

UNIVERSIDAD DE VALLADOLID
MÁSTER UNIVERSITARIO
Ingeniería Informática



TRABAJO FIN DE MÁSTER

**Estudio de métodos de interpolación
espacial y diseño y creación de
herramienta de visualización de mapas
interpolados con variables atmosféricas**

Realizado por **Mario Abad Diaz**

Universidad de Valladolid
20 de septiembre de 2024

Tutores: Jesús Vegas
Ramiro González

AGRADECIMIENTOS

Quiero expresar mi especial agradecimiento a David Mateos Villán por su ayuda y total disposición a la hora de afrontar este Trabajo de Fin de Master sirviendo como un gran apoyo para todo el proyecto.

Resumen

El espesor óptico de aerosoles (AOD) es un parámetro crítico para entender el impacto y la concentración de aerosoles en la atmósfera. Los aerosoles, que incluyen polvo, hollín y otras partículas, pueden generarse de procesos naturales como las erupciones volcánicas o por procesos humanos como las emisiones industriales. Estas partículas tienen un impacto significativo en el clima, tiempo, calidad del aire y salud pública. Un alto valor de AOD indica una mayor presencia de aerosoles, que puede reducir la visibilidad, afectar a la salud e influenciar en el clima al modificar la radiación de la luz solar.

En este contexto se hace necesario monitorizar los valores de AOD. Sin embargo, no existe una gran cantidad de datos. Es por ello que en este documento se investiga sobre métodos de interpolación para predecir valores en lugares sin mediciones. Finalmente se desarrolla una aplicación en Python capaz de utilizar los métodos de interpolación investigados que han obtenido los mejores resultados para crear visualizaciones de los mapas de interpolación con la herramienta.

Descriptores

Espesor óptico de aerosoles, AOD, Python, Interpolación, Kriging, Herramienta de visualización, AERONET

Abstract

Aerosol Optical Depth (AOD) is a critical parameter for the understanding of the impact and the concentration of aerosols in the atmosphere. Aerosols include dust, soot and other particles which can be generated by natural processes such as volcanic eruptions or by human processes such as industrial emissions. These fine particles have a high impact in climate, weather, air quality and public health. A high AOD measurement indicates a greater presence of aerosols, which can reduce visibility, affect health and modify the climate by blocking sun radiation.

In this context, the monitoring of AOD values is made necessary. However, there is a lack of data available. Is thus, the reason of this document being dedicated to the investigation of interpolation methods to predict values of AOD in emplacements without real measurements. Finally, it is developed a Python application which uses the interpolation methods investigated with the best results to create a tool which can interpolate and show a view of the interpolated maps.

Keywords

Aerosol Optical Depth, AOD, Python, Interpolation, Kriging, Visualization tool, AERONET

Índice general

AGRADECIMIENTOS	I
Índice general.....	IV
Índice de figuras.....	VII
Índice de ecuaciones	VIII
Índice de tablas	IX
.....	1
1: Introducción y Objetivos	1
1.1 Introducción	1
1.2 Aerosol Optical Depth (AOD).....	2
1.3 Aerosol Robotic Network (AERONET)	3
1.4 Objetivos del proyecto	4
2: Planificación del proyecto	5
2.1 Planificación.....	5
2.2 Materiales y presupuesto.....	7
2.2.1 Recursos técnicos.....	7
2.2.2 Recursos humanos.....	8
2.2.3 Presupuesto final	9
.....	10

3: Estado del arte	10
3.1 Introducción a la interpolación espacial.....	10
3.1.1 Métodos de interpolación no geoestadísticos	10
3.1.1.1 Nearest Neighbour (NN).....	11
3.1.1.2 Inverse Distance Weighting (IDW).....	11
3.1.2 Métodos de interpolación geoestadísticos	12
3.1.2.1 Simple Kriging (SK).....	14
3.1.2.2 Ordinary Kriging (OK)	15
3.1.2.3 Universal Kriging (UK)	15
3.1.2.4 Cokriging (CK).....	15
3.2 Análisis de artículos.....	16
3.2.1 Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature.....	16
3.2.2 Spatial Analysis of Air Temperature in the Qinghai-Tibet Plateau	17
3.2.3 A Comparison of Spatial Interpolation Techniques in Temperature Estimation	18
3.2.4 Interpolation Techniques for Climate Variables.....	19
3.2.5 Discusión	20
3.3 Selección de métodos de interpolación	21
.....	24
4: Técnicas y herramientas	24
4.1 Lenguaje utilizado y descripción de librerías seleccionadas	24
4.1.1 Pykrige	24
4.1.2 Tfinterpy	25
4.1.3 Geostatspy	26
4.1.4 Gstatsim.....	27
4.1.5 Arcpy.....	28
4.1.6 IDW.....	29
4.2 Métrica	30
4.3 Comparación de librerías.....	31
4.3.1 Medición Pykrige.....	31
4.3.2 Medición Geostatspy.....	32
4.3.3 Medición Gstatsim	34

4.3.4 Medición IDW	35
4.3.5 Comparativa general	37
.....	38
5: Aspectos relevantes del desarrollo del proyecto.....	38
5.1 Diseño de la interfaz	38
5.1.1 Herramienta de diseño de interfaces	38
5.1.2 Diseño de la interfaz	39
5.2 Implementación	43
.....	48
6: Conclusiones y Líneas de trabajo futuras	48
6.1 Conclusiones	48
6.2 Líneas futuras	49
.....	50
BIBLIOGRAFÍA.....	50

Índice de figuras

Ilustración 1 - Foto junio 2020 - tormenta de polvo sahariano [2]	2
Ilustración 2 - Cimel CE318-T [6]	3
Ilustración 3 - Diagrama de Gantt con la programación seguida.....	6
Ilustración 4 - Polígonos de Thiessen [12]	11
Ilustración 5 - Ejemplo variograma con Nugget, Sill y Range	13
Ilustración 6 - Ejemplos de 4 modelos de variograma comúnmente usados (a) esférico; (b) exponencial; (c) lineal; (d) Gaussiano	14
Ilustración 7 - Árbol de decisión para la elección de un método de interpolación [18].....	23
Ilustración 8 - Resultado de la mejor ejecución obtenida con la librería Pykrige.....	32
Ilustración 9- Resultado de la mejor ejecución obtenida con la librería Geostatspy.	34
Ilustración 10 - Resultado de la mejor ejecución obtenida con la librería Gstatsim.....	35
Ilustración 11 - Resultado de la mejor ejecución obtenida con la función IDW.	36
Ilustración 12 - Herramienta Qt Designer.....	39
Ilustración 13 - División de parámetros y ventana de visualización	40
Ilustración 14 - Visualización de mapa interpolado	41
Ilustración 15 - Interfaz pestaña Database	41
Ilustración 16 - Interfaz pestaña Data.....	42
Ilustración 17 - Interfaz pestaña Map.....	42
Ilustración 18 - Diagrama de clases	44
Ilustración 19 - Diagrama de funciones 1.....	46
Ilustración 20 - Diagrama de funciones 2.....	47

Índice de ecuaciones

Ecuación 1 - Inverse Distance Weighted.....	12
Ecuación 2 - Ecuación de semivarianza entre dos puntos	12
Ecuación 3 - Ecuación de semivarianza a partir de datos.....	13
Ecuación 4 - Ecuación Simple Kriging	14
Ecuación 5 - Ecuación Ordinary Kriging	15
Ecuación 6 - Ecuación Cokriging.....	16
Ecuación 7 - Cálculo de error absoluto (MAE)	30

Índice de tablas

Tabla 1 - Coste de recursos técnicos	8
Tabla 2 - Coste de recursos humanos	9
Tabla 3 - Resumen de métodos de interpolación utilizados y elegidos en los artículos analizados	21
Tabla 4 - Resultados librería PyKrige	32
Tabla 5 - Resultados librería GeostatsPy	33
Tabla 6 - Resultados librería Gstatsim	34
Tabla 7 - Resultados interpolación con IDW	36
Tabla 8 - Comparativa librerías e IDW	37

1: Introducción y Objetivos

En este capítulo se realizará una pequeña introducción al problema que se trata en este proyecto. A continuación, se explicará en detalle que es el espesor óptico de aerosoles y como se realizan las mediciones de esta medida. Finalmente, se establecen los objetivos que se persiguen a lo largo del trabajo y las restricciones con las que se cuenta de partida.

1.1 Introducción

El espesor óptico de aerosoles (AOD) es un parámetro crítico para entender el impacto y la concentración de aerosoles en la atmósfera. Los aerosoles, que incluyen polvo, hollín y otras partículas, pueden generarse de procesos naturales como las erupciones volcánicas o por procesos humanos como las emisiones industriales. Estas partículas tienen un impacto significativo en el clima, tiempo, calidad del aire y salud pública. Un alto valor de AOD indica una mayor presencia de aerosoles, que puede reducir la visibilidad, afectar a la salud e influenciar en el clima al modificar la radiación de la luz solar.

A pesar de la importancia del AOD, no existe una gran cantidad de mediciones precisas. Las mediciones desde el suelo como las que realiza AERONET proporcionan medidas precisas, pero son limitadas espacialmente ya que no pueden medir amplias zonas. Por el contrario, las mediciones realizadas por satélite, como las realizadas con MODIS, proporcionan datos de amplias zonas de medición, pero sufren los problemas de contar con vacíos temporales o de ciertas localizaciones debido a las nubes y a limitaciones de su órbita.

Los métodos de interpolación son una solución potente para paliar los problemas de discontinuidad en las mediciones de AOD. Al utilizar los datos disponibles para predecir valores en las zonas no medidas, la interpolación mejora la completitud de los datos. Existen varias técnicas de interpolación, algunas de ellas serán descritas y probadas a lo largo de este proyecto.

Alineado con lo expuesto, este Trabajo de Fin de Máster se dedicará a la investigación de métodos de interpolación estudiando el estado del arte utilizado para resolver problemas similares de interpolación espacial en el campo de la meteorología con el fin de proponer varios

métodos de interpolación interesantes para la resolución del problema planteado en este TFM. Finalmente se desarrollará un programa en Python para la interpolación de los mapas de AOD y su visualización permitiendo comparativas.

1.2 Aerosol Optical Depth (AOD)

El espesor óptico de aerosoles o Aerosol Optical Depth (AOD) es una medida de la extinción de la radiación solar debida a las partículas en suspensión en la atmósfera llamadas aerosoles como lo son el polvo, el humo, la arena... El AOD nos permite saber cuánta luz solar está siendo disipada o absorbida por estas partículas. Es una magnitud adimensional relacionado con la cantidad de aerosol que se encuentra en una columna vertical de la atmósfera desde el punto de medición. Un valor de 0.01 indica que la atmósfera está muy limpia mientras que un valor de 0.4 indica que hay condiciones muy brumosas [1]. Por ejemplo en la Ilustración 1 se puede apreciar una tormenta de arena salida del desierto del Sahara.

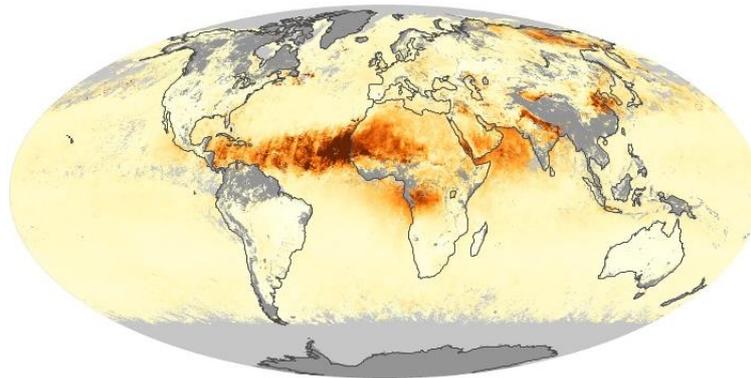


Ilustración 1 - Foto junio 2020 - tormenta de polvo sahariano [2]

La medida de AOD es ampliamente reconocida como un indicador crítico para comprender la calidad del aire de una región gracias a su capacidad de cuantificar el aerosol cargado en la atmósfera. Actualmente existe un interés por diseñar nuevos satélites y algoritmos que puedan mejorar las mediciones de AOD para cuantificar las concentraciones de partículas de una mejor forma y así poder hacer estudios de calidad del aire, estudios de cambio climático y evaluaciones de la salud ambiental. Las tecnologías actuales no permiten estudios de salud muy precisos a escala urbana [3].

1.3 Aerosol Robotic Network (AERONET)

La red robótica de aerosoles o Aerosol Robotic Network (AERONET) es la red de medidas de aerosoles a nivel mundial y por eso son los datos que usaremos en este proyecto. El programa AERONET está establecido por la NASA y PHOTONS expandido por otras redes y colaboradores de agencias estatales, institutos, universidades e individuos. Lleva proporcionando mediciones más de 25 años, actualmente se encuentra en la versión 3 del proceso. Esta versión cuenta con un algoritmo automático que permite hacer el control de calidad de los datos antes de su publicación casi en tiempo real. [4]

Las mediciones realizadas desde el suelo son consideradas las más precisas a la hora de medir AOD a cambio de tener una menor cobertura espacial. Cuentan con una calibración que permite obtener como resultado mediciones muy precisas. Mientras tanto las mediciones realizadas desde otras plataformas como los satélites tienen una mayor cobertura espacial a cambio de tener una mayor incertidumbre. Los satélites tienen múltiples problemas como determinar el AOD cuando tiene valores muy elevados, adyacencia de nubes, reflectancia de la superficie o topografía muy variante. AERONET se compone de más de fotómetros operados autónomamente utilizados para medir la irradiancia de haz directo colimado por el sol y la radiación direccional del cielo [5]. En la Ilustración 2 se puede ver un dispositivo utilizado en AERONET para medir AOD



Ilustración 2 - Cimel CE318-T [6]

El Grupo de Óptica Atmosférica de la Universidad de Valladolid (GOA-UVa) ha desarrollado una herramienta llamada CÆLIS con el objetivo de facilitar el manejo de los datos y la investigación de los componentes atmosféricos. Uno de sus proyectos llamado RIMA-

AERONET con el cual proporcionan datos en tiempo real del estado de fotómetros, documentar los fallos y diagnósticos de instrumentos y proporcionar acceso a los datos de AERONET. Es a través de la base de datos de CÆLIS que se van a obtener los datos utilizados en el proyecto.[7]

1.4 Objetivos del proyecto

Este proyecto se plantea en el ámbito de realizar un trabajo de fin de máster por lo que los objetivos son los planteados por los profesores del Grupo de Óptica Atmosférica. El proyecto tiene dos objetivos principales:

Objetivo-1: el primer objetivo es el de realizar una investigación del estado del arte de los métodos de interpolación disponibles en el entorno científico de Python. Esto comprende un análisis detallado de los métodos de interpolación disponibles al mismo tiempo que un análisis de las librerías en Python que sean aplicables al problema de interpolación de mapas de espesor óptico de aerosoles (AOD). El objetivo es valorar la precisión de estos métodos, su eficiencia y el grado de adaptación al problema que se plantea prestando una especial atención al rendimiento temporal obtenido.

Objetivo-2: el segundo objetivo es diseñar e implementar un programa que utilice los métodos de interpolación más eficientes encontrados en el objetivo anterior. Este programa será utilizado para interpolar datos de espesor óptico de aerosoles (AOD) creando mapas que representen visualmente la información proporcionada y la predicha. Siendo el objetivo de estos mapas que se puedan utilizar para estudios de AOD en casos de uso reales de monitorización e investigación atmosférica.

Para realizar el proyecto se cuenta con una serie de restricciones que vienen dadas por la naturaleza de ser un proyecto realizado para un trabajo de fin de máster y debido al entorno para el que se destina el resultado del proyecto:

Restricción-1: El proyecto debe ser un código realizado en el lenguaje de programación Python debido al entorno en el que se trabaja en el departamento.

Restricción-2: El alcance y duración del proyecto deben limitarse a la carga de trabajo establecida por la guía docente de la asignatura Trabajo Fin de Máster (150 horas).

2: Planificación del proyecto

En este apartado se describirá la planificación que se ha realizado para garantizar que la entrega del proyecto se realice durante el curso 2023/2024, conteniendo un diagrama con la división del tiempo y las fases en las que se divide el proyecto. Finalmente se describirá el presupuesto utilizado para el proyecto, teniendo en cuenta tanto los recursos técnicos como los humanos

2.1 Planificación

La planificación de este proyecto se plantea con el objetivo de ser entregada en una convocatoria del curso 2023/2024 y teniendo en cuenta la restricción de horas descrita en el apartado anterior. Para planificar la ejecución del proyecto se ha utilizado una metodología de cascada. Se ha considerado que este es el modelo más apropiado para un proyecto de esta naturaleza.

El modelo de cascada consiste en la planificación y ejecución de las distintas fases del proyecto de una forma lineal y secuencial apoyándose en las fases anteriores del proyecto. La utilización del modelo de cascada para un trabajo de este tipo proporciona un enfoque claro y estructurado idóneo para asegurarse de que todos los aspectos del proyecto están planificados y son ejecutados. Cada fase se construye sobre la anterior reduciendo el riesgo de que aparezcan problemas generados por el planteamiento en las últimas fases del proyecto. Este modelo es particularmente efectivo en trabajos de investigación con requisitos claramente definidos como es este trabajo de fin de máster.

Los actores identificados para la elaboración de este proyecto son por una parte los profesores e investigadores y por otra parte el alumno que ha seleccionado el proyecto. Sus roles son los siguientes:

- Profesores e investigadores: su papel consiste en realizar el planteamiento y la propuesta del proyecto, planteando los objetivos y realizando un seguimiento periódico. Al mismo tiempo tienen el papel de clientes que determinan la validez de la solución encontrada.
- Alumno: el papel del alumno es el de realizar la investigación, trabajo y memoria siguiendo las directrices proporcionadas por los profesores e investigadores preocupándose por que se alcancen los objetivos propuestos a tiempo.

Para la planificación es necesario aclarar que, a diferencia de la mayoría de los trabajos de fin de máster, el proyecto realizado durante la estancia en un grupo de investigación no guarda relación alguna con este trabajo por lo que el proyecto parte totalmente de cero y se cuentan únicamente con las horas dedicadas a la asignatura de Trabajo de Fin de Máster.

Este proyecto se asignó durante julio de 2023 pero ante el parón de verano y la falta de tiempo del alumno durante los siguientes meses no se comenzó a trabajar en él dedicando el tiempo necesario hasta abril de 2024 en donde se comenzó el proyecto prácticamente de cero por lo que para la planificación no se tendrá en cuenta el periodo anterior.

Durante el proyecto se siguió una planificación que se puede dividir en 3 fases diferenciadas a parte del parón que se produjo durante el mes de agosto debido a las vacaciones tanto de los profesores como del alumno.

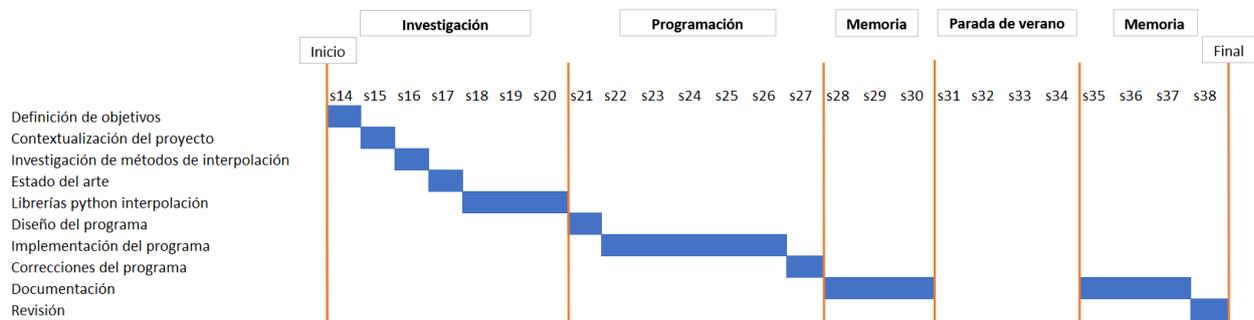


Ilustración 3 - Diagrama de Gantt con la programación seguida

En el diagrama de la Ilustración 3 se pueden diferenciar las 3 fases, estando la fase de Memoria interrumpida por la parada de verano, por las divisiones realizadas por las líneas verticales naranjas. Las fases de desarrollo del proyecto se desglosan en los siguientes subprocesos:

Fase 1 (investigación):

- Definición de objetivos

- Contextualización del proyecto
- Investigación de métodos de interpolación
- Estado del arte
- Librerías Python de interpolación

Fase 2 (Programación):

- Diseño del programa
- Implementación del programa
- Correcciones del programa

Fase 3 (Memoria):

- Documentación
- Revisión

2.2 Materiales y presupuesto

El presupuesto asociado al proyecto se divide en dos apartados: recursos técnicos y recursos humanos. Por un lado, en los recursos técnicos se considerarán aquellos sistemas, herramientas y objetos necesarios para la realización del proyecto. Por otro lado, dentro de los recursos humanos se considerará el coste asociado a los trabajadores.

2.2.1 Recursos técnicos

Para la elaboración de este proyecto se ha hecho uso de un ordenador portátil con las siguientes características:

- Procesador: Intel Core i5-1145G7 11th Gen @2.60GHz
- RAM: 8GB DDR4 3200 MHz
- Disco: SSD 250GB

El coste de este ordenador portátil es de aproximadamente 850€ y su vida útil se puede estimar a 4 años. Adicionalmente se precisa de una conexión a internet para realizar la investigación además de las consultas a la base de datos que se realizan en remoto. Se ha utilizado una conexión de 1Gb simétrico, el precio de una conexión de estas características es de aproximadamente 35€ mensuales.

En cuanto a las herramientas y software utilizado, para desarrollar el proyecto se ha utilizado un conjunto básico de herramientas para el desarrollo de software en Python. Se

han utilizado las siguientes:

- Microsoft Windows 11
- Herramientas Office 365
- Python
- Visual Studio Code
- Librerías Python gratuitas

La licencia de Windows 11 y la licencia de las herramientas de Office 365 normalmente tienen un coste asociado, pero en este caso el ordenador portátil incluye de fábrica este software sin un coste adicional. Las herramientas de programación utilizadas tienen un coste 0 ya que se ha optado por elegir las gratuitas.

	Coste	Porcentaje de uso mensual	Meses	Total
Ordenador	850 €	2,08%	4,5	79,68 €
Internet	35 €	100%	5	175 €
Windows 11	0 €	100%	4,5	0 €
Office 365	0 €	100%	4,5	0 €
Python	0 €	100%	4,5	0 €
Visual Studio Code	0 €	100%	4,5	0 €
Librerías Python	0 €	100%	4,5	0 €
				254,68 €

Tabla 1 - Coste de recursos técnicos

2.2.2 Recursos humanos

Para la realización de un proyecto como el planteado, es necesario contar con el perfil de un programador y el de un investigador. Estos perfiles en España tienen un sueldo medio de 28.200€ anuales [8] para el programador y de un investigador es de 25.277€ mensuales [9] . Adicionalmente se debe tener en cuenta el pago de la Seguridad Social por la contratación de trabajadores, este coste supone aproximadamente un 30% adicional al salario bruto Suponiendo que se ha dedicado un número igual de horas en cada semana de las que se compone la planificación se obtiene la siguiente tabla:

	Salario por hora	Horas totales	Total bruto	Total SS	Total
Programador	14,68 €	51	748,68 €	224,61 €	973,29 €
Investigador	13,17 €	99	1.303,86 €	391,16 €	1.695,01 €
					2668,30

Tabla 2 - Coste de recursos humanos

2.2.3 Presupuesto final

En vista de los totales recogidos en las tablas, el presupuesto final obtenido para el proyecto es de 2922,98 €.

3: Estado del arte

Este apartado tiene como objetivo realizar una introducción a los métodos de interpolación espacial explicando los métodos fundamentales. A continuación, se realiza un análisis de cuatro artículos sobre interpolación espacial en el campo de meteorología. Por último, se realiza la selección de los métodos de interpolación a utilizar en el estudio.

3.1 Introducción a la interpolación espacial

En el campo de análisis numérico se denomina interpolación espacial a la obtención de nuevos puntos partiendo del conocimiento de un conjunto de puntos [10]. Este tipo de estimación es particularmente útil para generar datos continuos en problemas temporales o espaciales. En la ciencia atmosférica las mediciones se realizan en puntos concretos del mapa, pero se desea conocer los valores de la variable medida en todo el mapa. La interpolación es esencial para crear estos modelos de datos continuos a partir de los puntos medidos. Esto permite crear mapas de variables continuas como precipitaciones, temperatura o AOD dando una ventaja a la hora de tomar decisiones.

Existen múltiples métodos que han sido creados a lo largo de la historia, este apartado se centrará en realizar una recopilación de los métodos más comúnmente utilizados en estudios ambientales, explicando brevemente en qué consisten. Los métodos de interpolación espacial se pueden clasificar en 2 tipos: métodos no geoestadísticos o deterministas y métodos geoestadísticos.

3.1.1 Métodos de interpolación no geoestadísticos

En este apartado se van a describir dos métodos de interpolación no geoestadísticos. Estos métodos se caracterizan por ser deterministas, es decir, crean superficies a partir de puntos medidos, basándose en el grado de similaridad a otros puntos cercanos o en el grado de suavizado [11].

3.1.1.1 Nearest Neighbour (NN)

El Método de Nearest Neighbour (NN), o vecino más cercano, predice el valor de un punto sin muestrear basándose en el valor del punto muestreado más cercano. Al dibujar la bisectriz perpendicular de las líneas entre los puntos muestreados se forman polígonos de Thiessen (ver Ilustración 4). Esto produce un polígono por punto muestreado, quedando el punto en el centro, de forma que dentro del polígono todos los puntos están más cerca de su punto muestreado que de cualquier otro punto muestreado que pertenecerá a otro polígono. De este modo todos los puntos no muestreados que se quieran estimar pertenecerán a un polígono y obtendrán el valor del punto muestreado más cercano.

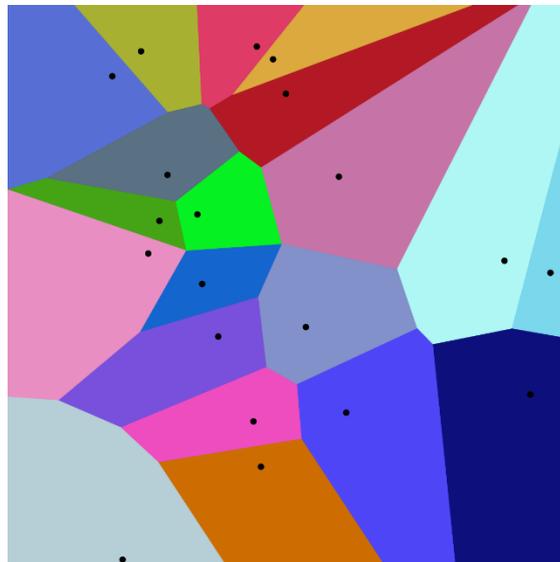


Ilustración 4 - Polígonos de Thiessen [12]

3.1.1.2 Inverse Distance Weighting (IDW)

Inverse Distance Weighting (IDW) o distancia inversa ponderada estima los valores de puntos sin muestrear utilizando una combinación lineal de valores en puntos muestreados ponderando con una función inversa de la distancia desde el punto de interés a los puntos muestreados. Este método asume que los puntos muestreados más cercanos al punto no muestreado que se quiere estimar son más similares que los puntos más lejanos y por ello tienen más peso.

$$\lambda_i = \frac{1/d_i^p}{\sum_{i=1}^n 1/d_i^p}$$

Ecuación 1 - Inverse Distance Weighted

Donde d_i es la distancia entre x_0 y x_i , p es un parámetro de potencia y n representa el número de puntos muestreados utilizados para la estimación. El mayor factor que afecta a la precisión de IDW es el parámetro de potencia. El peso del punto x_i disminuye al aumentarse la distancia, especialmente con valores del parámetro de potencia altos. De esta forma los puntos muestreados más cercanos tomarán más peso e influenciarán más la estimación.

La elección del parámetro de potencia es arbitraria. La elección más popular de p es 2 resultando en la utilización del método de inversa del cuadrado de la distancia. A medida que p aumenta solamente los valores muy próximos al valor estimado influenciarán en la predicción.

3.1.2 Métodos de interpolación geoestadísticos

En este apartado se van a describir 4 métodos de interpolación geoestadísticos. Estos métodos se caracterizan por ser deterministas, es decir, crean superficies a partir de puntos medidos, basándose en el grado de similaridad a otros puntos cercanos o en el grado de suavizado [11].

Antes de comenzar a describir los métodos e interpolación geoestadísticos es importante conocer el concepto de semivarianza [13]. La semivarianza (γ) de Z entre dos puntos se define con la Ecuación 2.

$$\gamma(x_i, x_0) = \gamma(h) = \frac{1}{2} \text{var}[Z(x_i) - Z(x_0)]$$

Ecuación 2 - Ecuación de semivarianza entre dos puntos

Donde h es la distancia entre el punto x_i y x_0 siendo $\gamma(h)$ el variograma. Una representación de $\gamma(h)$ respecto de h se conoce como el variograma experimental. Este variograma muestra varias características importantes que se pueden ver representadas en la Ilustración 5:

- **Nugget:** un valor positivo de $\gamma(h)$ cercano a 0 que muestra el error residual proveniente del error inherente a las herramientas de medición y de variaciones más pequeñas al intervalo en el que se toman las medidas.
- **Range:** valor de h en el que los valores del semivariograma se estabilizan.
- **Sill:** Valor del modelo del semivariograma en el que se alcanza el “Range”.

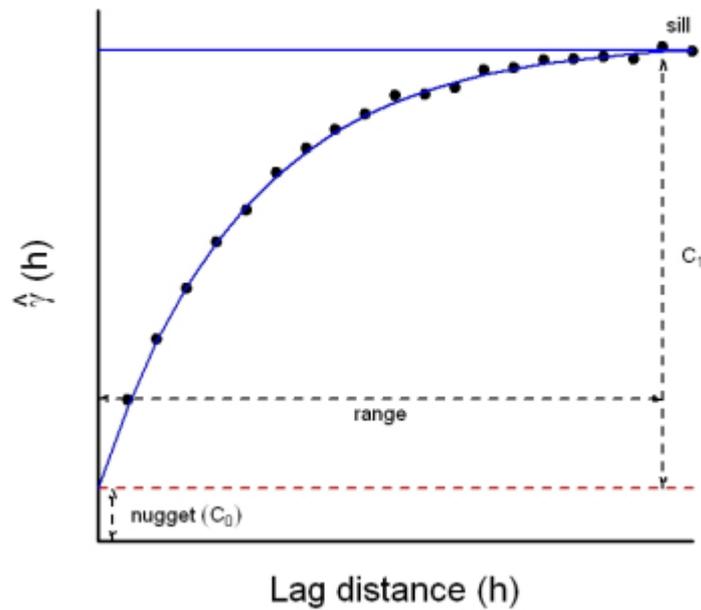


Ilustración 5 - Ejemplo variograma con Nugget, Sill y Range

En los métodos de interpolación geoestadísticos los semivariogramas representan la autocorrelación espacial entre los puntos muestreados.

La semivarianza puede ser estimada a partir de los datos utilizando la Ecuación 3.

$$\hat{\gamma}(h) = \frac{1}{2n} \sum_{i=1}^n (z(x_i) - z(x_i + h))^2$$

Ecuación 3 - Ecuación de semivarianza a partir de datos

Donde n es el número de pares de puntos muestreados separados por la distancia h. El modelado del variograma es muy importante para la interpolación espacial. Cuatro modelos comúnmente utilizados pueden ser vistos en la Ilustración 6.

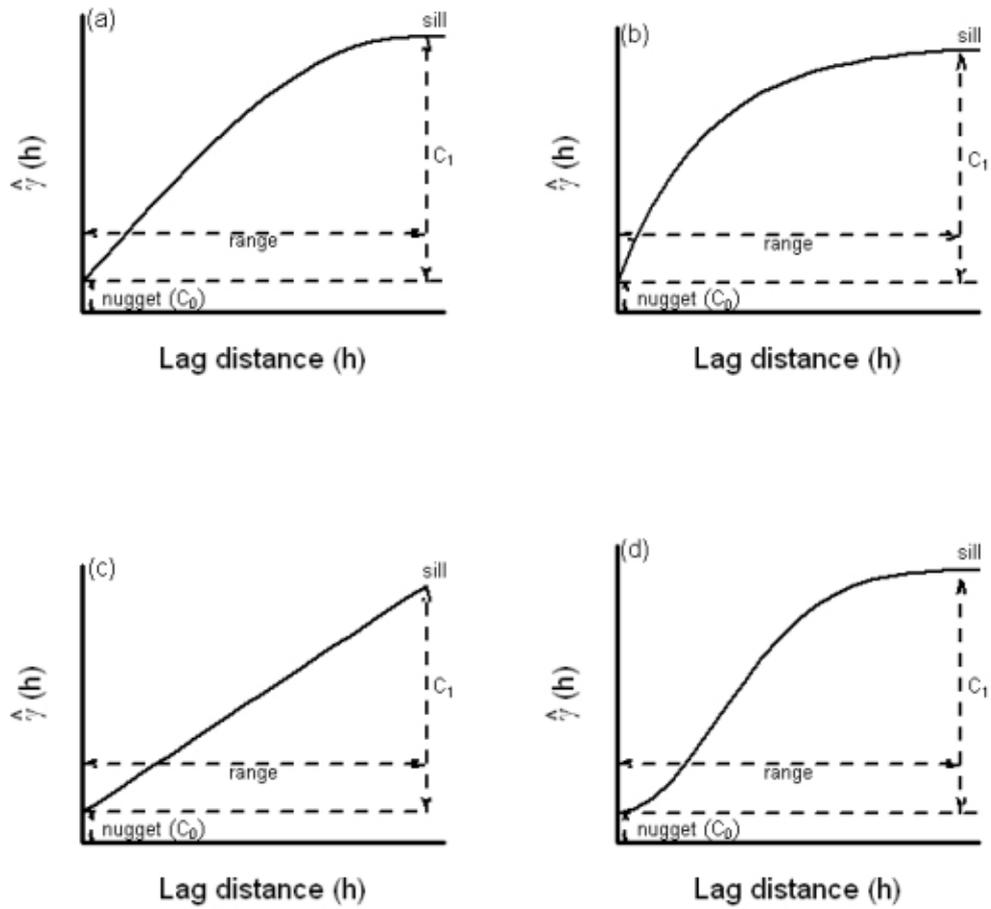


Ilustración 6 - Ejemplos de 4 modelos de variograma comúnmente usados (a) esférico; (b) exponencial; (c) lineal; (d) Gaussiano

3.1.2.1 Simple Kriging (SK)

La estimación de *Simple Kriging* (SK) está basada en la ecuación básica de varianza, dando como resultado la Ecuación 4.

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) + \left[1 - \sum_{i=1}^n \lambda_i \right] \mu$$

Ecuación 4 - Ecuación Simple Kriging

Donde μ es una media estacionaria conocida. El parámetro μ se da por constante en todo el rango de datos y es calculado como la media de los datos. SK es utilizado para estimar los residuos de este valor μ dado en una primera instancia y por ello se le conoce como “kriging con una media conocida”. El número de puntos utilizado para calcular la estimación en la ecuación es determinado por el rango de influencia del semivariograma utilizado.

3.1.2.2 Ordinary Kriging (OK)

El método de *Ordinary Kriging (OK)* es similar al *Simple Kriging (SK)* con la diferencia de que *OK* hace estimaciones utilizando una media no conocida para cada ventana de interpolación. Utiliza la Ecuación 5

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$$

Ecuación 5 - Ecuación Ordinary Kriging

La ecuación es similar a la que utiliza *SK* con la diferencia de que $[1 - \sum_{i=1}^n \lambda_i] = 0$ al asegurarse que la suma de los pesos $[\sum_{i=1}^n \lambda_i] = 1$.

3.1.2.3 Universal Kriging (UK)

El *Universal Kriging (UK)*, también conocido como Kriging con una tendencia, es una extensión del *Ordinary Kriging* que incorpora un componente de tendencia determinista en los modelos espaciales. A diferencia de *OK* que assume una media constante, este método utiliza una media en función de las coordenadas espaciales.

3.1.2.4 Cokriging (CK)

El cokriging es una forma de interpolación que a diferencia de *SK*, *OK* o *UK*, necesita la tener disponible información adicional. Es necesario conocer los valores de variables secundarias para cada localización muestreada para poder utilizar *cokriging*. Este método de interpolación aprovecha información secundaria, que no es necesariamente exhaustiva, para obtener una correlación cruzada entre la variable primaria y las variables secundarias utilizando la Ecuación 6.

$$\widehat{Z}_1(x_0) - \mu_1 = \sum_{i_1=1}^{n_1} \lambda_{i_1} [Z_1(x_{i_1}) - \mu_1(x_{i_1})] + \sum_{j=2}^{n_y} \sum_{i_j=1}^{n_j} \lambda_{i_j} [Z_j(x_{i_j}) - \mu_j(x_{i_j})]$$

Ecuación 6 - Ecuación Cokriging

En donde μ_1 es una media estacionaria conocida de la variable primaria, $Z_1(x_{i_1})$ es el dato de la variable primaria en el punto i_1 , $\mu_1(x_{i_1})$ es la media de muestras en la ventana de búsqueda, n_1 es el numero de puntos muestreados en la ventana de búsqueda para el punto x_0 utilizado para hacer la estimación, λ_{i_1} es la ponderación seleccionada para minimizar la estimación de la varianza de la variable primaria, n_y es el numero de variables secundarias, n_j es el número de variables secundarias j -ésimas en la ventana de búsqueda, λ_{i_j} es el peso asignado al elemento i_j -ésimo de la variable secundaria j -ésima, $Z_j(x_{i_j})$ es el valor en punto i_j -ésimo de la variable j -ésima y por último $\mu_j(x_{i_j})$ es la media de las muestras de la variable j -ésima en la ventana de búsqueda.

Además, al igual que en el kriging, existen métodos de interpolación de cokriging con variaciones similares pudiendo encontrar *Simple Cokriging (SCK)* o *Ordinary Cokriging (OCK)*

3.2 Análisis de artículos

En este apartado se analizan varios documentos estado del arte sobre interpolación espacial en el campo de la meteorología. Se ha elegido este campo debido a que el AOD se puede considerar que actúa como una variable meteorológica al estar afectada por la circulación y el movimiento de las grandes masas de aire a nivel global.

3.2.1 Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature

La primera propuesta analizada [14] tiene como objetivo resolver el problema de la calidad de la información y sobre todo de la cobertura alcanzada por los medidores al ser estaciones dispersas en el espacio. Para tratar de resolver este problema se emplean métodos de interpolación basándose en la temperatura y en la precipitación anual del área del valle del Ebro, lugar en el que esta información es de gran utilidad para la agricultura y la administración de

agua.

Para realizar las predicciones se utilizan métodos globales (modelos de regresión utilizando variables geográficas y topográficas como predictores y tendencias en superficies de diferente orden), métodos locales (polígonos de Thiessen, distancia inversa ponderada (IDW) y splines), métodos geoestadísticos (ordinary kriging, simple kriging, directional kriging, block kriging, universal kriging y co-kriging) y métodos mezclados que combinan métodos, globales y locales o geoestadísticos.

Para el estudio se utilizan datos de 380 estaciones de precipitaciones y de 222 estaciones de temperatura entre los años 1950 y 2000, datos proporcionados por el Instituto Nacional de Meteorología. Para la validación de los datos se utiliza Root Mean Square Error (RMSE) o raíz del error cuadrático medio y Mean Absolute error (MAE) o error absoluto medio.

Los resultados de este estudio muestran que en el caso de los mapas de precipitaciones tanto los métodos globales como los locales proporcionan unos resultados poco reales dando como resultado un error estadístico (RMSE y MAE) elevados, quedando los métodos geoestadísticos como los más precisos. En el caso de los mapas de temperatura los métodos que muestran el menor error estadístico (RMSE y MAE) son IDW seguido de cerca por los métodos de kriging.

El estudio finaliza afirmando que la conclusión que se debe obtener es que es necesario que se prueben distintos métodos de interpolación para cada mapa concreto ya que dependiendo de una gran cantidad de variables un método u otro será el más apropiado para cada caso.

3.2.2 Spatial Analysis of Air Temperature in the Qinghai-Tibet Plateau

La segunda propuesta analizada [15] tiene como objetivo estimar la distribución espacial de la temperatura del aire media en la meseta de Qinghai-Tibet. La temperatura del aire superficial es una medida importante para los estudios climáticos.

En este estudio se utilizan cuatro métodos de interpolación diferentes, dos de ellos pueden incorporar variables secundarias que pueden ayudar a mejorar las predicciones dadas las condiciones de irregularidad del terreno de la zona utilizada. Los dos métodos de una variable elegidos son uno no geoestadístico, IDW y otro geoestadístico, ordinary kriging. Los métodos con variable secundaria son cokriging y una mezcla de ordinary kriging con una corrección del efecto de la altitud en la temperatura.

Los datos utilizados para este estudio son proporcionados por 173 estaciones meteorológicas localizadas en la meseta de Qinghai-Tibet y alrededores. Estos datos son del rango de años de 1961 a 1990 para así cumplir con los requerimientos de la World Meteorological Organization

(WMO) para establecer una base climática fiable. Adicionalmente se cuenta con un modelo de elevación digital (DEM) de la zona.

Para evaluar los resultados de este estudio se han utilizado dos técnicas. La primera consiste en un análisis subjetivo. Es muy importante y a veces es la única forma de evaluar los resultados de la interpolación en una región vacía. Es necesario que las personas que hacen esta evaluación sean expertas en la zona geográfica. El segundo método utilizado es el de la varianza de las interpolaciones pese a no ser un método absoluto y no estar ligado a los datos utilizados.

Los resultados de este estudio muestran que con el uso de IDW los resultados obtenidos están lejos de la realidad ya que hay zonas del mapa con valores muy lejanos a la realidad por los cambios de altura. Con ordinary kriging los resultados mejoran al verse que este método si comprende mejor las transiciones de los cambios de temperatura. En el caso de cokriging, al incorporar la variable de altitud los resultados mejoran y obtiene mejores predicciones. Sin embargo, al utilizar datos con una altitud limitada, las predicciones en zonas de mayor altitud no son muy acertadas. Los mejores resultados se obtienen con el uso del método combinado de ordinary kriging con la correlación del efecto de la altitud en la temperatura al ajustarse mejor a los cambios de altura.

La conclusión obtenida en el estudio es que el problema de interpolar la temperatura del aire en Qinghai-Tibet no puede ser resuelto utilizando simplemente interpolación con una única variable como en IDW y ordinary kriging. Es necesario utilizar métodos que incorporen variables secundarias, en este caso que incluyan la altura para poder obtener resultados más precisos que reflejen mejor la realidad de la zona de interés.

3.2.3 A Comparison of Spatial Interpolation Techniques in Temperature Estimation

Este documento [16] tiene como objetivo realizar una comparación entre métodos de interpolación para la estimación de temperaturas ya que hasta este momento pocos estudios se habían centrado en este problema. Los objetivos principales son ver la eficacia de cada método de interpolación para predecir temperatura y determinar si algún atributo temporal o espacial afecta especialmente a los métodos de interpolación.

Los métodos de interpolación seleccionados para este estudio se dividen en geoestadísticos y no geoestadísticos. En el grupo de los geoestadísticos se encuentran inverse distance squared (IDS) y optimal inverse distance (ODA) o distancia inversa optima, trend surface analysis o análisis de superficie de tendencia, polynomial regression o regresión polinomial, cubic splining o esplines cúbicos y lapsed rate method o método basado en tasa de caída adiabática. Por otra parte, como método geoestadístico se utiliza el kriging y el cokriging.

Se utilizan los datos de Norte América dividiéndolo en dos regiones, este y oeste. Se toman los datos de temperatura máxima y de temperatura mínima en una franja de 10 años con escalas temporales diarias, estacionales y anuales.

Cada método de interpolación se compara utilizando mean absolute error (MAE) o error absoluto medio y mean squared error (MSE) error cuadrático medio. Para realizar la comparación de estadísticas obtenidas se utiliza validación cruzada de forma que un punto se retira del conjunto de datos, con el método de interpolación se predice el valor en el punto retirado y finalmente se compara el resultado obtenido con el valor real del punto.

Los resultados muestran que cuando la varianza de los datos es baja todos los métodos de interpolación obtienen unos resultados similares, pero según la varianza de la temperatura va incrementando se puede observar cómo comienzan a destacar dos métodos de interpolación, ODA y kriging.

La conclusión extraída en este estudio es que, para el conjunto de datos dado, el método de interpolación que obtiene el menor MAE es optimal inverse distance seguido de kriging. A pesar de obtener el menor MAE, ODA no es el método de interpolación visiblemente más plausible, en este aspecto el kriging es superior al presentar mapas más precisos a primera vista.

3.2.4 Interpolation Techniques for Climate Variables

El objetivo de este documento [17] es realizar una comparación de métodos de interpolación para paliar el hecho de que las mediciones meteorológicas se toman desde unos pocos puntos en el mapa por lo que no en todas las localizaciones existe una estación meteorológica de la que se puedan obtener datos reales. Este estudio tiene como objetivo encontrar el método o métodos de interpolación más adecuados para predecir precipitaciones y temperatura máxima.

Para este estudio se han considerado tres métodos de interpolación: inverse distance weighted averaging (IDWA) o promedio ponderado del inverso de la distancia, thin plate smoothing splines o suavizado de splin de placa delgada y cokriging.

Los datos utilizados provienen del estado de Jalisco, Méjico. Es un área de aproximadamente 20.000 km^2 que cuenta con unas 200 estaciones que han generado los datos utilizados en este estudio entre 1940 y 1990. Adicionalmente se cuenta con un modelo digital de elevación basado en celdas de 1 km^2 que se utiliza como variable secundaria para el método de interpolación de cokriging.

La precisión de los métodos de interpolación se mide utilizando validación cruzada. Para ello se dejan 25 estaciones meteorológicas fuera del método de interpolación (20% del total de

estaciones) y se compara el valor generado con el método de interpolación elegido con el real y se hace una diferencia porcentual de los valores.

Los resultados muestran que el cokriging es el método de interpolación más preciso cuando la variable secundaria tiene una alta correlación con las precipitaciones, esto es durante el verano. Mientras tanto el resto del año el suavizado de splin de placa delgada obtiene unos resultados más precisos. En cambio, para la temperatura, al mantener una mayor correlación con la altitud, el método de interpolación de cokriging proporciona mejores resultados.

La conclusión que se obtiene en este documento es que el uso de splines y de cokriging es más recomendado que el de IDWA porque proporcionan una predicción del error que describen la calidad de la predicción. El uso de los splines es más recomendado que el de cokriging ya que en los meses en los que no hay una alta correlación de la altura con los valores medidos obtiene unas predicciones más precisas. Por último, se señala que todas las técnicas usadas mejorarían con más estaciones meteorológicas realizando mediciones.

3.2.5 Discusión

Se presenta en esta sección una breve discusión de los trabajos analizados en donde se comparan los trabajos estudiados por los métodos de interpolación utilizados, la tipología de los datos utilizada, las métricas utilizadas para valorar el mejor método de interpolación y por último el método o los métodos de interpolación elegidos. Esta comparación se muestra en la Tabla 3.

Documento	Métodos de interpolación	Métricas	Método de interpolación elegido
4.3.1	Polígonos de Thiessen, IDW, splines, ordinary kriging, simple kriging, directional kriging, block kriging, universal kriging y cokriging	RMSE y MAE	IDW y kriging
4.3.2	IDW, ordinary kriging, cokriging y ordinary kriging con correlación	Análisis subjetivo y varianza	Ordinary kriging con correlación
4.3.3	IDS, ODA, trend surface analysis, lapsed rate method, kriging y cokriging	MAE, MSE y validación cruzada	ODA y kriging
4.3.4	IDWA, cokriging y thin plate smoothing splines	Validación cruzada	Splines y kriging

Tabla 3 - Resumen de métodos de interpolación utilizados y elegidos en los artículos analizados

3.3 Selección de métodos de interpolación

La selección de un método de interpolación adecuado para los datos disponibles es crucial, pero no es una tarea sencilla. El rendimiento de los métodos de interpolación depende de muchos factores por lo que no hay una elección directa que funcione para todos los casos. Se deben considerar varios factores para la elección del método de interpolación, la naturaleza de los datos de la variable primaria, tamaño de la muestra, distribución, disponibilidad de variables secundarias entre otros factores que son importantes a la hora de elegir el método de interpolación más adecuado. Incluso la disponibilidad del software y los recursos computacionales son factores a tener en cuenta.

De los trabajos analizados en el anterior apartado se pueden sacar varias conclusiones que es recomendable tener en cuenta a la hora de escoger el método de interpolación:

- 1- Se debe probar con más de un método de interpolación ya que no existe un método de interpolación que sea el mejor en todos los casos.
- 2- Tener variables secundarias resulta de gran ayuda y generalmente mejora la precisión de los métodos de interpolación.
- 3- Existe un cierto factor subjetivo a la hora de elegir el método de interpolación ya que los métodos deben ser visiblemente plausibles y en ocasiones un criterio de medición es

un ojo experto.

- 4- Dependiendo del momento del año del que sean los datos hay ocasiones en las que es mejor un método que otro.
- 5- Tanto los métodos de interpolación geoestadísticos como los no geoestadísticos y sus combinaciones pueden resultar ser el método de interpolación óptimo para el conjunto de datos.
- 6- Se deben tener en cuenta la disponibilidad del software con los métodos de interpolación y los recursos computacionales disponibles.

Visto que la elección de los métodos de interpolación es una decisión importante al igual que dificultosa, es conveniente consultar el mayor número de fuentes posibles para tener la mejor información posible para tomar la decisión. Es por ello que se decide consultar adicionalmente una revisión de métodos de interpolación aplicados en ciencias ambientales [18] como apoyo. En esta revisión se hace una comparación exhaustiva entre 38 métodos de interpolación. En esta comparación se hacen descripciones y caracterizaciones de todos los métodos para finalmente proponer un árbol de decisión para elegir un método de interpolación apropiado para el conjunto de datos disponible.

Siguiendo el árbol de decisión propuesto en este estudio, el cual se puede ver en la Ilustración 7, se obtiene la siguiente propuesta:

- 1 los datos muestran una estructura espacial o un variograma no lineal.
 - 2 estimación de una variable continua.
 - 3 no se cuenta con información de variables secundarias.
 - 4 media global conocida
 - 5 Simple Kriging (SK)

La propuesta resultante es el uso de simple kriging como método de interpolación. Siguiendo las conclusiones obtenidas del análisis de los trabajos del anterior apartado se llega a la conclusión de que se deben utilizar métodos de interpolación tanto de kriging como no geoestadísticos.

1	Data or residuals show spatial structure or a non-linear variogram	
2	Estimation of continuous variable	
3	No information of secondary variables available	
4	Global mean known	SK
4*	Global mean unknown and using local means	OK
3*	Information of secondary variables available	
5	Global mean known	
6	Secondary variable is only categorical	
7	Stratification	SKWS
7*	Non-stratification	SKIm
6*	Secondary variable is not only categorical	
8	Stratification	SCKWS
8*	Non-stratification	
9	Sparse samples of secondary variable and multiple samples in search window	SCK
9*	Dense samples of secondary variable and single sample in search window	SCCK
5*	Global mean unknown and using local means	
10	Secondary information available for each point being estimated	
11	Spatial trend is apparent and only coordinates available	UK
11*	Other secondary variable available	
12	An apparent global relation with the secondary variable	SKIm
12*	The relation is not so apparent	KED
10*	Secondary information not available for each point being estimated	
13	Secondary variables including a categorical variable	
14	Only a categorical variable available	
15	Multiple samples in search window	OKWS
15*	Dense samples of secondary variable and single sample in search window	OCCK
14*	Other secondary information available	OCKWS
13*	Secondary variables without categorical variable	
16	Sparse samples of secondary variable and multiple samples in search window	
17	Many secondary variables and PCA needed	PCK
17*	PCA not needed to reduce the number of secondary variables	
18	Avoid negative weights and artificially limiting the effect of secondary variable	SOCK
18*	Accept above two drawbacks	OCK
16*	Dense samples of secondary variable and single sample in search window	OCCK
2*	Estimation of categorical variable or uncertainty assessment	IK & its variants
1*	Data or residuals show no spatial structure or linear variogram or sample size is too small to derive a reliable variogram	
19	No secondary variables available	
20	Abrupt estimation acceptable	
21	Using single sample for estimation	NN
21*	Using multiple samples for estimation	
22	Using three samples for estimation	TIN
22*	Using more than three natural neighbour samples for estimation	NaN
20*	Abrupt estimation unacceptable	
23	Using more than three natural neighbour samples weighted by area	NaN
23*	Using nearest several samples weighted by distance	IDW
19*	Secondary variable available	
24	Using information of coordinates	
25	Only coordinates information used with inexact estimation	
26	Using nearby samples	Splines & LTS
26*	Using all samples	TSA
25*	May use other variables with exact or inexact estimation	TPS
24*	Not using information of coordinates	
27	Only categorical secondary information available	
28	Only one variable available	CI
28*	Multiple variables available	RT
27*	Continuous secondary information available	
29	Univariate or multiple secondary information	LM
29*	Require multiple secondary information	RT

Ilustración 7 - Árbol de decisión para la elección de un método de interpolación [18]

4: Técnicas y herramientas

Este apartado se dedica a la descripción de las librerías seleccionadas de entre las librerías de interpolación disponibles en Python. Por último, se realiza una comparativa entre los resultados obtenidos por cada método de interpolación en cada librería utilizando una métrica que se describe en el capítulo

4.1 Lenguaje utilizado y descripción de librerías seleccionadas

El lenguaje que se va a utilizar en el programa es Python, viene dado como requerimiento del proyecto por lo que el programa se tendrá que adaptar a él. Python cuenta con una serie de ventajas que hacen que trabajar en este lenguaje sea una tarea agradable. Cuenta con una gran variedad de librerías para todo tipo de funcionalidades teniendo la mayoría de ellas un gran soporte de la comunidad. Es un lenguaje versátil y compatible con las plataformas principales (Linux, Windows y macOS).

En Python hay disponibles varias librerías dedicadas a la interpolación espacial que cuentan con métodos de interpolación de kriging entre las que destacan: Pykrige, Tfinterpy, GeostatsPy, Gstatsim y Arcpy.

4.1.1 Pykrige

Esta librería [19] está diseñada para la interpolación utilizando métodos geoestadísticos, más concretamente kriging y algunas de sus variantes. Cuenta con métodos tanto 2D como 3D por lo que es capaz de realizar interpolaciones tanto para geoestadística como para problemas más

complejos que requieran interpolación en tres dimensiones como las utilizadas en los radares.

Los algoritmos de interpolación disponibles son todos geoestadísticos: ordinary kriging tanto 2D como 3D, universal kriging tanto 2D como 3D, regression kriging y classification kriging.

La versión más reciente disponible es la 1.7.2 publicada en mayo de 2024. Es compatible con versiones más recientes que Python 3.5. Entre sus dependencias se encuentran las librerías numpy y scipy.

La instalación se realiza con el comando:

```
pip install pykrige
```

Para hacer uso de la interpolación con los métodos ordinary kriging y universal kriging se deben utilizar los siguientes comandos:

```
pykrige.ok.OrdinaryKriging(x, y, z, variogram_model)
pykrige.ok.UniversalKriging(x, y, z, variogram_model)
```

Estos métodos cuentan con más parámetros de los utilizados que se pueden encontrar en la documentación adjunta, pero estos son los necesarios para que funcionen:

- x (array): Coordenadas en el eje X de los puntos
- y (array): Coordenadas en el eje Y de los puntos
- z (array): Valor de la variable en los puntos
- variogram_model (string): Especifica el tipo de variograma a utilizar, se puede elegir entre: linear, power, gaussian, spherical, exponential y hole-effect. Por defecto se utiliza el tipo lineal.

4.1.2 Tfinterpy

La librería Tfinterpy (TensorFlow-based Interpolation Python package) [20] es una librería diseñada para reducir el tiempo de ejecución, Previamente se habían creado librerías que consiguen este mismo fin utilizando OpenMP o CUDA aprovechándose del paralelismo y de la aceleración por GPU, desafortunadamente estas librerías no se hicieron código abierto. Por el contrario, Tfinterpy es una librería de código abierto basada en TensorFlow por lo que puede aprovechar la aceleración por GPU para reducir los tiempos de ejecución. Adicionalmente cuenta con métodos de interpolación tradicionales.

Esta librería contiene como implementación de método de interpolación no geoestadístico IDW

y como métodos geoestadísticos simple kriging y ordinary kriging.

La última versión disponible de la librería es la 1.1.1 de noviembre de 2022. Es compatible con todas las versiones de Python3. Como dependencias, necesita que estén instaladas las librerías Numpy, Pandas, TensorFlow, Keras, Scipy, VTK y PyQt5.

La instalación se realiza con el gestor de librerías de Python por medio del siguiente comando:

```
pip install tfinterpy
```

4.1.3 Geostatspy

La librería Geostatspy [21] es una implementación en Python de la conocida librería GSLIB programada en Fortran. Esta librería ha sido creada por un profesor de la Universidad de Stanford con el objetivo de evolucionar y pasar a utilizar Python para todo el trabajo que realiza el grupo de investigación y los alumnos en el campo de la geoestadística.

El paquete cuenta con métodos de interpolación geoestadística de kriging tanto en 2D como en 3D. Adicionalmente proporciona herramientas de visualización y de transformación de datos.

La versión más reciente de esta librería es la 0.0.71 de agosto de 2024. Para utilizar esta librería es necesario utilizar Python de versión al menos 3.7.10 debido a las dependencias. La lista de dependencias es la siguiente: numpy, pandas, numba, scipy, matplotlib, pyplot, tqdm y statsmodels.

La instalación de la librería se realiza con el siguiente comando:

```
pip install geostatspy
```

Para hacer uso de la interpolación con el método de kriging se debe hacer uso del siguiente comando:

```
geostats.kb2d(df, xName, yName, zName, por_min,  
             por_max, nx, xmn, xsiz, ny, ymn, ysiz, nxdis, nydis,  
             ndmin, ndmax, radius, ktype, skmean, vario)
```

Este método cuenta con los siguientes parámetros:

- df (dataframe): Dataframe con tres columnas nombradas con las coordenadas x, y además del valor de los puntos z.

- xName (string): Nombre de la columna de la coordenada x en el dataframe.
- yName (string): Nombre de la columna de la coordenada y en el dataframe.
- zName (string): Nombre de la columna z con valor de los puntos en el dataframe.
- por_min (double): Valor mínimo de la columna z que se tendrá en cuenta.
- por_max (double): Valor máximo de la columna z que se tendrá en cuenta.
- nx, xmn, xsiz (double): Definición de la malla en el eje x.
- ny, ymn, ysiz (double): Definición de la malla en el eje y.
- nxdis, nydis (double): Número de puntos de discretización por bloque, si establecido a 1 se realiza point kriging.
- ndmin, ndmax (int): Número de puntos mínimo y máximo que se tendrán en cuenta en cada bloque de kriging.
- radius (int): máximo radio de búsqueda isotrópico.
- ktype (int): tipo de kriging a utilizar, si establecido a 0 simple kriging y si establecido a 1 ordinary kriging.
- Skmean (double): media utilizada en caso de establecer simple kriging como método de interpolación
- vario (GSLIB.variogram): variograma utilizado para la interpolación.

4.1.4 Gstatsim

Gstatsim [22] es un paquete diseñado específicamente para la interpolación geoestadística y la simulación. Está inspirado en recursos geoestadísticos de licencia abierta como GeostatsPy y SciKit-GStat. El objetivo de las funciones implementadas es tratar con grandes datasets realizando validaciones.

El paquete cuenta con métodos de interpolación geoestadística de kriging en sus variantes simple kriging y ordinary kriging. Adicionalmente cuenta con funciones de visualización y análisis de variogramas.

La versión disponible más reciente de esta librería es la 1.0.6 de agosto de 2023. La versión de Python mínima no se especifica, se ha probado con Python 3.10.11 y ha funcionado correctamente. La lista de dependencias es la siguiente: numpy, pandas, scipy, tqdm, sklearn, matplotlib y SciKit-GStat.

La instalación de la librería se realiza con el siguiente comando:

```
pip install gstatsim
```

Para hacer uso de la interpolación geoestadística con el método de kriging se debe hacer uso

del siguiente comando:

```
gs.Interpolation.okrige(Pred_grid_xy, df, xName, yName, zName, k, vario, rad, quiet)
```

Esta función cuenta con los siguientes parámetros:

- Pred_grid_xy (array): array de 2 dimensiones que contiene las coordenadas x e y de cada punto que va a ser interpolado.
- df (dataframe): dataframe con las columnas x, y, z con nombre en las que se encuentran respectivamente coordenada x, coordenada y y valor en el punto.
- xName (string): Nombre de la columna de la coordenada x en el dataframe.
- yName (string): Nombre de la columna de la coordenada y en el dataframe.
- zName (string): Nombre de la columna z con valor de los puntos en el dataframe.
- k (int): Número de puntos máximo a tener en cuenta para la interpolación.
- vario (array): Array con los siguientes parámetros de un variograma: azimuth, nugget, major_range, minor_range, sill y variogram_model
- rad (int): máximo radio de búsqueda isotrópico.
- quiet (bool): false para mostrar datos de la interpolación, true para que no muestre ningún dato en línea de comandos

4.1.5 Arcpy

Arcpy [23] es una librería que representa un conjunto de elementos propios de ArcGIS que completan la librería de Python para dar acceso al manejo y explotación de información espacial. Esta librería es un producto licenciado que tiene un coste asociado si se quieren utilizar los métodos de interpolación que proporciona en la librería ArcGIS PRO [24].

Esta librería proporciona métodos de interpolación no geoestadísticos como: IDW, Natural Neighbor o Spline. También cuenta con métodos de interpolación geoestadísticos de kriging de tipo universal kriging y ordinary kriging.

La versión más reciente disponible de ArcGIS PRO es la 3.3 lanzada en 2024 e incluye todos los paquetes necesarios para su funcionamiento.

ArcGIS PRO es un software que se descarga de la web de ArcGIS [25].

Un ejemplo de interpolación con kriging que se podría realizar con esta librería es:

```
Kriging(dataframe, zName, KrigingModel(semivariogramType, lagSize,
majorRange, partialSill, nugget))
```

Esta función cuenta con los siguientes parámetros:

- dataframe (dataframe): dataframe con las columnas x, y, z con nombre en las que se encuentran respectivamente coordenada x, coordenada y y valor en el punto.
- zName (string): Nombre de la columna de la coordenada z en el dataframe.
- KrigingModel (kriging model): hay dos modelos de kriging: UniversalKrigingModel y OrdinaryKrigingModel. Ambos toman como parámetros el tipo de semivariograma a utilizar (semivariogramType), el tamaño del lag (lagSize), range máximo (majorRange), partialSill y el nugget.

4.1.6 IDW

Para la utilización de IDW se utiliza una función genérica fácilmente implementable:

```
def idw(self, lon, lat, values, lon_grid, lat_grid, power):
    lon = np.asarray(lon)
    lat = np.asarray(lat)
    values = np.asarray(values)
    lon_grid = np.asarray(lon_grid)
    lat_grid = np.asarray(lat_grid)

    interpolated_values = np.zeros((len(lat_grid), len(lon_grid)))

    for i, y in enumerate(lat_grid):
        for j, x in enumerate(lon_grid):
            distances = np.sqrt((lon - x)**2 + (lat - y)**2)
            weights = 1 / distances**power
            normalized_weights = weights / np.sum(weights)
            interpolated_values[i, j] = np.sum(values *
            normalized_weights)

    return interpolated_values
```

Esta función cuenta con los siguientes parámetros:

- lon (Array): Array con los valores de longitud.
- lat (Array): Array con los valores de latitud.
- values (Array): Array con los valores que corresponden a longitud y latitud.

- lon_grid (Array): Array con los valores de longitud para el array de interpolación.
- lat_grid (Array): Array con los valores de latitud para el array de interpolación.
- power (float, optional): Parametro de potencia para interpolación. Potencia 2 por defecto.

4.2 Métrica

En esta sección se describe la metodología elegida para evaluar y comparar los métodos de interpolación de las librerías encontradas. Con el fin de obtener predicciones adecuadas al problema que se pretende abordar, se ha decidido escoger como métrica de evaluación el error absoluto (MAE), que se calcula con la Ecuación 7.

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}$$

Ecuación 7 - Cálculo de error absoluto (MAE)

De esta forma es una media aritmética de los errores medios $|e_i| = |y_i - x_i|$, donde y_i es la predicción y x_i es el valor real. Los valores obtenidos con esta métrica son siempre positivos y mejores cuanto más se aproximen a 0.

La métrica de MAE se va a calcular utilizando la librería *sklearn.metrics* que proporciona una gran variedad de métricas entre las que se encuentra la función *mean_absolute_error* que se utiliza de la siguiente forma:

```
mean_absolute_error(eval_values, eval_interpolated_values)
```

Esta función recibe los siguientes parámetros:

- eval_values (array): Array que contiene los valores reales a comparar.
- eval_values_interpolated (array): Array que contiene los valores calculados por el método de interpolación a comparar.

Para realizar la medición de esta métrica se va a emplear validación cruzada de forma que se puedan comparar los valores reales medidos en algunas estaciones meteorológicas con los valores estimados por los métodos de interpolación. La validación cruzada se va a realizar dejando un 20% de las estaciones fuera del método de interpolación. Esta operación se repetirá 10 veces para obtener una media para evitar el caso de que se dé la casualidad de sacar de la interpolación los puntos que puedan ser más diferentes por tener valores mucho más altos o bajos que los de los puntos cercanos.

4.3 Comparación de librerías

Para conocer el rendimiento que tiene cada librería se va a realizar una serie de ejecuciones de los métodos de interpolación utilizando para la medición de su eficacia la métrica descrita en la anterior sección. Además, se valorará el tiempo de ejecución que toma cada una de las ejecuciones de cada modelo de interpolación. Se utilizarán distintos parámetros para obtener resultados más precisos y ver el efecto de variar estos parámetros en el tiempo de ejecución y en la precisión de los métodos de interpolación. Los parámetros elegidos para variar en cada librería dependerán de los disponibles de entre ellos se seleccionarán los que más impacto tendrán en los resultados.

La selección de parámetros se ha realizado teniendo en cuenta el requisito de que la ejecución debe ser relativamente rápida. Para realizar esta comparativa se utilizarán los datos obtenidos de la base de datos CAELIS. Se ha utilizado el mapa del mundo completo con 220 estaciones ya que también se ha probado a hacer con una región más pequeña, Península Ibérica, y los resultados no han sido demasiado representativos al solamente contar con datos de 10 estaciones meteorológicas.

La selección de los parámetros Nugget, Sill y Range del variograma seleccionado se ha realizado modelando un variograma con los datos disponibles para la interpolación.

Las librerías Arcpy y TFinterpy han quedado fuera de la comparativa al haber tenido problemas para su instalación:

- TFinterpy: ha presentado problemas para su instalación debido a una incompatibilidad de la versión de una librería necesaria para su uso con la versión de Python utilizada.
- Arcpy: al ser de pago el software de las funciones de interpolación no se ha podido utilizar esta librería.

4.3.1 Medición Pykrige

Los 5 mejores resultados obtenidos por tener el MAE menor utilizando la librería Pykrige se pueden observar en la Tabla 4.

Kriging type	nPoints	variogramType	MAE	Time (s)
Ordinary	200	exponential	0,29167607	5,03700924
Universal	200	exponential	0,29167607	6,19509864
Ordinary	100	exponential	0,29308786	3,19063663
Universal	100	exponential	0,29308786	5,3810029
Ordinary	50	exponential	0,29314871	4,13443971

Tabla 4 - Resultados librería PyKrige

En la Tabla 4 de resultados se puede ver como el tipo de kriging en esta librería no es un parámetro demasiado relevante ya que se van alternando posiciones en el top dependiendo de los otros parámetros. El número de puntos es uno de los parámetros clave, cuantos más puntos representados mejores resultados se han obtenido. El radio de influencia no ha tenido mucha importancia ya que al igual que el tipo de kriging, se ha ido alternando puestos en el top dependiendo del resto de parámetros. Por último, el tipo de variograma claramente más preciso ha resultado ser el exponencial para este caso de uso. La mejor ejecución se puede observar en la Ilustración 8.

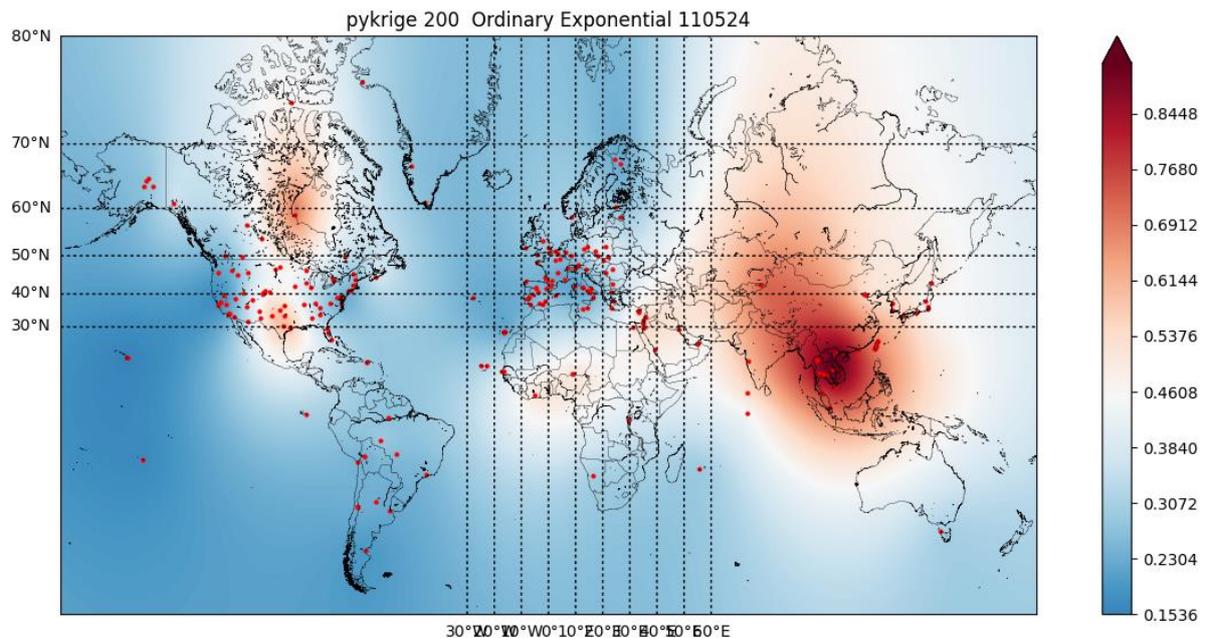


Ilustración 8 - Resultado de la mejor ejecución obtenida con la librería Pykrige.

4.3.2 Medición Geostatspy

Los 10 mejores resultados obtenidos por tener el MAE menor utilizando la librería Geostatspy

se pueden ver en la siguiente Tabla 5.

Kriging type	nPoints	radius	variogramType	MAE	Time (s)
Simple	50	50	gaussian	0,24977293	6,24135184
Simple	50	10	gaussian	0,24978563	5,96192718
Simple	200	50	gaussian	0,2600206	0,98523164
Simple	200	10	gaussian	0,26122108	6,19509864
Simple	50	10	spherical	0,2666427	7,82132053
Simple	50	50	spherical	0,26946545	1,17372847
Simple	100	50	gaussian	0,27084543	5,3810029
Simple	50	10	exponential	0,27098438	5,81670594
Simple	100	10	Gaussian	0,27127253	0,64072537
Simple	50	50	exponential	0,27517989	6,9763186

Tabla 5 - Resultados librería GeostatsPy

En la Tabla 5 de resultados se puede ver como pese a haber utilizado tanto ordinary kriging como simple kriging, todos los resultados dentro del top 10 son utilizando simple kriging. Otro dato a tener en cuenta es que ni el número de puntos ni el radio parece afectar significativamente a los resultados. El tipo de variograma que mejores resultados ha obtenido es el Gaussian, siendo el spherical el segundo y exponential el último. Todos los tiempos de ejecución se han mantenido por debajo de los 10 segundos por lo que se pueden considerar ejecuciones rápidas. La mejor ejecución se puede observar en la Ilustración 9.

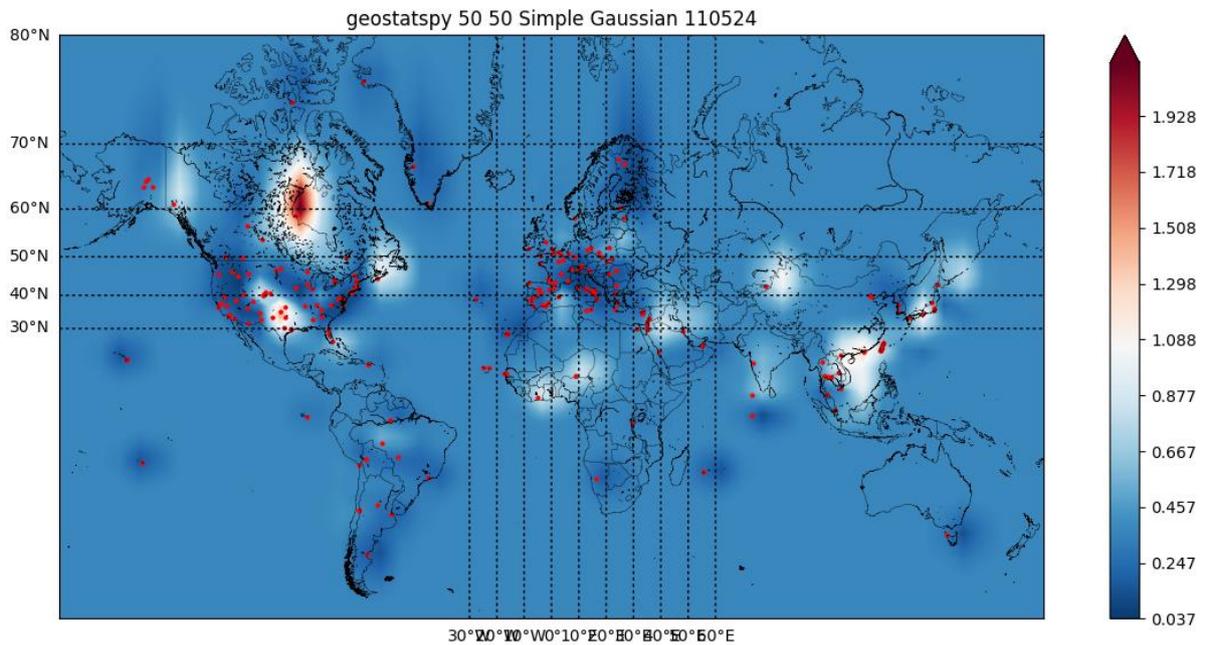


Ilustración 9- Resultado de la mejor ejecución obtenida con la librería Geostatspy.

4.3.3 Medición Gstatsim

Los 10 mejores resultados obtenidos por tener el MAE menor utilizando la librería Gstatsim se pueden consultar en la siguiente Tabla 6.

Kriging type	nPoints	radius	variogramType	MAE	Time (s)
Ordinary	200	10	gaussian	0,27874319	5,05903265
Ordinary	200	50	gaussian	0,27874319	1,82002994
Simple	200	10	gaussian	0,27874319	6,21712205
Simple	200	50	gaussian	0,27874319	1,00725505
Ordinary	100	10	gaussian	0,28009239	3,21266004
Ordinary	100	50	gaussian	0,28009239	4,12163636
Simple	100	10	gaussian	0,28009239	0,66274878
Simple	100	50	gaussian	0,28009239	5,40302631
Ordinary	50	10	gaussian	0,28015054	4,15646312
Ordinary	50	50	gaussian	0,28015054	15,0850815

Tabla 6 - Resultados librería Gstatsim

En la Tabla 6 de resultados se puede observar como el tipo de kriging no ha sido un parámetro

que haya afectado demasiado al resultado. El número de puntos ha sido el segundo parámetro que más ha contribuido al resultado. El radio es un parámetro que claramente tampoco ha tenido mucho que ver con el resultado obtenido. Sin duda el parámetro que más ha afectado es el tipo de variograma elegido, el Gaussian es claramente el óptimo para este problema. Los tiempos de ejecución han sido reducidos siendo en el peor de los casos de aproximadamente 15 segundos. La mejor ejecución se puede observar en la Ilustración 10.

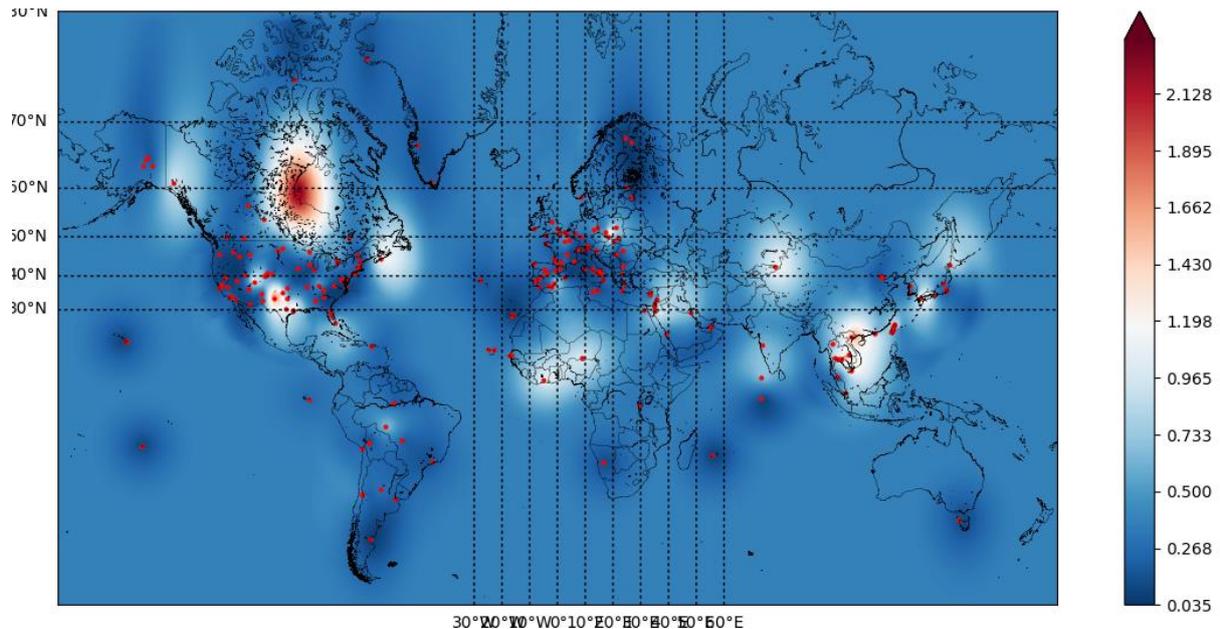


Ilustración 10 - Resultado de la mejor ejecución obtenida con la librería Gstatsim.

4.3.4 Medición IDW

Los 10 mejores resultados obtenidos por tener el MAE menor utilizando la función IDW se pueden consultar en la siguiente Tabla 7.

Power	nPoints	MAE	Time (s)
3	100	0,18683531	0,31266141
2	100	0,1869655	0,1961937
2	5000	0,18888904	457,135501
4	100	0,19514683	0,31368208
2	1000	0,19619516	18,0876081
3	5000	0,20200795	599,420278
2	500	0,20491021	4,32479644
3	1000	0,20751027	28,9071269
8	100	0,21172641	0,38758636
4	5000	0,21193741	619,492183

Tabla 7 - Resultados interpolación con IDW

En la Tabla 7 se puede ver como el parámetro de power o potencia es el que más ha afectado a los resultados obtenidos, siendo 3 la potencia óptima seguida de 2. El número de puntos ha sido también importante, aunque a diferencia de lo que se piensa normalmente, menos ha sido más. Los tiempos de ejecución se mantienen contenidos a excepción de las ejecuciones de más de 500 puntos que superan los tiempos aceptables de unos pocos segundos. La mejor ejecución se puede observar en la Ilustración 11.

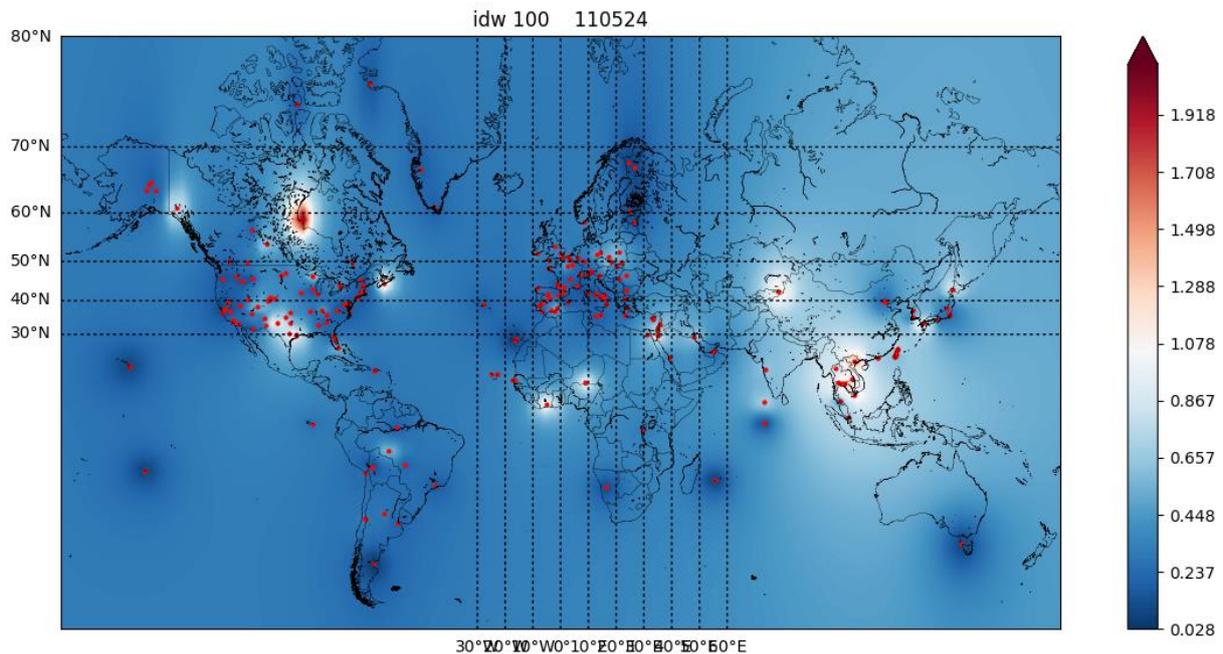


Ilustración 11 - Resultado de la mejor ejecución obtenida con la función IDW.

4.3.5 Comparativa general

En esta subsección se va a realizar una comparativa de los resultados obtenidos por todas las librerías y funciones medidas. En la Tabla 8 se muestra el mejor resultado obtenido por cada librería/función.

Posición	Librería/función	MAE	Tiempo (s)
1	IDW	0,18683531	0,31266141
2	Geostatspy	0,24977293	6,24135184
3	Gstatsim	0,27874319	5,05903265
4	Pykrige	0,29167607	5,03700924

Tabla 8 - Comparativa librerías e IDW

A la vista de estos resultados, bajo el criterio del error absoluto medio parece claro que IDW es el método de óptimo para este caso. Sin embargo, cabe destacar que como se aprendió del estudio [15] es importante que se realice una valoración subjetiva visual, es en este punto que Pykrige pese a haber obtenido la peor puntuación con el error absoluto medio, destaca frente a los otros métodos de interpolación ya que el mapa creado es de un carácter diferente. Pykrige ha creado un mapa que tiene transiciones entre zonas con un alto valor de AOD y zonas con un valor bajo mientras que el resto de los métodos de interpolación parece que se han limitado a pintar el mapa de un color más neutro. Es por ello que debe tomarse muy en cuenta pese a sus resultados de error absoluto medio.

5: Aspectos relevantes del desarrollo del proyecto

Como se ha comentado al inicio del documento uno de los objetivos principales que se pretende realizar con este trabajo es el de crear un programa capaz de realizar interpolaciones y visualizarlas en una interfaz. Inicialmente el objetivo era poder visualizar únicamente un mapa de interpolación al mismo tiempo, pero durante el desarrollo del programa se consideró interesante tener múltiples visualizaciones simultáneas para así poder comparar mapas de interpolación, ya bien sea para comparar métodos de interpolación o los resultados obtenidos en distintas fechas.

5.1 Diseño de la interfaz

En este apartado se describe la herramienta de diseño de interfaces utilizada para crear la interfaz del programa y se describe el diseño de la interfaz creada.

5.1.1 Herramienta de diseño de interfaces

La interfaz utilizada está diseñada utilizando la herramienta QtDesigner. Es una popular herramienta de Qt que originalmente está diseñado para C++ y es ampliamente utilizada para aplicaciones en las que es necesario crear una interfaz. En este caso se utiliza Pyside2 que es la versión de la librería Qt en Python. Esta librería cuenta con una guía que enseña a utilizarla desde cero [26].

La herramienta QtDesigner es un software compuesto de una interfaz gráfica que permite editar ficheros “.ui” Ilustración 12. Estos ficheros son compilados y convertidos a ficheros “.py” para

que Python sea capaz de leerlos y entenderlos.

Las interfaces Qt están compuestas de elementos que llaman widgets, estos elementos pueden ser botones, campos de texto, campos de fecha y todo tipo de elementos que compone una interfaz.

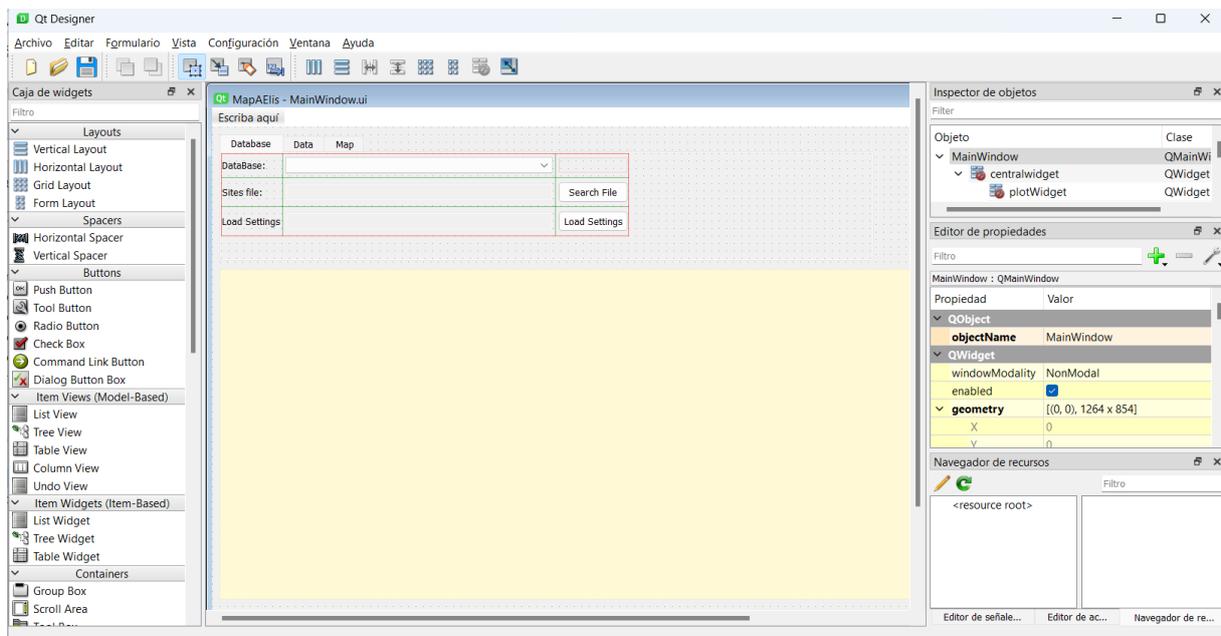


Ilustración 12 - Herramienta Qt Designer

Esta herramienta permite crear la interfaz de la aplicación deseada arrastrando los widgets desde la barra izquierda hasta el panel central en el que se encuentra la previsualización de la interfaz creada. En la barra izquierda se pueden editar los nombres y características de los widgets pudiendo así darle el nombre deseado en el código y modificar propiedades como el tamaño exacto en píxeles que se desea o si se quiere que aparezca visible por defecto entre otras características modificables.

5.1.2 Diseño de la interfaz

Para el diseño de la interfaz se ha partido de la interfaz básica que se proporcionó junto con el código. Esta interfaz tiene como objetivo mostrar mapas de interpolación pudiendo elegir la base de datos de la que se obtienen los datos, la fecha y otras características de los datos y por último la proyección deseada y el tipo de interpolación a utilizar. La interfaz está dividida en 2 partes claramente diferenciadas, la ventana de visualización del mapa de interpolación en la parte inferior y las pestañas con la selección de parámetros en la parte superior como se puede

ver en la Ilustración 13.

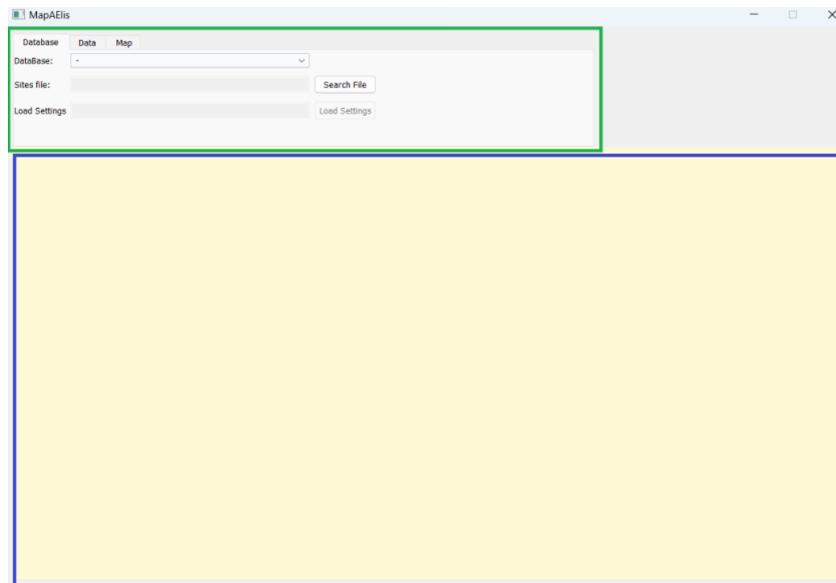


Ilustración 13 - División de parámetros y ventana de visualización

En la parte inferior de la Ilustración 13 se encuentra el cuadro en el que se representan los mapas de interpolación. Los mapas de interpolación se representan con una barra a su derecha con la escala de color respecto al valor de AOD. Si se representa un único mapa de interpolación se dejan disponibles unas herramientas para poder realizar ampliaciones en el mapa, mover el mapa y así cambiar la zona que se está viendo y guardar la imagen entre otras funciones. Se puede ver un ejemplo en la Ilustración 14. En el caso de que se represente más de un mapa de interpolación, las herramientas para modificar la visualización desaparecen ya que en caso de dejarlas la vista no queda limpia. Se pueden representar hasta un máximo de 8 mapas de interpolación al mismo tiempo.

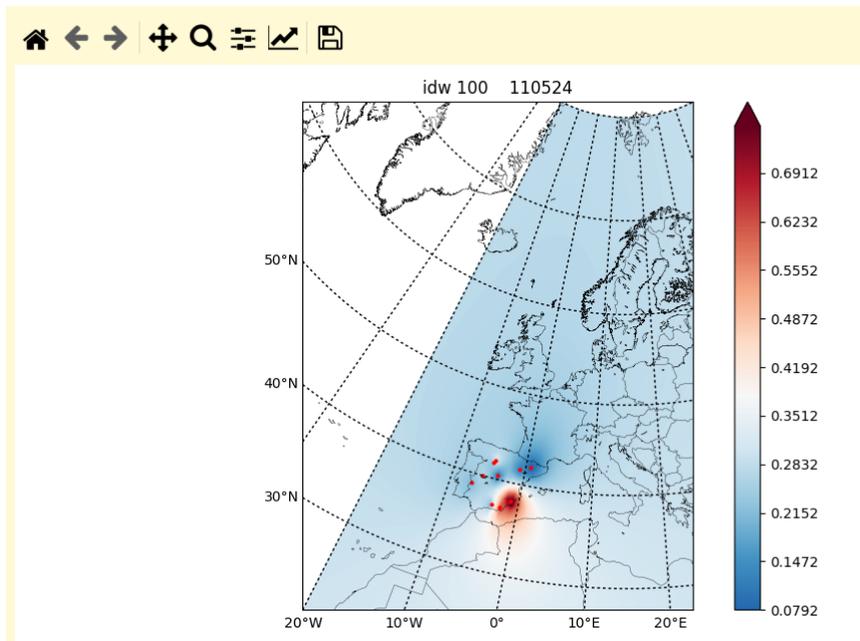


Ilustración 14 - Visualización de mapa interpolado

En la parte superior de la Ilustración 13, al ser una gran cantidad de parámetros y campos que se deben modificar y al tener en cuenta que el tamaño de la ventana abierta debe adaptarse a pantallas de tamaños variados, se ha decidido dividir la interfaz superior en 3 pestañas para poder compactar el tamaño de la interfaz: Database, Data y map.

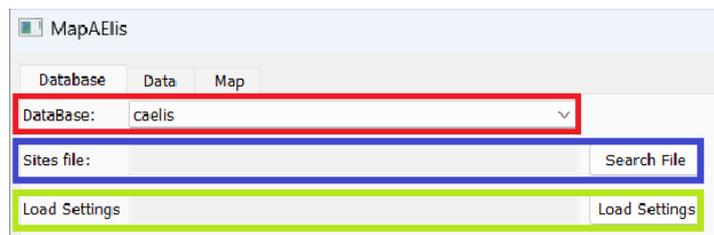


Ilustración 15 - Interfaz pestaña Database

En la pestaña “Database” que se puede ver en la Ilustración 15 se pueden encontrar los siguientes campos:

- DataBase (En rojo): campo desplegable para seleccionar la base de datos de la que se desean obtener los datos. Como opciones se presentan “Caelis” y “Polos”. Hasta que este desplegable no tiene seleccionada una opción no se desbloquean los desplegables siguientes.
- Sites file (En azul): este campo cuenta con un botón “Search File” que abre el explorador de archivos para seleccionar un fichero que contiene las localizaciones de las que se

quiere obtener los datos para realizar la interpolación. Una vez elegido el fichero se muestra la ruta en la barra.

- Load Settings (En verde): con esta opción se pretende agilizar la tarea de visualizar mapas de interpolación que se consultan habitualmente. Al pulsar en el botón de Load Settings se abre el explorador de archivos para seleccionar un fichero “.yaml” con una configuración preestablecida para que se configure automáticamente en las siguientes pestañas de la interfaz pudiendo modificar en él todos los campos generales a los métodos de interpolación.



Ilustración 16 - Interfaz pestaña Data

En la pestaña Data, que se puede ver en la Ilustración 16, se pueden encontrar los siguientes campos:

- Date (En Rojo): en este campo se puede introducir la fecha de la que se quieren consultar los datos en la base de datos. También se pueden utilizar las flechas situadas en la parte derecha del campo para ir modificando los campos.
- Level (En Azul): desplegable que permite seleccionar el nivel de altura respecto al suelo de las mediciones que se quieren utilizar en la interpolación.
- wln (En Verde): desplegable que permite seleccionar la longitud de onda de la luz en la que se han tomado las mediciones de AOD.

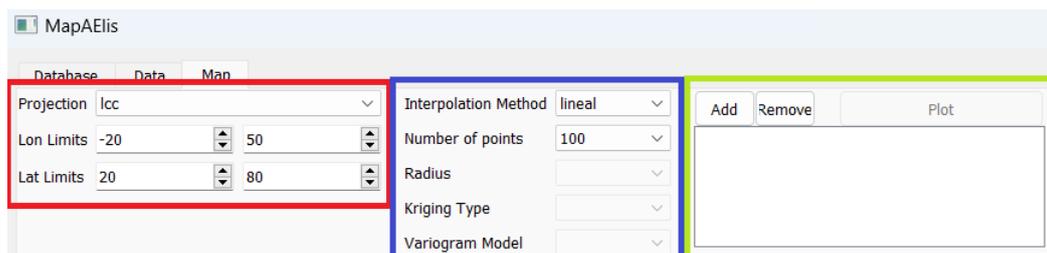


Ilustración 17 - Interfaz pestaña Map

La pestaña Map, que se puede ver en la Ilustración 17, está dividida en tres secciones, en la primera (en rojo) se modifican los parámetros de la proyección:

- Projection: Desplegable que contiene una multitud de proyecciones disponibles para poder observar el mapa de interpolación desde distintas perspectivas.

- Lon Limits: Este apartado consta de dos campos que permiten introducir valores y modificar el introducido con las flechas situadas a la derecha de cada campo. El de la izquierda es para el límite de longitud inferior y el derecho para el límite de latitud superior.
- Lat Limits: Este apartado consta de dos campos que permiten introducir valores y modificar el introducido con las flechas situadas a la derecha de cada campo. El de la izquierda es para el límite de latitud inferior y el derecho para el límite de latitud superior.

En la segunda sección (en azul) se puede seleccionar el método de interpolación y modificar sus parámetros:

- Interpolation method: Desplegable en el que se selecciona la librería/función a utilizar para la interpolación de entre los 5 disponibles. En función de la librería/función elegida se desbloquean los siguientes campos de esta columna que son aplicables para tal método de interpolación.
- Number of points: Desplegable para seleccionar el número de puntos a utilizar en la interpolación.
- Radius: Desplegable para seleccionar el radio a utilizar en la interpolación.
- Kriging type: Desplegable para seleccionar el tipo de kriging a utilizar en la interpolación. Se muestran los métodos disponibles para la librería/función seleccionada en Interpolation method.
- Variogram type: Desplegable para seleccionar el tipo de variograma a utilizar en la interpolación. Se muestran los métodos disponibles para la librería/función seleccionada en Interpolation method.

En la tercera sección (en verde) es donde se pueden añadir las vistas que se desean representar:

- Add: Botón para añadir a la lista de vistas una vista con la selección actual de parámetros del método de interpolación.
- Remove: Botón para eliminar la vista seleccionada de la lista de vistas.
- Plot: Botón para comenzar la ejecución de los métodos de interpolación seleccionados en la lista de vistas.
- Lista de vistas: Lista conteniendo la lista de vistas a representar, se pueden seleccionar las vistas que se desean para eliminar. Todas las vistas disponibles en esta lista serán ejecutadas y representadas hasta un máximo de 8.

5.2 Implementación

Para desarrollar el programa en Python con la interfaz descrita en el anterior apartado implementado los métodos de interpolación elegidos se realiza el desarrollo de un programa.

El código está dividido en 3 carpetas que contienen 8 ficheros de código Python, en la Ilustración 18 se puede ver cómo están estructurados los ficheros en las carpetas y las relaciones que tienen.

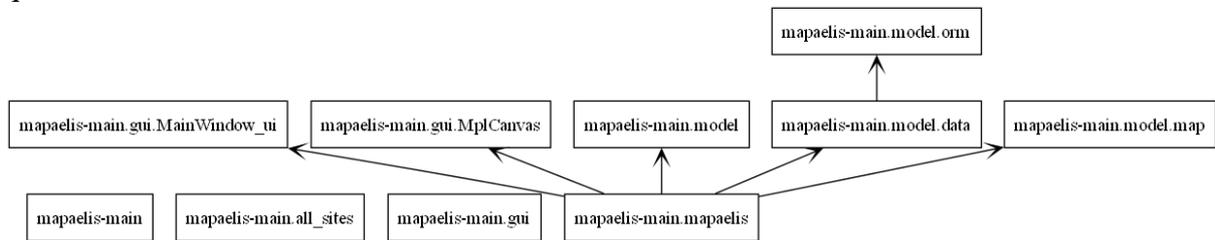


Ilustración 18 - Diagrama de clases

A continuación, se va a realizar un breve resumen de las funcionalidades contenidas en cada fichero, para un mayor detalle de las funciones disponibles y de la implementación realizada se recomienda consultar el código [27].

All_sites.py: fichero que contiene los nombres de todos los sitios disponibles en las bases de datos como fuente de datos para realizar las interpolaciones

Mapaelis.py: fichero principal que se ejecuta para arrancar el programa. Contiene funciones para inicializar y utilizar la interfaz a su vez que llamadas a la clase *data.py* para obtener los datos necesarios de las bases de datos para las interpolaciones y llamadas a la clase *map.py* para ejecutar los métodos de interpolación.

MainWindow_ui.py: fichero generado por QtDesigner a partir del fichero MainWindow.ui. Contiene el código necesario para generar la interfaz deseada con los nombres adecuados para cada elemento.

MplCanvas.py: fichero que contiene la función necesaria para imprimir en un widget de la interfaz una representación de un mapa de interpolación.

Data.py: fichero encargado de realizar la consulta a la base de datos para obtener los datos necesarios para los mapas que se quieren interpolar

Orm.py: fichero complementario a data.py, contiene métodos auxiliares para las consultas a la base de datos.

Map.py: fichero encargado de ejecutar todos los métodos de interpolación, contiene todos los códigos con los métodos de interpolación.

En la siguiente página se pueden encontrar las imágenes Ilustración 19 e Ilustración 20 que contienen todas las funciones que se pueden encontrar en el código de cada uno de los ficheros junto con las variables.

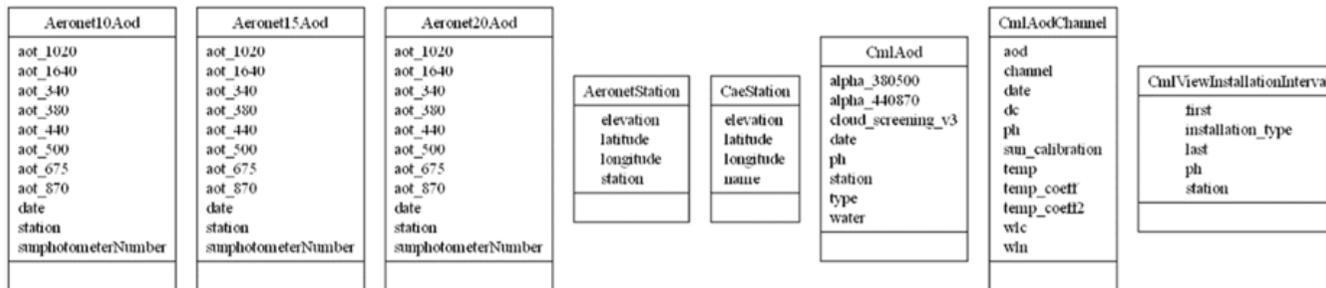


Ilustración 19 - Diagrama de funciones 1

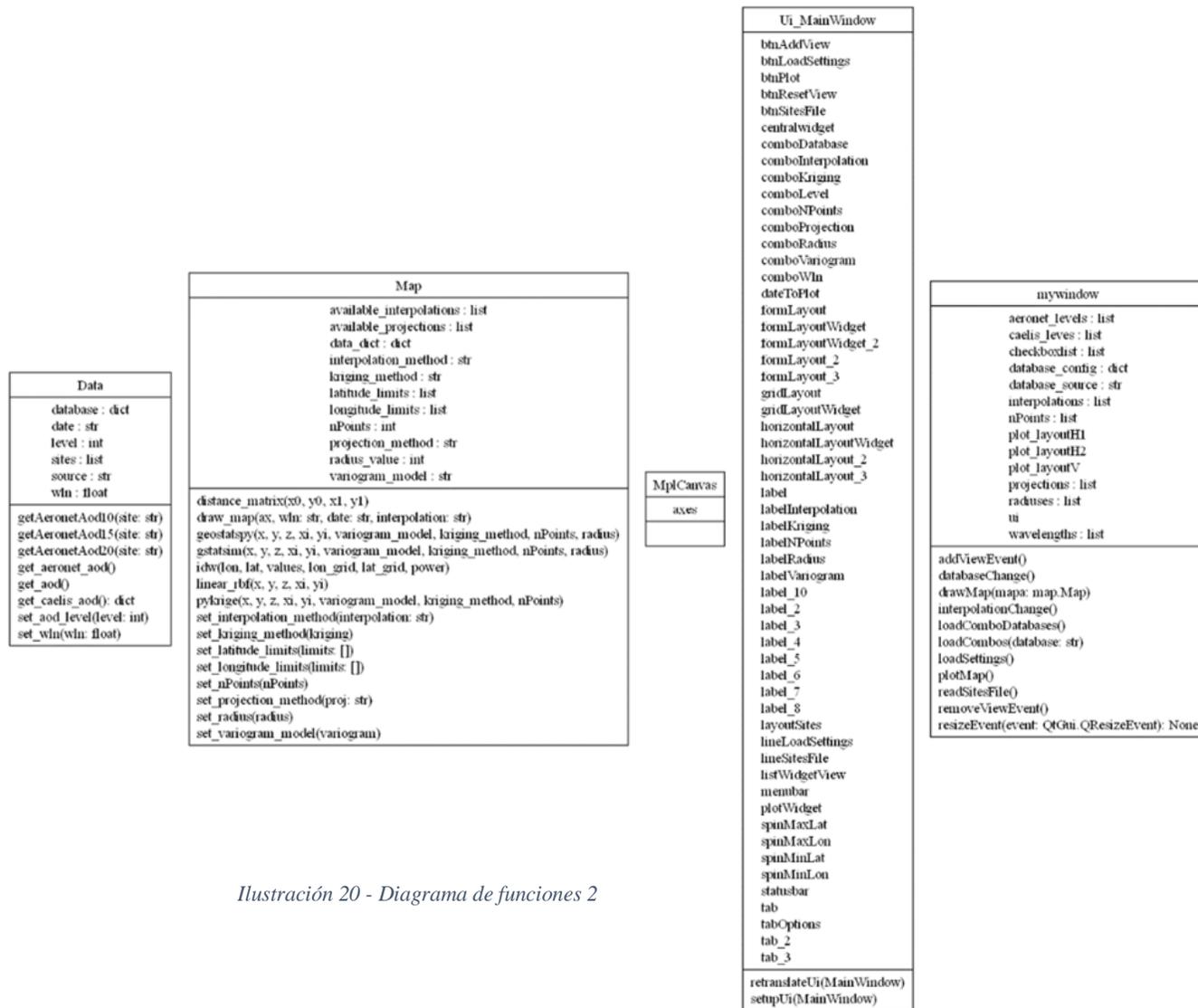


Ilustración 20 - Diagrama de funciones 2

6: Conclusiones y Líneas de trabajo futuras

Se presentan en este último capítulo las conclusiones que se han obtenido del desarrollo del mismo y de los resultados. Adicionalmente se proponen una serie de reflexiones/conclusiones personales que se sacan de la realización de este proyecto. Finalmente se proponen posibles líneas de trabajo para continuar y ampliar este proyecto.

6.1 Conclusiones

En primer lugar, es importante analizar el grado en el que se han cumplido los objetivos planteados al inicio del documento. El primer objetivo de realizar una investigación sobre el estado del arte de los métodos de interpolación disponibles en Python para mapas de AOD se ha resuelto en los capítulos 3 y 4 primero estudiando el contexto del trabajo para ponerse en situación y después investigando sobre distintos métodos de interpolación para finalizar realizando un estado del arte de los documentos de estudios que hacen uso de estos métodos de interpolación. El segundo objetivo de programar un software en Python que permita realizar mapas de interpolación con los datos de distintas mediciones de AOD y visualizar los resultados, todo en una interfaz. Este objetivo se ha completado en los capítulos 5 y 6 realizando una descripción de las librerías para posteriormente compararlas utilizando una métrica y seleccionar las más aptas para realizar el objetivo. Finalmente se ha diseñado e implementado una interfaz en Python capaz de mostrar mapas de interpolación utilizando métodos de interpolación de 5 fuentes distintas.

A la vista de lo comentado anteriormente se puede decir que los objetivos quedan alcanzados, cumpliendo con las restricciones planteadas.

La elección de la planificación en cascada ha sido acertada para la realización de este trabajo al haber podido completar las fases una por una de forma secuencial en el tiempo

previsto.

Personalmente se sacan varias conclusiones/reflexiones de este proyecto que pueden ayudar a la realización de proyectos similares en el futuro:

- No existe un método de interpolación que sea óptimo para todos los casos, es cierto que hay métodos de interpolación que usualmente funcionan mejor en cierto tipo de problemas, pero se deben realizar pruebas con varios ya que los resultados pueden sorprender.
- A la vista de los resultados presentados en el apartado de comparación de librerías, pese a que visualmente los mapas sean plausibles, se saca la conclusión de que para mejorar la precisión de los métodos de interpolación es necesario contar con más datos, ya bien sea de más estaciones meteorológicas o bien con más variables meteorológicas o del entorno para complementar la información.
- La métrica del espesor óptico de aerosoles tiene un papel muy importante a la hora de identificar anomalías en el aire. Puede ser un detector de cambio climático y de contaminación por lo que es importante que se siga investigando en la materia para contar con la mayor cantidad de información posible.

6.2 Líneas futuras

Se plantean, a continuación, algunas posibles líneas de trabajo futuro que no han podido abordarse en el presente proyecto:

- Como se ha comentado anteriormente, a la vista de los resultados de precisión de los métodos de interpolación, sería interesante probar en el futuro con métodos de interpolación con más de una variable como el cokriging ya que en los estudios analizados parecen prometedores.
- Para mejorar la precisión de las predicciones y para obtener más datos reales sería una gran ayuda contar con más estaciones meteorológicas capaces de medir el AOD en ese punto.

BIBLIOGRAFÍA

- [1] «ESRL Global Monitoring Laboratory - Global Radiation and Aerosols». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://gml.noaa.gov/grad/surfrad/aod/>
- [2] «Aerosol Optical Depth». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: https://earthobservatory.nasa.gov/global-maps/MODAL2_M_AER_OD
- [3] X. Wei, N.-B. Chang, K. Bai, y W. Gao, «Satellite remote sensing of aerosol optical depth: advances, challenges, and perspectives», *Critical Reviews in Environmental Science and Technology*, vol. 50, n.º 16, pp. 1640-1725, ago. 2020, doi: 10.1080/10643389.2019.1665944.
- [4] «Aerosol Robotic Network (AERONET) Homepage». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://aeronet.gsfc.nasa.gov/>
- [5] D. M. Giles *et al.*, «Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements», *Atmospheric Measurement Techniques*, vol. 12, n.º 1, pp. 169-209, ene. 2019, doi: 10.5194/amt-12-169-2019.
- [6] «CE318-T - Sun Sky Lunar Multispectral Photometer | Cimel Electronique». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://www.cimel.fr/solutions/ce318-t/>
- [7] «CÆLIS». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://www.caelis.uva.es/>
- [8] «Sueldo: Programador en España 2024 | Glassdoor». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: https://www.glassdoor.es/Sueldos/programador-sueldo-SRCH_KO0,11.htm
- [9] «Salario para Investigador en España - Salario Medio». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://es.talent.com/es/salary?job=investigador>
- [10] «Interpolation», *Wikipedia*. 18 de junio de 2024. Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://en.wikipedia.org/w/index.php?title=Interpolation&oldid=1229664148>
- [11] «Deterministic methods for spatial interpolation—ArcGIS Pro | Documentation». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en:

- <https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/deterministic-methods-for-spatial-interpolation.htm>
- [12] «Polígonos de Thiessen», *Wikipedia, la enciclopedia libre*. 8 de agosto de 2024. Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Pol%C3%ADgonos_de_Thiessen&oldid=161754726
- [13] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*. Oxford University Press, 1997.
- [14] S. Vicente-Serrano, M. Saz-Sánchez, y J. Cuadrat, «Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature», *Clim. Res.*, vol. 24, pp. 161-180, 2003, doi: 10.3354/cr024161.
- [15] X. Li, G. Cheng, y L. Lu, «Spatial Analysis of Air Temperature in the Qinghai-Tibet Plateau», *Arctic, Antarctic, and Alpine Research*, vol. 37, n.º 2, pp. 246-252, may 2005, doi: 10.1657/1523-0430(2005)037[0246:SAOATI]2.0.CO;2.
- [16] F. C. Collins, «A comparison of spatial interpolation techniques in temperature estimation», nov. 1995, Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <http://hdl.handle.net/10919/38139>
- [17] A. D. Hartkamp, A. Stein, y J. W. White, «Interpolation Techniques for Climate Variables».
- [18] J. Li y A. D. Heap, «Spatial interpolation methods applied in the environmental sciences: A review», *Environmental Modelling & Software*, vol. 53, pp. 173-189, mar. 2014, doi: 10.1016/j.envsoft.2013.12.008.
- [19] «PyKriging Documentation».
- [20] Z. Chen y B. Zhong, «TFInterpy: A high-performance spatial interpolation Python package», *SoftwareX*, vol. 20, dic. 2022, doi: 10.1016/j.softx.2022.101229.
- [21] M. Pycrz *et al.*, *GeostatsPy Python Package: Open-source Spatial Data Analytics and Geostatistics*. (14 de septiembre de 2024). Jupyter Notebook. Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://github.com/GeostatsGuy/GeostatsPy>
- [22] E. MacKie, M. Field, L. Wang, Z. Yin, N. Schoedl, y M. Hibbs, *GStatSim*. (20 de octubre de 2022). Zenodo. doi: 10.5281/ZENODO.7230276.
- [23] «An overview of the Interpolation toolset—ArcGIS Pro | Documentation». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/an-overview-of-the-interpolation-tools.htm>
- [24] «Install ArcPy—ArcGIS Pro | Documentation». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://pro.arcgis.com/en/pro-app/latest/arcpy/get-started/installing-arcpy.htm>
- [25] «Descargar ArcGIS Pro—ArcGIS Pro | Documentación». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://pro.arcgis.com/es/pro-app/latest/get-started/download-arcgis-pro.htm>
- [26] «Create your first PySide2 app in Qt Designer». Accedido: 18 de septiembre de 2024. [En línea]. Disponible en: <https://www.pythonguis.com/tutorials/pyside-first-steps-qt-designer/>

[27] «marioabaddiaz/mapaelis-tfm». Accedido: 18 de septiembre de 2024. [En línea].
Disponibile en: <https://github.com/marioabaddiaz/mapaelis-tfm>