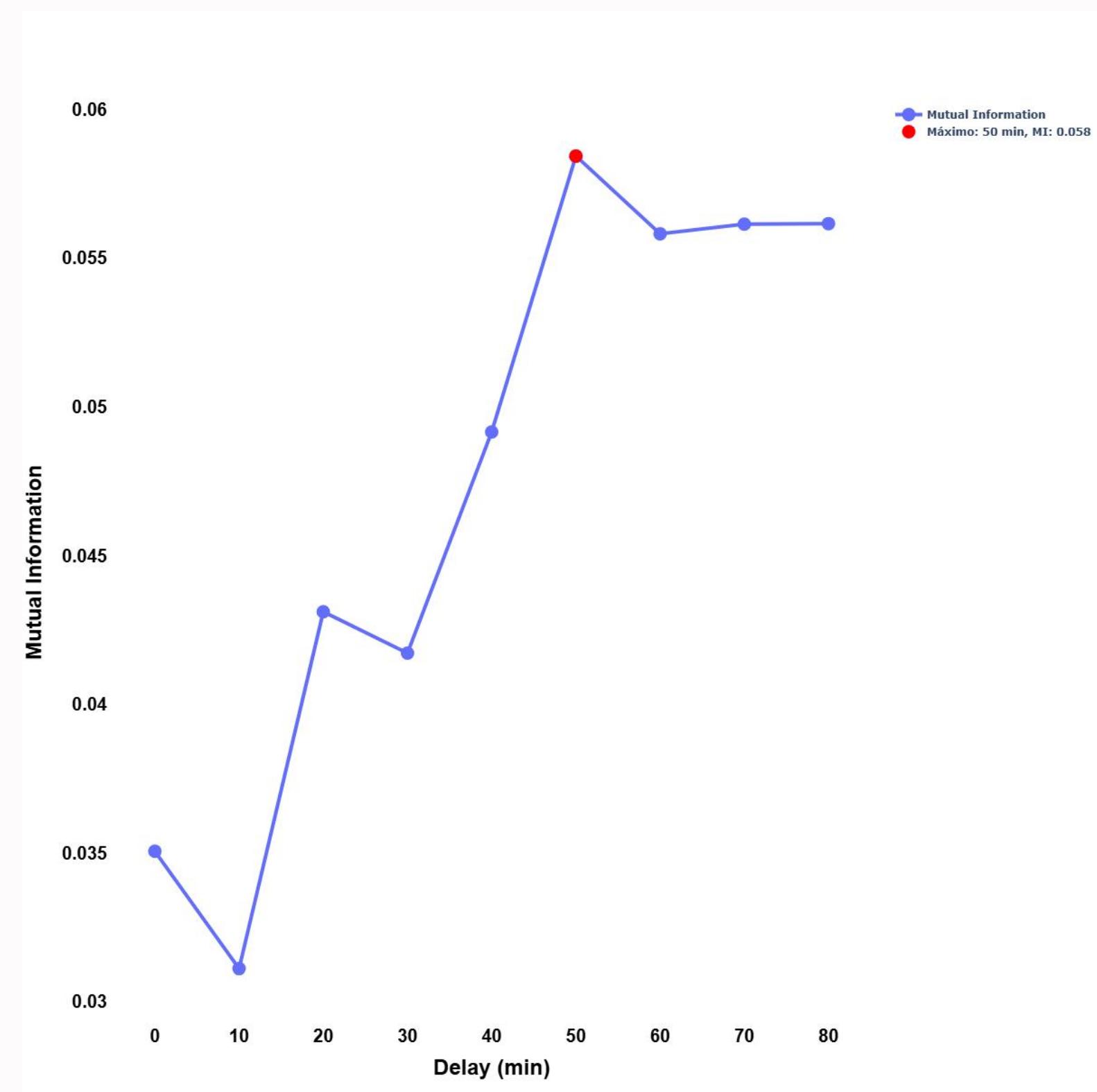
# Correlación Cruzada



Delay (min)	SIL	SND	AISP
0	0.2544	0.2709	0.1441
10	0.3480	0.3070	0.1471
20	0.4641	0.3566	0.1520
30	0.5656	0.4372	0.1590
40	0.6282	0.5174	0.1661
50	0.6384	0.5204	0.1751
60	0.5996	0.4989	0.1925
70	0.5308	0.4745	0.2231
80	0.4696	0.4683	0.2532



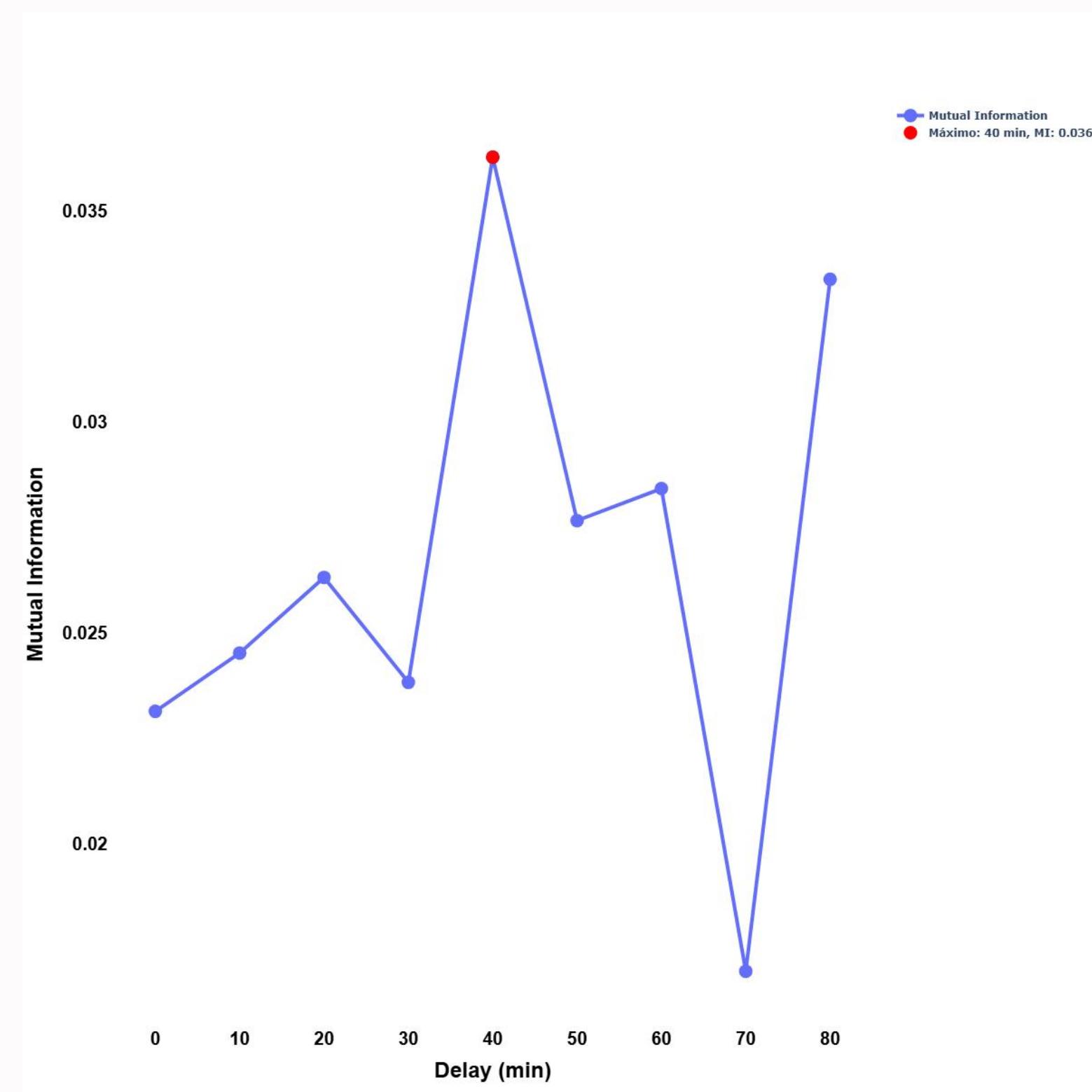
# Información Mutua



Delay (min)	SIL	SND	AISP
0	0.0350	0.0231	0.0146
10	0.0311	0.0254	0.0152
20	0.0431	0.0263	0.0128
30	0.0417	0.0238	0.0163
40	0.0491	0.0362	0.0217
<b>50</b>	<b>0.0584</b>	0.0276	0.0198
60	0.0558	0.0284	0.0207
70	0.0561	0.0169	0.0225
80	0.0561	0.0333	0.0340



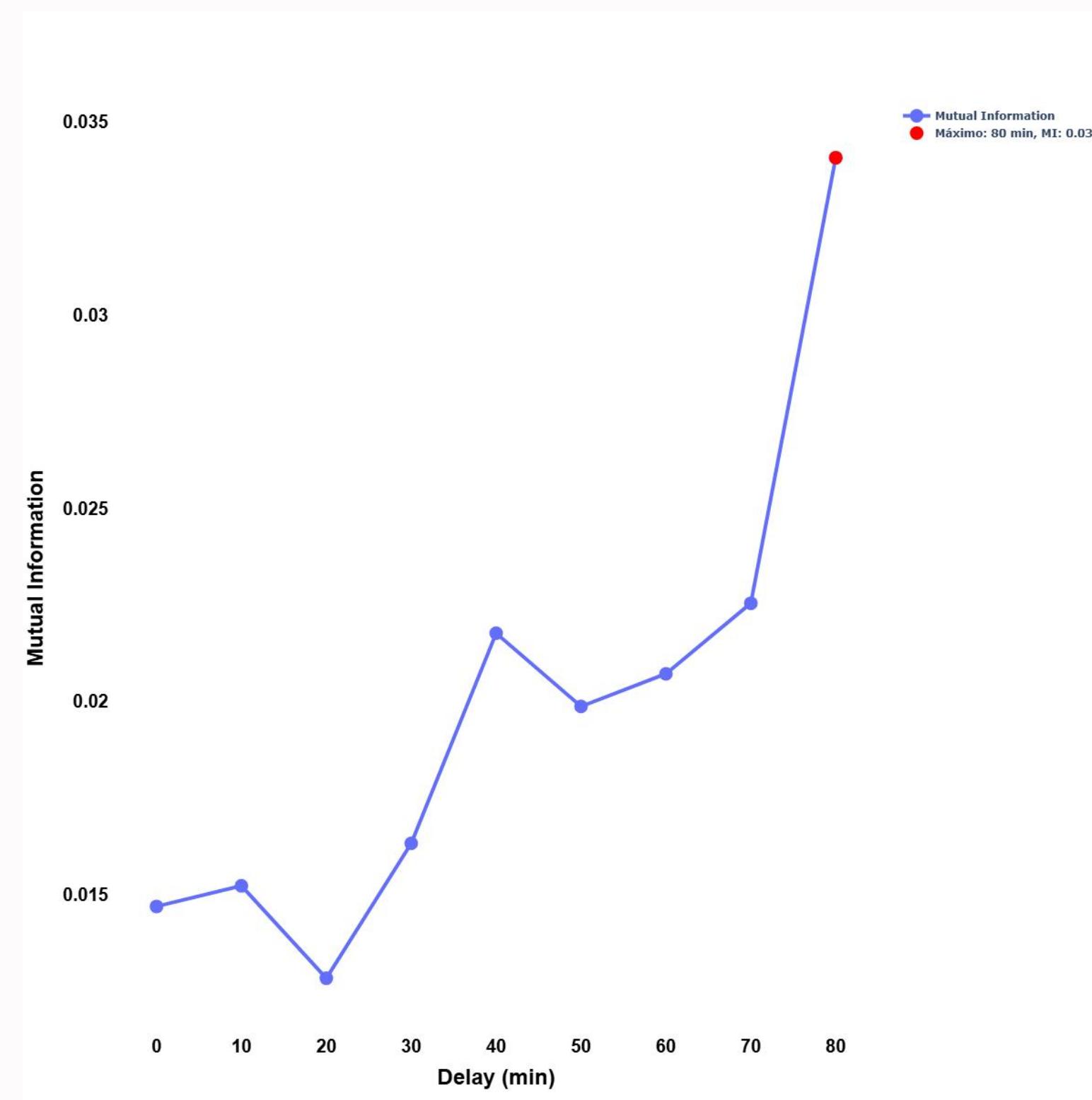
# Información Mutua



Delay (min)	SIL	SND	AISP
0	0.0350	0.0231	0.0146
10	0.0311	0.0254	0.0152
20	0.0431	0.0263	0.0128
30	0.0417	0.0238	0.0163
<b>40</b>	<b>0.0491</b>	<b>0.0362</b>	0.0217
50	0.0584	0.0276	0.0198
60	0.0558	0.0284	0.0207
70	0.0561	0.0169	0.0225
80	0.0561	0.0333	0.0340



# Información Mutua



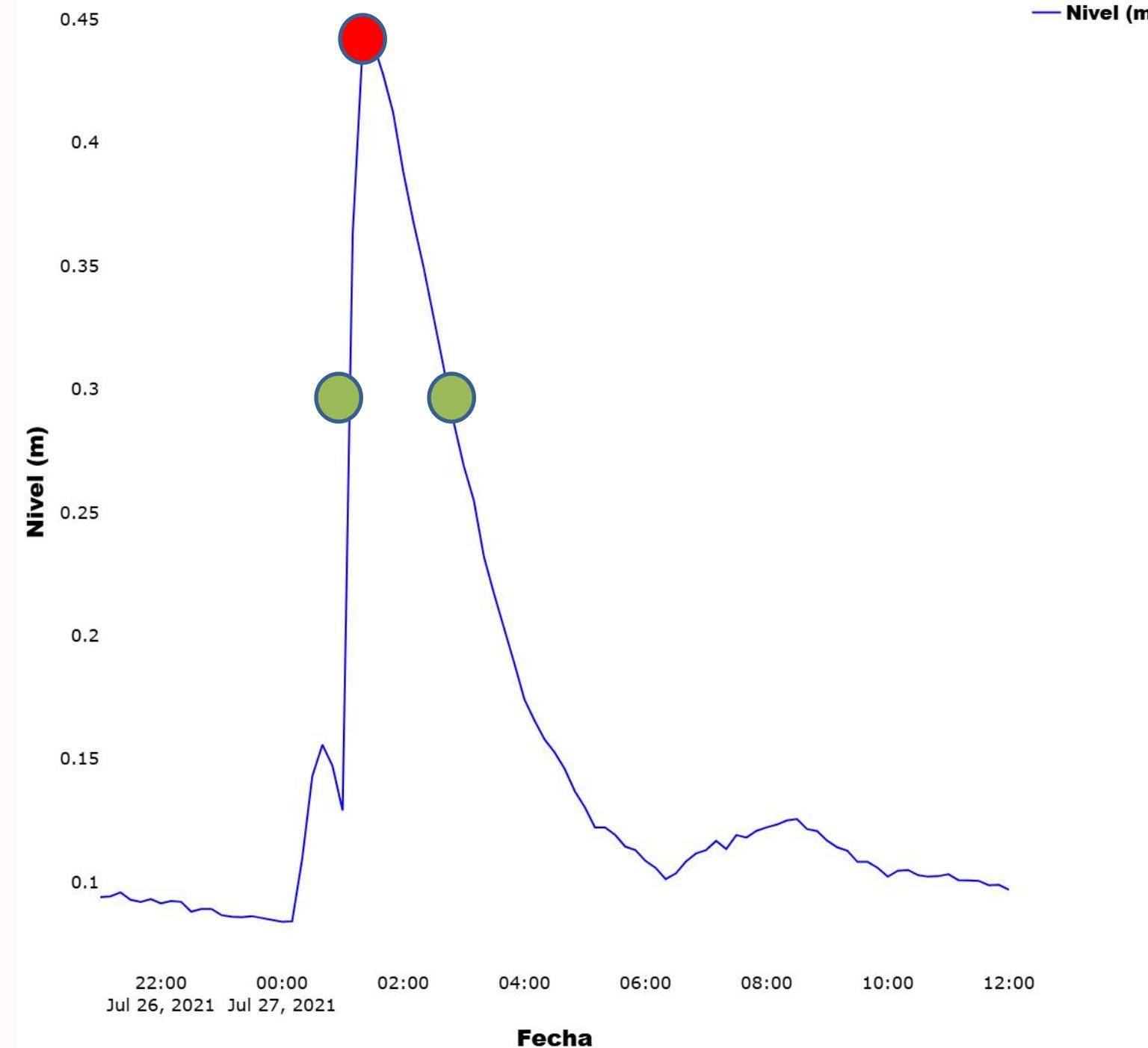
Delay (min)	SIL	SND	AISP
0	0.0350	0.0231	0.0146
10	0.0311	0.0254	0.0152
20	0.0431	0.0263	0.0128
30	0.0417	0.0238	0.0163
40	0.0491	0.0362	0.0217
50	0.0584	0.0276	0.0198
60	0.0558	0.0284	0.0207
70	0.0561	0.0169	0.0225
<b>80</b>	<b>0.0561</b>	<b>0.0333</b>	<b>0.0340</b>

- Características Relevantes
- Definir Posibles Datos de Entradas y Salidas

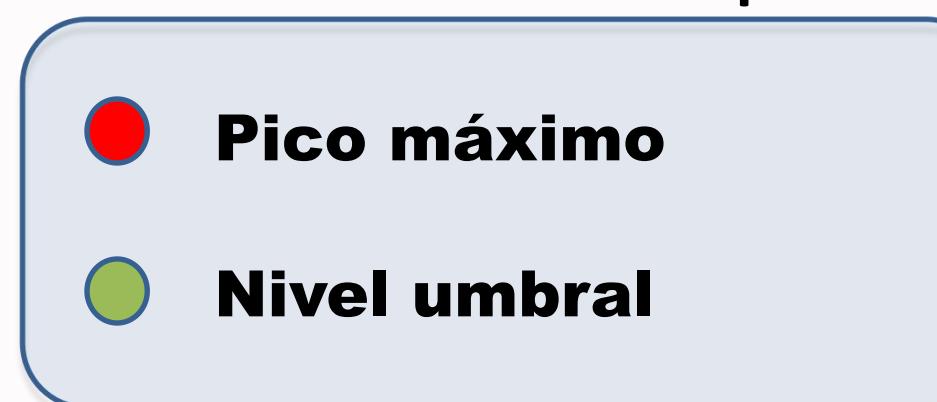


## Extracción de Características

# Eventos

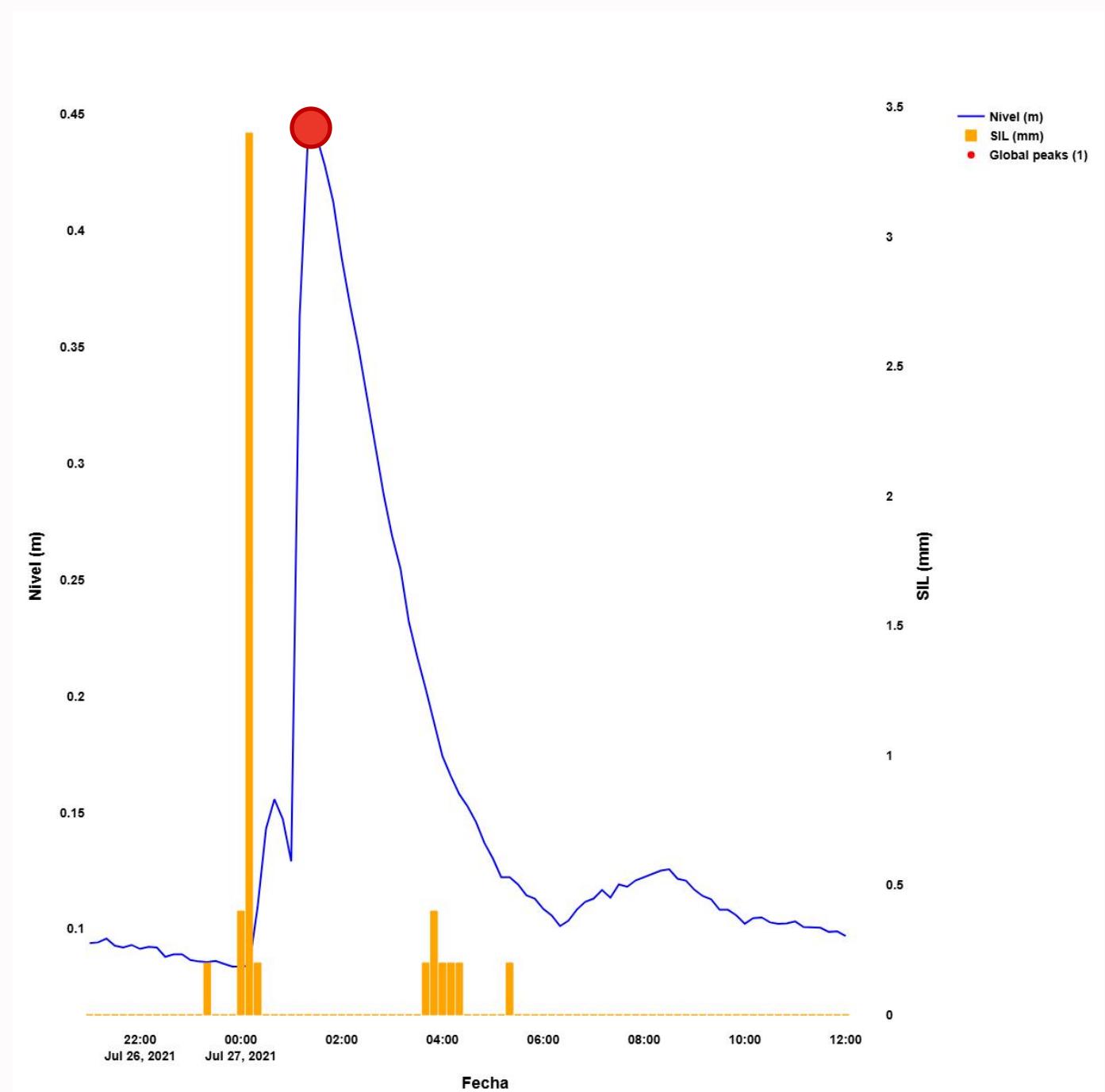
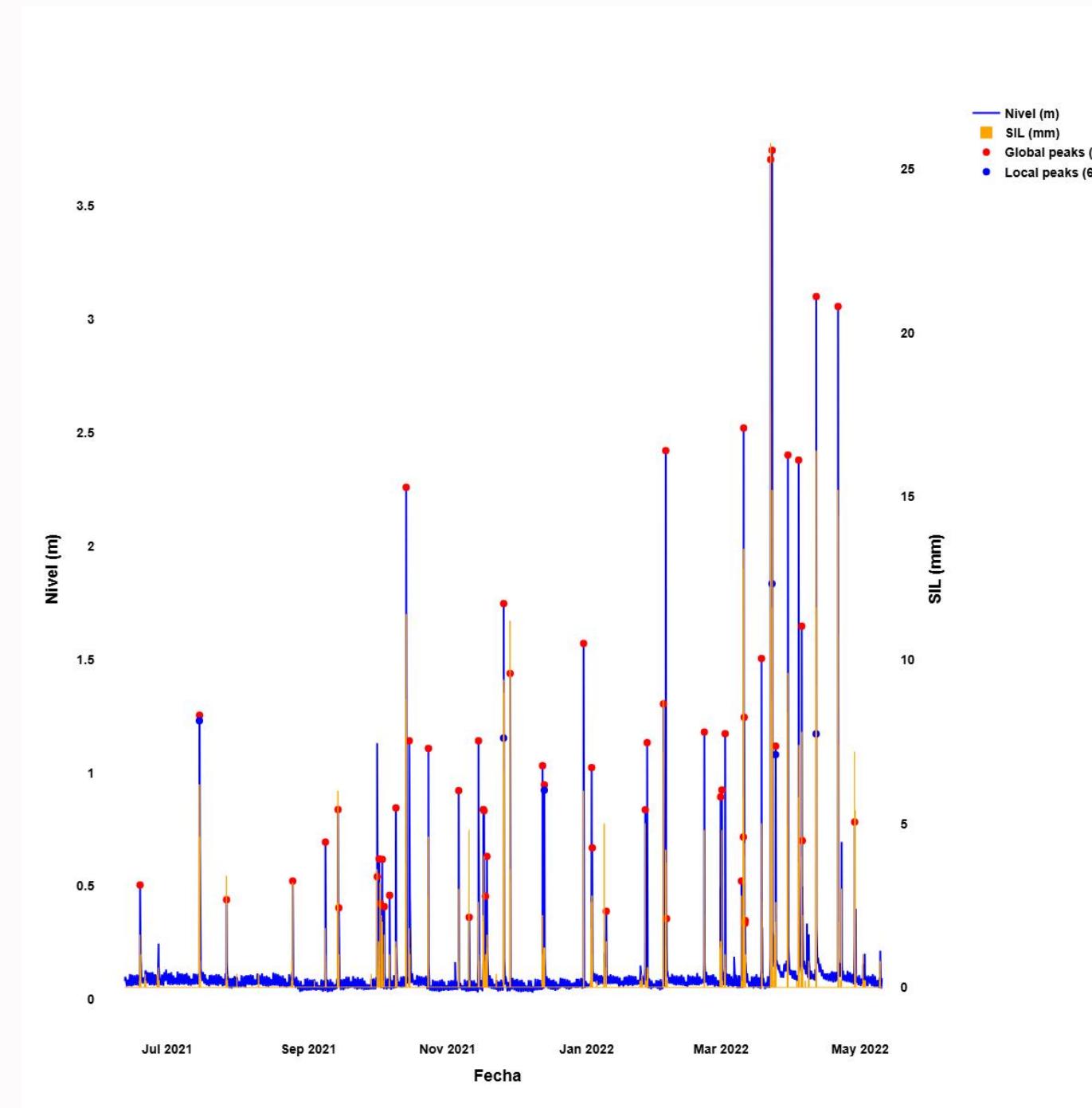


- **Evento:** Niveles que supera un nivel umbral
- **Nivel umbral:** Nivel establecido como significativo
- **Inicio de un evento:** Fecha en la que el nivel aumenta superando el nivel umbral
- **Fin de un evento:** Fecha en la que el nivel desciende superando el nivel umbral





# Picos de Nivel con SIL



- 58 picos de nivel picos globales, implica 58 eventos

- Zoom

- Definir el Alcance mínimo esperado
- Plantear y Establecer Criterios



**Objetivos y  
Estratégias**



# Objetivo y Criterios

Desarrollar un modelo de machine learning funcional para implementar un sistema de alerta temprana de inundaciones

## Criterios establecidos

- **Nivel umbral = 0.3 metros**
- Colección de precipitaciones hasta 80 minutos antes de pico global
- Colección de precipitación acumulada en diferentes ventanas hasta 80 minutos antes del pico global
- **Niveles críticos = [0.5, 1, 1.5, 2] metros**



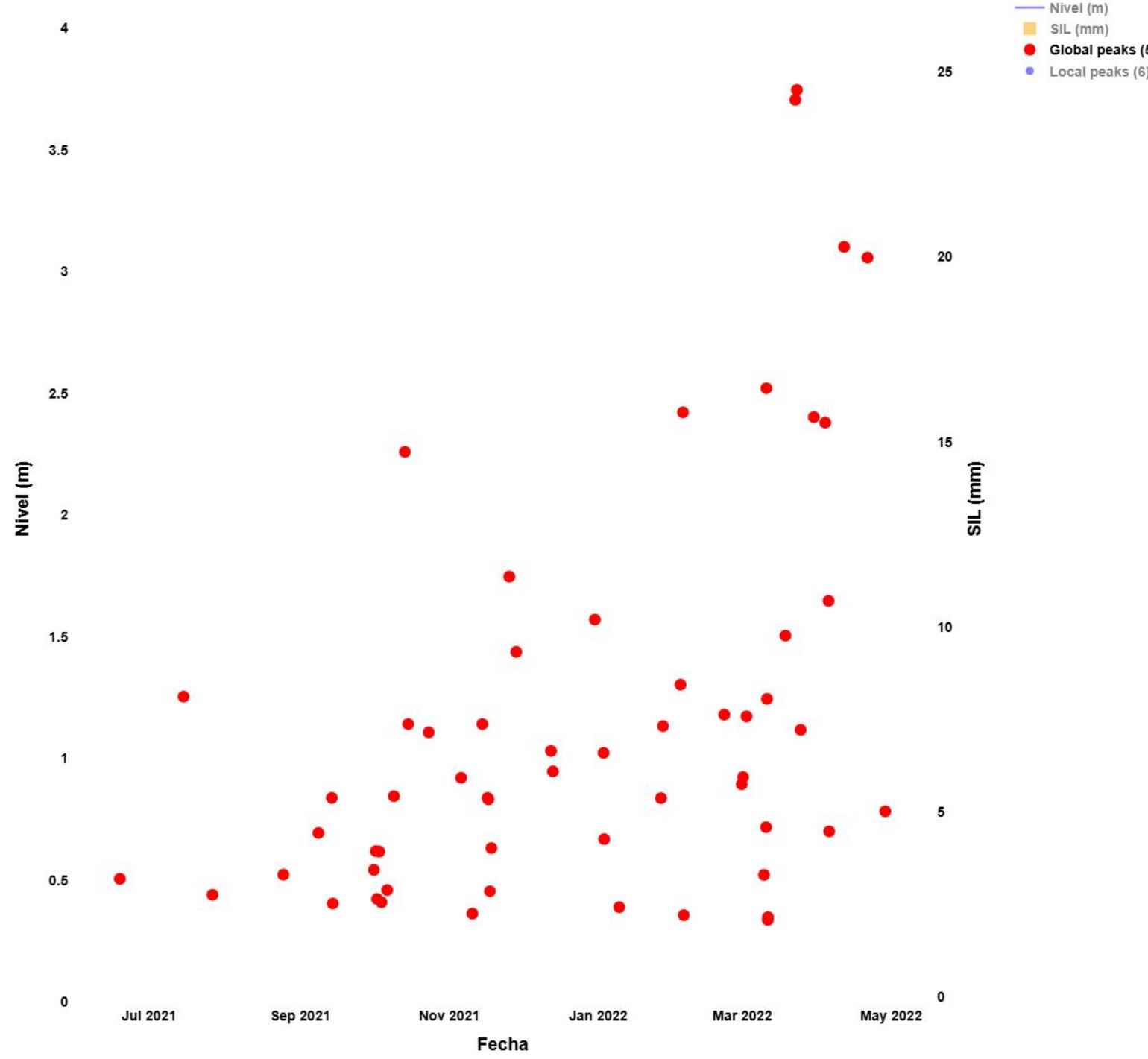
- Predicción de Picos
- Predicción de Eventos Críticos



**Selección de  
Modelos**



# Predicción de picos

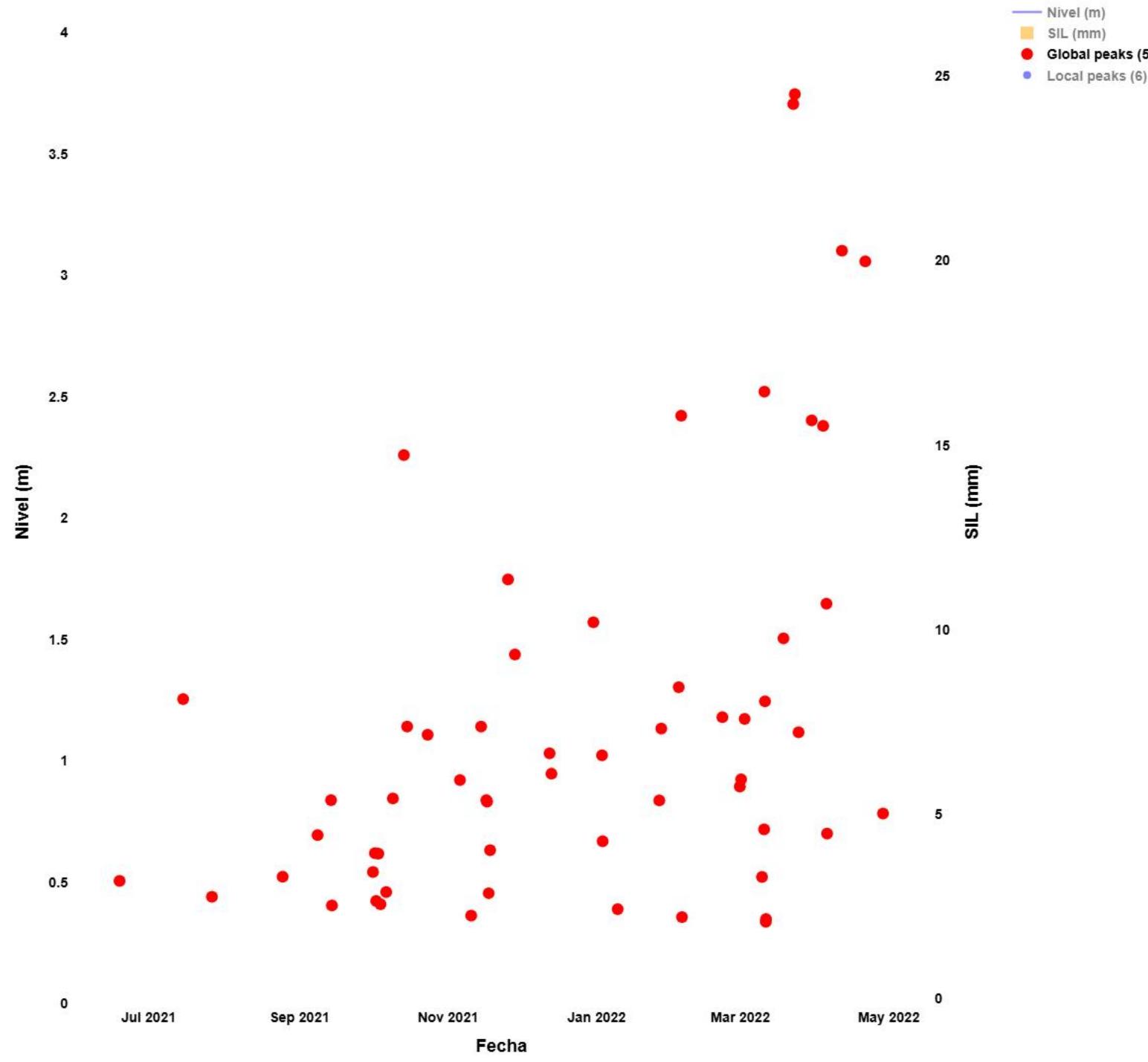


## Modelo de Regresión Lineal

- Por la simplicidad y cantidad de los datos
- Como referencia para modelos de más complejos



# Predicción de eventos críticos



## Modelo Support Vector Machine

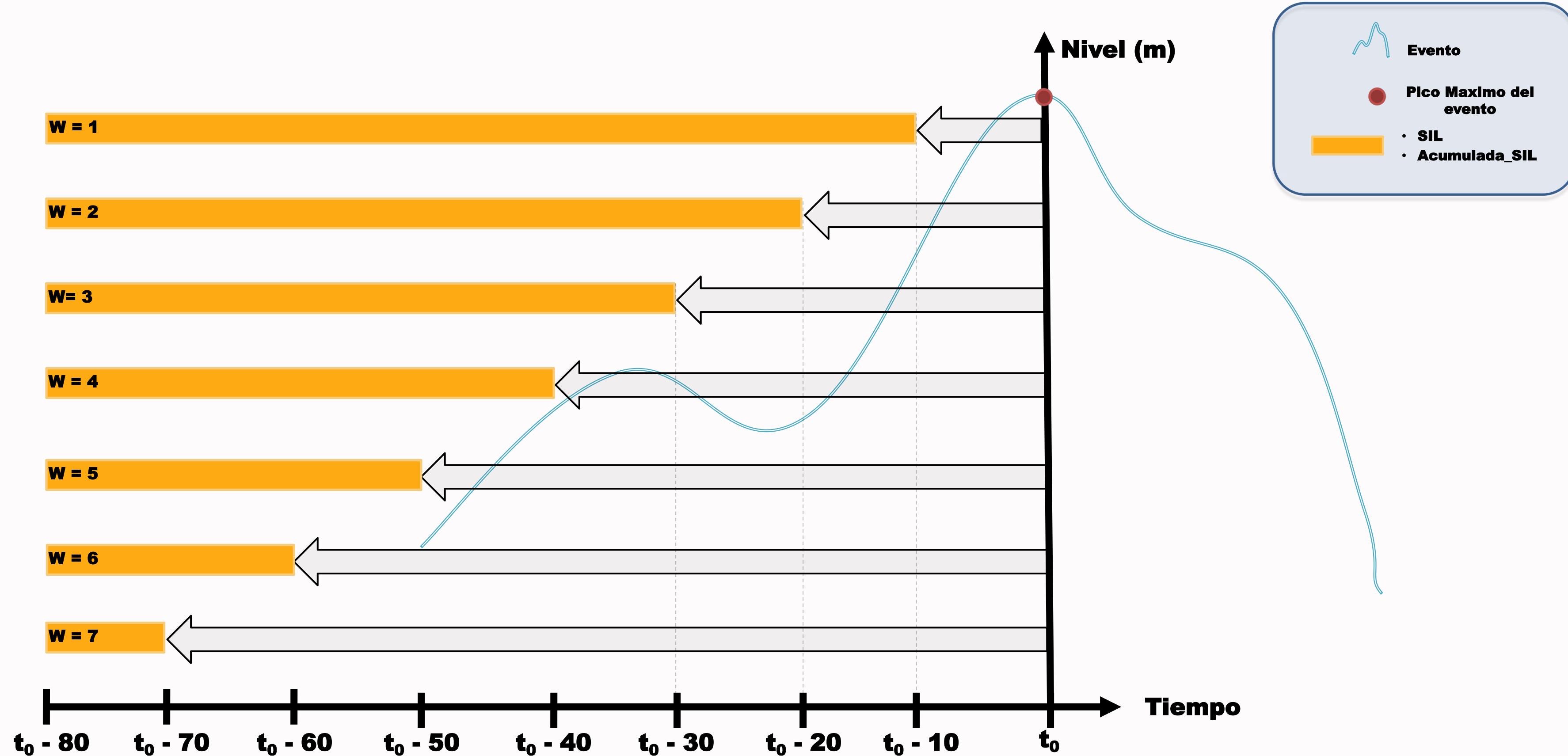
- Por el manejo entre las posibles clases en datos desbalanceados
- Capacidad de capturar relaciones no lineales

- Selección de datos de entradas para cada modelo
- Visualización esquemática de las ventanas de contexto para cada modelo



## Selección de Features

# Modelo de Regresión Lineal



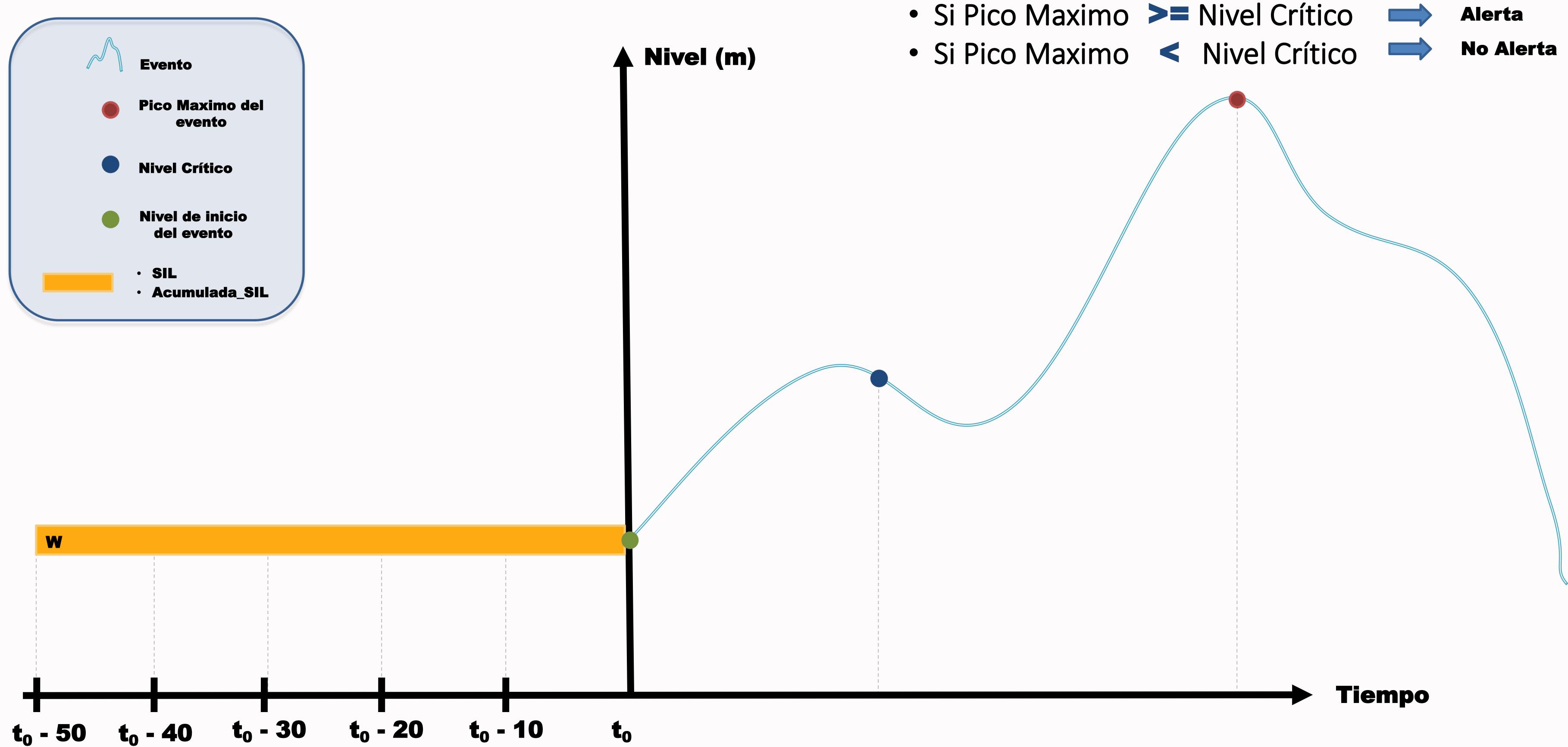


# Modelo de Regresión Lineal

- Para evaluar la capacidad de generalización del modelo se utilizará **5-fold cross validation**.
  - ✓ **R<sup>2</sup>** (Coefficient of Determination)
  - ✓ **RMSE** (Root Mean Squared Error)
  - ✓ **MAPE** (Mean Absolute Percentage Error)



# Modelo de SVM





- Busqueda de los Mejores Features



**Optimización  
de Modelos**



# Busqueda de los mejores features

## Modelo de Regresión Lineal

- Todas las posibles combinaciones de los features para cada ventana

$$N(w) = 2^{(15-2(w-1))-1}$$

$$(w = 1, 2, \dots, 7)$$

## Modelo Support Vector Machine

- Todas las posibles combinaciones para cada ventana

$$N(w) = 2^{9-2(w-1)-1}$$

$$(w = 0, 1, \dots, 5)$$

- Tablas comparativas
- Capacidad de generalización de los modelos



**Evaluación y  
Desempeño de  
los Modelos**

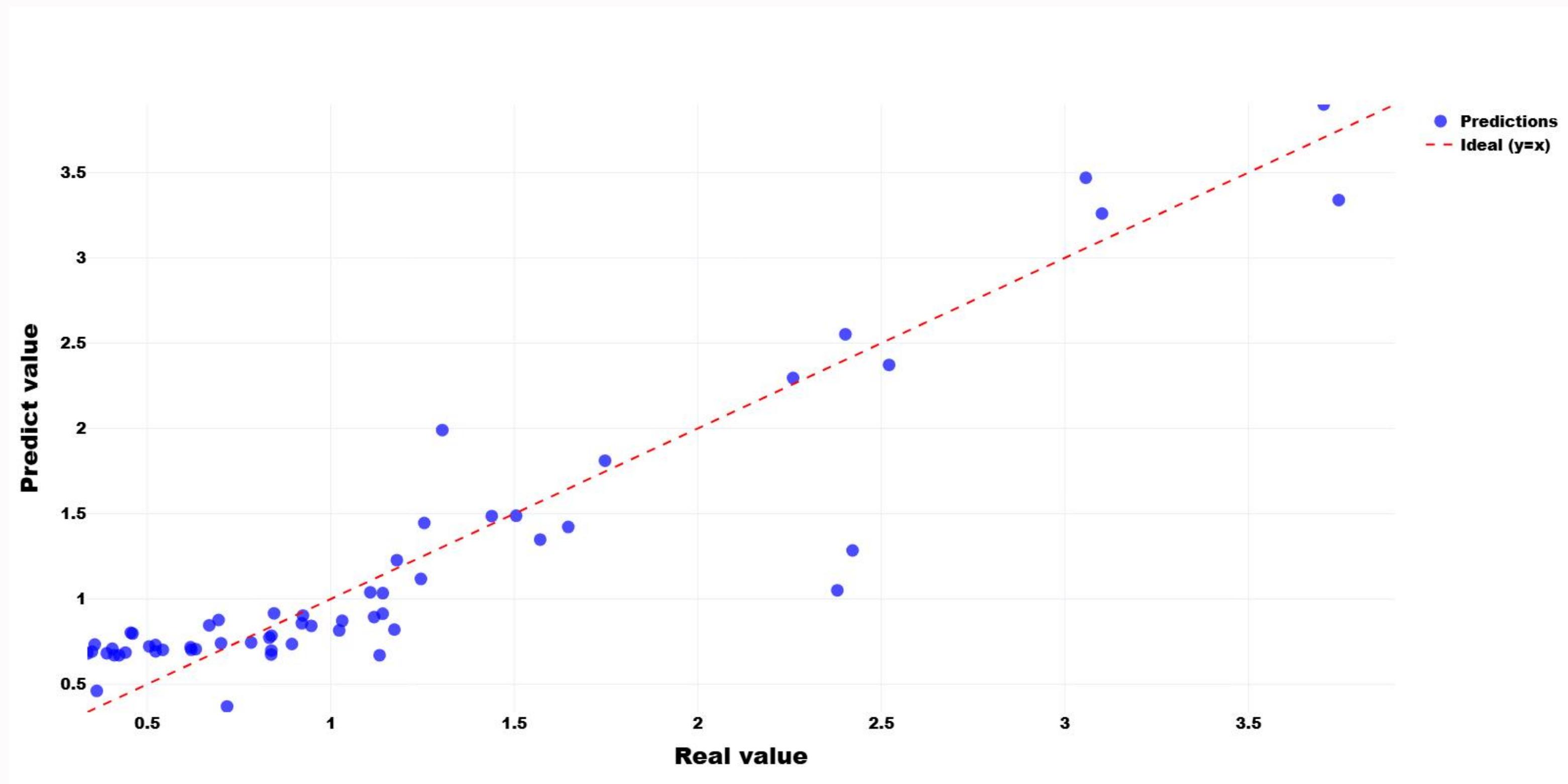


# Modelo de Regresión Lineal

<b>Experimento</b>	<b>W</b>	<b>R2</b>	<b>RMSE</b>	<b>MAPE</b>
1	1	0.8412	0.3304	0.2761
	2	0.8443	0.3272	0.2807
	3	0.8470	0.3243	0.2819
	4	0.8209	0.3509	0.3127
	5	0.6139	0.5152	0.3672
	6	0.0478	0.8090	0.7372
	7	0.0194	0.8210	0.7639



# Modelo de Regresión Lineal



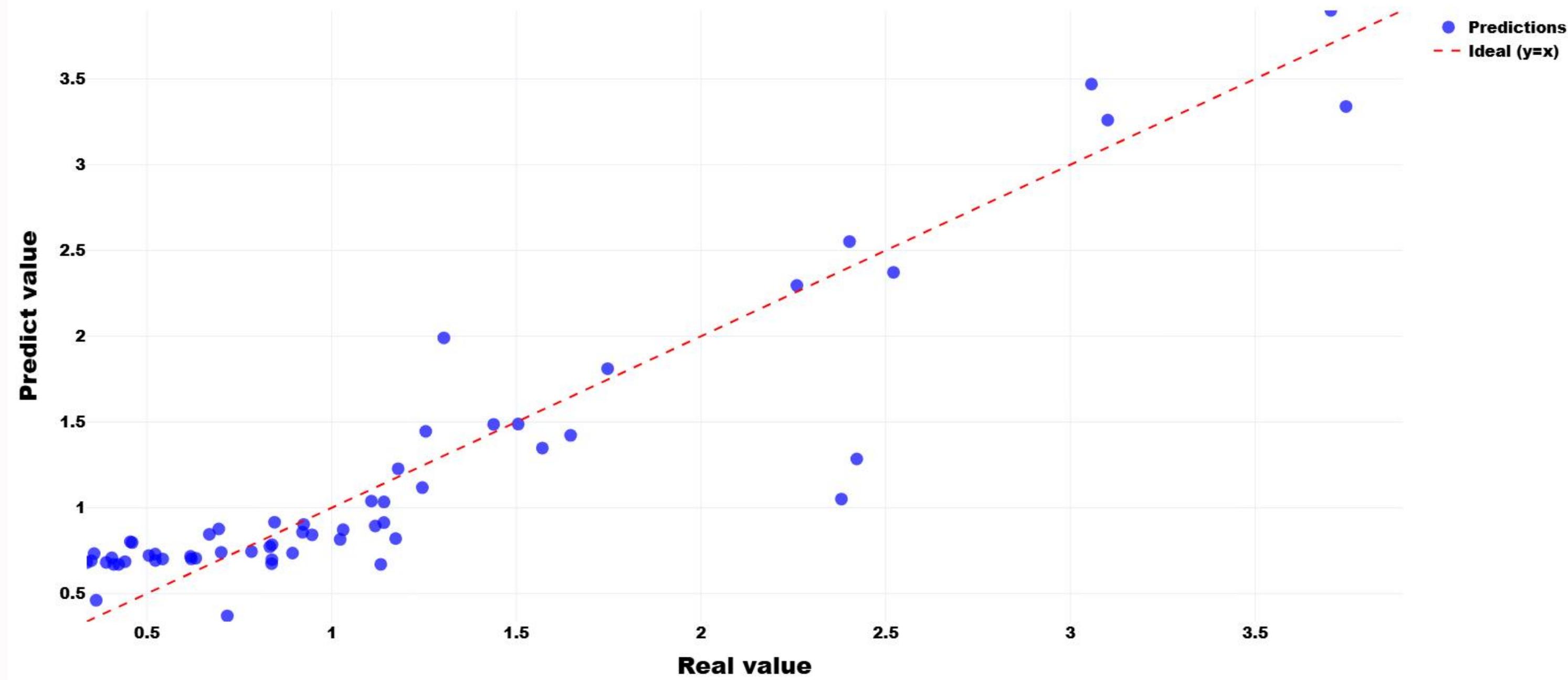


# Modelo de Regresión Lineal

<b>Experimento</b>	<b>W</b>	<b>RMSE</b>	<b>R2</b>	<b>MAPE</b>
2	1	0.3304	0.8412	0.2761
	2	0.3272	0.8443	0.2807
	3	0.3243	0.8470	0.2819
	4	0.3509	0.8209	0.3127
	5	0.5152	0.6139	0.3672
	6	0.8090	0.0478	0.7372
	7	0.8210	0.0194	0.7639



# Modelo de Regresión Lineal



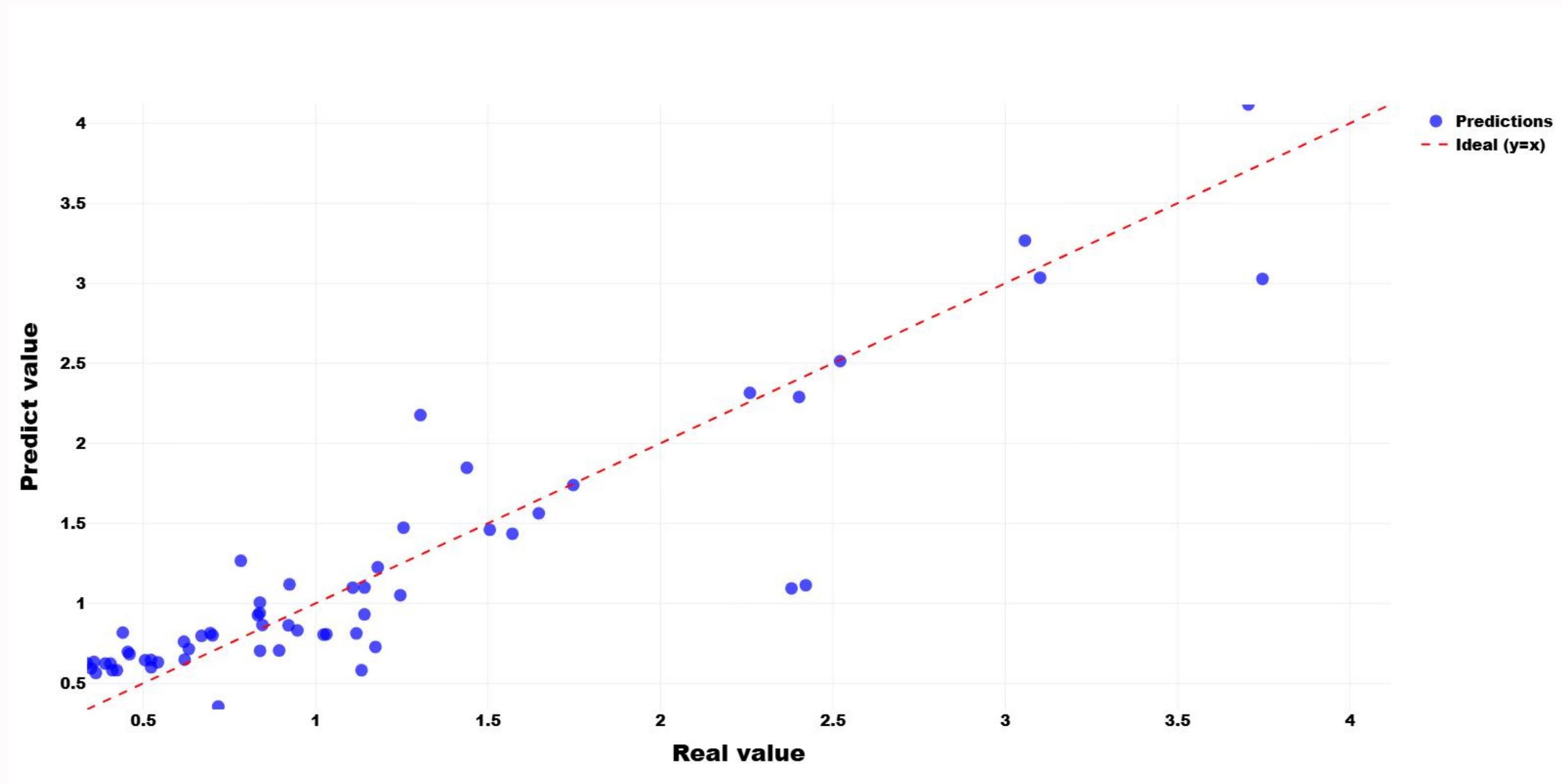


# Modelo de Regresión Lineal

<b>Experimento</b>	<b>W</b>	<b>MAPE</b>	<b>R2</b>	<b>RMSE</b>
3	1	0.2647	0.8153	0.3563
	2	0.2695	0.8301	0.3417
	3	0.2693	0.8193	0.3525
	4	0.2965	0.7400	0.4228
	5	0.3672	0.6139	0.5152
	6	0.6614	-0.1165	0.8760
	7	0.7638	-0.0302	0.8415



# Modelo de Regresión Lineal

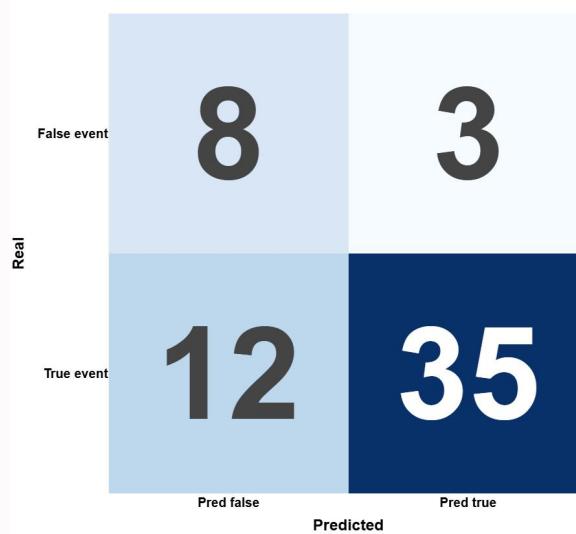




# Modelo Support Vector Machine

- Nivel crítico = 0.5 metros

Experimento	w	f1	precision	recall	FN
1	0	0.8095	0.9189	0.7234	13
	1	0.8235	0.9211	0.7447	12
	2	0.7952	0.9167	0.7021	14
	3	0.6842	0.8966	0.5532	21
	4	0.4762	0.9375	0.3191	32
	5	0.4194	0.8667	0.2766	34

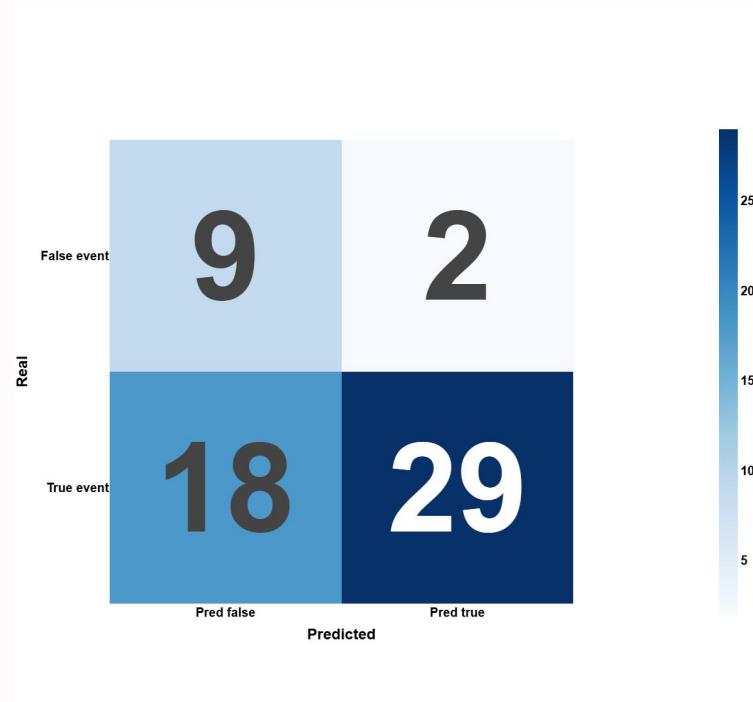




# Modelo Support Vector Machine

- Nivel crítico = 1 metros

Experimento	w	f1	precision	recall	FN
2	0	0.7778	0.7500	0.8077	5
	1	0.8235	0.8400	0.8077	5
	2	0.7692	0.7692	0.7692	6
	3	0.6000	0.8571	0.4615	14
	4	0.6364	0.5250	0.8077	5
	5	0.6567	0.5366	0.8462	4

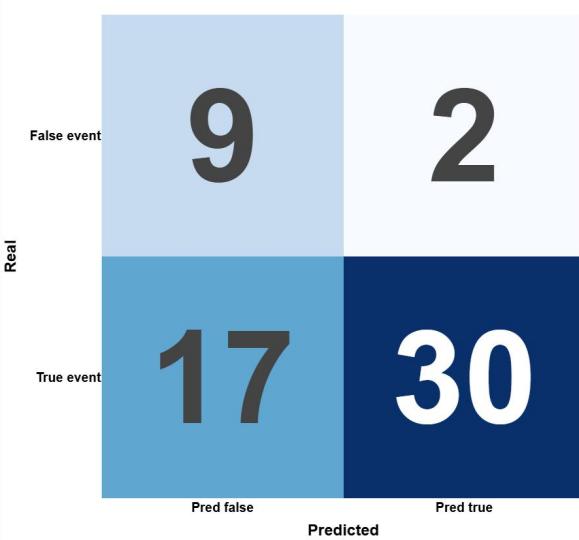




# Modelo Support Vector Machine

- Nivel crítico = 1.5 metros

Experimento	w	f1	precision	recall	FN
3	0	0.7097	0.6111	0.8462	2
	1	0.7097	0.6111	0.8462	2
	2	0.6875	0.5789	0.8462	2
	3	0.4583	0.3143	0.8462	2
	4	0.4444	0.2927	0.9231	1
	5	0.4783	0.3333	0.8462	2





# Modelo Support Vector Machine

- Nivel crítico = 2 metros

Experimento	w	f1	precision	recall	FN
4	0	0.5600	0.4375	0.7778	2
	1	0.6400	0.5000	0.8889	1
	2	0.5385	0.4118	0.7778	2
	3	0.3636	0.2286	0.8889	1
	4	0.2909	0.1739	0.8889	1
	5	0.3810	0.2424	0.8889	1

