

# Aumento de resolución de datos a través de métodos de Machine Learning

Matías Palma Manterola

Universidad de La Frontera

17 de octubre de 2022



UNIVERSIDAD  
DE LA FRONTERA

# Tabla de Contenidos

## 1 Introducción

## 2 Fundamentos Teóricos

## 3 Metodología

## 4 Resultados



# Contexto y descripción del problema

- Almacén de aguas subterráneas.
  - Calidad.
  - Cantidad.
  - Ampliamente distribuidas.
- Sobreexplotación.
  - Disminución del nivel.
- Monitoreo de cuencas.
- Satélite GRACE.
  - Estimación del agua total presente.



# Satélite GRACE

- *Gravity Recovery And Climate Experiment.*
- Satélites que orbitan sobre la Tierra.
  - 200 Km de distancia entre ellos.
- Captar movimientos de las aguas.
  - Contrarrestar:
    - Sequías.
    - Inundaciones.
    - Socavones.
- Mediciones a baja resolución.

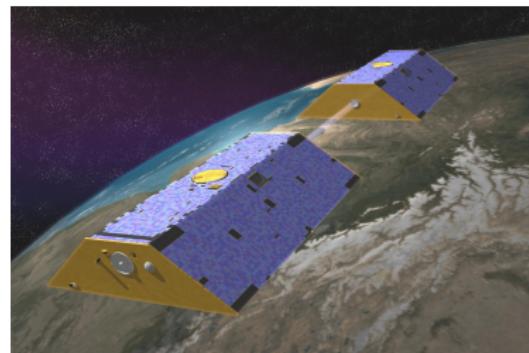


Figura 1: Misión GRACE lanzada el año 2002.

# Anomalías TWS

- *Terrestrial Water Storage*
- Agua almacenada sobre y debajo de la superficie.
  - Agua de dosel.
  - Ríos y lagos.
  - Humedad de suelo.
  - Aguas subterráneas.

## Anomalía TWS

$$TWS_t = \frac{TWS_t - \mu}{\delta}, \quad \mu, \delta \in \mathbb{R}$$



## *Downscaling*

- Inferir información de alta resolución.
    - Datos satelitales.
    - Imágenes.
  - Enfoques dinámicos o estadísticos.
    - Meteorología.
    - Climatología
    - Teledetección

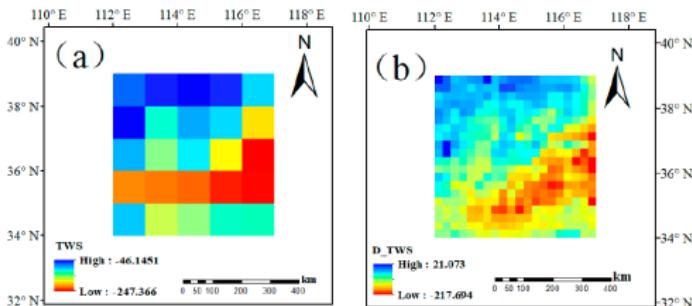


Figura 2: Reducción de escala.

# Objetivo general y específicos

## Objetivo general

- Implementar un aumento de resolución a datos satelitales proporcionados por GRACE a través de un modelo predictivo para obtener productos grillados en Chile a alta resolución.

## Objetivos específicos

- Definir la metodología.
- Recolectar datos necesarios.
- Preprocesar los datos.
- Implementar y entrenar el modelo predictivo.
- Validar predicciones en base a observaciones *in situ*.

# Tabla de Contenidos

1 Introducción

2 Fundamentos Teóricos

3 Metodología

4 Resultados



UNIVERSIDAD  
DE LA FRONTERA

# Productos grillados

## Píxel

- Unidad básica de una imagen.
- Cada píxel contiene canales de información
  - RGB.
  - Números reales ( $\mathbb{R}$ ).
  - Etiquetas de clase.

## Ráster

- Conjunto de píxeles ordenados.
  - Filas.
  - Columnas.
- El valor del píxel se asocia al centroide del mismo.



# Árboles de decisión

## Árboles de decisión

- Algoritmo de aprendizaje automático del tipo supervisado.
- Segmentación recursiva del espacio de predicción en regiones regulares.
- Minimiza la suma de cuadrados de cada región (regresión).

## Random Forest

- Conjunto de árboles de decisión.
  - *Bootstrap sample*.
  - Selección aleatoria de características.
- Contrarresta la inestabilidad de los árboles.



# Índice de Correlación de Pearson

- Busca cuantificar la similitud lineal en las tasas de crecimiento.

## Índice de Correlación

Sean  $S_1$  y  $S_2$  dos series temporales con observaciones  $u_1, \dots, u_p$  y  $v_1, \dots, v_p$ , respectivamente, entonces

$$R^2(S_1, S_2) = \frac{\sum_{j=1}^p \sum_{i=1}^p (u_i - \bar{u}_j)(v_i - \bar{v}_j)}{\sqrt{\sum_{i=1}^p (u_i - \bar{u})^2} \sqrt{\sum_{i=1}^p (v_i - \bar{v})^2}} \in [-1, 1]$$



# Tabla de Contenidos

## 1 Introducción

## 2 Fundamentos Teóricos

## 3 Metodología

## 4 Resultados



# Metodología adoptada

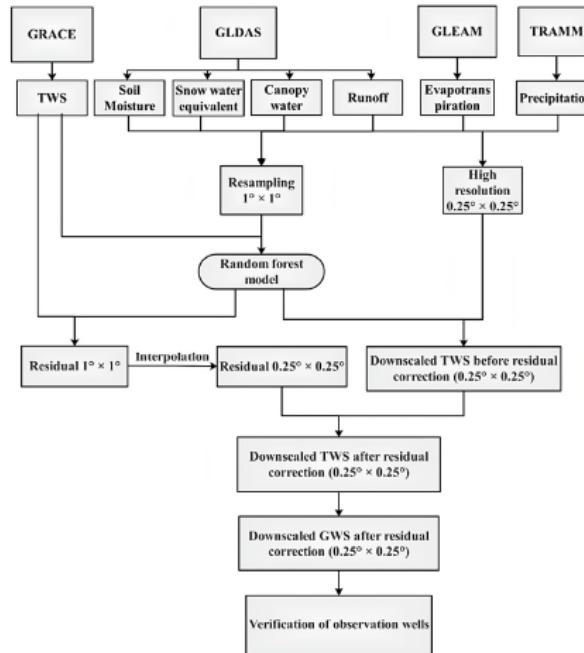


Figura 3: Metodología utilizada en (Chen et al., 2019), pág. 7

# Dominio espacio-temporal

- Territorio continental chileno.
  - Longitud desde  $66.5^{\circ}\text{E}$  hasta  $77.5^{\circ}\text{E}$ .
  - Latitud  $16.5^{\circ}\text{S}$  hasta  $56.75^{\circ}\text{S}$
- Aumento de resolución de productos grillados desde abril del año 2002 hasta febrero del año 2017.

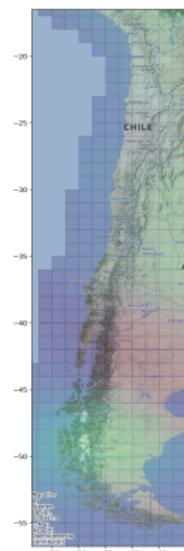


Figura 4: Datos de grage para abril del año 2002 en el territorio continental chileno.



UNIVERSIDAD  
DE LA FRONTERA

# GRACE TWS

- Anomalías TWS.
- Datos globales mensuales.
- Resolución a  $1^\circ$  ( $\sim 110$  Km).
- lwe\_thickness: grosor de agua líquida equivalente.

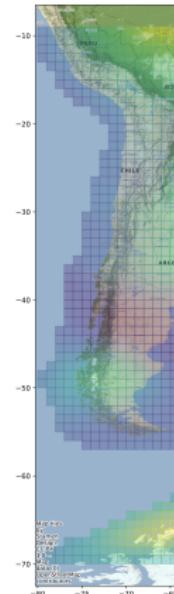


Figura 5: Datos obtenidos desde la base de datos [OPeNDAP](#).



UNIVERSIDAD  
DE LA FRONTERA

# ERA5-Land

- Representa la evolución de variables terrestres.
  - Humedad de suelo.
  - Nieve equivalente en agua.
  - Evapotranspiración.
  - Agua de desel.
  - Escorrentía.
- Combina datos de modelos con observaciones.
- Datos globales mensuales
- Resolución de  $0.1^\circ$  ( $\sim 9$  Km)

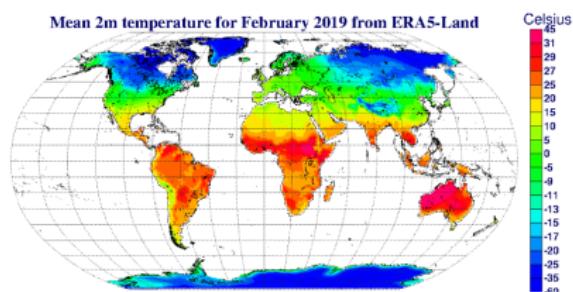


Figura 6: Temperatura promedio mensual obtenida de la base de datos [Copernicus](#).

# (CR)<sup>2</sup>-met

- Información meteorológica
  - Precipitación.
  - Temperatura.
- Datos muestreados diariamente para el territorio continental.
- Modelos estadísticos para traducir diversos componentes.
- Resolución de 0.05° ( $\sim 5$  Km)

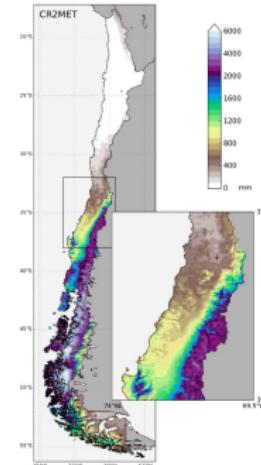


Figura 7: Precipitación acumulada anual promedio provenientes de la base de datos [\(CR\)<sup>2</sup>-met](#).



UNIVERSIDAD  
DE LA FRONTERA

# Dirección General de Aguas

- Observaciones de pozos de bombeo.
- Mediciones diarias para la zona centro y norte del país.
- Mide profundidad (m) del pozo.
  - Mayor profundidad implica menor cantidad de agua disponible.

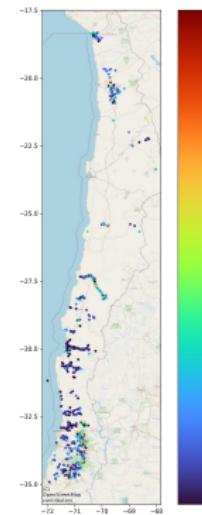


Figura 8: Observaciones de pozos provenientes de la DGA proporcionados por (CR)<sup>2</sup>.



# Conjunto de entrenamiento y de prueba

- Alta variabilidad climatológica en el territorio.
- Estratificación de datos.
  - Disminución de sesgo.
  - Predicciones coherentes respecto a la naturaleza de las variables.



Figura 9: Macrozonas de Chile.



UNIVERSIDAD  
DE LA FRONTERA

# Tabla de Contenidos

1 Introducción

2 Fundamentos Teóricos

3 Metodología

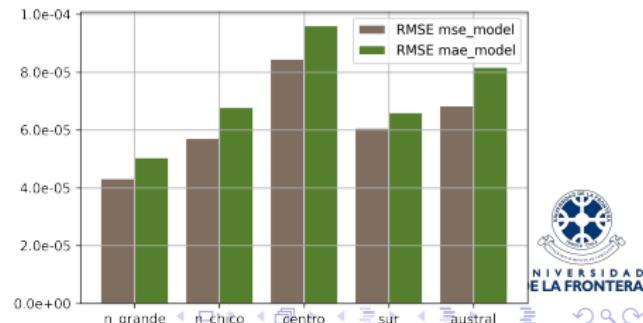
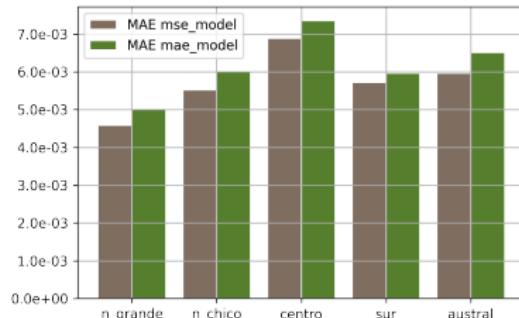
4 Resultados



UNIVERSIDAD  
DE LA FRONTERA

# Métricas de validación

Modelo/Métrica de eval.	Error medio absoluto (MAE)	Raíz del error cuadrático medio (RMSE)
Mín. error cuadrático medio	$5,69 \times 10^{-3}$	$6,23 \times 10^{-5}$
Mín. error medio absoluto	$6,15 \times 10^{-3}$	$7,25 \times 10^{-5}$



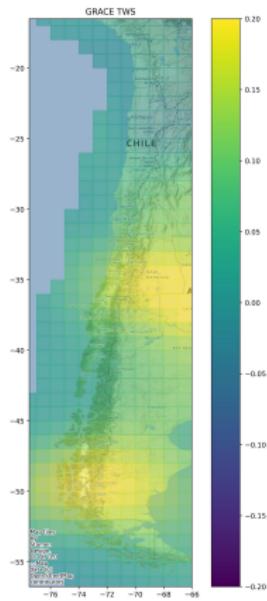


Figura 12: Productos grillados a alta y baja resolución para el mes de septiembre del año 2002.

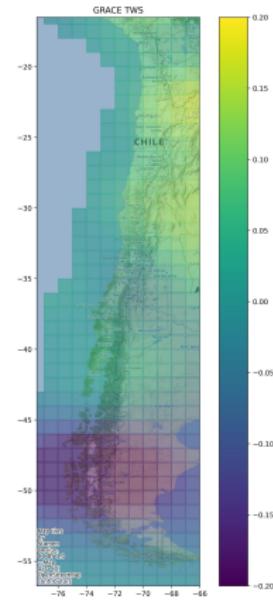
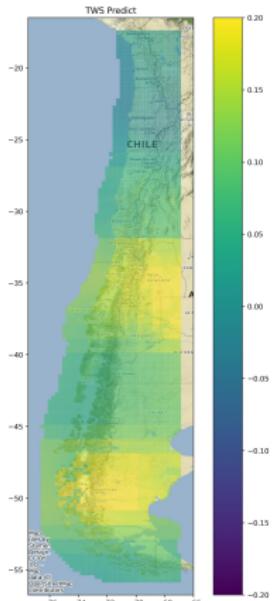
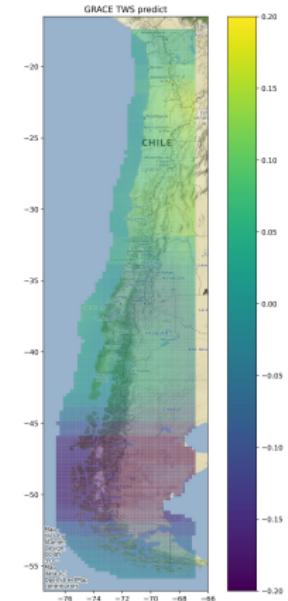


Figura 13: Productos grillados a alta y baja resolución para el mes de abril del año 2017.



UNIVERSIDAD  
DE LA FRONTERA

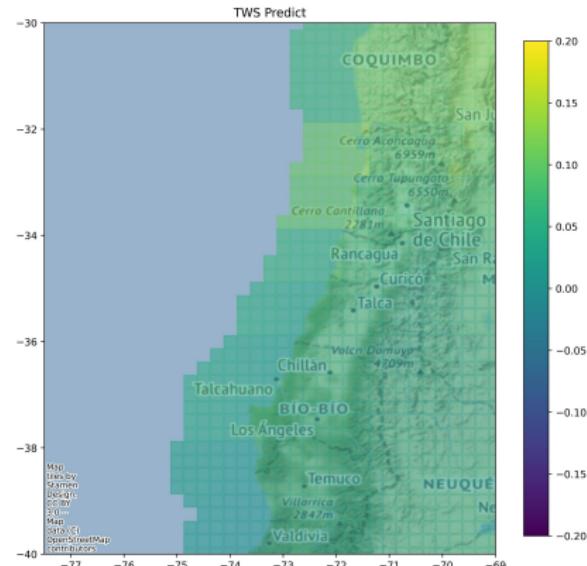
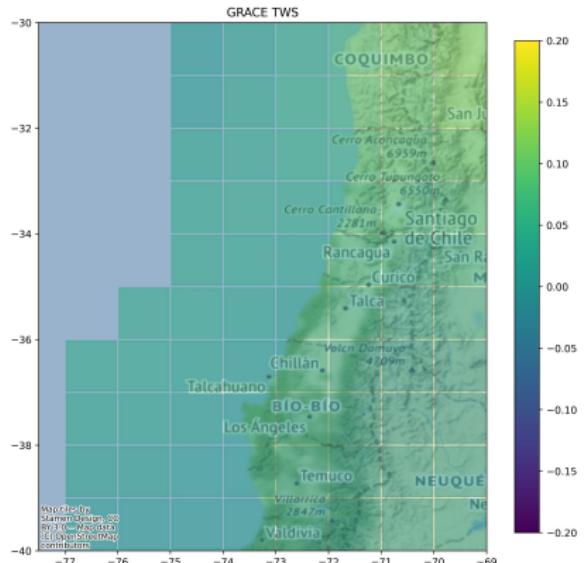


Figura 14: Parche a alta y baja resolución para la zona centro del país en el mes de junio del año 2016.

# Zona de validación

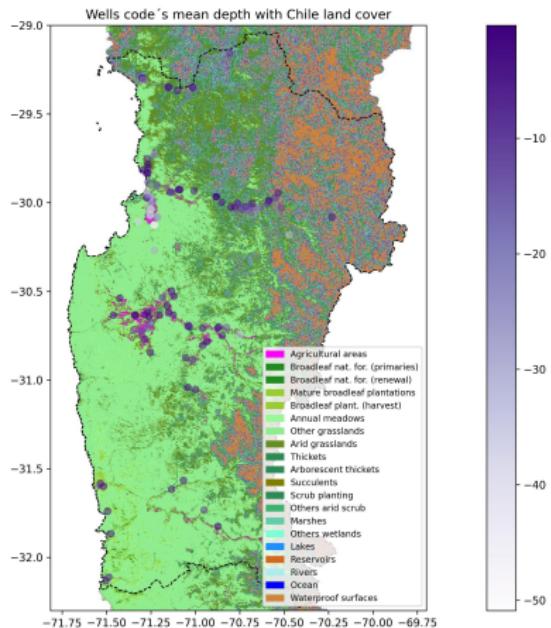
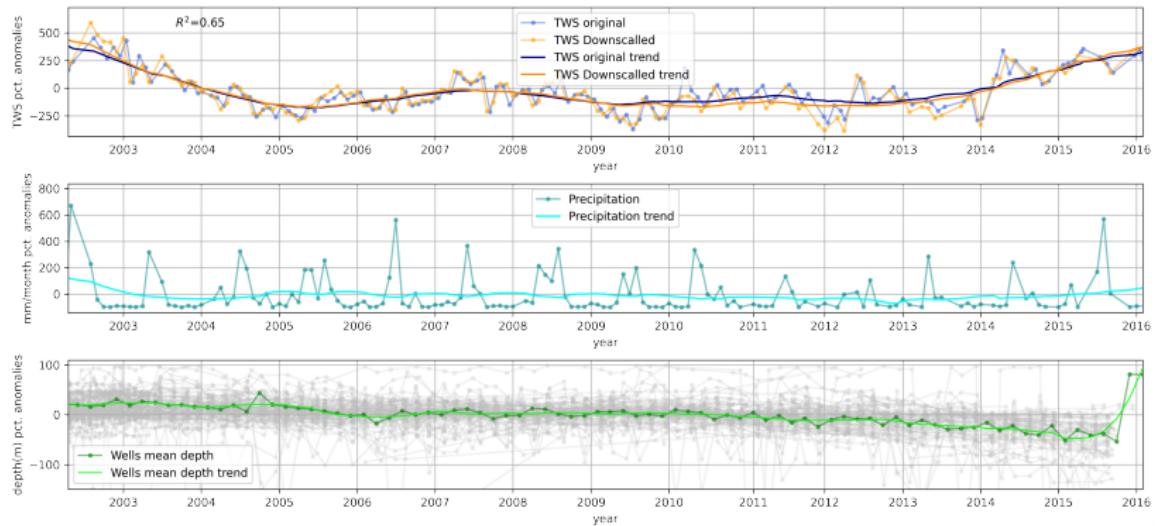


Figura 15: Región de Coquimbo con observaciones de pozos y zonas agrícolas.

# Series de tiempo

Data anomalies for Coquimbo region



UNIVERSIDAD  
DE LA FRONTERA

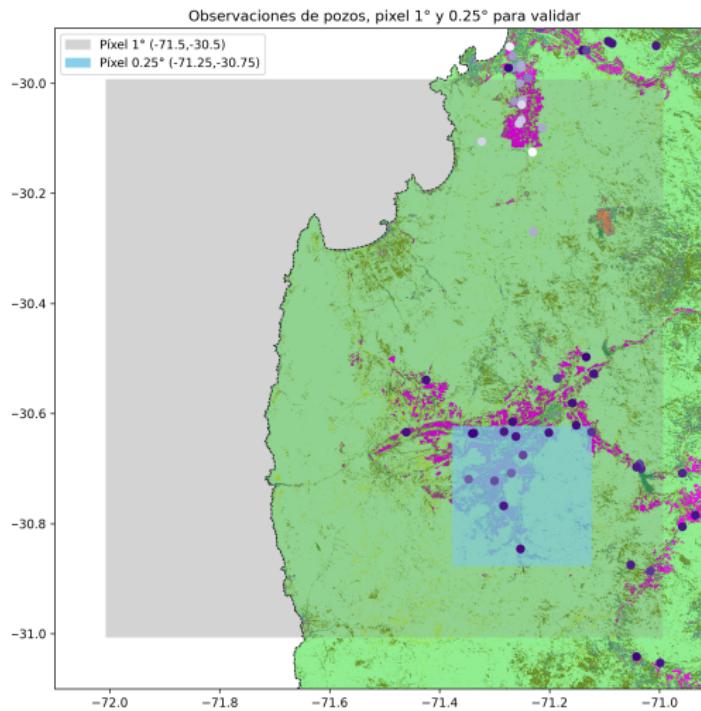
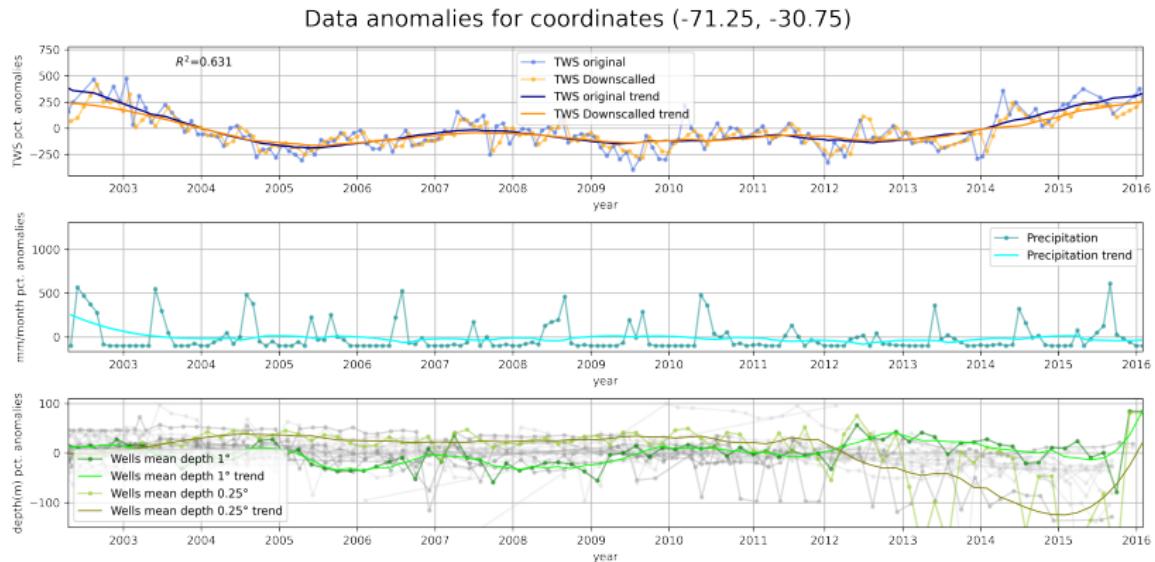


Figura 16: Píxel a alta y baja resolución para validar en zonas agrícolas con observaciones de Pozos.

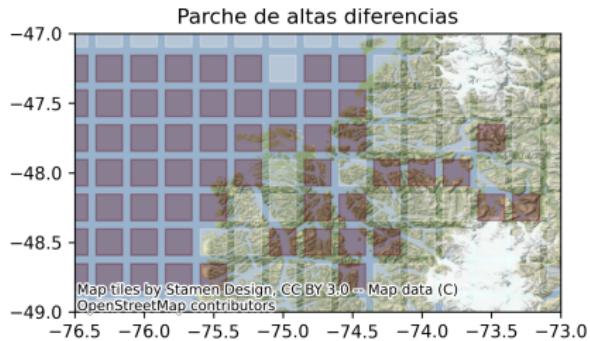
# Series de tiempo



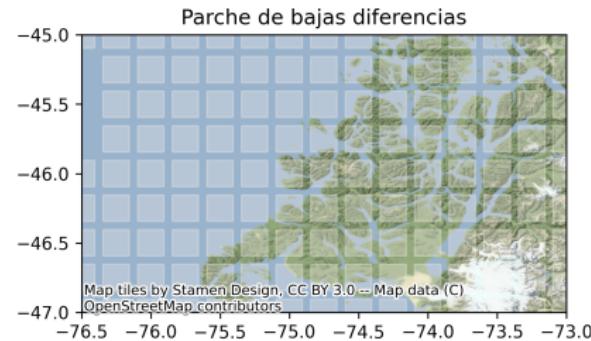
# Interpolación bilineal y Bosque aleatorio

- Diferencias absolutas entre promedio de píxeles.
  - Selección de píxeles con una diferencia mayor al 50 %.
    - 6.7 % del total de píxeles.
- Aislamiento de la zona en cuestión.





**Figura 17:** Parche a alta resolución donde la interpolación bilineal y la regresión usando *Random Forest* difieren en sobrermanera.



**Figura 18:** Parche a alta resolución donde la interpolación bilineal y la regresión usando *Random Forest* no difieren en demasía.

# Estudio de variabilidad

Parche	$\sigma^2$ CDO	$\sigma^2$ Random Forest
Altas diferencias	$1,1664 \times 10^{-6}$	$3,03805 \times 10^{-6}$
Bajas diferencias	$9,665881 \times 10^{-6}$	$9,351364 \times 10^{-6}$



# Estudio de variabilidad

Variable	Altas diferencias	Bajas diferencias
<b>Soil moisture</b>	0.0973583	0.1291611
<b>Evapotranspiration</b>	$7.9032 \times 10^{-7}$	$8.68624 \times 10^{-7}$
<b>Runoff</b>	$1.11489 \times 10^{-5}$	$1.5682 \times 10^{-5}$
<b>Precipitation</b>	6.109682	5.185894
<b>Canopy water</b>	$9.409 \times 10^{-9}$	$8.836 \times 10^{-9}$
<b>Snow water equivalent</b>	15.23432	9.04167
$\sigma^2$ promedio	3.57356	2.39279



# Conclusiones

- La metodología adoptada logra predecir de forma válida los resultados.
- Las variables explicativas son datos homogéneos y fiables de centros especializados.
- La validación de resultados mostró una alta variabilidad en zonas donde las variables explicativas también la presentaban.
- Con los productos grillados obtenidos es posible estudiar con más detalle y de manera más granular el comportamiento de las aguas subterráneas en Chile.

# Bibliografía

- Breiman, Leo. 2001. Random forests. *Machine learning*, 45(1), 5–32.
- Chen, Li, He, Qisheng, Liu, Kun, Li, Jinyang, & Jing, Chenlin. 2019. Downscaling of GRACE- derived groundwater storage based on the random forest model. *Remote Sensing*, 11(24), 2979.
- Hyndman, Rob J, & Athanasopoulos, George. 2018. *Forecasting: principles and practice*. OTexts.
- Kuhn, Max, Johnson, Kjell, et al. 2013. *Applied predictive modeling*. Vol. 26. Springer.
- Sammut, Claude, & Webb, Geoffrey I. 2011. *Encyclopedia of machine learning*. Springer Science & Business Media.



# Aumento de resolución de datos a través de métodos de Machine Learning

Matías Palma Manterola

Universidad de La Frontera

17 de octubre de 2022



UNIVERSIDAD  
DE LA FRONTERA