

Reducción de escala de datos satelitales a través del modelo de Bosque Aleatorio

Matías Palma Manterola

Universidad de La Frontera

Octubre, 2022



UNIVERSIDAD DE LA FRONTERA

Tabla de Contenidos

1 Introducción

2 Fundamentos Teóricos

3 Metodología

4 Resultados

5 Conclusiones



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.

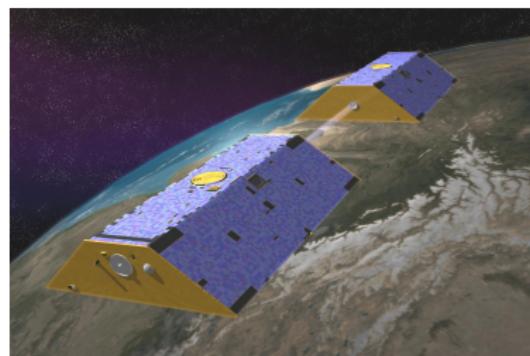


Figura 1: Misión GRACE lanzada el año 2002.



UNIVERSIDAD DE LA FRONTERA

Anomalías TWS

- *Terrestrial Water Storage*
- Agua almacenada sobre y debajo de la superficie.
 - Agua de desel.
 - 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

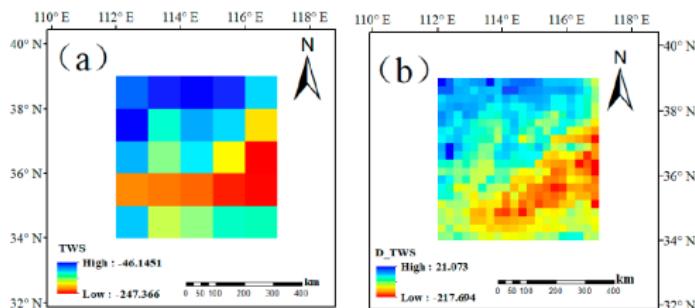


Figura 2: Reducción de escala.



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

5 Conclusiones



Árboles de decisión

- Algoritmo de aprendizaje automático del tipo supervisado.
- Segmentación recursiva del espacio de predicción en regiones regulares.
 - Media.
 - Moda.
- Posee ciertas desventajas
 - Inestabilidad.
 - Rendimiento predictivo menor al óptimal.
- Minimiza la suma de cuadrados de cada región.



Random Forest

- Conjunto de árboles de decisión.
 - *Bootstrap sample.*
- Contrarresta la inestabilidad de los árboles.
- Mejor generalización en la predicción.
- Menos susceptible al sobreajuste.



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

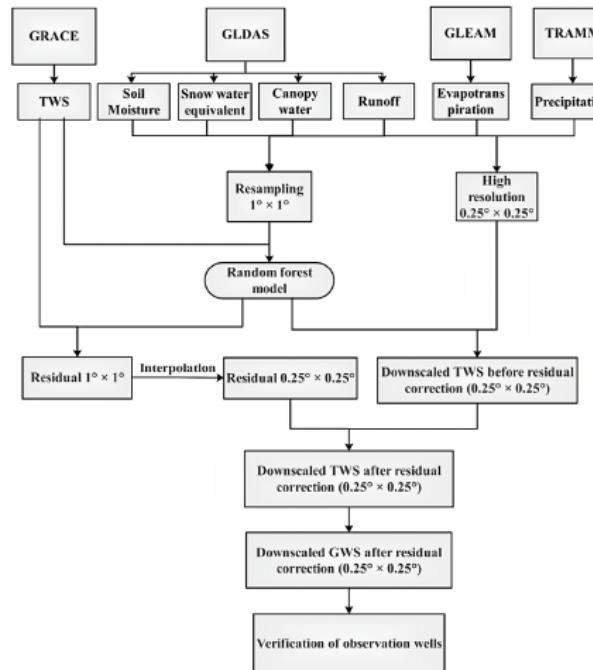
3 Metodología

4 Resultados

5 Conclusiones



Downscaling of GRACE-Derived Groundwater Storage Based on the Random Forest Model



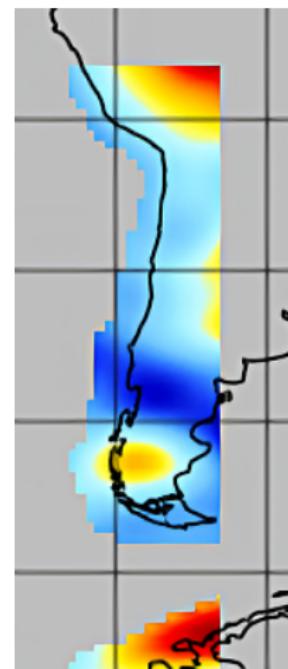
Dominio espacio-temporal

- Territorio continental chileno.
 - Longitud desde 63.5°E hasta 79.5°E.
 - Latitud 16.5°S hasta 69.5°S
- Aumento de resolución de productos grillados desde abril del año 2002 hasta febrero del año 2017.



GRACE TWS

- Anomalías TWS.
- Datos globales mensuales.
- Resolución a 1° (~ 110 Km).
- lwe_thickness: grosor de agua líquida equivalente.



ERA5-Land

- Representa la evolución de variables terrestres.
 - Humedad de suelo.
 - Nieve equivalente en agua.
 - Evapotranspiración.
 - Agua de dosel.
 - Escorrentía.
- Combina datos de modelos con observaciones.
- Datos globales mensuales
- Resolución de 0.1° (~ 9 Km)



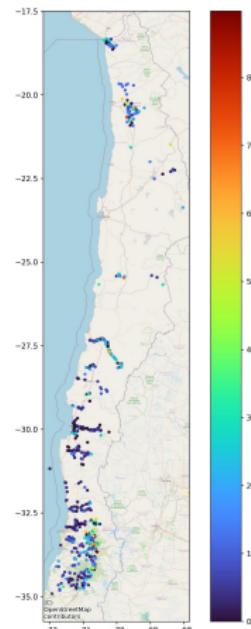
(CR)²-met

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



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.



Remuestreo espacial



Remuestreo temporal



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 3: Macrozonas de Chile.



Implementación del modelo

```
# model with the mse as the loss function
mse_model = RandomForestRegressor(n_estimators = 200,
                                   criterion = "mse",
                                   max_depth = None,
                                   max_features = 'auto',
                                   oob_score = False,
                                   n_jobs = -1,
                                   random_state = 73
)
mse_model = mse_model.fit(X_train.drop(['geometry'], axis = 1), y_train)
```

Figura 4: Minimización de error cuadrático medio.

```
# model with the mae as the loss function
mae_model = RandomForestRegressor(n_estimators = 200,
                                   criterion = "mae",
                                   max_depth = None,
                                   max_features = 'auto',
                                   oob_score = False,
                                   n_jobs = -1,
                                   random_state = 73
)
mae_model = mae_model.fit(X_train.drop(['geometry'], axis = 1), y_train)
```

Figura 5: Minimización de error medio absoluto.



Tabla de Contenidos

1 Introducción

2 Fundamentos Teóricos

3 Metodología

4 Resultados

5 Conclusiones



Métricas de validación

Modelo RF	MAE	RMSE
Minimización mse	$5,69 \times 10^{-3}$	$6,23 \times 10^{-5}$
Minimización mae	$6,15 \times 10^{-3}$	$7,25 \times 10^{-5}$

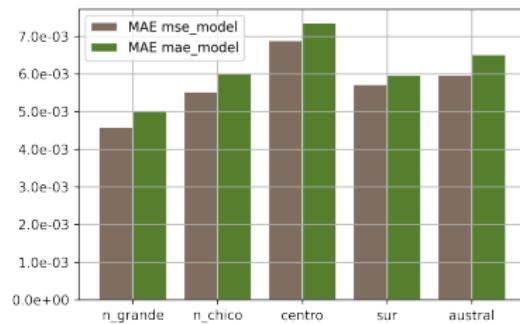


Figura 6: Error medio absoluto

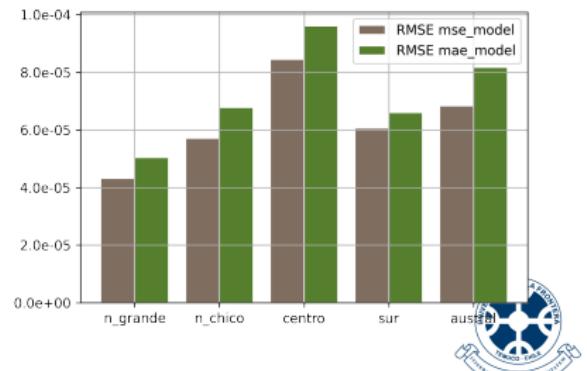
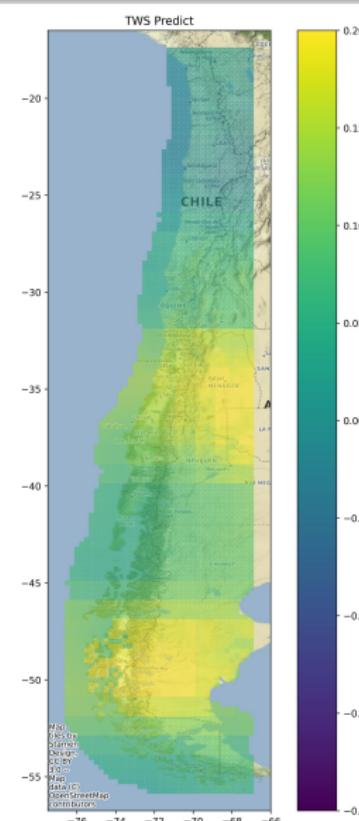
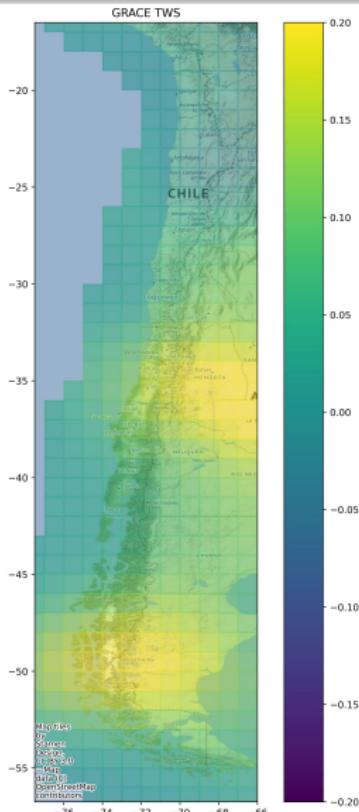
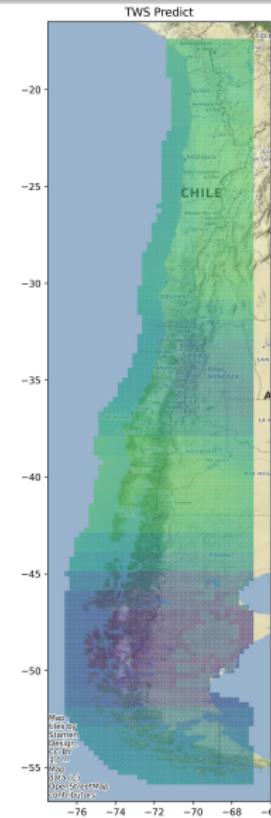
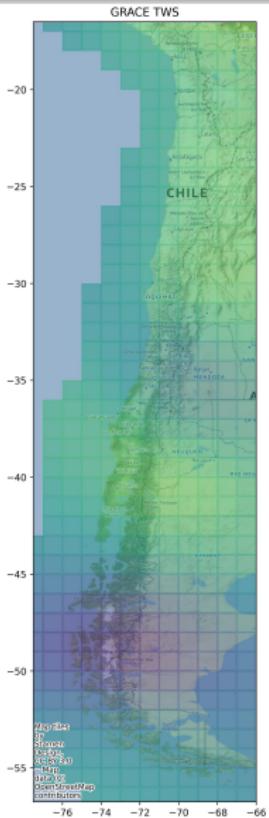


Figura 7: Raíz del error cuadrático medio

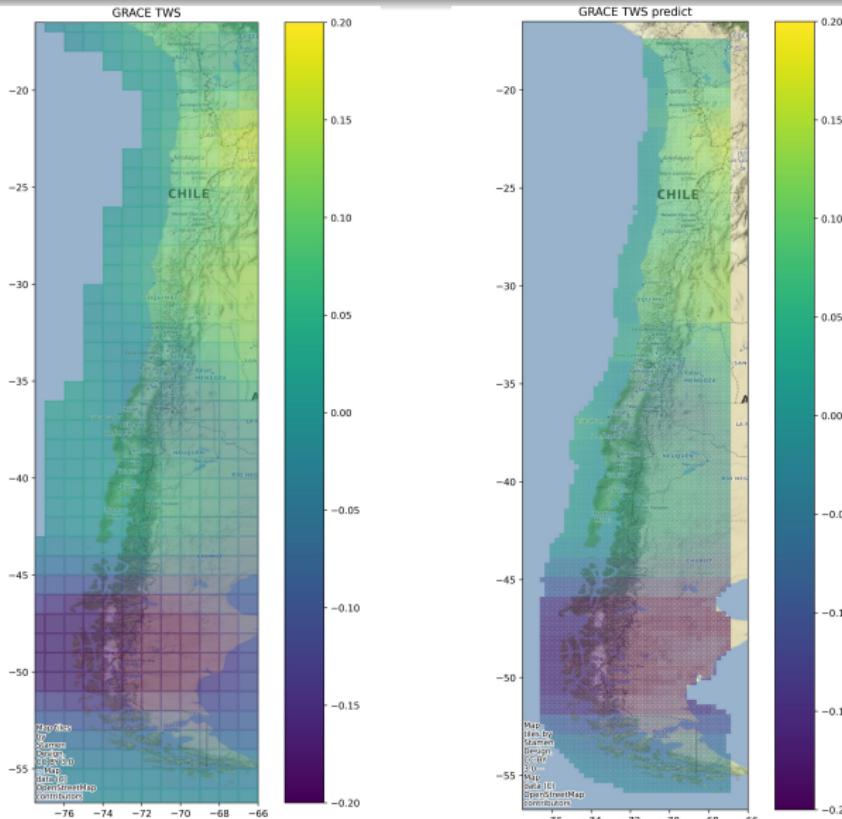




UNIVERSIDAD DE LA FRONTERA

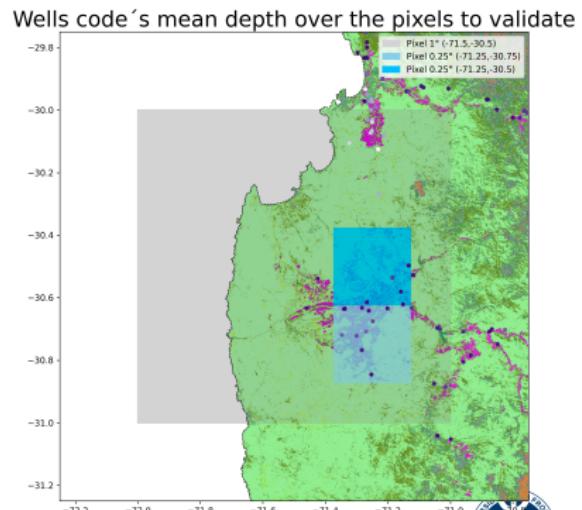
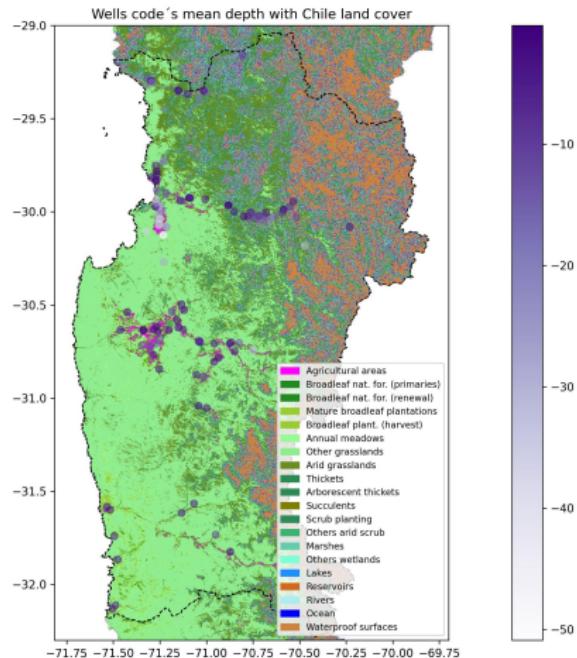


VERSIDAD DE LA FRONTERA



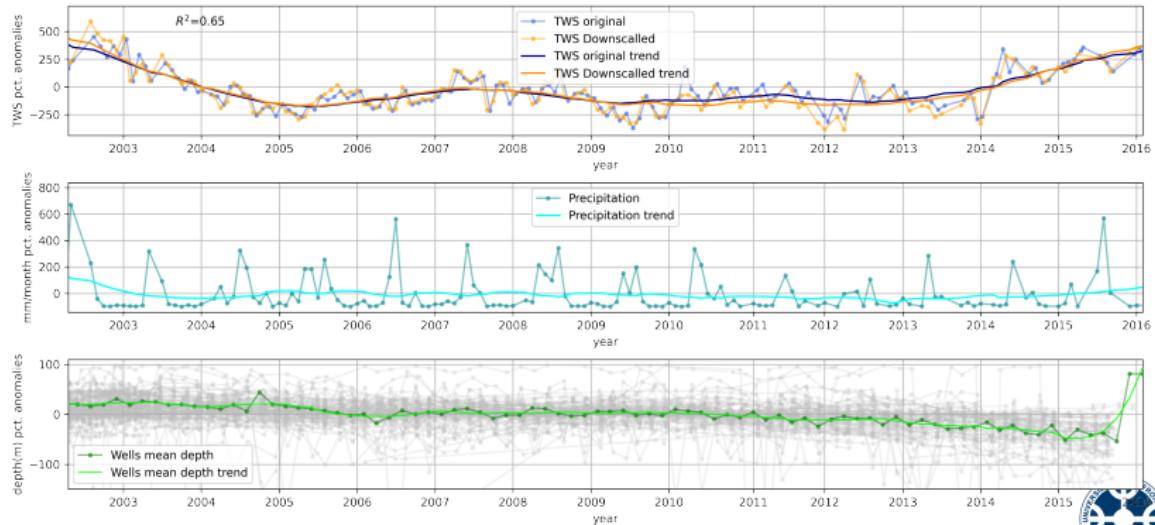
VERSIDAD DE LA FRONTERA

Zona de validación

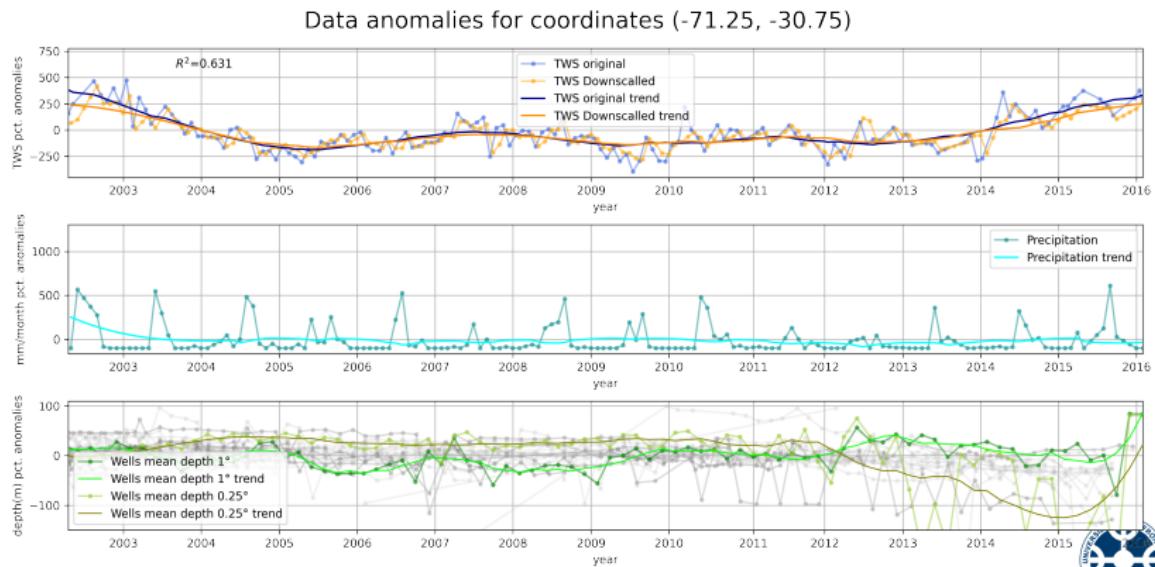


Series de tiempo

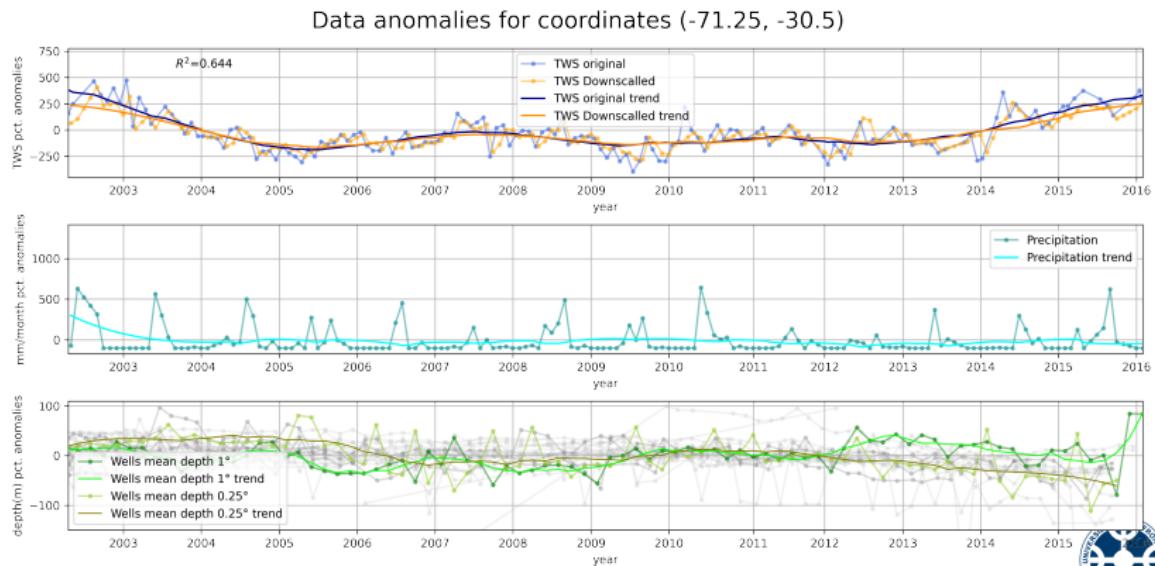
Data anomalies for Coquimbo region



Series de tiempo



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.

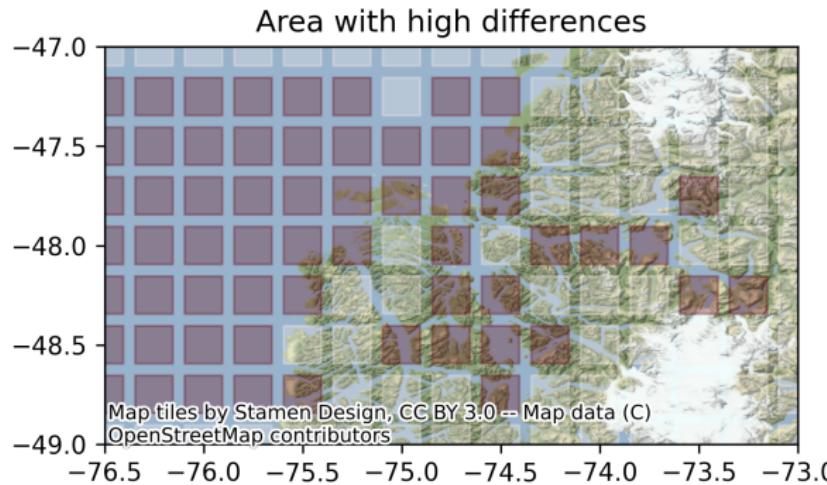


Tabla de Contenidos

1 Introducción

2 Fundamentos Teóricos

3 Metodología

4 Resultados

5 Conclusiones



