ICAI - Universidad Pontificia Comillas

ESTIMACIÓN DE LA CONTAMINACIÓN A NIVEL DE CÓDIGO POSTAL EN LA COMUNIDAD DE MADRID, UTILIZANDO DATOS ESPACIALES

Autor: Rafael Gómez-Aparici Vega

Tutor: Manuel Alejandro Betancourt Odio

MADRID, mayo 2023

Resumen

La contaminación atmosférica es uno de los desafíos ambientales más urgentes y preocupantes en el mundo actual. Sus efectos nocivos sobre la salud humana y el medio ambiente son ampliamente reconocidos, lo que ha generado un creciente interés en abordar este problema desde diferentes enfoques científicos y tecnológicos. Si se quiere estudiar los efectos que tiene la contaminación, tanto como afectan los distintos contaminantes en la salud humana, como el efecto que tienen distintos niveles de concentración de estos, uno de los pasos más importantes es poder conocer de forma precisa la distribución espacial que presentan cada uno de estos contaminantes dentro de un mismo territorio. Es por esto por lo que las administraciones públicas han puesto énfasis en la instalación de redes de medición para poder realizar el seguimiento de las regiones más densamente pobladas. Sin embargo, estas redes no son capaces de proporcionar datos a un nivel muy detallado ya que el número de estaciones que serían necesarias para poder llegar a este nivel de detalle sería demasiado elevado y resultaría impráctico tanto a nivel logístico, ya que en las grandes ciudades no se dispone de emplazamientos en los que resulte fácil instalar este tipo de infraestructuras, como a nivel económico, ya que se necesita un gran desembolso no solo en su construcción, sino también en su operación y mantenimiento. En este sentido se pretende suplir esta necesidad creando un modelo que sea capaz de estimar los niveles de contaminación de cada uno de los contaminantes a nivel de código postal a partir de los datos recogidos por la red de medición existente. Este estudio se ha centrado en la elaboración de un modelo válido para la Comunidad de Madrid a partir de los datos proporcionados por la red de calidad del aire de esta región, obteniendo un modelo para la estimación a nivel de código postal dentro de esta.

Palabras Clave

Contaminación atmosférica, salud, interpolación, distancia inversa, kriging.

Abstract

Air pollution is one of the most urgent and worrying environmental challenges in the world today. Its harmful effects on human health and the environment are widely recognized, which has generated a growing interest in addressing this problem from different scientific and technological approaches. If we want to study the effects of pollution, both how different pollutants affect human health and the effect of different levels of concentration of these pollutants, one of the most important steps is to precisely know the spatial distribution of each of these pollutants within the same territory. This is why public administrations have emphasized the installation of measurement networks to monitor the most densely populated regions. However, these networks are not able to provide data at a very detailed level because the number of stations that would be necessary to reach this level of detail would be too high and it would be impractical both logistically, since large cities do not have sites where it is easy to install this type of infrastructure, and economically, since it requires a large outlay not only in its construction, but also in its operation and maintenance. In this sense, the aim is to make up for this need by creating a model capable of estimating the pollution levels of each of the pollutants at the postal code level from the data collected by the existing measurement network. This study has focused on the development of a valid model for the Community of Madrid from the data provided by the air quality network of this region, obtaining a model for the estimation at the zip code level within this region.

Key Words

Air pollution, health, interpolation, inverse distance, kriging.

Índice

Introducción

La contaminación atmosférica es uno de los desafíos ambientales más urgentes y preocupantes en el mundo actual. Sus efectos nocivos sobre la salud humana y el medio ambiente son ampliamente reconocidos, lo que ha generado un creciente interés en comprender y abordar este problema desde diferentes enfoques científicos y tecnológicos. Se trata de un problema global que contribuye a la aparición prematura de enfermedades respiratorias, cardiovasculares y otros problemas de salud.

En este contexto, el conocimiento de la distribución de los distintos niveles de contaminación a través de una misma región se ha convertido en una herramienta clave para comprender la magnitud de este fenómeno en áreas urbanas densamente pobladas ya que la distribución espacial de la contaminación puede variar significativamente dentro de un mismo territorio, debido a diversos factores como la actividad industrial, el tráfico vehicular, la geografía y las condiciones climáticas locales.

La falta de información detallada y confiable sobre la calidad del aire en áreas específicas dificulta la identificación de problemas de contaminación, la toma de decisiones informadas y la implementación de medidas efectivas para mitigar sus efectos perjudiciales. Es por esto por lo que la medición precisa y actualizada de los niveles de contaminación se ha convertido en una preocupación central en el campo de la gestión ambiental y la salud pública.

Enfrentar este desafío requiere contar con datos precisos y actualizados sobre los niveles de contaminantes en distintos puntos geográficos. Tradicionalmente, la recopilación de esta información se ha basado en estaciones de monitoreo fijas y limitadas en número, lo que conlleva a una representación incompleta de la realidad y una falta de detalle en las mediciones. Esto deja lagunas significativas en la comprensión de la verdadera magnitud y distribución espacial de la contaminación.

En los últimos años, sin embargo, se han desarrollado nuevas técnicas y metodologías para obtener datos de contaminación de forma más precisa y completa. Entre estos destacan los modelos de simulación computacional que han permitido obtener mediciones más detalladas y en tiempo real, cubriendo áreas más extensas y proporcionando una imagen más completa de los niveles de contaminación en distintos puntos del territorio.

Estas simulaciones se realizan a través de métodos de interpolación utilizando los datos de estaciones de medición disponibles, a partir de los cuales es posible realizar una buena estimación de la contaminación pudiéndose generar mapas de contaminación más completos y detallados, proporcionando estimaciones fiables en áreas donde no se dispone de mediciones directas pudiendo llegar incluso a estimaciones a nivel de código postal en una región.

La hipótesis que se plantea al problema de investigación es:

Es posible mejorar la información disponible actualmente sobre los niveles de contaminación atmosférica a través de la implementación de un modelo basado de métodos de estimación espacial.

A partir de dicho problema, el objetivo general de la investigación es:

Realizar un análisis sobre los tipos de modelos más utilizados, e implementar el que mejor se ajuste a las necesidades de la región de estudio para poder ser utilizado para obtener datos de contaminación a nivel de código postal.

Para el logro de dicho objetivo se han planteado los siguientes objetivos específicos:

I. Estudiar algunos de los métodos comúnmente utilizados para la estimación de contaminación atmosférica, evaluando las fortalezas e inconsistencias que ofrece cada uno para poder elegir cuales llevar a la práctica.
II. Elegir los métodos de interpolación que se van a estudiar y realizar la implementación de estos.
III. Evaluar la calidad y la precisión de los resultados obtenidos para cada uno de los modelos implementados.
IV. Comparar los distintos modelos entre sí para evaluar cual es más oportuno aplicar según la circunstancia.

Con lo expuesto anteriormente este trabajo se va a articular en cuatro capítulos en los que se va a ir desarrollando el estudio.

En el primer capítulo se expone la relevancia del tema del que trata la investigación exponiendo cuales son los problemas generados por la contaminación atmosférica tanto a nivel internacional como nacional y más en profundidad en la Comunidad de Madrid de cuya su red de medición se han extraído los datos para la elaboración de este trabajo.

En el segundo capítulo se presentan las distintas metodologías de estimación existentes, cuales se han decido estudiar en este trabajo y los motivos por lo que se ha hecho, así como un desarrollo más en profundidad del funcionamiento matemático de estos.

En el tercer capítulo se describe como se ha llevado a cabo la obtención y limpieza de los datos, la implementación de los modelos elegidos, las consideraciones realizadas para el estudio y los resultados obtenidos con cada uno de ellos.

Finalmente, en el cuarto capítulo se exponen las conclusiones a las que se ha llegado con este estudio, así como una evaluación de los resultados obtenidos por cada uno de los modelos estudiados.

Capítulo I: Efectos medioambientales de la contaminación a nivel internacional y de España

Niveles de contaminación atmosférica

En España la contaminación atmosférica es un problema ambiental significativo que afecta la calidad de vida de las personas y tiene repercusiones en la salud humana y en el medio ambiente en general. La contaminación atmosférica se refiere a la presencia de sustancias tóxicas y contaminantes en el aire, que son emitidos por fuentes naturales y actividades humanas.

Según informes y estudios realizados en años anteriores (Ballester., 2005), algunas de las principales fuentes de contaminación atmosférica en España incluyen el transporte, las industrias, la generación de energía y la calefacción residencial. Los vehículos a motor, especialmente los diéseles, emiten una cantidad significativa de contaminantes, como partículas en suspensión y dióxido de nitrógeno (NO_2) (Querol., 2008).

Las actividades industriales, como la producción de acero, la generación de energía a partir de combustibles fósiles y la quema de residuos, también contribuyen a la contaminación del aire principalmente en forma de partículas en suspensión (PM_{10} y $PM_{2.5}$) que junto con condiciones meteorológicas desfavorables que impiden la dispersión vertical de los contaminantes pueden contribuir a la acumulación de estas durante un largo periodo de tiempo y que se den situaciones como la que se observa en la figura 1.1 (AEMET., 2022) donde se llegaron a alcanzar peligrosos niveles de ambos en toda la meseta central así como en el sureste peninsular.

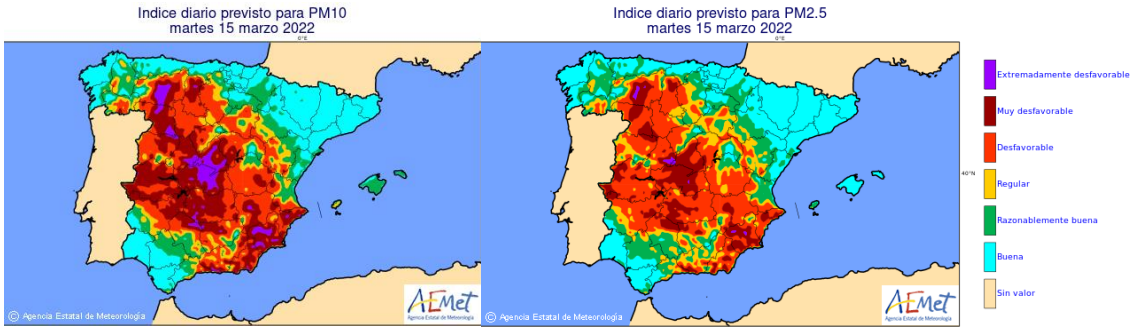

Figura 1.1. Mapa de niveles de partículas en suspensión 15 de marzo de 2022

Según datos de la Agencia Europea de Medio Ambiente, en informes anteriores se ha identificado que varias ciudades españolas, como Madrid, Barcelona y Valencia, han experimentado niveles preocupantes de contaminación atmosférica (AEMA., 2017). Los contaminantes atmosféricos, como el dióxido de nitrógeno y las partículas en suspensión, pueden tener efectos negativos en la salud, especialmente en personas con enfermedades respiratorias y cardiovasculares (Muñoz y Quiroz., 2007). Además, la contaminación atmosférica también puede contribuir al cambio climático y tener impactos en los ecosistemas y la biodiversidad.

Para abordar este problema, se han implementado diversas medidas y regulaciones en España. Estas incluyen la promoción del uso de vehículos eléctricos, la mejora de la eficiencia energética en la industria, la adopción de energías renovables y la implementación de planes de calidad del aire en áreas urbanas. Sin embargo, es necesario continuar trabajando en la reducción de la contaminación atmosférica y en la concienciación de la población sobre la importancia de la calidad del aire y sus efectos en la salud.

En los últimos años, la contaminación atmosférica se ha convertido en un problema ambiental significativo en muchas ciudades, ya que la densidad de producción de gases contaminantes se encuentra muy relacionada con la densidad de población y las ciudades son focos con alta densidad poblacional especialmente en un país como España caracterizado por tener un gran interior muy poco poblado y un reducido número de focos con una alta concentración de población, entre los que se incluye Madrid. La calidad del aire en la capital española ha sido motivo de preocupación debido a la presencia de diversos contaminantes, como el dióxido de nitrógeno (NO_2) y las partículas en suspensión (PM_{10} y $PM_{2.5}$).

Según informes (Ecologistas en Acción, 2022), Madrid ha experimentado niveles de contaminación atmosférica que superan los límites legales establecidos por la Unión Europea. En particular, el tráfico urbano es una de las principales fuentes de contaminación en la ciudad. Los vehículos que funcionan con motores diésel emiten altas cantidades de NO_2, un gas tóxico que puede tener efectos adversos para la salud humana y el medio ambiente.

La contaminación por partículas también ha sido una preocupación en Madrid. Las partículas en suspensión, especialmente las $PM_{2.5}$, son partículas pequeñas que pueden penetrar en los pulmones y causar problemas respiratorios y cardiovasculares. Las fuentes de estas partículas incluyen el tráfico, la industria, la calefacción residencial y las condiciones meteorológicas desfavorables que atrapan los contaminantes en la atmósfera.

La calidad del aire en Madrid ha llevado a la implementación de medidas para reducir la contaminación. Se han establecido protocolos de actuación que incluyen restricciones al tráfico, especialmente durante episodios de alta contaminación. Además, se ha fomentado el uso de transporte público y se han promovido políticas para incentivar el uso de vehículos eléctricos y la mejora de la eficiencia energética de los edificios.

Efecto de los niveles de contaminación en la salud y en la biodiversidad

La contaminación atmosférica se refiere a la presencia de sustancias nocivas en el aire que respiramos. Estas sustancias pueden ser de origen natural o antropogénico, como la quema de combustibles fósiles, la emisión de gases de escape de vehículos y la liberación de productos

químicos industriales. Según la Organización Mundial de la Salud (OMS), la contaminación del aire es responsable de alrededor de 4 millones de muertes prematuras al año en todo el mundo (OMS., 2021). Esto es debido a que uno de los mayores efectos de la contaminación atmosférica es en la salud humana. Las partículas finas en el aire pueden penetrar en los pulmones y el sistema circulatorio, lo que puede provocar enfermedades respiratorias, enfermedades cardiovasculares y cáncer. Y esto es algo que afecta también al continente europeo, según un estudio de la Agencia Europea de Medio Ambiente, la contaminación del aire provoca 400,000 muertes prematuras cada año en Europa (AEMA., 2018).

A nivel de España, la problemática de la contaminación del aire también afecta de manera desproporcionada a las zonas con mayor densidad de población y concentración de industrias. Esta disparidad se puede observar claramente en la figura 1.2 (Jiménez y Gil., 2018), en la que se revela que las áreas urbanas densamente pobladas y las regiones industriales presentaron niveles más altos de mortalidad prematura durante los años 2000-2009 a causa de partículas PM_{10}, con lo que se puede intuir que como se esperaba las concentraciones de este tipo de partículas fueron superiores en esas zonas, con las consecuencias para la salud que eso conllevó.

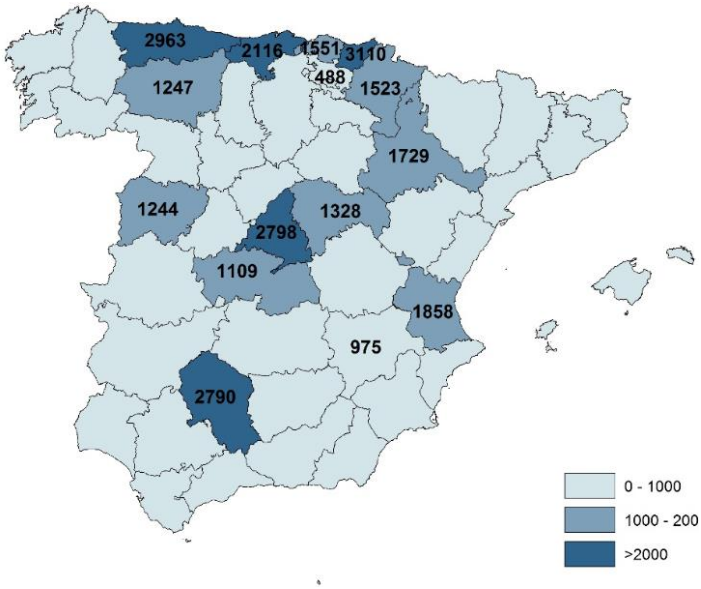

Figura 1.2. Número de muertes atribuibles a la contaminación por partículas PM_{10} en el periodo 2000-2009.

Además de afectar a la salud humana, la contaminación atmosférica también tiene un impacto significativo en la biodiversidad y los ecosistemas. La lluvia ácida, por ejemplo, es un tipo de contaminación atmosférica que puede dañar la flora y fauna. La lluvia ácida se forma cuando los óxidos de nitrógeno y sulfuro se mezclan con el agua en la atmósfera, formando ácido sulfúrico

y ácido nítrico. Estos ácidos pueden depositarse en la superficie terrestre, acidificando el suelo y el agua, lo que puede matar plantas y animales.

En resumen, la contaminación atmosférica tiene un impacto significativo en la salud y la biodiversidad. La reducción de las emisiones de contaminantes es esencial para proteger la salud humana y el medio ambiente. Los gobiernos y las empresas tienen la responsabilidad de tomar medidas para reducir la contaminación atmosférica y promover una economía más sostenible.

Situación actual de medidores en la Comunidad de Madrid

El primer paso a la hora de conocer los efectos de la contaminación y poder tomar medidas para atajarla consiste en conocer los niveles en los distintos puntos de una región, por ello en un mundo cada vez más preocupado por la calidad del medio ambiente, la estimación de los niveles de contaminación se ha convertido en una herramienta indispensable para comprender y abordar los desafíos ambientales. Sin embargo, no siempre es posible contar con estaciones de medición en todos los puntos de un territorio ya sea por la extensión de este, como por los elevados costes de instalación y mantenimiento. Por ello, exploraremos la utilidad de poder estimar los niveles de contaminación en ausencia de estaciones de medición generalizadas, destacando cómo esta capacidad puede brindar beneficios significativos para la protección y conservación del medio ambiente.

La estimación de los niveles de contaminación desempeña un papel fundamental en la planificación urbana y la conservación del medio ambiente. Al tener una idea precisa de los niveles de contaminación en diferentes áreas, los urbanistas y los planificadores pueden diseñar ciudades y comunidades más sostenibles, evitando la construcción de infraestructuras sensibles en zonas altamente contaminadas. Además, esta información también puede ser utilizada para establecer áreas protegidas y políticas de conservación, garantizando la preservación de ecosistemas frágiles y la biodiversidad.

Red de Calidad del Aire de la Comunidad de Madrid:

La Comunidad de Madrid cuenta con una red de medidores de contaminación atmosférica distribuidos en diferentes puntos de la región que proporcionan datos en tiempo real sobre los niveles de contaminación atmosférica en las diferentes ubicaciones.

Esta red de monitoreo de la calidad del aire está gestionada por el Área de Calidad del Aire de la Dirección General de Medio Ambiente y Sostenibilidad de la Comunidad de Madrid que se encarga de monitorizar los niveles de diversos contaminantes atmosféricos, incluyendo dióxido de nitrógeno (NO_2), partículas en suspensión (PM_{10} y $PM_{2.5}$), ozono (O_3), entre otros.

Los datos recopilados por los medidores se utilizan para evaluar la calidad del aire y tomar medidas en caso de que se superen los límites establecidos. Además, la información se utiliza para informar al público y a las autoridades competentes sobre la situación de la contaminación atmosférica en la región.

Actualmente la Red de Calidad del Aire de la Comunidad de Madrid que está formada por la propia red del ayuntamiento de Madrid (figura 1.4) y 24 estaciones distribuidas entre zonas urbanas rurales e industriales en los diferentes municipios de la comunidad como se muestra en la figura 1.3 (CAM., 2023).

Figura 1.3. Localización de las estaciones de la Red de Calidad del Aire

En cuanto al municipio de Madrid este tiene su propia Red Automática de Vigilancia de la Calidad Atmosférica del Ayuntamiento de Madrid que consta de 25 estaciones desplegadas por diversas zonas de la ciudad como se muestra en la figura 1.4 (CAM., 2023).

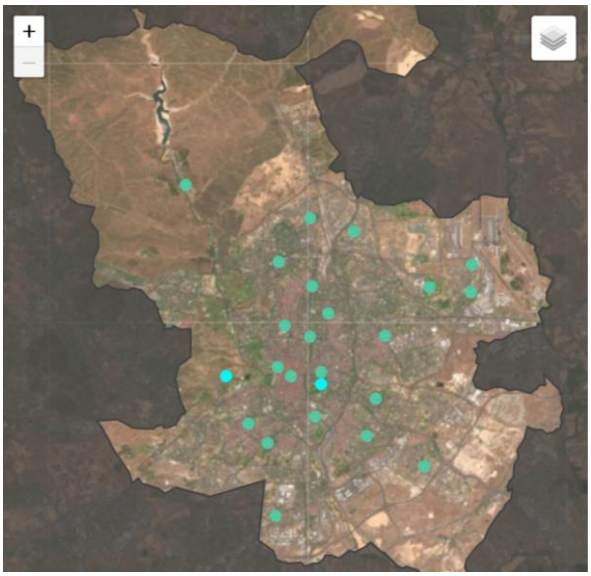

Figura 1.4. Localización actual de las estaciones de medición en el municipio de Madrid

Estas estaciones permiten estimar la calidad del aire en puntos concretos, pero al no encontrarse homogéneamente repartidas tanto a nivel de la ciudad como de la comunidad existe un sesgo a la hora de determinar los niveles de unas zonas frente a otras (C.P.D., 2012) ya que a la hora de publicar los datos solo se publican los niveles por estación (Madrid Salud., 2023) dejando a las zonas que se encuentren alejadas de las estaciones sin conocer sus niveles reales de contaminación. Es por esto por lo que existe la necesidad de realizar una interpolación espacial para generar información fiable a lo largo del tiempo pudiéndose así obtener datos a nivel de código postal en todo el territorio.

La estimación de los niveles de contaminación en áreas donde no existen estaciones de medición permite obtener una visión más completa del panorama general de la contaminación ambiental. Al combinar datos limitados de estaciones de monitoreo con técnicas de interpolación y modelos predictivos, se pueden generar mapas y modelos tridimensionales que representen las áreas no monitoreadas. Esto proporciona información valiosa sobre los patrones y las tendencias de la contaminación en todo el territorio, permitiendo identificar las zonas más afectadas y establecer estrategias de mitigación de manera más precisa.

La capacidad de estimar los niveles de contaminación es fundamental para la toma de decisiones informadas en política ambiental. Los gobiernos y las autoridades locales pueden utilizar estos datos estimados para identificar las áreas con mayores problemas de contaminación y priorizar las acciones necesarias. Además, los modelos de estimación pueden ayudar a predecir escenarios futuros de contaminación y evaluar el impacto de diferentes medidas de control. Esto permite desarrollar estrategias de mitigación más efectivas y dirigidas, optimizando los recursos disponibles.

También es crucial contar con estas estimaciones para proteger la salud pública. La exposición a contaminantes atmosféricos y ambientales puede tener efectos adversos en la salud de las personas, incluyendo enfermedades respiratorias, cardiovasculares y cáncer. Al estimar los niveles de contaminación en áreas no monitoreadas, se pueden identificar posibles puntos críticos de exposición y tomar medidas preventivas para minimizar los riesgos. Además, esta información estimada puede ser utilizada para educar y concienciar a la población sobre los peligros de la contaminación y promover cambios en el comportamiento individual y colectivo.

Capítulo II: Metodología para la estimación de partículas contaminantes del aire a nivel de código postal

Este capítulo presenta los métodos más utilizados para estimar los niveles de contaminación ambiental, tratando las limitaciones que presentan cada uno de ellos y evaluando cuales aportan mayores beneficios a la hora de estimar los niveles de partículas contaminantes a nivel de código postal. A partir de esto se explicarán más en profundidad las metodologías que se elijan para llevar posteriormente a la implementación.

Métodos de estimación más utilizados para estimar niveles de contaminación

Existen varios métodos de estimación de la contaminación ambiental que se utilizan actualmente, entre ellos:

1. Monitoreo directo: Este método consiste en medir los contaminantes en tiempo real utilizando instrumentos de monitoreo. Se pueden realizar mediciones en la atmósfera, en el agua o en el suelo. Los datos obtenidos son precisos y confiables, pero suelen ser costosos y requieren equipo especializado.
2. Modelos matemáticos: Estos modelos utilizan información de la calidad del aire y factores ambientales, como la velocidad del viento y la temperatura, para estimar la concentración de contaminantes en una determinada área. Este método es útil para prever la contaminación futura y puede ser menos costoso que el monitoreo directo.
3. Muestreo pasivo: Este método utiliza dispositivos que no requieren energía para tomar muestras de contaminantes durante un período determinado. Luego, las muestras se analizan en un laboratorio para determinar la cantidad de contaminantes presentes. Este método es útil para evaluar la contaminación a largo plazo y en áreas remotas, pero puede no ser tan preciso como el monitoreo directo.
4. Índices de calidad del aire: Los índices de calidad del aire son herramientas que utilizan valores de contaminación medidos y modelados para proporcionar información simplificada sobre la calidad del aire en un área determinada. Estos índices se utilizan comúnmente para informar al público sobre la calidad del aire y tomar decisiones sobre la salud y la seguridad pública.
5. Bioindicadores: Los bioindicadores son organismos vivos que se utilizan para medir la calidad del aire, el agua y el suelo. Los bioindicadores pueden ser plantas, animales o microorganismos y su presencia, ausencia o cambios en su estado de salud pueden indicar la presencia de contaminantes en el medio ambiente. Este método puede ser útil para monitorear la contaminación a largo plazo y en áreas remotas.

Modelos Matemáticos para la estimación de datos espaciales

Los modelos matemáticos son herramientas potentes para estimar la concentración de contaminantes en el medio ambiente. Estos modelos utilizan una serie de ecuaciones matemáticas que relacionan la cantidad de contaminantes emitidos, las condiciones meteorológicas y otros factores ambientales para predecir la concentración de contaminantes en una determinada área.

Para desarrollar un modelo matemático, se deben recopilar datos precisos sobre las fuentes de contaminación, las condiciones meteorológicas, la topografía y la calidad del aire en una determinada área. Estos datos se utilizan para crear una representación matemática de la situación, que luego se puede utilizar para simular diferentes escenarios de contaminación.

Los modelos matemáticos se utilizan para prever la contaminación futura, identificar áreas de alta contaminación y evaluar el impacto de las medidas de control de la contaminación. Por ejemplo, un modelo matemático podría utilizarse para prever la cantidad de contaminantes emitidos por una fábrica y su impacto en la calidad del aire en una determinada área. Esto permitiría a los reguladores evaluar si se necesitan medidas de control adicionales para proteger la salud pública y el medio ambiente.

Algunos de los modelos matemáticos más comunes utilizados para estimar la contaminación incluyen modelos de dispersión atmosférica, modelos de transporte y modelos de calidad del aire. Cada modelo tiene sus propias fortalezas y limitaciones y se utiliza en función de las necesidades específicas del proyecto y los datos disponibles. En general, los modelos matemáticos son herramientas importantes para evaluar la contaminación y tomar decisiones informadas sobre la gestión de la calidad del aire.

Modelos de estimación de la contaminación mediante interpolación

La interpolación es un método ampliamente utilizado para estimar la contaminación ambiental en áreas donde no se han tomado mediciones directas. La técnica se basa en la recopilación de datos de diferentes puntos del área de estudio para estimar la concentración de contaminantes en lugares donde no se han tomado mediciones directas. Este método es especialmente útil para identificar patrones de contaminación y áreas de alto riesgo.

Uno de los modelos más comunes de interpolación es la interpolación inversa ponderada (IIP), que estima la concentración de contaminantes en un punto de interés como una combinación lineal de los valores de medición de los puntos más cercanos, ponderados por su distancia. La IIP se ha utilizado con éxito para estimar la concentración de dióxido de nitrógeno (NO_2) en áreas urbanas de China, donde la contaminación del aire es un problema importante. En este estudio, se encontró que la IIP era más precisa que otros métodos de interpolación, como el de kriging ordinario.

El kriging es otro modelo común de interpolación, que utiliza un modelo de regresión espacial para estimar la concentración de contaminantes en una determinada ubicación. La kriging ha sido utilizada para estimar la concentración de partículas finas ($PM_{2.5}$) en áreas urbanas de Estados Unidos. En este estudio, se encontró que la kriging tenía una precisión aceptable para estimar la concentración de $PM_{2.5}$, pero que la precisión variaba según el área de estudio y la calidad de los datos.

Otro de los usos de la interpolación es su utilización para crear mapas de concentración de contaminantes en un área determinada, lo que puede ayudar a identificar patrones de contaminación y áreas de alto riesgo. Estos mapas se pueden utilizar para tomar decisiones informadas sobre la gestión de la calidad del aire y para evaluar el impacto de las medidas de control de la contaminación. En un estudio reciente, se utilizó la IIP para crear mapas de concentración de NO_2 en áreas urbanas de América Latina. En este estudio, se encontró que la IIP era una técnica efectiva para estimar la concentración de NO_2 y que los mapas creados eran útiles para identificar áreas de alto riesgo y priorizar las medidas de control de la contaminación.

A pesar de las ventajas de la interpolación, también tiene limitaciones. En particular, la precisión de la interpolación depende en gran medida de la calidad y la cantidad de datos disponibles. Además, la interpolación puede ser sensible a la elección de los parámetros del modelo y a la distribución espacial de los puntos de medición.

En conclusión, la interpolación es un método útil para estimar la contaminación ambiental en áreas donde no se han tomado mediciones directas. La elección del método de interpolación adecuado y la calidad de los datos disponibles son importantes para garantizar una estimación precisa de la concentración de contaminantes.

Método de la distancia inversa

El método de la distancia inversa es una técnica comúnmente utilizada para la estimación de la contaminación mediante interpolación. Este método se basa en la premisa de que los valores medidos en puntos cercanos tienen una mayor influencia en la estimación de la concentración de contaminantes en un punto de interés.

El método de la distancia inversa se basa en el principio de que los puntos de medición más cercanos a un punto de interés tienen una mayor influencia en la estimación de la concentración de contaminantes en ese punto. La técnica asigna un peso a cada punto de medición en función de su distancia al punto de interés. Los puntos más cercanos reciben un peso mayor, mientras que los puntos más alejados tienen un peso menor en la estimación.

Una de las ventajas del método de la distancia inversa es su simplicidad y facilidad de implementación. No requiere suposiciones específicas sobre la distribución espacial de los datos y no necesita ajustar un modelo matemático complicado. Además, puede proporcionar resultados satisfactorios cuando los datos de medición están disponibles en una densidad razonable alrededor del punto de interés.

Sin embargo, el método de la distancia inversa también tiene algunas limitaciones importantes. La precisión de la estimación depende en gran medida de la del grado de influencia que se les otorgue a los puntos más cercanos. Un valor incorrecto puede resultar en una subestimación o sobreestimación de la concentración de contaminantes. La elección de este parámetro requiere un análisis cuidadoso y puede variar según el contexto y las características del problema.

Otra limitación del método de la distancia inversa es que no considera la variabilidad espacial de los datos y asume que los puntos más cercanos tienen una influencia igualmente importante. Sin embargo, esto puede no ser cierto en todos los casos, especialmente cuando la distribución espacial de los puntos de medición es irregular o existen barreras físicas que afectan la dispersión de los contaminantes.

A pesar de estas limitaciones, el método de la distancia inversa sigue siendo ampliamente utilizado en la estimación de la contaminación mediante interpolación, especialmente cuando se cuenta con datos de medición adecuados y se aplica con precaución. Es una herramienta útil para obtener una estimación aproximada de la concentración de contaminantes en lugares donde no se han tomado mediciones directas.

Método de Kriging

El kriging, es un método estadístico utilizado para estimar valores desconocidos en ubicaciones no muestreadas a partir de observaciones realizadas en puntos cercanos. El método se basa en el análisis geoestadístico, que considera la variabilidad espacial de un fenómeno y utiliza la información de la vecindad para hacer predicciones.

La teoría básica del kriging se basa en el principio de la estacionariedad, que supone que las propiedades del fenómeno en estudio son consistentes en todo el dominio. En el contexto de la estimación de la contaminación, esto implica que los niveles de contaminantes tienen una estructura espacial que puede ser modelada y utilizada para predecir valores en ubicaciones no muestreadas.

El kriging se divide en dos etapas principales: el cálculo de semivariogramas y la interpolación de los valores desconocidos. El semivariograma es una función que describe la variabilidad espacial de los datos y proporciona información sobre la relación entre la distancia y la diferencia

de los valores observados. Se calcula a partir de la variabilidad de los datos en diferentes distancias y direcciones, y su forma característica se utiliza para seleccionar el modelo de variograma adecuado.

Una vez que se ha modelado el semivariograma, se utiliza para interpolar los valores desconocidos. El kriging considera tanto la información espacial de los datos observados como la información proporcionada por el modelo de variograma para realizar estimaciones en ubicaciones no muestreadas. Las estimaciones obtenidas mediante kriging no solo se basan en los valores observados más cercanos, sino también en la estructura espacial de los datos y la incertidumbre asociada a la estimación.

El método de kriging ha demostrado ser útil en la estimación de la contaminación en diversas áreas, como la calidad del aire, la contaminación del agua y la contaminación del suelo. Por ejemplo, se utilizó el kriging ordinario para estimar la concentración de $PM_{2.5}$ (partículas finas en suspensión) en una ciudad altamente industrializada. Los resultados mostraron que el kriging proporcionaba estimaciones precisas de la contaminación atmosférica en ubicaciones no muestreadas, lo que permitía identificar áreas con altos niveles de contaminantes y orientar estrategias de control de la contaminación.

El método de kriging se ha consolidado como una técnica efectiva para la estimación de la contaminación mediante interpolación. Su capacidad para capturar la estructura espacial de los datos y considerar la incertidumbre asociada a las estimaciones lo convierte en una herramienta valiosa en el campo de la ciencia ambiental. Sin embargo, es importante destacar que la precisión de las estimaciones de contaminación mediante kriging depende de la calidad de los datos y de la correcta modelización del variograma.

El uso del método de kriging en la estimación de la contaminación ofrece numerosas ventajas, como la capacidad de generar mapas de contaminantes en ubicaciones no muestreadas y la posibilidad de evaluar la incertidumbre asociada a las estimaciones. Esto permite a los científicos y responsables de la toma de decisiones tener una visión más completa de los problemas ambientales y tomar medidas efectivas para abordarlos.

Desarrollo matemático

Índices y Conjuntos:

- o: Índice utilizado para señalar los datos referidos al punto en donde el valor de contaminación se está estimando.
- i: Subíndice utilizado para señalar los puntos cuyos datos de contaminación están siendo utilizados para realizar la estimación.
- j: Subíndice utilizado para señalar un punto de medición diferente al i cuando se comparan datos de dos estaciones de medición diferentes.
- k: Subíndice utilizado para referirse a los distintos tipos de contaminantes medidos en las estaciones de la red de calidad del aire.
- P: Conjunto de partículas contaminantes del aire con el elemento típico p_k, $p \in P$
- E: Conjunto de estaciones de medición de calidad del aire con elemento típico e_i, $e \in E$
- E_p: Conjunto de estaciones de medición de la calidad del aire que miden la partícula p_k, $E_p \subseteq E \ / e \in E_p$

La estimación mediante interpolación consiste darle una determinada importancia o "peso" a los puntos en donde se tiene información para hallar el valor en otro punto en el cual se desconoce, tal y como se observa en la fórmula (2, 3). La diferencia entre el método de la distancia inversa y el de kriging se encuentra en procedimiento que se sigue para asignar estos pesos.

En la interpolación por medio de la distancia inversa se calculan los pesos o ponderaciones inversas de los valores conocidos en función de su distancia al punto de estimación de la forma que se explica a continuación.

Sea s_o un punto con coordenadas (x, y) cuyo valor de contaminación $\hat{z}(s_o)$ de la partícula p_k se quiere estimar y sean $z(s_i)$ los valores de contaminación medidos para dicha partícula por el conjunto de estaciones E_p que miden la partícula en los puntos s_i cercanos a s_o, supongamos que tenemos n estaciones de medición con valores conocidos ($z(s_1)$, $z(s_2)$, ..., $z(s_n)$) con sus respectivas coordenadas espaciales (x_1, y_1), (x_2, y_2), ..., (x_n, y_n).

La distancia $d(s_o, s_i)$ entre el punto cuyo valor de contaminación estamos estimando s_o y cada punto con valor conocido s_i se puede calcular utilizando la ecuación (2.1) de la distancia euclidiana.

$$d(s_o, s_i) = \sqrt{(x - x_i)^2 + (y - y_i)^2} \qquad (2.1)$$

Posteriormente, se aplica una función de ponderación inversa a las distancias para calcular el peso (w_i) que se le otorgará a cada valor conocido $z(s_i)$ en función de su distancia a s_o como se muestra en (2.2). Estos pesos (w_i) indican la importancia relativa que se le va a dar a cada valor conocido a la hora de realizar la estimación en el punto s_o.

$$w_i = \frac{1}{(d(s_o, s_i))^p} \qquad (2.2)$$

Para este cálculo se utiliza el parámetro *p* el cual controla el grado de influencia que se les otorga a los puntos más cercanos frente a los más lejanos. Siempre que p sea mayor que 1, cuanto más cercano se encuentre un punto conocido, mayor será el peso que se le otorgue y a medida que aumente el valor de *p* mayor será la importancia otorgada a los puntos más próximos frente a los más remotos. Mientras que para valores de p menores que 1 se otorgará mayor relevancia a los puntos más lejanos.

Una vez obtenidos los pesos el siguiente paso consiste en normalizar los valores obtenidos como se muestra en (2.3), de este modo el valor λ_i que se obtiene es una representación de que porcentaje del valor final representará cada valor conocido.

$$\lambda_i = \frac{w_i}{w_1+w_2+\cdots+w_n} \tag{2.3}$$

Finalmente, se realiza la estimación del valor desconocido $\hat{z}(s_o)$ asignando su peso a cada uno de los valores conocidos utilizando los pesos normalizados como se muestra en (2.4).

$$\hat{z}(s_o) = \sum_{i=1}^{n} \lambda_i * z(s_i) \tag{2.4}$$

La estimación que se obtiene es el resultado de realizar una combinación lineal de los valores conocidos ponderados por los pesos inversos de la distancia.

En cuanto a la estimación a partir del modelo de kriging, esta se basa en el análisis geoestadístico y la teoría de la estadística espacial. Existen diferentes tipos de kriging, siendo los más comunes el kriging ordinario y el kriging universal. El kriging ordinario, que es el que se va a utilizar, asume estacionariedad local, lo que implica que las estimaciones se basan únicamente en los puntos cercanos. Por otro lado, el kriging universal tiene en cuenta la estacionariedad global, lo que permite utilizar información de puntos más lejanos y mejorar la precisión de las estimaciones.

Al igual que en el método de la distancia inversa se asume que las predicciones a realizar ($\hat{z}(s_o)$) van a consistir en combinaciones lineales de los datos de los que se dispone ($z(s_i)$) como se muestra en (2.5) estando la diferencia en la forma de obtener los pesos normalizados (λ_i).

$$\hat{z}(s_o) = \sum_{i=1}^{n} \lambda_i * z(s_i) \tag{2.5}$$

Para hallar estos pesos normalizados con el método de kriging el primer paso es calcular el semivariograma experimental γ, que es una función que describe la variabilidad espacial de los datos. El semivariograma mide la semivarianza entre pares de puntos en función de la distancia que los separa calculándose como se muestra en (2.6)

$$\gamma(\mathrm{d}(s_o, s_i)) = \frac{Var(z(s_o)-z(s_i))}{2} \tag{2.6}$$

Donde γ es el valor del semivariograma para una distancia determinada, Var es la varianza y $z(s_i)$ es el valor de la contaminación en el punto determinado siendo s_o el punto en el que se quiere realizar la estimación y s_i el resto de los puntos con datos conocidos.

Una vez que se ha calculado el semivariograma experimental se selecciona un semivariograma teórico que se va a utilizar para realizar la estimación, los semivariogramas son gráficos de las semivarianzas en función de la distancia a un punto, a partir de una cierta distancia la semivarianza dejará de aumentar estabilizándose en un valor igual a la varianza media, a esta región se la llama silo o patamar y la distancia entre el inicio y el comiendo del silo se denomina rango. Existen 4 modelos fundamentales de semivariogramas (Viera y González., 2002).

Semivariograma esférico: Es uno de los modelos más utilizados y se caracteriza por tener un crecimiento inicial rápido y alcanzar un límite superior o silo. A medida que la distancia aumenta, el semivariograma se estabiliza en el silo, su fórmula es la que aparece en (2.7).

$$\gamma\big(\mathrm{d}(s_o, s_i)\big) = c + (silo - c) * \left(\frac{3*\mathrm{d}(s_o,s_i)}{2*rango} - \frac{\mathrm{d}(s_o,s_i)^3}{2*rango^3}\right), \text{ para } 0 \leq \mathrm{h} \leq \text{range}$$

$$\gamma\big(\mathrm{d}(s_o, s_i)\big) = silo, \text{ para } \mathrm{h} > \text{rango} \qquad (2.7)$$

Donde $\gamma\big(\mathrm{d}(s_o, s_i)\big)$ es el valor del semivariograma a una determinada distancia h, c es el valor del semivariograma en a una distancia = 0, también conocido como "efecto pepita". Silo es el valor máximo que alcanza el semivariograma (límite superior), siendo el rango la distancia a la cual el semivariograma alcanza el silo y define el alcance de la correlación espacial.

Semivariograma exponencial: En este modelo, el semivariograma aumenta exponencialmente a medida que la distancia entre los puntos se incrementa. No alcanza un límite superior, lo que significa que la variabilidad continúa aumentando a medida que aumenta la distancia calculándose como se muestra en (2.8).

$$\gamma\big(\mathrm{d}(s_o, s_i)\big) = c + (silo - c) * (1 - e^{\frac{-h^2}{rango}}) \qquad (2.8)$$

Semivariograma gaussiano: El semivariograma gaussiano se caracteriza por una curva simétrica y suave. Alcanza el sill rápidamente y se estabiliza a medida que la distancia se aleja del origen. Es útil para describir fenómenos que exhiben una estructura de correlación espacial suave y continua y se calcula como se muestra en (2.9).

$$\gamma\big(\mathrm{d}(s_o, s_i)\big) = c + (silo - c) * (1 - e^{\frac{-h^2}{rango}}), \text{ para } \mathrm{d}(s_o, s_i) < \text{rango}$$

$$\gamma\big(\mathrm{d}(s_o, s_i)\big) = c + (silo - c), \text{ para } \mathrm{d}(s_o, s_i) > \text{rango} \qquad (2.9)$$

Semivariograma lineal: En este modelo, el semivariograma aumenta linealmente a medida que la distancia aumenta. No alcanza un límite superior y no muestra saturación. Es útil para describir fenómenos con una variabilidad constante a medida que aumenta la distancia. Es el más simple de todos calculándose tal y como se muestra en (2.10).

$$\gamma\big(\mathrm{d}(s_o, s_i)\big) = c + (silo - c) * (\frac{\mathrm{d}(s_o, s_i)}{rango}) \quad (2.10)$$

Una vez seleccionado el semivariograma teórico que se va a utilizar se ajustan los parámetros (c, silo y rango) para minimizar la diferencia entre el semivariograma experimental y el teórico escogido. Con esto se resuelve la ecuación de kriging (2.11) donde los subíndices "i" y "j" denotan diferentes puntos muestreados, obteniendo así los pesos normalizados λ_i necesarios para realizar la estimación. Donde n es el número de observaciones y m es el multiplicador de Lagrange utilizado para la minimización de las restricciones.

$$\sum_{i=1}^{n} \lambda_i * \gamma\left(\mathrm{d}\big(s_i, s_j\big)\right) + m = \gamma\big(\mathrm{d}(s_o, s_i)\big)$$

$$\sum_{i=1}^{n} \lambda_i = 1 \quad (2.11)$$

El desarrollo matemático del kriging implica el cálculo del semivariograma, el ajuste de un modelo de variograma y la utilización de la estructura espacial para realizar estimaciones en ubicaciones no muestreadas. El método busca obtener estimaciones precisas considerando la información espacial y la incertidumbre asociada a las estimaciones.

Capítulo III: Aplicación y Resultados

Implementación

Para la estimación de los niveles de contaminación en la Comunidad de Madrid se van a utilizar los datos de los últimos 10 años (desde el 1 de enero de 2012 hasta el 31 de diciembre de 2021) de las 47 estaciones repartidas por la Comunidad de Madrid. Conociendo los valores de contaminación por hora para cada uno de los 12 distintos contaminantes medidos y la localización de estos puestos de medida.

La implementación del modelo de interpolación tanto mediante la distancia inversa como por el método de kriging se realizará mediante programas en MATLAB. Una vez que se hayan implementado los modelos de interpolación en MATLAB se evaluarán las estimaciones obtenidas.

Para el método de la distancia inversa se experimentará con diferentes valores del parámetro p. Esto nos permitirá observar cómo afecta la elección de este parámetro a los resultados obtenidos tanto en precisión como en carga computacional, para poder modularlo y obtener el valor que produzca una estimación más fiable. Los valores de p escogidos han sido el 1, 2 y 3, ya que los valores de p usualmente escogidos para este tipo de interpolación suelen estar entre 1 y 3, además de ser 2 el más común (Gotway., 1996).

Otro parámetro cuyo impacto se va a estudiar es el valor del radio alrededor del punto donde se está estimando a partir del cual se descartan las mediciones y no se tienen en cuenta para la estimación. Esto provoca que las estaciones fuera de ese radio se consideren demasiado lejanas como para tener relevancia a la hora de calcular la estimación.

Para poder evaluar los resultados obtenidos se elegirán 2 estaciones de medida de manera aleatoria para utilizar como “estaciones test” en cuyas coordenadas se estimarán los valores de contaminación durante los 10 años de los que disponemos datos utilizando las medidas del resto de estaciones, estos valores obtenidos se compararán con las propias medidas de estas 2 estaciones para poder evaluar el error cometido por cada modelo. Se ha decidido utilizar 2 estaciones como test para poder asegurar que las evaluaciones obtenidas son consistentes y que no reflejan un sesgo local.

Es importante precisar que no todas las estaciones toman medidas para todos los tipos de contaminantes, por esto se tendrá que discriminar y solo estimar valores para los contaminantes para los que la estación “test” tenga medidas ya que de lo contrario no tendríamos datos con los que comparar las estimaciones para obtener su nivel de error. También hay que tener en cuenta que solo se puede utilizar el conjunto de estaciones que tomen medidas de ese tipo de contaminantes para el conjunto de “training”.

Para medir la precisión de los modelos se utilizarán 2 estimadores, el primero será el error cuadrático medio (ECM), calculado como se muestra en (3.1) donde $\hat{Y}_i$ es cada uno de los valores estimados por el modelo para cada hora en las coordenadas de la estación test e Y_i es cada uno de los respectivos valores registrados por dicha estación.

$$ECM = \frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}{n} \quad (3.1)$$

El segundo estimador que se utilizará será el de autocorrelación para comprobar que la estructura de datos obtenida guarda relación con la muestreada, es decir, si en los datos de la estación que se está usando como "test" se observa que existe una correlación ya sea positiva o negativa entre los valores de distintos contaminantes en los valores estimados se debe respetar esa correlación.

Para crear la estimación de cada modelo se realizará la estimación un número elevado de veces utilizando cada vez un grupo diferente de "estaciones training", y se realizará la media de las estimaciones realizadas para asegurar de esta forma que la estimación es consistente. El grupo de training que se utilizará en cada iteración será elegido de manera aleatoria entre todas las estaciones que no sean "test". Para el porcentaje de estaciones que serán elegidas como "training" en cada ocasión se evaluarán utilizando tanto el 85% como el 70% para comprobar que no se comete ni "overfitting" ni "underfitting" en ninguno de los modelos (Foody, McCulloch & Yates., 1995).

Una vez obtenidos tanto el error cuadrático medio como la autocorrelación de cada uno de los modelos se podrá comparar tanto su precisión y consistencia como sus tiempos de procesamiento para así poder obtener conclusiones respecto a las bondades de los distintos modelos, así como valorar cual escoger según las necesidades surgidas.

Análisis de los resultados

Una vez implementados los modelos se puede estudiar la validez de cada uno de ellos, para ello se ha estimado, con cada una de las configuraciones estudiadas, los niveles de contaminación de hora a hora de los 10 años de los que se disponían datos y se ha hallado el error cuadrático medio (ECM) para el conjunto de estimaciones por tipo de contaminante para cada una de las distintas configuraciones.

Se ha elegido como puntos en los que realizar el análisis la estación de Plaza de España, ya que se encuentra dentro de la red del propio ayuntamiento de Madrid y por tanto en una zona con una elevada cantidad de estaciones cercanas, la estación de Colmenar Viejo, que se encuentra fuera de la ciudad de Madrid pero en una zona con una densidad media de estaciones y por último la estación de Guadalix de la sierra, que se encuentra en la parte más exterior de la Comunidad de Madrid y por tanto en una zona con una baja densidad de estaciones de medición cercanas.

Resultados por el método de la distancia inversa

En primer lugar, se analizará cómo se comporta el método de la distancia inversa, para la estación de Plaza de España (tabla 3.1).

Contaminantes	Plaza de España					
Dióxido de azufre	4.58	5.20	5.13	5.14	5.20	5.03
Dióxido de nitrógeno	25.78	25.64	25.57	25.61	25.65	25.48
Monóxido de carbono	0.22	0.25	0.25	0.25	0.25	0.24
Monóxido de nitrógeno	48.96	55.92	55.14	55.22	55.64	53.52
Óxidos de nitrógeno	83.45	89.55	88.91	89.06	89.39	87.62
Porcentaje (%)	85	70	85	70	85	70
Exponente (p)	1	1	2	2	3	3
Tiempo (s)	1.74	1.70	1.87	1.79	1.95	1.86

Tabla 3.1. ECM por el método de la distancia inversa para la estación de Plaza de España

Se puede observar que los datos se comportan de una forma muy consistente, ya que el error cuadrático medio apenas varía según el porcentaje de estaciones que se seleccione para realizar la estimación y tampoco lo hace según modifiquemos el exponente p. Esto se explica ya que la estación de medición de Plaza de España se encuentra situada en una parte central del municipio de Madrid tal y como se aprecia en la figura 1.4, esto hace que haya un elevado número de estaciones cercanas y haya poca variación en las estimaciones según el número de estaciones que se utilicen.

En segundo lugar, se analizará los datos obtenidos para la estación de medición de Guadalix de la Sierra (tabla 3.2).

Contaminantes	Guadalix de la Sierra					
Dióxido de nitrógeno	21.96	5.31	6.27	6.30	1.58	4.07
Monóxido de nitrógeno	20.76	4.95	5.84	5.87	1.34	3.59
Óxidos de nitrógeno	42.49	10.07	11.75	12.37	2.87	7.54
Ozono	13.40	3.11	3.59	3.78	0.89	2.38
Partículas en suspensión < PM10	6.70	1.59	1.78	1.83	0.45	1.23
Porcentaje (%)	85	70	85	70	85	70
Exponente (*p*)	1	1	2	2	3	3
Tiempo (s)	1.91	1.73	2.01	1.81	2.18	2.05

Tabla 3.2. ECM por el método de la distancia inversa para la estación de Guadalix de la Sierra

En esta ocasión se observa que los datos ya no se comportan de una forma tan consistente ya que el error cuadrático varía según la configuración que utilicemos. Principalmente se observa una gran diferencia entre el error cometido al utilizar un 70% de estaciones frente a utilizar un 85% cuando se utiliza un exponente p=1 o p=3 mejorando las estimaciones al reducir el número de datos utilizados en el primer caso y empeorándolo en el segundo. Esto guarda relación con la situación en la que se encuentra la estación de Guadalix de la Sierra ya que está situada en una zona rural de la periferia de la Comunidad de Madrid como se observa en la figura 1.4.

Esta localización conlleva que existan pocas estaciones de medición cercanas, lo que produce que cuando el exponente es bajo (p=1) la media de estimaciones genere mejores resultados cuantas menos estaciones se seleccionen ya que un exponente bajo reduce la importancia que se le otorga a las estaciones cercanas, pero al seleccionar un menor número de estaciones, de media se reduce el peso de las estaciones lejanas al ser más y resultar por tanto excluidas en mayor número. Utilizando la configuración contraria (p=3), la media de estimaciones genera mejores resultados cuantas mas estaciones se seleccione ya que se otorga una mayor importancia a las estaciones cercanas y cuanto mayor número de estaciones se seleccione mas probable es que estas estaciones se encuentren incluidas. En cuanto al caso intermedio (p=2), se observa que el error obtenido es indistinto del número de estaciones que se seleccionen ya que un efecto compensa al otro.

Por último, se analizan los resultados de la estación de medición situada en Colmenar Viejo contenidos en la tabla 3.3.

Contaminantes	Colmenar Viejo					
Dióxido de nitrógeno	9.61	4.59	0.40	2.39	2.33	1.26
Monóxido de carbono	0.18	0.09	0.01	0.05	0.06	0.03
Monóxido de nitrógeno	13.64	6.61	0.57	3.28	3.28	1.74
Óxidos de nitrógeno	21.59	10.45	0.88	5.13	5.00	2.82
Ozono	11.78	5.60	0.47	2.85	3.05	1.82
Partículas en suspensión < PM10	5.47	2.63	0.21	1.28	1.40	0.82
Porcentaje (%)	85	70	85	70	85	70
Exponente (p)	1	1	2	2	3	3
Tiempo (s)	1.57	1.51	1.59	1.53	1.61	1.57

Tabla 3.3. ECM por el método de la distancia inversa para la estación de Colmenar Viejo

En esta ocasión se observa que los datos tampoco se comportan de una forma consistente ya que el error cuadrático varía según la configuración que utilicemos. Principalmente se observa una gran diferencia entre el error cometido al utilizar un 70% de estaciones frente a utilizar un 85% en esta ocasión en todas las configuraciones. La estación de Colmenar Viejo también se encuentra situada en la periferia, pero no en una zona tan rural como la del caso anterior como se observa en la figura 1.4 por lo que tiene un mayor número de estaciones a una distancia cercana.

Se observa para el caso en el que el exponente tiene un valor de p=1 y p=3 se obtiene una estimación mejor con un menor número de estaciones, esto indica que se está produciendo una sobreestimación debido a que alguna de las estaciones a pesar de estar próxima a la estación de colmenar viejo tiene unos resultados suficientemente distintos a esta, esto probablemente se deba a que se trate de 2 estaciones que estén próximas en términos de coordenadas pero que exista una diferencia notable de altura entre ambas o una barrera que separe los niveles de contaminación de una con la otra como una montaña o un bosque, la forma en la que estos factores influyen ha la hora de estimar los niveles de contaminación atmosférica es algo que se digno de investigarse en otro estudio.

Atendiendo a los resultados obtenidos utilizando un exponente de p=2, se observa que son los que menor error producen además de nuevo se observa cómo se reduce el error a medida que aumenta la cantidad de datos utilizada.

En cuanto a los tiempos de procesamiento se observa que estos son independientes del lugar donde se esté realizando la estimación, las pequeñas diferencias se deben a que en cada localización se han estimado los niveles de contaminantes distintos, ya que en cada lugar solo se han estimado los niveles de los contaminantes que mide la estación que se encuentra en esa localización puesto que con el resto de los contaminantes no se disponen de datos con los que comparar. Al haber

estimado contaminantes diferentes en cada lugar el número de datos utilizado en cada caso es distinto ya que existe un número de estaciones diferente que mide cada tipo de contaminante.

Aún así todas las estimaciones se han realizado en un tiempo de entre 1,6 y 2,1 segundos que es el tiempo tomado de media por iteración para estimar los datos de una contaminante hora a hora de todos los días entre los años 2012 y 2021 que son los años cuyos datos se han utilizado para el estudio.

En cuanto a las diferencias en tiempo de procesamiento dependiendo de la configuración seleccionada se observa que aumentan ligeramente los tiempos de procesamiento cuando se utilizan más estaciones para la predicción, lo cual es lógico ya que tiene más datos que procesar, sin embargo, no aumentan de forma lineal ya que cuando se utiliza un 85% de las estaciones frente al 70% se estarían utilizando un 21% de datos adicionales pero el tiempo de procesamiento sólo aumenta un 4% de media lo cual es positivo ya que implica que aumentar el número de estaciones poco efecto en el coste computacional. En cuanto a las diferencias según el exponente utilizado se observa que a medida que crece el exponente aumenta ligeramente el tiempo de procesamiento ya que a mayor exponente mayor es la complejidad del cálculo a realizar.

Resultados por el método de kriging

Los resultados obtenidos utilizando el método de la distancia kriging se pueden ver en la tabla 3.4 en la que se muestran los estimadores obtenidos para las tres estaciones utilizando una estimación utilizando un semivariograma esférico.

Contaminantes	Plaza de España		Guadalix de la Sierra		Colmenar Viejo	
Dióxido de azufre	4.91	5.08				
Dióxido de nitrógeno	25.76	25.52	10.55	3.95	10.56	3.74
Monóxido de carbono	0.22	0.24			0.21	0.08
Monóxido de nitrógeno	49.43	54.39	7.45	2.31	5.32	1.98
Óxidos de nitrógeno	84.21	88.22	18.94	7.10	12.72	4.72
Ozono			3.41	3.96	1.17	2.76
Partículas en suspensión < PM10			5.07	2.17	4.99	1.79
porcentaje (%)	85	70	85	70	85	70
tiempo (s)	12.11	10.66	12.32	11.76	13.43	12.96

Tabla 3.4. ECM por el método de kriging para cada una de las estaciones

Se observa que al igual que con el método de la distancia inversa los resultados son mas consistentes para la estación de Plaza de España que para las otras estaciones, ya que tanto para la estación de Guadalix de la Sierra como la situada en Colmenar Viejo el error cuadrático medio se menor cuantas menos estaciones de medición se seleccionen esto apunta a que se está produciendo “overfitting” ya que se empeoran los resultados a medida que se añaden más datos

por lo que habría que estudiar si se podría mejorar este método imponiendo algún tipo de restricción a la hora de seleccionar cuantas estaciones se seleccionan a la hora de crear el modelo.

Esto hace que, pese a ser un modelo más complejo que el de la distancia inversa los resultados en lugares con alta densidad de estaciones sean igual de fiables, pero en zonas con densidad baja los errores sean mayores además de poco consistentes según el número de datos seleccionados.

En cuanto al tiempo de procesamiento se observa que es mayor cuanto mayor sea el número de estaciones de medición utilizados para modelar la contaminación. Siendo este tiempo de ente 10,6 y 13,4 segundos para estimar los datos de una contaminante hora a hora de todos los días entre los años 2012 y 2021 que son los años cuyos datos se han utilizado para el estudio.

Se han modelado las estimaciones producidas para los niveles de contaminación de óxidos de nitrógeno para el día 1 de enero del año 2012 utilizando cada el modelo de estimación por la distancia inversa (figura 4.1) con un exponente de p=2, ya que este es el que más consistencia y fiabilidad ha demostrado y para el modelo de kriging (figura 4.2).

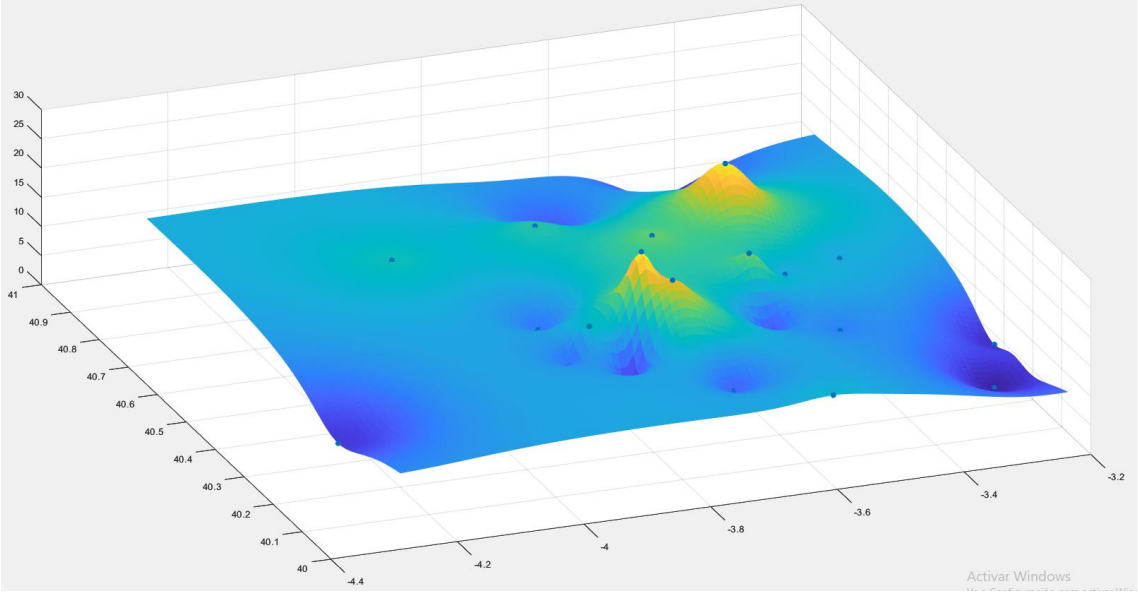

Figura 4.1. Mapa de Calor de los niveles de NO_x creado por el modelo de la distancia inversa

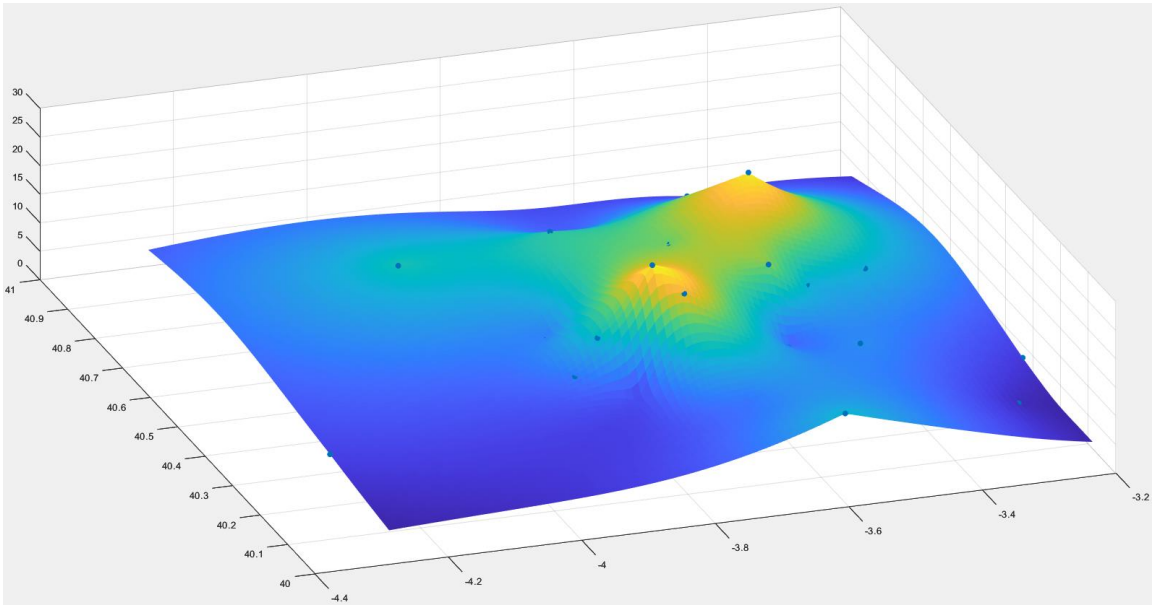

Figura 4.2. Mapa de Calor de los niveles de NO_x creado por el modelo de krigging

Los ejes x e y representan las coordenadas de la región de Madrid, el eje z representa la estimación del nivel de NO_x medio para ese día y los puntos indican la situación de las estaciones que miden este contaminante. Se puede observar claramente como en la parte central, que es donde se encuentra la ciudad de Madrid es donde se concentran el mayor número de estaciones de medición, así como los niveles más altos de contaminante.

Capítulo IV: Conclusiones

El presente trabajo analizó la posibilidad de crear modelos capaces de estimar los niveles de contaminación de distintos tipos de contaminantes a nivel de código postal a partir de datos espaciales proporcionados por la red de calidad del aire de la Comunidad de Madrid.

Uno de los principales hallazgos es la diferencia de comportamientos que presenta el modelo de la distancia inversa según los parámetros utilizados para realizar la estimación llegando a la conclusión de que la mejor opción es utilizar un exponente con un valor de p=2, ya que este exponente garantiza una mayor consistencia en las predicciones, tanto a la hora de realizarlas en zonas poco rurales como en zonas densamente pobladas, puesto que presenta pocas variaciones según el número de estaciones de medición utilizadas para realizar la estimación.

Al contrario que cuando se utilizan exponentes superiores como p=1 que al dar menos importancia a las estaciones cercanas depende de que el número de estas no sea muy inferior al número de estaciones a una mayor distancia ya que de lo contrario el peso de las lejanas al ser superior en número acaba llevando a una estimación incorrecta. En cuanto a la utilización de un exponente más elevado, como p=3, nos encontramos con el problema de sobreestimación, lo cual implica que si se da el caso de que existe una estación próxima con un valor anómalo, ya sea debido a que se encuentra a una diferencia de altitud con el resto o porque se encuentre en una zona singular en cuanto a niveles de contaminación como un parque natural, o simplemente porque tenga un fallo a la hora de realizar las mediciones, ese valor tomará una elevada importancia a la hora de calcular la estimación siendo más difícil que sea compensado por el resto de mediciones por lo que con un solo valor singular crea unos error de predicción muy elevados lo cual es un riesgo que se quiere evitar.

En cuanto a la estimación a través del modelo de kriging se ha encontrado que a pesar de ser más complejo de implementar y de calcular, ya que requiere de aproximadamente 8 veces más de tiempo de computación, no llega a ofrecer unos resultados más precisos que los obtenidos por el método de la distancia inversa, tanto en las zonas con más datos donde sus estimaciones son igual de fiables, como en las zonas donde se dispone de menos estaciones de medición, ya que aquí presenta una gran variabilidad según las estaciones cercanas al punto estimado puesto que este modelo tiende a sobreestimar cuando se cuentan con pocos datos.

Por lo tanto, se ha obtenido que es posible realizar estimaciones de forma fiable a nivel de código postal para distintos tipos de contaminante y que la mejor manera para ello es a través del método de la distancia inversa ya que ofrece los datos más fiables y consistentes y con una menor carga computacional, además existe la posibilidad de realizar estimaciones en tiempo real ya que el tiempo de procesamiento es reducido (1,2 segundos para estimar todas las horas durante 10 años) por lo que sería viable su implementación en los distintos portales de información de cara al

público existentes actualmente en la Comunidad de Madrid para permitir que se pueda obtener información de forma personalizada.

Por último, como futuras líneas de investigación se podría estudiar que impacto tiene el terreno a en la precisión de las estimaciones realizadas ya que se ha en este estudio se ha tomado la Comunidad de Madrid como si fuera un terreno homogéneo mientras que existen barreras naturales que impiden la distribución uniforme de la contaminación atmosférica. Otra posible línea de investigación se podría estudiar las diferencias de altitud que existen entre las localizaciones de las distintas estaciones de medición de la red de calidad del aire de la comunidad de Madrid y como impactan estas en las medidas obtenidas de los diferentes tipos de contaminantes puesto que cada uno poseerá una densidad diferente lo que podría alterar su distribución.

Bibliografía

AEMET, 2022. Figura 1 [Mapa de niveles de partículas en suspensión 15 de marzo de 2022]. Disponible en: www.aemet.es/prediccion/calidad_del_aire/indice_diario_calidad_del_aire

OMS-Organización Mundial de la Salud. (22 de septiembre del 2021). Ambient (outdoor) air pollution. Disponible en: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

AEMA. Air quality in Europe-2018 report. European Environmental Agency. Report No 12/2018, ISBN: 978-92-9213-990-2, Luxemburg: Publications Office of the European Union. 2018. Disponible en: https://www.eea.europa.eu/publications/air-quality-in-europe-2018.

Ballester, F. (2005). Contaminación atmosférica, cambio climático y salud. Revista Española de salud pública, 79, 159-175.

Querol, X. (2008). Calidad del aire, partículas en suspensión y metales. Revista Española de Salud Pública, 82, 447-454.

European Environment Agency. (2017). Air quality in Europe - 2017 report. Recuperado de https://www.eea.europa.eu/publications/air-quality-in-europe-2017

Ecologistas en Acción. (2022). Informe de calidad del aire Madrid 2021. Recuperado de https://www.ecologistasenaccion.org/wp-content/uploads/2022/01/informe-calidad-aire_madrid-2021.pdf

Díaz Jiménez, J., y Linares Gil, C. (2018). Impacto de la contaminación atmosférica sobre la mortalidad diaria a corto plazo en España.

Muñoz, A. M., Paz, J. J., & Quiroz, C. M. (2007). Efectos de la contaminación atmosférica sobre la salud de adultos que laboran en diferentes niveles de exposición. Revista Facultad Nacional de Salud Pública, 25(2), 85-94.

CAM, Consejería de medio ambiente, ordenación del territorio y sostenibilidad, 2023. Disponible en: http://gestiona.madrid.org/azul_internet/html/web/3.htm?ESTADO_MENU=3

Gotway C.A., Ferguson R.B., Hergert G.W., Peterson, T.A. 1996. Comparison of Kriging and Inverse Distance Methods for mapping soil parameters. Soil Science Society of American Journal 60:1237-1247.

Viera, M. A. D., & González, R. C. (2002). Geoestadística aplicada. Instituto de Geofísica, UNAM-Instituto de Geofísica y Astronomía, CITMA. Cuba.

Foody, G. M., McCulloch, M. B., & Yates, W. B. (1995). The effect of training set size and composition on artificial neural network classification. International Journal of Remote Sensing, 16(9), 1707-1723.

DE, C. P. D. (2012). Cambios espaciales y temporales en la contaminación por dióxido de nitrógeno en el municipio de Madrid (2001-2011). DJJCM, 29, 127.

Madrid Salud. (s.f.). En tiempo real: datos de calidad del aire. Recuperado de https://airedemadrid.madrid.es/portales/calidadaire/es/Bases-de-datos-y-publicaciones/Bases-de-datos-de-calidad-del-aire/En-tiempo-real/?vgnextfmt=default&vgnextchannel=650a89e859517710VgnVCM1000001d4a900aRCRD

jBuist (2023). kriging(x,y,z,range,sill) MATLAB Central File Exchange. Recuperado de (https://www.mathworks.com/matlabcentral/fileexchange/57133-kriging-x-y-z-range-sill)

Simone Fatichi (2023). Inverse Distance Weight , MATLAB Central File Exchange. Recuperado de (https://www.mathworks.com/matlabcentral/fileexchange/24477-inverse-distance-weight.

Anexos

I Código Principal

TFG.m

```
%Nombre Partículas
particulas = ["Benceno" "Dióxido de azufre" "Dióxido de nitrógeno"
"Hidrocarburos no metánicos" "Hidrocarburos totales" "Monóxido de carbono"
"Monóxido de nitrógeno" "Óxidos de nitrógeno" "Ozono" "Partículas en
suspensión < PM10" "Partículas en suspensión < PM2,5" "Tolueno"];
%particulas = ["Ozono"];
particulasn = erase(regexprep(particulas,'(ó)','o')," ");
particulasn = regexprep(particulasn,'Ó','O');
particulasn = regexprep(particulasn,'á','a');
particulasn = regexprep(particulasn,'í','i');
particulasn = erase(particulasn, "<");
particulasn = erase(particulasn, ",");

%Nombre Estaciones
estaciones = ["ALCALA_DE_HENARES_OK" "ALCOBENDAS_OK" "ALCORCON_OK"
"ALGETE_OK" "ARANJUEZ_OK" "ARGANDA_DEL_REY_OK" "ARTURO_SORIA_OK"
"AVDA__RAMON_Y_CAJAL_OK" "BARAJAS_PUEBLO_OK" "BARRIO_DEL_PILAR_OK"
"CASA_DE_CAMPO_OK" "CASTELLANA_OK" "COLLADO_VILLALBA_OK" "COLMENAR_VIEJO_OK"
"COSLADA_OK" "CUATRO_CAMINOS_OK" "EL_ATAZAR_OK" "EL_PARDO_OK"
"ENSANCHE_DE_VALLECAS_OK" "ESCUELAS_AGUIRRE_OK" "FAROLILLO_OK"
"FUENLABRADA_OK" "GETAFE_OK" "GUADALIX_DE_LA_SIERRA_OK" "JUAN_CARLOS_I_OK"
"LEGANES_OK" "MAJADAHONDA_OK" "MENDEZ_ALVARO_OK" "MORATALAZ_OK" "MOSTOLES_OK"
"ORUSCO_DE_TAJUNA_OK" "PARQUE_DEL_RETIRO_OK" "PLAZA_CASTILLA_OK"
"PZA__DE_ESPANA_OK" "PZA__DEL_CARMEN_OK" "PZA__ELIPTICA_OK"
"RIVAS_VACIAMADRID_OK" "SANCHINARRO_OK" "TORREJON_DE_ARDOZ_OK"
"TRES_OLIVOS_OK" "URB__EMBAJADA_OK" "VALDEMORO_OK" "VALLECAS_OK"
"VILLA_DEL_PRADO_OK" "VILLAREJO_DE_SALVANES_OK" "VILLAVERDE_OK"];

%Años
yearinicio = 2012;
yearfinal = 2021;

%Leer datos
tic
contaminantes = leerdatos(particulas,particulasn,estaciones);
tiempo_leer_datos = toc;

%Definir p
potencias = [1];
porcentajes = [100];
n = 5;
n_estaciones_test = 3;

%Crossvalidation
n_potencias = length(potencias);
n_porcentajes = length(porcentajes);
estaciones_test = randsample(estaciones, n_estaciones_test);
for i=1:n_estaciones_test
    estacion_test = estaciones_test(i);
    %HallarMedidas
    for x=1:length(particulas)
        datos = contaminantes.(sprintf('%s',particulasn(x))).datos;
        datos = removevars(datos,{'MAGNITUD' 'DESCRIPCIONMAGNITUD'
'CODIGO_NACIONAL' 'LATITUD' 'LONGITUD'});
```

```
medidas.(sprintf('%s',erase(estacion_test,"_OK"))).(sprintf('%s',particulasn(
x))) = datos(any(datos.NOMBRE_ESTACION == erase(estacion_test,"_OK"),2),:);
    end
    particulas_t = transpose([particulas, 'porcentaje (%)', 'potencia',
'tiempo (s)']);
    for k=1:n_potencias
        p = potencias(k);
        for j=1:n_porcentajes
            por_test = porcentajes(j);
            %HallarMediaEstimaciones
            tic
                media = Crossvalidation(contaminantes, medidas, estaciones,
estacion_test, particulas, particulasn, yearinicio, yearfinal, p, por_test,
n);
            tiempo_medias = toc;

            %HallarError
            error = [];
            estacion_test = erase(estacion_test,"_OK");
            for x=1:length(particulas)
                A =
removevars(medidas.(sprintf('%s',erase(estacion_test,"_OK"))).(sprintf('%s',p
articulasn(x))), {'NOMBRE_ESTACION'});
                B = media.(sprintf('%s',erase(particulasn(x)," ")));
                B = array2table(B,"VariableNames",{'ANO','MES','DIA' 'H01'
'H02' 'H03' 'H04' 'H05' 'H06' 'H07' 'H08' 'H09' 'H10' 'H11' 'H12' 'H13' 'H14'
'H15' 'H16' 'H17' 'H18' 'H19' 'H20' 'H21' 'H22' 'H23' 'H24'});

                B = array2table(B,"VariableNames",{'ANO','MES','DIA' 'H01'});
                error(x,1) = mean_squared_error(A,B);
            end
            particulas_t = [particulas_t, transpose([transpose(error),
por_test, p, tiempo_medias])];
        end
    end
    %Escribir
    directoriodatos = append("Resultados/", estacion_test,".xlsx");
    writematrix(particulas_t, directoriodatos, "FileType", "spreadsheet");
end
```

II Código para leer y procesar los datos

leerdatos.m

```
function contaminante =
leerdatos(particulas,particulasn,estaciones,estaciones_test)
    n_est = length(estaciones);
    n_par = length(particulas);

    for i=1:n_par
        contaminante.(sprintf('%s',erase(particulasn(i)," "))).datos = [];
    end

    for i=1:n_est
        directoriodatos = append("DCONTAM/",estaciones(i),".xlsx");
        datos =
rmmissing(readtable(sprintf('%s',directoriodatos),'Range','A:AG'));
        for j=1:n_par
```

```
            contaminante.(sprintf('%s',particulasn(j))).datos =
[contaminante.(sprintf('%s',particulasn(j))).datos;
datos(strcmp(datos.DESCRIPCIONMAGNITUD, particulas(j)),:)];

        end
    end

    if nargin >= 4
        n_est = length(estaciones_test);
        for i=1:n_est
            directoriodatos = append("DCONTAM/",estaciones_test(i),".xlsx");
            datos = readtable(sprintf('%s',directoriodatos),'Range','A:AG');
            for j=1:n_par

contaminante.(sprintf('%s',particulasn(j))).(sprintf('%s',estaciones_test(i))
) = removevars((datos(strcmp(datos.DESCRIPCIONMAGNITUD,
particulas(j)),:)),{'MAGNITUD' 'DESCRIPCIONMAGNITUD' 'CODIGO_NACIONAL'
'NOMBRE_ESTACION' 'LATITUD' 'LONGITUD'});
            end
        end

    end
end
```

III Código utilizado para realizar el "crossvalidation"

Crossvalidation.m

```
function media = Crossvalidation(contaminantes, medidas, estaciones,
estacion_test, particulas, particulasn, yearinicio, yearfinal, p, por_test,
n)

    %Asignar Training y Test
    n_est = length(estaciones);
    estaciones = setdiff(estaciones, estacion_test);
    n_estaciones_training = round(n_est*por_test/100);
    estacion_test = erase(estacion_test,"_OK");

    %Puntos
    coordinates = readcoordinates(contaminantes, estacion_test, particulasn);
    puntox = coordinates(1,1);
    puntoy = coordinates(1,2);
    punto = [puntox,puntoy];

    %Interpolar
    for k=1:n
        estaciones_training = randsample(estaciones, n_estaciones_training);

        for i=1:length(particulas)
            if
size(medidas.(sprintf('%s',estacion_test)).(sprintf('%s',particulasn(i))), 1)
> 0

contaminantes.(sprintf('%s',particulasn(i))).(sprintf('resultados_%s_%s',esta
cion_test,string(k))) =
interpolar(contaminantes.(sprintf('%s',particulasn(i))).datos, yearinicio,
yearfinal, punto, estaciones_training, p);
            end
        end
```

```
    end

    %Media
    dias = (yearfinal - yearinicio + 1);
    dias = dias*365 + round(dias/4);
    for i=1:length(particulas)
        media.(sprintf('%s',erase(particulasn(i)," "))) = zeros([dias 27]);
    end
    for k=1:n
        for i=1:length(particulas)
            if
size(medidas.(sprintf('%s',estacion_test)).(sprintf('%s',particulasn(i))), 1)
> 0
                media.(sprintf('%s',erase(particulasn(i)," "))) =
media.(sprintf('%s',erase(particulasn(i)," "))) +
contaminantes.(sprintf('%s',particulasn(i))).(sprintf('resultados_%s_%s',esta
cion_test(1),string(k)));
            end
        end
    end
    for i=1:length(particulas)
        media.(sprintf('%s',erase(particulasn(i)," "))) =
media.(sprintf('%s',erase(particulasn(i)," ")))/n;
    end
end
```

mean_squared_error.m

```
function error = mean_squared_error(A,B)
    %A = table2array(A);
    A.ANO(A.ANO < 2000) = A.ANO(A.ANO < 2000) + 2000;
    fechas_tabla1 = A(:, 1:3);
    fechas_tabla2 = B(:, 1:3);
    fechas_comunes = intersect(fechas_tabla1, fechas_tabla2, 'rows');
    A = A(ismember(fechas_tabla1, fechas_comunes), :);
    B = B(ismember(fechas_tabla2, fechas_comunes), :);
    A = A(:, 4:27);
    B = B(:, 4:27);
    a = table2array(A);
    b = table2array(B);
    a = mean(a,2);
    b = mean(b,2);
    error=mse(a,b);
end
```

IV Código para realizar la interpolación día a día

interpolar.m

```
function resultados =
interpolar(datos,yearinicio,yearfinal,punto,estaciones_training, p)
    if nargin >=5
        estaciones_training = erase(estaciones_training,"_OK");
        datos = datos(any(datos.NOMBRE_ESTACION == estaciones_training,2),:);
    end
    datos = removevars(datos,{'MAGNITUD' 'DESCRIPCIONMAGNITUD'
'CODIGO_NACIONAL' 'NOMBRE_ESTACION'});
    dias = (yearfinal - yearinicio + 1);
    dias = dias*365 + round(dias/4);
```

```
    resultados = zeros([dias 27]);
    l = 1;
    for i=yearinicio:yearfinal
        datosyear = datos(or(datos.ANO == i, datos.ANO == (i-2000)),:);
        for j=1:12
            datosmes = datosyear((datosyear.MES == j),:);
            for k=1:31
                datosdia = datosmes((datosmes.DIA == k),:);
                datosdia = datosdia{:,:};
                datosdia = datosdia(all(~isnan(datosdia),2),:);
                resultadosdia = [i j k interpolardia(datosdia,punto, p)];
                resultadosdia =
resultadosdia(all(~isnan(resultadosdia),2),:);
                if size(resultadosdia,1) > 0
                    resultados(l,:) = resultadosdia;
                    l = l + 1;
                end
            end
        end
    end
end
```

interpolardia.m

```
function resultadosdia = interpolardia(datosdia,punto, p)

    X1 = datosdia(:,1); %Longitudes
    X2 = datosdia(:,2); %Latitudes
    F =  datosdia(:,6:end); %Contaminación
    F = mean(F,2);

    if isempty(F)>0
        resultadosdia = NaN([1 24]);
    else
        %resultadosdia = idw([X1,X2], F, punto, p);
        resultadosdia = kriging_crossval([X1,X2], F, punto);
    end

end
```

V Código para realizar la interpolación por el método de la distancia inversa

Idw.m

```
function Fint = idw(X0,F0,Xint,p,rad,L)
    if nargin < 6
        L = 2;
        if nargin < 5
            rad = inf; %Radio
            if nargin < 4
                p = 2; %Normalización
            end
        end
    end

    N = size(X0,1); %Numero inputs
    M = size(X0,2); %Dimensiones
    Q = size(Xint,1); %Numero outputs
```

```
    Fint = zeros(Q,size(F0,2));
    for i = 1:Q
        DeltaX = X0 - repmat(Xint(i,:),N,1);
        DabsL = zeros(size(DeltaX,1),1);
            for j = 1:M
                DabsL = DabsL + abs(DeltaX(:,j)).^L;
            end
        Dmat = DabsL.^(1/L);
        Dmat(Dmat==0) = eps;
        Dmat(Dmat>rad) = inf;

        W = 1./(Dmat.^p);
        Fint(i,:) = sum(W.*F0)/sum(W);
    end
end
```

VI Código para realizar la Interpolación por el método de kriging

riging.m

```
function elevations = kriging(X0,F0, Xint,range,sill)

%input
x = X0(:,1);
y = X0(:,2);
elevations = zeros(1,size(F0,2));
if nargin < 2
    error('Error. Input at least PointsX, PointsY and PointsElev')
end
if ~exist('range','var')
    range = 26440.092;
end
if ~exist('sill','var')
    sill = 62583.893;
end
%Calculating trend
xx = x(:);
yy = y(:);
PointsX = x;
PointsY = y;
N = length(xx);
O = ones(N,1);
L = length(PointsX);
for k = 1:size(F0,2)
    z = F0(:,k);
    zz = z(:);
    C = [xx yy O]\zz;
    PointsElev = z - (C(1).*x + C(2).*y +C(3));
    S = size(PointsElev);
    if S(1) > S(2)
        PointsElev = PointsElev.';
    end
    %Kriging

    %Construir matrix de redundancia
    ReduMatrix = zeros(L,L);
    for i = 1:L
        for j = 1:L
```

```
            Distance = sqrt((PointsX(i) - PointsX(j))^2 + (PointsY(i) -
PointsY(j))^2);
            if Distance > range
                ReduMatrix(i, j) = sill;
            else
                ReduMatrix(i, j) = sill*((3*Distance./(2*range)) -
1/2*(Distance ./range).^3);
            end
        end
    end

    ProxVector = zeros(L, 1);
    gridXcor = Xint(1);
    gridYcor = Xint(2);
    for a = 1:L
        Distance = sqrt((PointsX(a) - gridXcor)^2 + (PointsY(a) -
gridYcor)^2);
       if Distance > range
        ProxVector(a) = sill;
       else
        ProxVector(a) = sill*((3*Distance./(2*range))-1/2*(Distance
./range).^3);
       end
    end

    Weights = ReduMatrix \ ProxVector;
    XYElev = PointsElev * Weights;
    XYElev = XYElev + ((C(1) * gridXcor) + (C(2) * gridYcor) + C(3) );
    elevations(1,k) = XYElev;
end
end
```