UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA
CAMPUS D'ALCOI

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Politécnica Superior de Alcoy

Utilización de Herramientas Estadísticas para el Análisis del Tráfico y la Contaminación Ambiental en Valencia

Trabajo Fin de Grado

Grado en Ingeniería Informática

AUTOR/A: Guillem Amat, Fernando Manuel

Tutor/a: Pérez Bernabeu, Elena

Cotutor/a: Camacho García, Andrés

CURSO ACADÉMICO: 2023/2024

Resumen:

El Trabajo Final de Grado analiza la evolución de la contaminación y el tráfico en Valencia, con énfasis en el NOx y el tráfico en la Plaza del Ayuntamiento. Destaca el impacto de la pandemia de COVID-19 en 2020, que redujo la actividad vehicular y los niveles de contaminación. Se investiga el efecto de medidas urbanas, como la peatonalización, en la calidad del aire. El objetivo es informar políticas para mejorarla. Se emplean modelos de aprendizaje automático para predecir la evolución del NOx, evaluando su impacto.

Palabras clave

Contaminación, Tráfico, Evaluación de datos, Análisis estadístico, Peatonalización, NOx, Regresión, Forecasting, Machine Learning, LSTM, ARIMA, SARIMA

Summary:

The Final Degree Project analyzes the evolution of pollution and traffic in Valencia, with emphasis on NOx and traffic in the Plaza del Ayuntamiento. It highlights the impact of the COVID-19 pandemic in 2020, which reduced vehicular activity and pollution levels. The effect of urban measures, such as pedestrianization, on air quality is investigated. The objective is to inform policies for improvement. Machine learning models are used to predict the evolution of NOx, evaluating its impact.

Keywords:

Pollution, Traffic, Data Evaluation, Statistical Analysis, Pedestrianization, NOx, Regression, Forecasting, Machine Learning, LSTM, ARIMA, SARIMA

Resum:

Este Treball Final de Grau analitza l'evolució de la contaminació i el trànsit a València, amb èmfasi en el NOx i el trànsit a la Plaça de l'Ajuntament entre 2018 i 2023. Destaca el canvi provocat per la pandèmia de COVID-19 el 2020, amb una reducció notable del trànsit i la contaminació. S'investiguen els efectes de mesures urbanes com la pacificació de carrers. L'estudi busca oferir una visió completa de la contaminació a València i analitzar diversos models per predir l'evolució del NOx sense la influència de la pandèmia ni les mesures urbanes.

Paraules clau:

Contaminació, Trànsit, Avaluació de dades, Anàlisi estadística, Vianants, NOx, Regressió, Pronòstic, Aprenentatge automàtic, LSTM, ARIMA, SARIMA

Índice general

Dedicatoria

Me gustaría mucho poder dedicar este trabajo final a mis familiares más cercanos: a mis padres, que tanto me ayudaron cuando era necesario, y a mi hermana, que siempre sabe sacar tiempo para estar con la familia sin dejar por ello nunca de ganarse su camino. Tampoco quisiera olvidar nunca a muchas amistades, tanto recientes, en el periodo universitario y en el campo del ciclismo, como anteriores de toda la vida, por quienes he mantenido mi confianza en momentos más complicados.

Agradecimientos

Aunque nunca podré listar a tantos como quisiera, me gustaría al menos agradecer la oportunidad de presentar este trabajo a muchas personas, tanto familiares y amistades de toda la vida, como otros tantos profesores a lo largo de toda mi formación.

Se podría llegar a obviar la influencia de la educación primaria y secundaria pero, en mi opinión, ese tiempo es crucial, no sólo en el contexto académico, sino para marcarnos una base firme de educación y ética desde la que proyectar nuestros objetivos de futuro.

En una siguiente etapa, ya en el bachiller científico, encontré algunas dificultades y pérdidas de interés por algunas materias impartidas, pero procurando cumplir con las que me parecían menos alejadas de mi perspectiva de futuro en el campo de la ingeniería. Siempre agradeceré a Delfino Arnedo Valero por sus cursos de Física y Química, que me recordaron mi vocación recibiendo una formación con una aplicación más tangible.

De mi periodo de formación profesional en el campo eléctrico y automático, me gustaría mencionar principalmente a Camilo, mi profesor de Electrotecnia, que a sus setenta y tantos años seguía tan alegre y agradable como un niño y era capaz de despertar el interés por sus explicaciones. La suya fue una asignatura en la que encontré como todos grandes dificultades, pero que me permitió aprender a buscar mis propios caminos para resolver problemas a mi manera. Aprendí a formatear mi aprendizaje para que éste fuera aplicable aún con un mapa memoria incapaz de memorizar como un loro 80 fórmulas sin otra aplicación futura que vomitarlas en su examen sin aprender su motivo, valiéndome de la trigonometría y de su teorema del seno para resolver cualquiera de sus problemáticas con corrientes alternas trifásicas inductivas o capacitivas. También gustaría mencionar a Alfonso Llorens, profesor de asignaturas como Sistemas Eléctricos de Medida y Regulación, asignatura con la que aprendí a razonar matemáticamente la progresión coherente de los sistemas hacia un fin definido y su tanta extrapolación a cualquier otro ámbito en la vida, donde todo tiene sus antecedentes y sus consecuencias.

Respecto a mi periodo universitario, aunque no pueda referirme a tantos como quisiera, gustaría mencionar a algunos de los profesores que he tenido el gusto de cursar como por ejemplo: Romina del Rey Tormos, mi profesora de Fundamentos Físicos de la Informática en primer curso del grado, de quien aprendí lo tanto como me faltaba entender para razonar objetivamente cosas que podía llegar a creer conocer. Ignacio Miró Orozco, mi profesor de Tecnología de Computadores, cuya asignatura disfruté mucho, no sólo por ser un campo cercano a mis conocimientos eléctricos, sino por sus explicaciones y su formalismo con la puntualidad y la educación. Jordi Joan Llinares, mi profesor de asignaturas como Introducción a la Programa-

ción, donde pude entender lo que tan poco sabía en verdad con mis conocimientos de C, y de otras asignaturas posteriores más aplicadas como fueron Machine-Learning o Introducción a la Programación de Videojuegos, donde pude empezar a imaginar el casi infinito alcance de la informática en campos tan diversos. Elena Pérez, mi profesora de Estadística con quien aprendí a valorar la influencia de cualquier cosa sobre su todo, quien me propuso como candidato para la beca Alcoy Smart-City y una posterior colaboración con el equipo de Ángel Juan y por último también fue tutora de mi trabajo final de grado, ayudándome a aplicar en él muchos de los aprendizajes de estos años. Javier Esparza, profesor de asignaturas como Estructuras de Datos y Algoritmos, donde fue capaz de explicar muchos temas algorítmicos, reduciéndolos a explicaciones tanto más entendibles por nosotros. Adolfo Ferre, profesor de asignaturas como Introducción a los Sistemas Operativos y Computación Paralela, en las que con sus exámenes se nos exigió entender por completo los temarios desde su base, y no sólo desde una perspectiva de preguntas tipo en concreto. Rubén Pérez, profesor de asignaturas aplicadas como Programación de Soluciones Informáticas para Dispositivos Móviles y Visión por Computador, enseñándonos a aplicar la informática a casos tan cercanos a nosotros como una aplicación gráfica o razonar la propia visión en el transcurso del tiempo. Andrés Camacho, junto con Elena Pérez, me dieron la posibilidad de contribuir en 2023 en la cátedra Alcoy Smart-City y también me asesoró en el desarrollo de mi Trabajo Final de Grado. Ángel Juan que me permitió junto con Elena contribuir en una versión del algoritmo con el que gestionar el progreso más eficiente de trabajos. También querría mencionar a compañeros de estudios y amistades de mi periodo universitario como pudieron ser Daniel Gaspar, Leandro Albero, Adrián Sánchez, Natxo, Sara Sánchez, José Cantó y Raúl Mira, en mi primer año de universidad, Guillem Carrión, Carlos Muñóz, Isabel Ferri, Andrea Jordá, María Torres, Sergio Barrachina, Kevin Mora, Nicolás Camarena o Ángel Lleixá, en mi segundo y otros tantos como Miguel Lopez, Adelardo, Carlos Aliaga y David García entre otros tantos, en años posteriores. También quisiera mencionar otros compañeros que me ayudaron con mis últimos trabajos de la universidad, como fueron Xabier Martín, que me ayudó con el algoritmo de gestión de proyectos y Raquel Soriano, quien me ha ayudado con algunos puntos del desarrollo de mi trabajo final.

1. Motivación

Durante una beca de colaboración realizada en el departamento de Estadística e Investigación Operativa Aplicadas y Calidad de la Escuela Politécnica Superior de Alcoy en 2023 se analizaron los índices de contaminación ambiental de Alcoy con resultados interesantes. Se consideró que sería muy útil extender el estudio también a otras ciudades de diferentes características, eligiendo la ciudad de Valencia, que incluye nuevos escenarios de análisis, como una mayor área, una mayor población y la existencia de otras zonas, tanto industrial y residencial como portuaria. Este análisis en profundidad de la situación a nivel de contaminantes en la ciudad de Valencia se desarrolla como Trabajo Fin de Grado del Grado en Ingeniería Informática de la Universidad Politécnica de Valencia.

2. Objetivos

Este trabajo tiene como objetivo principal evaluar cómo el tráfico rodado afecta al impacto de los contaminantes ambientales en la ciudad de Valencia. Se centrará en el análisis del índice de NOx en la Plaza del Ayuntamiento durante los últimos años, empleando técnicas de Análisis de Datos y Machine Learning (Aprendizaje Automático) mediante el lenguaje de programación Python, ampliamente utilizado en el ámbito de la Ciencia de Datos. Por otro lado, también se evaluará el beneficio ambiental resultado de medidas como la peatonalización de la Plaza del Ayuntamiento de Valencia y otras iniciativas destinadas a reducir el impacto del tráfico en la calidad del aire en la ciudad.

Se evalúa el cumplimiento y/o la intención de cumplir con los Objetivos de Desarrollo Sostenible (ODS) por parte de la ciudad de Valencia. Se analiza cómo las políticas locales y las acciones implementadas por las autoridades municipales están alineadas con los ODS, especialmente aquellos relacionados con la salud y el bienestar (ODS 3), ciudades sostenibles (ODS 11), acción por el clima (ODS 13) y la vida de Ecosistemas Terrestres (ODS 15).

Se estudia también la posible implicación de la contaminación de la combustión del tráfico marítimo en la calidad del aire de Valencia, como indican en el artículo [Waqas et al., 2024].

Es fundamental considerar el impacto de las medidas socioeconómicas en la calidad del aire y cómo el aprendizaje automático puede contribuir a la predicción y mitigación de la contaminación atmosférica. El óxido de nitrógeno (NOx) es un contaminante atmosférico comúnmente asociado con la quema de combustibles fósiles, especialmente en motores de vehículos y en procesos industriales.

La presencia de NOx en la atmósfera puede tener varios impactos negativos en la salud humana y en el medio ambiente. Por ejemplo, puede contribuir a la formación de SMOG y ozono troposférico, así como a la lluvia ácida. Además, puede irritar los pulmones, aumentar el riesgo de problemas respiratorios y afectar negativamente a las personas con enfermedades respiratorias crónicas como el asma.

Por lo tanto, el NOx se utiliza como un indicador de la calidad del aire y la contaminación atmosférica. Los niveles altos de NOx en el aire pueden indicar la presencia de contaminación atmosférica y pueden desencadenar acciones regulatorias para controlar y reducir las emisiones de fuentes contaminantes [Subrahmanyam et al., 2023].

Se establece un entorno de trabajo en Python 3.12.2, importando los paquetes necesarios. Estas bibliotecas son herramientas populares en el ámbito de la Ciencia de Datos y el Aprendizaje Automático en Python y se utilizarán para procesar los datos y realizar análisis estadísticos y modelos de regresión.

A continuación se enumeran los paquetes usados así como una breve descripción:

- **os**: Proporciona funciones para interactuar con el sistema operativo, como acceder a archivos y directorios.
- **pandas**: Es una biblioteca de análisis de datos que proporciona estructuras de datos flexibles y herramientas para manipular y analizar conjuntos de datos.
- **seaborn**: Es una biblioteca de visualización de datos basada en matplotlib que proporciona una interfaz de alto nivel para crear gráficos estadísticos atractivos e informativos.
- **matplotlib**: Es una biblioteca de visualización de datos en 2D que produce figuras en una variedad de formatos y en diferentes entornos interactivos o no interactivos.
- **statsmodels**: Es una biblioteca que proporciona clases y funciones para la estimación de muchos modelos estadísticos diferentes, así como para realizar pruebas estadísticas y exploración de datos.
- **tensorflow.keras.**: Es una API de alto nivel para la construcción y entrenamiento de modelos de aprendizaje profundo. Integrada en TensorFlow, combina la versatilidad de Keras con la potencia de TensorFlow, ofreciendo una sintaxis intuitiva y una gran escalabilidad para el diseño, entrenamiento y despliegue de modelos de aprendizaje profundo. Esta integración facilita la experimentación y la producción de modelos en una variedad de escenarios de aplicación, proporcionando a los desarrolladores una herramienta poderosa y flexible.
- **itertools**: Es una biblioteca de Python que proporciona funciones para crear iteradores para bucles eficientes.
- **numpy**: Es una biblioteca fundamental para la computación científica en Python que proporciona un objeto de matriz multidimensional de alto rendimiento y herramientas para trabajar con estas matrices.

3. Impacto esperado

Los resultados de este estudio permitirán evaluar objetivamente la utilidad de políticas medioambientales aplicadas recientemente en Valencia para procurar cumplir con las normativas de los ODS relacionados, como la restricción de carriles abiertos al tráfico o la peatonalización de la Plaza del Ayuntamiento.

Se espera encontrar una gran influencia del tráfico portuario en la evolución de los contaminantes del estudio, gracias a la correlación de los periodos con vientos de origen costero.

Igualmente, sería deseable encontrar evidencias que demuestre objetivamente el valor de medidas ecológicas como las estudiadas, y que estas puedan servir como herramientas aplicables a nuevos escenarios, ya sea en Valencia u otras ciudades, promoviendo una mayor sostenibilidad de la Tierra.

4. Adaptación del trabajo a los ODS

Figura 1: Objetivos de Desarrollo Sostenible de las Naciones Unidas (ONU).

Los ODS son un conjunto de 17 objetivos interconectados adoptados por todos los Estados miembros de las Naciones Unidas en 2015 como parte de la Agenda 2030 para el Desarrollo Sostenible. Estos objetivos abordan los desafíos más apremiantes a los que se enfrenta el mundo, desde la erradicación de la pobreza hasta la protección del medio ambiente y el fomento de la paz y la justicia.

El propósito de los ODS es proporcionar un marco global para orientar las políticas y acciones hacia un futuro sostenible, donde se equilibren las necesidades económicas, sociales y

ambientales de las generaciones presentes y futuras. Los ODS reconocen la interconexión entre los diferentes aspectos del desarrollo humano y ambiental, y promueven enfoques integrados para abordar los desafíos globales.

Como se indica en [Lebrusán and Toutouh, 2021], es fundamental entender cómo las políticas locales, como Madrid Central, pueden contribuir a la mejora de la calidad del aire en las ciudades, lo cual está alineado con varios ODS, incluido el ODS 11 sobre ciudades sostenibles y el ODS 3 sobre salud y bienestar. Estudios como este resaltan la importancia de medidas específicas para abordar la contaminación atmosférica y sus efectos en la salud pública.

La aplicación de los ODS está ganando cada vez más relevancia a nivel internacional, con gobiernos, organizaciones internacionales, empresas y sociedad civil comprometidos en su implementación. El cumplimiento de los ODS es ampliamente reconocido como un indicador de progreso hacia un desarrollo sostenible y equitativo, y las acciones que contribuyen a alcanzar estos objetivos son valoradas y promovidas en todo el mundo.

En el contexto de este trabajo, se pueden identificar varios objetivos de desarrollo sostenible relevantes:

ODS 3: Salud y Bienestar: El análisis de la calidad del aire y su relación con la salud humana aborda directamente el ODS 3, que busca garantizar una vida saludable y promover el bienestar para todos en todas las edades. Al identificar los contaminantes atmosféricos y su impacto en la salud de la población, tu trabajo contribuye a concienciar sobre los riesgos ambientales y promover medidas para mejorar la calidad del aire y proteger la salud pública.

ODS 11: Ciudades y Comunidades Sostenibles: El estudio de la calidad del aire y el tráfico vehicular en áreas urbanas contribuye al ODS 11, que busca hacer que las ciudades y los asentamientos humanos sean inclusivos, seguros, resilientes y sostenibles. Al identificar los problemas ambientales en entornos urbanos y proponer soluciones para mejorar la calidad de vida en las ciudades, tu trabajo apoya los esfuerzos para construir comunidades más sostenibles y habitables.

ODS 13: Acción por el Clima: El análisis de la contaminación atmosférica y su relación con el tráfico vehicular también está relacionado con el ODS 13, que insta a tomar medidas urgentes para combatir el cambio climático y sus impactos. Al identificar las fuentes de emisiones contaminantes y evaluar su contribución al calentamiento global, tu trabajo proporciona información clave para desarrollar estrategias de mitigación y adaptación que ayuden a enfrentar el desafío del cambio climático.

ODS 15: Vida de Ecosistemas Terrestres: Aunque indirectamente, el análisis de la contaminación atmosférica también puede vincularse al ODS 15, que busca proteger, restaurar y promover un uso sostenible de los ecosistemas terrestres. Al evaluar los impactos ambientales de la actividad humana, incluido el tráfico vehicular, tu trabajo destaca la importancia de preservar la biodiversidad y los servicios de los ecosistemas para garantizar un medio ambiente saludable y sostenible para las generaciones futuras.

En resumen, este trabajo se alinea con los ODS al contribuir a la comprensión y aplicación de los desafíos ambientales y sociales, promoviendo acciones que conduzcan a un desarrollo sostenible en línea con la Agenda 2030 que aprobó la ONU en 2015 para alcanzar el Desarrollo Sostenible.

5. Resumen

5.1. Resumen

Este Trabajo Final de Grado se centra en analizar la evolución de la contaminación ambiental y el tráfico en Valencia, con énfasis en el contaminante NOx y el tráfico en la Plaza del Ayuntamiento, entre 2018 y 2023.

En 2020, se destaca un cambio significativo debido a la pandemia de COVID-19, que redujo drásticamente la actividad vehicular y los niveles de contaminación, ofreciendo una oportunidad para estudiar el impacto del tráfico en la calidad del aire. También se investiga el efecto de medidas urbanas, como la peatonalización de una calle importante cerca de la Plaza del Ayuntamiento, que resultó en una mejora en los niveles de contaminantes [Lebrusán and Toutouh, 2021].

Este trabajo busca proporcionar una visión integral de la contaminación en Valencia, explorando las relaciones entre los niveles de contaminantes y los factores que los afectan, con el objetivo de informar políticas efectivas para mejorar la calidad del aire en la ciudad.

Se analizan varios modelos de aprendizaje automático para predecir la evolución esperada del NOx en la Plaza del Ayuntamiento sin la influencia del COVID-19 y sin la peatonalización, evaluando el beneficio ambiental de estos eventos para predecir la evolución del NOx a partir de marzo de 2020 y compararlos con los datos reales influidos por el confinamiento y las medidas de peatonalización [Sanchez et al., 2021].

Como señalan en el artículo [Subrahmanyam et al., 2023], la calidad del aire es un aspecto crucial para la salud pública, y la predicción de la contaminación atmosférica es esencial para mitigar sus efectos.

5.2. Summary

This Final Degree Project focuses on analyzing the evolution of environmental pollution and traffic in Valencia, emphasizing the NOx pollutant and traffic in the Plaza del Ayuntamiento between 2018 and 2023.

In 2020, a significant change is highlighted due to the COVID-19 pandemic, which drastically reduced vehicular activity and pollution levels, offering an opportunity to study the impact of traffic on air quality. The effect of urban measures, such as the pedestrianization of a major street near the town hall square, which improved pollution levels, is also investigated [Lebrusán and Toutouh, 2021].

This work seeks to provide a comprehensive view of pollution in Valencia, exploring the relationships between pollutant levels and the factors that affect them, to inform effective policies to improve air quality in the city.

Several machine learning models are analyzed to predict the expected evolution of NOx in the town hall square without the influence of COVID-19 and pedestrianization, evaluating against it the environmental benefit of these events to predict the evolution of NOx from March 2020 and compare them with actual data influenced by confinement and pedestrianization measures [Sanchez et al., 2021].

As stated in the article [Subrahmanyam et al., 2023], air quality is a crucial aspect for public health, and predicting air pollution is essential to mitigate its effects.

5.3. Resum

Este Treball Final de Grau es centra en analitzar l'evolució de la contaminació ambiental i el trànsit a València, amb èmfasi en el contaminant NOx i el trànsit en la plaça de l'ajuntament, entre 2018 i 2023.

En 2020 es destaca un canvi significatiu a causa de la pandèmia de COVID-19, que va reduir dràsticament l'activitat vehicular i els nivells de contaminació, oferint una oportunitat per a estudiar l'impacte del trànsit en la qualitat de l'aire. També s'investiga l'efecte de mesures urbanes, com la vianalització d'un carrer important a prop de la Plaça de l'Ajuntament, que va resultar en una millora en els nivells de contaminants [Lebrusán and Toutouh, 2021].

Este treball busca proporcionar una visió integral de la contaminació a València, explorant les relacions entre els nivells de contaminants i els factors que els afecten, amb l'objectiu d'informar sobre polítiques efectives per millorar la qualitat de l'aire a la ciutat.

S'analitzen diversos models d'aprenentatge automàtic per predir l'evolució esperada del NOx en la Plaça de l'Ajuntament sense la influència del COVID-19 i sense la vianalització, avaluant enfront d'això el benefici ambiental d'aquests esdeveniments per predir l'evolució del NOx a partir de març de 2020 i comparar-los amb les dades reals influïdes pel confinament i les mesures de vianalització [Sanchez et al., 2021].

Com diuen a l'article [Subrahmanyam et al., 2023], la qualitat de l'aire és un aspecte crucial per a la salut pública, i la predicció de la contaminació atmosfèrica és essencial per mitigar els seus efectes.

6. Palabras clave

Contaminación, Tráfico, Evaluación de datos, Análisis estadístico, Peatonalización, NOx, Regresión, Forecasting, Machine Learning, LSTM, ARIMA, SARIMA

7. Prólogo

La contaminación atmosférica representa un desafío global debido a su impacto negativo en el medio ambiente y la salud humana, siendo el tráfico vehicular una de las principales causas en áreas urbanas como Valencia. El óxido de nitrógeno (NOx) es un contaminante preocupante, generado principalmente por vehículos motorizados, lo que ha llevado a implementar diversas estrategias para reducir sus emisiones. A pesar de estos esfuerzos, la contaminación atmosférica sigue siendo un problema crucial, subrayando la necesidad de evaluar constantemente las fuentes de contaminación y la eficacia de las medidas de mitigación aplicadas.

Diferentes estudios se han enfocado en los contaminantes ambientales, proporcionando conocimientos valiosos para comprender mejor este problema y desarrollar soluciones efectivas.

Con este trabajo se persigue confirmar la relación entre el tráfico de coches y la evolución desfavorable de los contaminantes ambientales en la ciudad de Valencia.

Se decide acotar el área de estudio a la Plaza del Ayuntamiento de Valencia, donde se dispone de todos los datos ambientales, contaminantes y de tráfico, para el periodo del estudio que comprende entre 2018 y 2020.

Para el análisis se recogerán históricos de contaminantes ambientales y del tráfico presente en el área y periodo del estudio. Por un lado, los datos ambientales recogidos proceden de diferentes sensores distribuidos por toda la ciudad de Valencia, que facilitan datos que pueden llegar a ser incluso diarios. Por otro lado, los datos del tráfico conseguidos escalan el tráfico medio mensual de los días laborales en multitud de calles distribuidas por toda Valencia.

Esta diferencia en la frecuencia de medida obliga a adaptar los datos con mayor frecuencia a la de los de menor, pese a perder resolución por ello, en el estudio, para poder analizarlos en una misma escala.

Este Trabajo Fin de Grado ha sido posible gracias a los conocimientos adquiridos a lo largo de todo el grado, con especial énfasis en asignaturas como Estadística y Machine Learning, las cuales me han permitido comprender los datos del caso y aplicar modelos de estudio estadístico para su resolución.

8. Índices y listas

8.1. Índices de tablas

8.2. Índices de figuras

8.3. Lista de abreviaturas

AQI (Air Quality Index)
ARIMA (Autoregressive Integrated Moving Average)
DL (Deep Learning)
EPSA (Escuela Politécnica Superior de Alcoy)
ICA (Índide de Calidad del Aire)
Lasso (Least Absolute Shrinkage and Selection Operator)
LSTM (Long Short-Term Memory)
MAE (Mean Absolute Error)
MAPE (Mean Absolute Porcentual Error):
ML (Machine Learning)
MSE (Mean Squared Error)
NOx (Óxidos de Nitrógeno)
ODS (Open Document Spreadsheet) Es un formato de archivo de hoja de cálculo utilizado por aplicaciones de software de oficina, como LibreOffice Calc y Apache OpenOffice Calc.
ONU (Organización de las Naciones Unidas)
PM2.5 (Partículas Menores de 2.5 micras de diámetro)
PM10 (Partículas Menores de 10 micras de diámetro)
RMSE (Root Mean Squared Error)
RSS (Residual Sum of Squares)
SARIMA (Seasonal Autoregressive Integrated Moving Average)
TFG (Trabajo Final de Grado)

9. Metodología

El presente Trabajo de Fin de Grado (TFG) en Ingeniería Informática se centra en el análisis de datos ambientales y de tráfico en la ciudad de Valencia, utilizando información obtenida de fuentes públicas como la web `https://valencia.opendatasoft.com`, desde la que se recuperaron históricos de la evolución de contaminantes del ambiente de Valencia, y la web `https://www.valencia.es/es/cas/movilidad/otras-descargas`, de donde se consiguen recuentos del tráfico de Valencia para los años del estudio.

Se descargan datos históricos ambientales de diferentes zonas de Valencia con el objetivo de profundizar en el análisis de sus contaminantes sectorizando la ciudad para permitirle un estudio con una mayor resolución.

En una segunda etapa, se realiza un análisis descriptivo de los datos de contaminantes y un estudio de BoxPlots mensuales de los contaminantes, incluyendo una explicación de la tendencia al alza en los meses más fríos. Se grafican también los valores mensuales de los contaminantes bajo estudio para analizar su evolución particular.

Se recopilan también datos históricos de la intensidad del tráfico de vehículos motorizados por toda Valencia, obteniendo estos de la web `https://www.valencia.es/es/cas/movilidad/otras-descargas`.

Tras filtrar estos datos, se analiza la evolución promedio mensual de las calles bajo estudio, permitiendo analizar la evolución de ambos conjuntos de datos en periodos de tiempo comunes y evaluar sus posibles correlaciones.

Después, se recurre a modelos de Machine-Learning para estimar una evolución del índice NOx de la plaza del ayuntamiento desde marzo del 2020 hasta finales del 2022 para calcular la reducción de contaminación resultante de la reducción del tráfico resultante del confinamiento de 2020 y de la peatonalización aplicada a la Plaza del Ayuntamiento de Valencia.

En resumen, este TFG aborda un análisis exhaustivo de los datos ambientales y de tráfico en Valencia, explorando su relación y posibles implicaciones para la calidad del aire en la ciudad.

Se considera también la posible influencia sobre la calidad del aire de Valencia, de los contaminantes debidos al tráfico del puerto.

10. Resultados

10.1. Estudio de las estaciones de datos

La ciudad de Valencia cuenta con multitud de puntos de medida de gases distribuidos por toda ella, pero no sabemos si presentan una distribución geográfica y periodos de tiempo adecuados para el estudio.

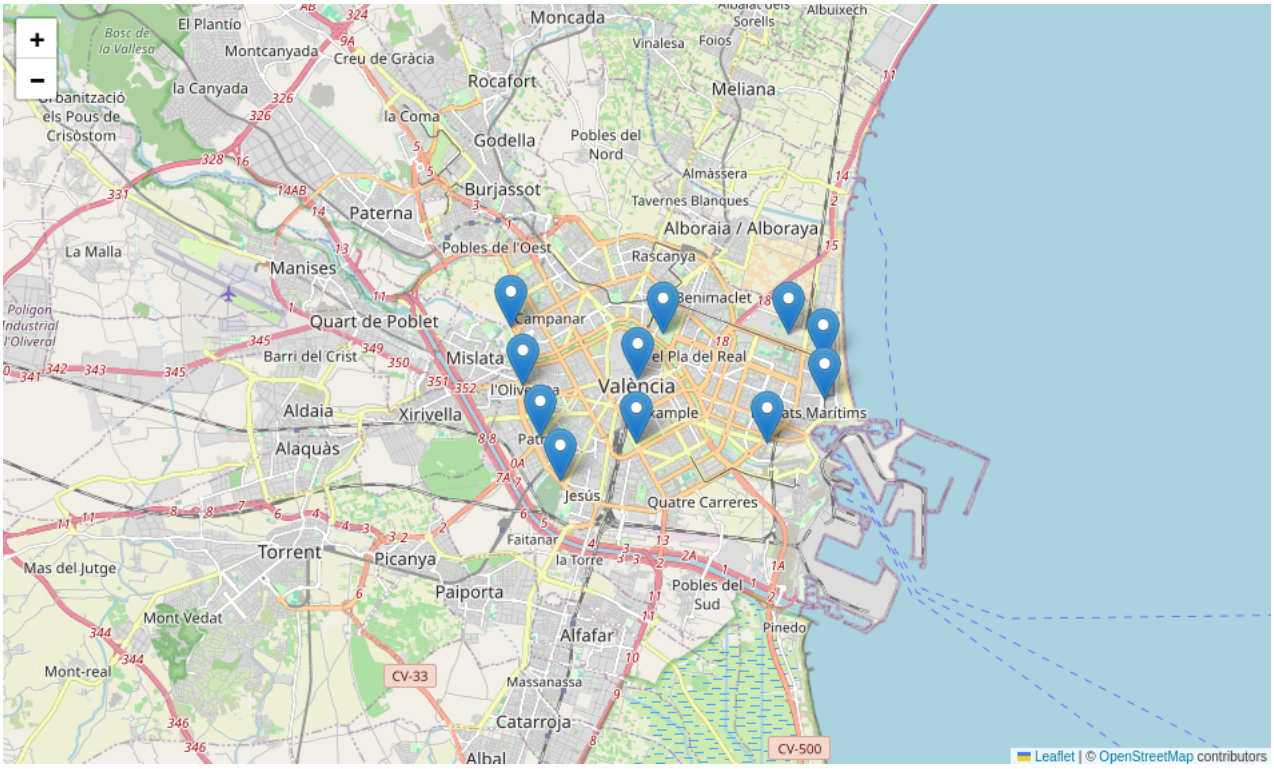


Figura 2: Lectura de las coordenadas de las estaciones de Valencia.

En este estudio, ante sus diferencias geográficas, respecto del mar o del centro urbano, y la disponibilidad heterogénea de datos entre las estaciones se ha decidido recortar esta multiplicidad de estaciones y centrar el estudio sobre la Plaza del Ayuntamiento, atendiendo a su tráfico y a sus estaciones de análisis ambiental en exclusiva para no solapar sus datos bajo los de otras estaciones de otras zonas diferentes de la ciudad.

10.2. Estudio de la disponibilidad de datos

Cada una de las estaciones presentan diferentes periodos en su recogida de datos y será necesario consultar todos ellos para razonar cómo aplicarlos al estudio.

A continuación se muestran los periodos con disponibilidad de datos de las estaciones utilizadas para el estudio, que son: 'Valencia_Centro', de la que extraeremos medidas del NOx y de 'Consellería_Meteo', de la que tomaremos valores ambientales.

Atendiendo a la disponibilidad de la evolución del NOx, acotaremos nuestra recogida de datos a la disponibilidad de estos datos de NOx, que sucede entre octubre del 2018 hasta acabar todo 2022.

Se incluyen en el anexo 12.1.2 las figuras de disponibilidad de datos del resto de las estaciones.

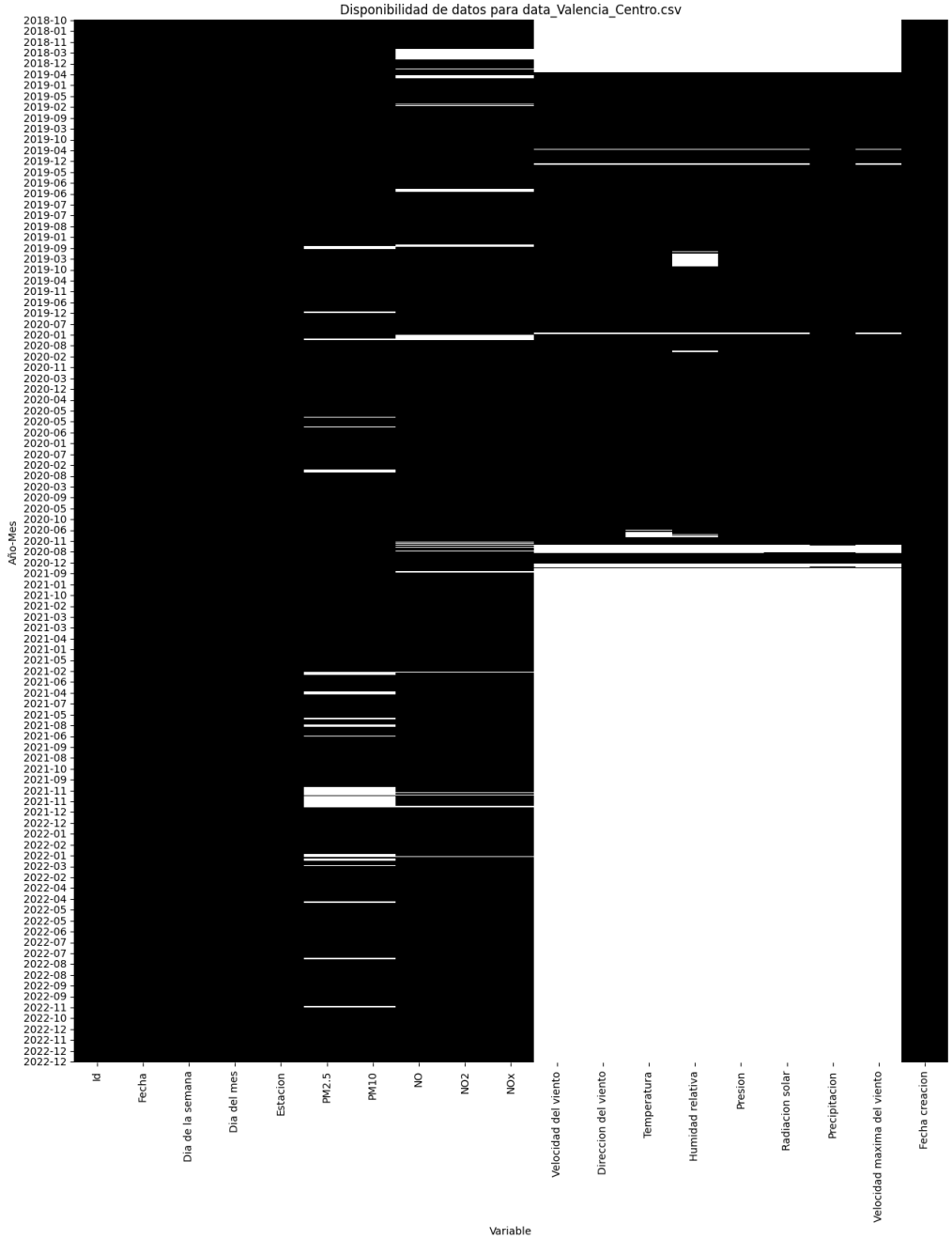


Figura 3: Disponibilidad de datos ambientales de la Plaza del Ayuntamiento.

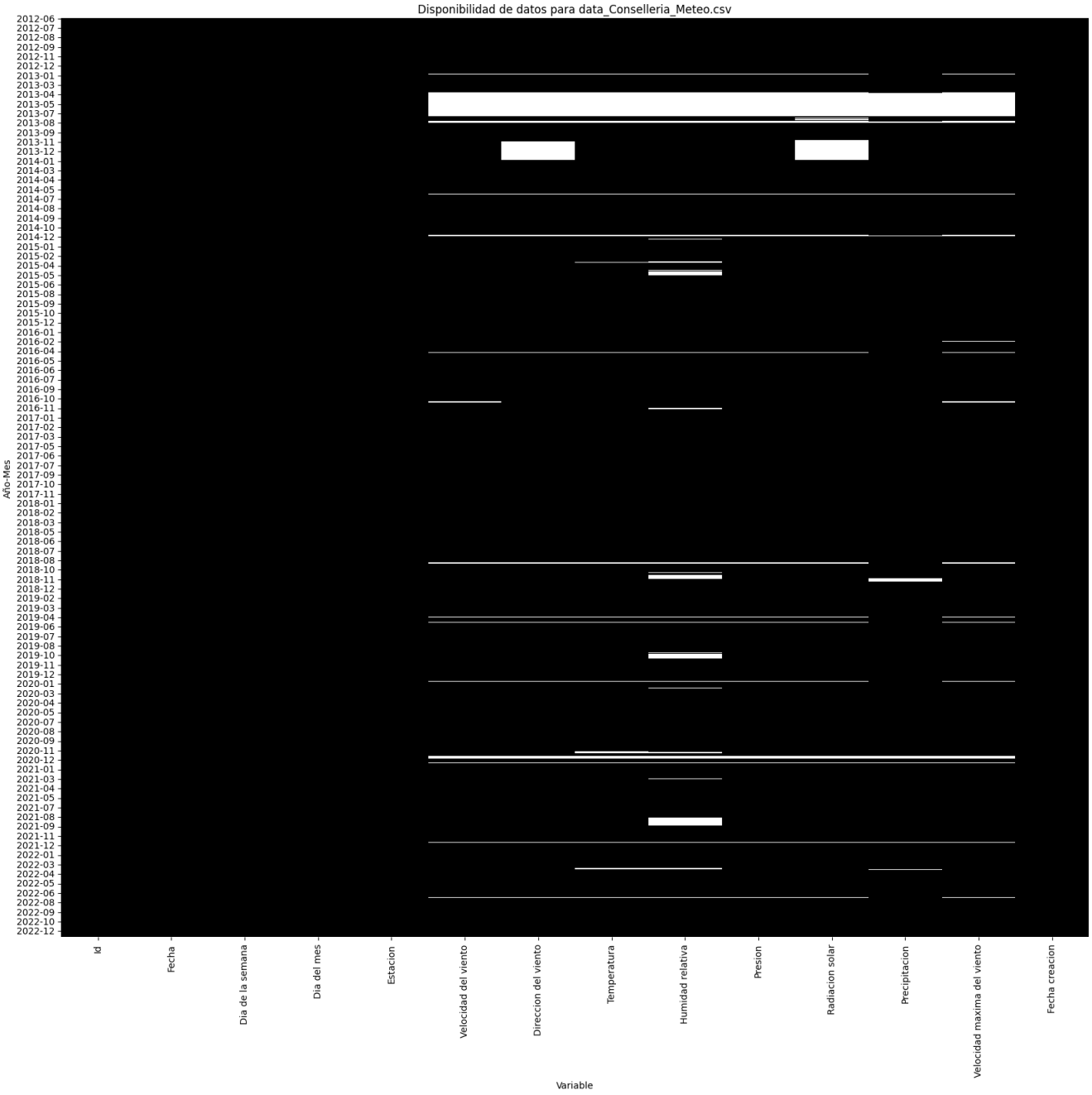


Figura 4: Disponibilidad de datos Consellería_Meteo.

10.3. Recopilación de datos de tráfico

El uso de datos públicos como imágenes de circuito cerrado de televisión (CCTV) de tráfico, información de los sensores de la ciudad inteligente de Seúl y datos meteorológicos son clave para estimar las concentraciones de partículas contaminantes PM2.5 y PM10.

Se han obtenido valores específicos del tráfico de muchas calles de Valencia para los años 2016-2024 con una resolución mensual del tráfico diario medio de los días laborables. Con

estos datos, tendremos la posibilidad de analizar objetivamente la evolución real del tráfico presente en la ciudad de Valencia, pudiendo buscar con ellos una correlación hacia los valores de contaminación ambiental recogidos en el mismo periodo.

El proceso de recopilación de datos se logró a través de la página web oficial del Ayuntamiento de Valencia: `https://www.valencia.es`, donde se publica esta información. Esta información publicada en bruto se recogerá en un archivo .ods con una página diferente para los datos de cada mes.

Como se indica en [Won et al., 2022], el empleo de tecnologías como las imágenes de CCTV y los datos de sensores urbanos puede ser fundamental para comprender y predecir la calidad del aire en entornos urbanos.

Figura 5: Datos del tráfico de Valencia para enero del 2018.

10.4. Análisis de datos

10.4.1 Análisis de datos ambientales

Se descargan las bases de datos con los históricos de contaminantes ambientales de Valencia. Para ello se selecciona el archivo imdss.csv publicado en la web que contiene registros de datos de tráfico de la Plaza del Ayuntamiento.

Se realiza una exploración inicial del archivo CSV para comprender su estructura y contenido. Se lleva a cabo la limpieza inicial de datos para corregir posibles inconsistencias, como valores faltantes o mal formateados.

Los artículos [Rhyu et al., 2024] y [Park et al., 2023] desarrollan diferentes métodos con que filtrar datos atípicos o faltantes que pudieran contaminar los datos objetivo. Se filtran los registros del archivo CSV y sus outliers para seleccionar únicamente los datos relevantes para la Plaza del Ayuntamiento. Esto implica identificar y extraer las mediciones de tráfico realizadas en la ubicación específica de la plaza.

10.4.2 Estudio de la influencia del viento del puerto sobre los datos ambientales

Para evaluar el estudio de la influencia de los contaminantes portuarios en el ambiente de la ciudad de Valencia se analiza el histórico de las direcciones del viento y de la evolución simultánea del NOx y se estudia la frecuencia con que un viento desde el puerto se refleja en un aumento del NOx de la ciudad debido a los motores del tráfico portuario.

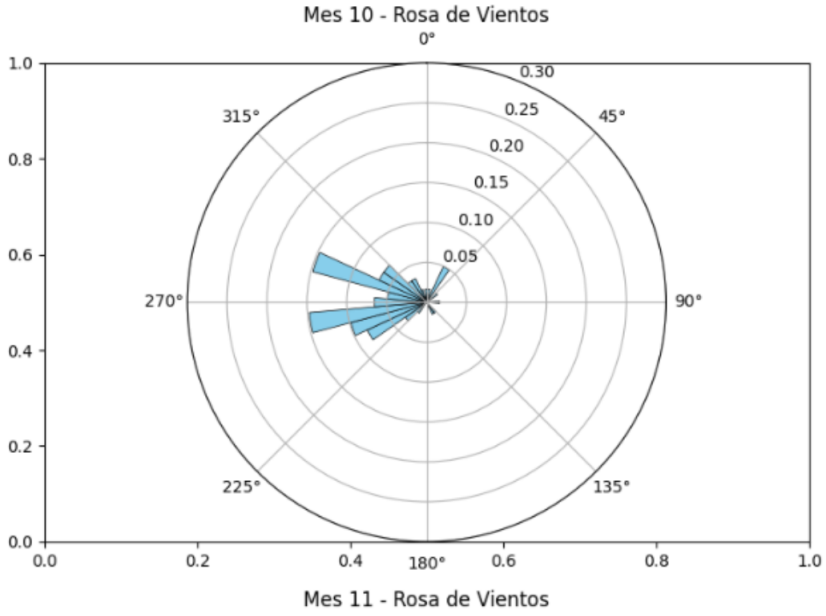


Figura 6: Acumulado de dirección de vientos

```
# Definir una función para filtrar outliers
def filter_outliers(series):
    if isinstance(series, pd.Series):
        # Convertir la serie al tipo de datos numérico
        series = pd.to_numeric(series, errors='coerce')
        Q1 = series.quantile(0.25)
        Q3 = series.quantile(0.75)
        IQR = Q3 - Q1
        lower_bound = Q1 - 1.5 * IQR
        upper_bound = Q3 + 1.5 * IQR
        return series.mask((series < lower_bound) | (series > upper_bound))
    else:
        return series
# Columnas a filtrar
columns_to_filter = ['NO', 'NO2', 'NOx', 'PM2.5', 'PM10', 'Ruido', 'C8H10',
                     'Velocidad del viento', 'Direccion del viento',
                     'Temperatura', 'Humedad relativa', 'Presion',
                     'Radiacion solar', 'Precipitacion',
                     'Velocidad maxima del viento']
# Filtrar outliers para cada conjunto de datos
for column in columns_to_filter:
    if column in t_plaz_ayu.columns:
        t_plaz_ayu[column] = filter_outliers(t_plaz_ayu[column])
    if column in data_valencia_centro.columns:
        data_valencia_centro[column] =
        filter_outliers(data_valencia_centro[column])
    if column in data_conselleria_meteo.columns:
        data_conselleria_meteo[column] =
        filter_outliers(data_conselleria_meteo[column])
```

Cuadro 1: Filtrado de outliers

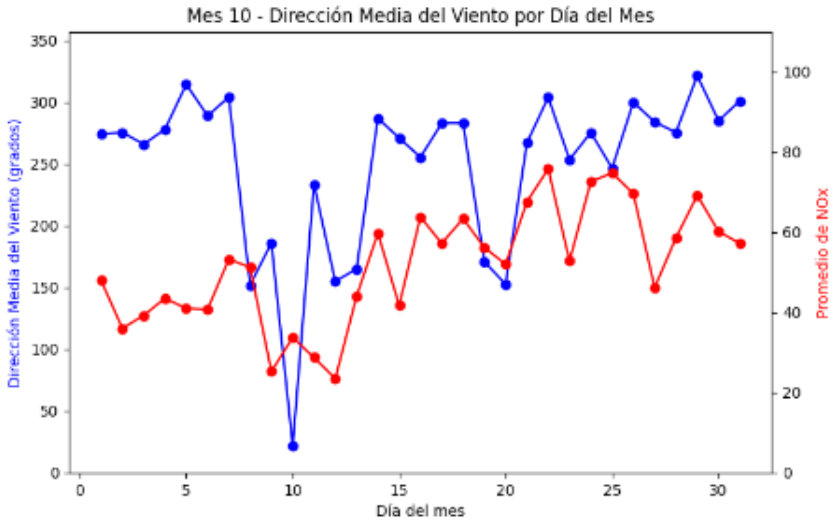


Figura 7: Cronograma con la evolución de la media semanal del NOx y vientos

En el cuadro 2 se muestra el código para el análisis de la evolución de esta dirección del viento con la respuesta simultánea del NOx para el mes de octubre de 2018. Se puede consultar en el anexo 12.1.2 el resto de figuras con la disponibilidad de datos específica de cada estación.

Atendiendo a que en los periodos de datos con una mayor acumulado de vientos de origen este no se recogen incrementos apreciables en los datos de contaminantes, entendemos que no podemos apreciar correlación entre los gases contaminantes del transporte del puerto y el índice NOx del centro de la ciudad.

```
# Cargar los datos desde el archivo CSV
data = pd.read_csv('data_Valencia_Centro.csv')
# Convertir la columna de fecha a tipo datetime
data['Fecha'] = pd.to_datetime(data['Fecha'])
# Filtrar los datos para obtener las observaciones donde la dirección del
viento es desde el este
data_east_wind = data[(data['Direccion del viento'] > 85)
& (data['Direccion del viento'] < 95)]
# Calcular el promedio de NOx para las observaciones donde el viento es desde
el este
avg_nox_east_wind = data_east_wind['NOx'].mean()
# Filtrar los datos para obtener las observaciones donde la dirección del viento
no es desde el este
data_non_east_wind =
data[(data['Direccion del viento'] < 85) | (data['Direccion del viento'] > 95)]
# Calcular el promedio de NOx para las observaciones donde el viento no es desde el est
avg_nox_non_east_wind = data_non_east_wind['NOx'].mean()
# Comparar los promedios de NOx para ambas condiciones
print("Promedio de NOx cuando el viento sopla desde el este:",
       avg_nox_east_wind)
print("Promedio de NOx cuando el viento no sopla desde el este:",
       avg_nox_non_east_wind)
```

Cuadro 2: Filtrado de datos

Condición del Viento	**Promedio de NOx**
Cuando el viento sopla desde el este	41.333333333333336
Cuando el viento no sopla desde el este	55.24848484848485

Cuadro 3: Promedio de NOx según la dirección del viento

10.4.3 Análisis de datos ambientales y de tráfico combinados

[Lebrusán and Toutouh, 2021] señalan que el uso de datos abiertos es esencial para evaluar el impacto de las medidas de mejora de la calidad del aire en áreas urbanas.

En la Fig. 8 se realiza una representación visual combinada que incluye un análisis descriptivo de boxplots mensuales con los datos de contaminantes ambientales, junto con la evolución del tráfico mensual en la Plaza del Ayuntamiento de Valencia. Esta representación conjunta facilita la evaluación de posibles relaciones entre ambos conjuntos de variables, permitiendo una comprensión orientativa de su interdependencia y comportamiento.

Durante el período analizado en la Tabla 2, se observa una gran disminución en el volumen de tráfico a partir de marzo de 2020, coincidiendo con las medidas de confinamiento implementadas debido a la pandemia de COVID-19. Este descenso en el tráfico se mantuvo en los meses siguientes y después se vio reforzado por la peatonalización de la Plaza del Ayuntamiento en mayo de 2021, que restringió aún más el acceso vehicular a la zona.

La reducción significativa en el tráfico motorizado durante el período de confinamiento y su mantenimiento posterior con la peatonalización de la Plaza del Ayuntamiento han tenido un impacto notable en los niveles de contaminantes ambientales. Se observa una disminución en la concentración de contaminantes atmosféricos durante este período, lo que sugiere una correlación entre la reducción del tráfico y la mejora de la calidad del aire en la zona estudiada.

Esta representación integrada en la Fig. 8 ofrece una visión completa de la interacción entre el tráfico urbano y la calidad del aire, destacando el papel crucial de las políticas de movilidad urbana en la mitigación de la contaminación atmosférica y la promoción de entornos más saludables y sostenibles.

La Fig. 8 muestra el drástico efecto del periodo de confinamiento sobre el tráfico de la Plaza del Ayuntamiento, y su posterior persistencia gracias a la peatonalización aplicada a la misma en mayo del 2021, que cancela el tráfico entrante a la plaza desde la Avenida de la Paz. Gracias a esta reducción de tráfico, en los meses posteriores se nota aún una cierta reducción del NOx (Óxidos de Nitrógeno) de la zona.

Como dicen en el artículo [Sanchez et al., 2021], las medidas de reducción del tráfico y peatonalización pueden tener un impacto significativo en las emisiones de contaminantes atmosféricos, como el NOx, lo que subraya la importancia de estas políticas en la mejora de la calidad del aire urbano.

10.5. Modelos de estudio de la correlación entre los datos

10.5.1 Lectura de los documentos con datos que estimar

En los artículos [Datta et al., 2024], [Rodríguez-García et al., 2023], [Liu et al., 2024], [Türkoğlu et al., 2024] se resalta que el análisis de correlación y visualización de datos es fundamental para entender las relaciones entre las variables ambientales y la contaminación del aire, lo que puede informar sobre la elaboración de modelos predictivos más precisos y efectivos.

Se hace un análisis completo de correlación entre diversas variables y la variable objetivo 'NOx' para escalar así la dependencia de ésta para con los demás datos conocidos. Para este análisis se utilizan bibliotecas como Pandas, Seaborn y Matplotlib para llevar a cabo estas tareas y visualizaciones.

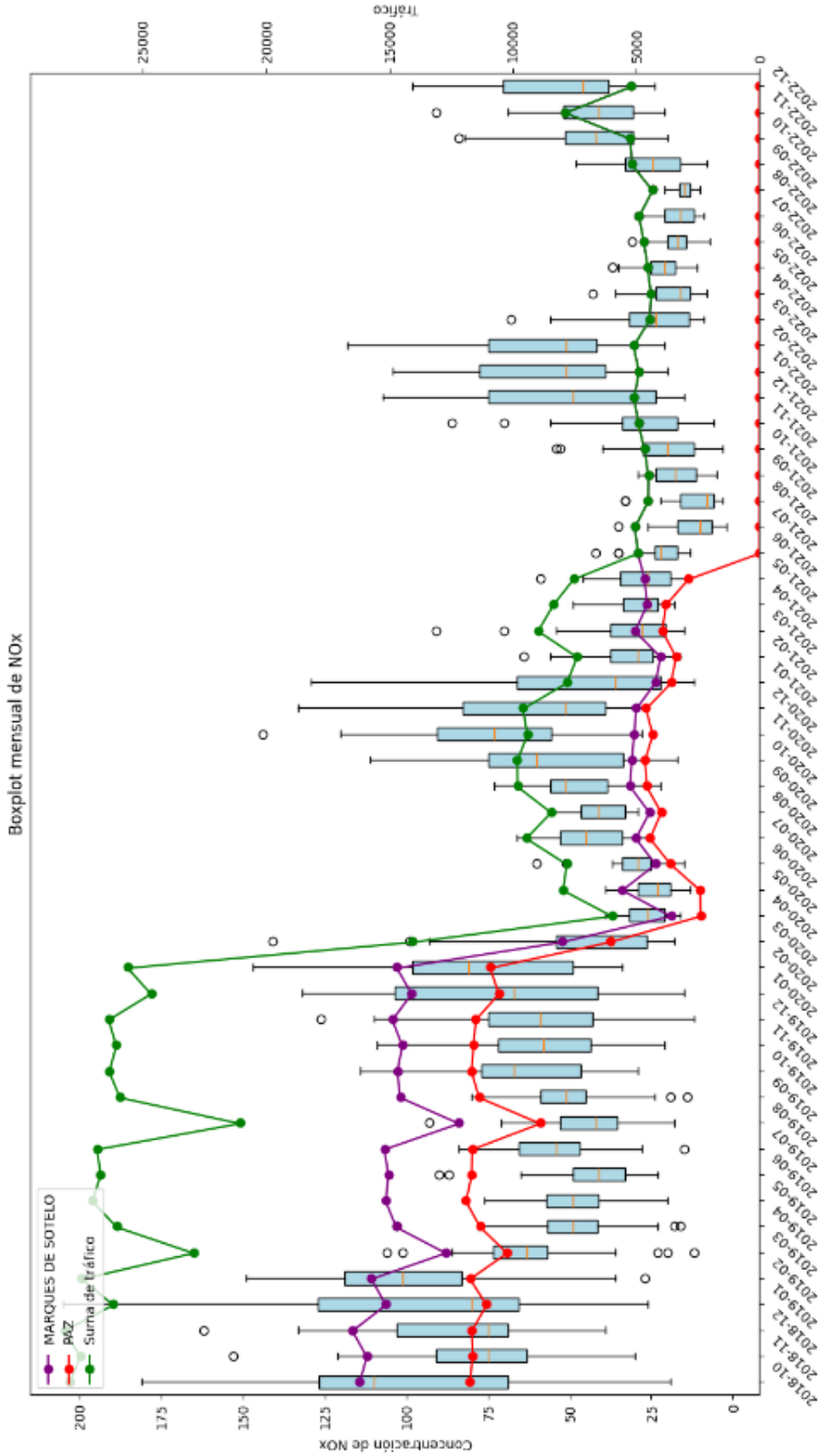


Figura 8: Evolución del tráfico entrante a la Plaza del Ayuntamiento de valencia y su impacto ambiental

Para la preparación de los datos se leen varios archivos CSV que contienen datos de diferentes fuentes y se realizan algunas manipulaciones en los datos, como el formateo de fechas, la aplicación de filtros por un rango de fechas específico y la creación de una nueva columna en uno de los DataFrames.

Para el análisis de correlación se calcula la correlación entre diferentes variables, centrándose especialmente en la variable 'NOx' y se imprimen las 10 columnas más influyentes en el valor de 'NOx' para cada mes.

Por último, para la visualización de los datos se generan gráficos de dispersión para explorar la relación entre la variable 'NOx' y las variables predictoras más correlacionadas, incluyendo estos gráficos una línea de regresión lineal para visualizar la tendencia recogida.

Se leen los datos con los que trabajar, de tres fuentes diferentes: El recuento del tráfico de las calles de valencia publicado en su web (t_plaz_ayu), los datos de contaminación ambiental de una estación central de Valencia (data_valencia_centro) y los datos ambientales recogidos por otra estación del centro de Valencia (data_conselleria_meteo).

```
# Leer datos

t_plaz_ayu = pd.read_csv('t_plaz_ayu.csv')
data_valencia_centro = pd.read_csv('data_Valencia_Centro.csv')
data_conselleria_meteo = pd.read_csv('data_Conselleria_Meteo.csv')
```

Cuadro 4: Lectura de los datos

Se filtran los datos para el periodo de interés con las medias mensuales de entre semana de 2018 a 2022 (Cuadro 5).

```
data_valencia_centro['Fecha'] = pd.to_datetime(data_valencia_centro['Fecha'])
data_valencia_centro = data_valencia_centro[
(data_valencia_centro['Fecha'] <= '2022-12-31')]

data_valencia_centro_nox = data_valencia_centro[['Fecha', 'NOx']]
data_conselleria_meteo['Fecha'] =
        pd.to_datetime(data_conselleria_meteo['Fecha'])

data_conselleria_meteo = data_conselleria_meteo[
(data_conselleria_meteo['Fecha'] <= '2022-12-31')]

t_plaz_ayu['Fecha'] = pd.to_datetime(t_plaz_ayu['Date'])
t_plaz_ayu = t_plaz_ayu[(t_plaz_ayu['Fecha'] <= '2022-12-31')]

# Rellena los valores faltantes con cero en las columnas 'PAZ' y
'MARQUES DE SOTELO'
t_plaz_ayu['PAZ'].fillna(0, inplace=True)
t_plaz_ayu['MARQUES DE SOTELO'].fillna(0, inplace=True)

# Genera el nuevo DataFrame con la columna adicional \textit{t_plaza}
sumando 'PAZ' y 'MARQUES DE SOTELO'
t_plaz_ayu['t_plaza'] = t_plaz_ayu['PAZ'] + t_plaz_ayu['MARQUES DE SOTELO']

# Filtra solo los días de entre semana (lunes a viernes)
data_valencia_centro =
data_valencia_centro[data_valencia_centro['Fecha'].dt.dayofweek < 5]
data_conselleria_meteo =
data_conselleria_meteo[data_conselleria_meteo['Fecha'].dt.dayofweek < 5]
t_plaz_ayu = t_plaz_ayu[t_plaz_ayu['Fecha'].dt.dayofweek < 5]

# Guarda el DataFrame resultante en un nuevo archivo CSV llamado total_traf.csv
t_plaz_ayu.to_csv('total_traf.csv', index=False)
```

Cuadro 5: Filtrado de fechas

```
# Formatea las fechas
t_plaz_ayu.rename(columns={'Date': 'Fecha'}, inplace=True)
t_plaz_ayu['Fecha'] = pd.to_datetime(t_plaz_ayu['Fecha'])
data_valencia_centro['Fecha'] = pd.to_datetime(data_valencia_centro['Fecha'])
data_conselleria_meteo['Fecha'] =
        pd.to_datetime(data_conselleria_meteo['Fecha'])
```

Cuadro 7: Formateado de fechas

En el Cuadro 6 se muestra el código para filtrar las columnas de interés.

```
data_valencia_centro = data_valencia_centro.drop(columns=['PM1','O3', 'SO2',
        'CO', 'NH3', 'C7H8',
       'C6H6', 'Ruido', 'C8H10', 'Velocidad del viento',
       'Direccion del viento', 'Temperatura', 'Humidad relativa', 'Presion',
       'Radiacion solar', 'Precipitacion', 'Velocidad maxima del viento',
       'As (ng/m³)', 'Ni (ng/m³)', 'Cd (ng/m³)', 'Pb (ng/m³)', 'B(a)p (ng/m³)',
       'Fecha creacion', 'Fecha baja'])
```

Cuadro 6: Filtrado de columnas de interés

Se formatean las fechas de los diferentes documentos para trabajarlos en conjunto, el código se muestra en el Cuadro 7.

```
# Agrupa por mes y calcular la media de cada variable para cada mes
data_valencia_centro_monthly_mean =
data_valencia_centro.groupby(data_valencia_centro['Fecha'].
                                dt.to_period('M')).mean()
# Resetea el índice para tener una columna de fecha
data_valencia_centro_monthly_mean.reset_index(inplace=True)
# Renombra la columna de fecha
data_valencia_centro_monthly_mean.rename(columns={'Fecha': 'Mes'},inplace=True)
# Muestra el DataFrame resultante
print(data_valencia_centro_monthly_mean)
# Extrae las columnas relevantes de data_valencia_centro.csv
data_valencia_centro_nox = data_valencia_centro[['Fecha', 'NOx']]
```

Cuadro 8: Media mensual

Se calcula la media de cada variable para cada mes, como se puede ver en los Cuadros 8 y 9.

Mes	Id	Día del mes	PM1	PM2.5	PM10	NO	NO2	NOx	O3	C6H6	CO
2019-01	28907.0	16.0	NaN	18.903226	34.709677	34.444444	NaN	NaN	NaN	NaN	NaN
2019-02	29143.0	14.5	NaN	26.464286	43.285714	29.920000	NaN	NaN	NaN	NaN	NaN
2019-03	29379.0	16.0	NaN	21.193548	36.000000	17.033333	NaN	NaN	NaN	NaN	NaN
2019-04	29623.0	15.5	NaN	14.133333	23.033333	14.100000	NaN	NaN	NaN	NaN	NaN
2019-05	29867.0	16.0	NaN	12.677419	22.677419	13.290323	NaN	NaN	NaN	NaN	NaN
2019-06	30111.0	15.5	NaN	11.300000	20.233333	11.160000	NaN	NaN	NaN	NaN	NaN
2019-07	30355.0	16.0	NaN	10.870968	17.129032	13.677419	NaN	NaN	NaN	NaN	NaN
2019-08	30603.0	16.0	NaN	10.225806	16.516129	11.870968	NaN	NaN	NaN	NaN	NaN
2019-09	30847.0	15.5	NaN	11.160000	19.320000	14.000000	NaN	NaN	NaN	NaN	NaN
2019-10	31091.0	16.0	NaN	14.129032	25.387097	20.258065	NaN	NaN	NaN	NaN	NaN
2019-11	31335.0	15.5	NaN	6.466667	13.466667	20.866667	NaN	NaN	NaN	NaN	NaN
2019-12	31579.0	16.0	NaN	13.655172	20.344828	21.870968	NaN	NaN	NaN	NaN	NaN
2020-01	31842.0	16.0	NaN	23.551724	34.275862	28.521739	NaN	NaN	NaN	NaN	NaN
2020-02	32112.0	15.0	NaN	23.793103	34.931034	28.793103	NaN	NaN	NaN	NaN	NaN
2020-03	32382.0	16.0	NaN	12.645161	17.580645	14.645161	NaN	NaN	NaN	NaN	NaN
2020-04	32656.5	15.5	NaN	10.433333	11.866667	7.100000	NaN	NaN	NaN	NaN	NaN
2020-05	32931.0	16.0	NaN	5.103448	9.034483	6.548387	NaN	NaN	NaN	NaN	NaN
2020-06	33205.5	15.5	NaN	7.655172	13.827586	7.366667	NaN	NaN	NaN	NaN	NaN
2020-07	33480.0	16.0	NaN	9.741935	23.290323	10.322581	NaN	NaN	NaN	NaN	NaN
2020-08	33759.0	16.0	NaN	5.333333	7.296296	9.032258	NaN	NaN	NaN	NaN	NaN
2020-09	34033.5	15.5	NaN	4.400000	4.933333	11.600000	NaN	NaN	NaN	NaN	NaN
2020-10	34308.0	16.0	NaN	3.806452	4.161290	13.322581	NaN	NaN	NaN	NaN	NaN
2020-11	34582.5	15.5	NaN	6.366667	6.566667	19.037037	NaN	NaN	NaN	NaN	NaN
2020-12	34857.0	16.0	NaN	4.225806	4.548387	21.285714	NaN	NaN	NaN	NaN	NaN

Cuadro 9: Datos mensuales

10.5.2 Estudio de las correlaciones entre predictores y variable de respuesta (NOx)

```
# Une DataFrames
df_concatenado =
pd.concat([data_valencia_centro_nox, data_conselleria_meteo], ignore_index=True)
# Extrae las columnas relevantes de t_plaz_ayu.csv
t_plaz_ayu_subset = t_plaz_ayu[['Fecha', 't_plaza']]
# Unir df_concatenado con t_plaz_ayu_subset
df_concatenado =
    pd.merge(df_concatenado, t_plaz_ayu_subset, on='Fecha', how='left')
# Elimina la columna \textit{Id} si existe en df_concatenado
if 'Id' in df_concatenado.columns:
    df_concatenado.drop('Id', axis=1, inplace=True)
# Agrupa por mes y calcular la media para cada grupo
df_mes = df_concatenado.groupby(pd.Grouper(key='Fecha', freq='M')).mean()
# Estándariza los valores a tanto por 1
df_mes = df_mes.div(df_mes.max())
# Guarda el DataFrame en un archivo CSV con una fila por mes
df_mes.to_csv('plaza_periodo_2018_2022.csv')
## Lista variables más correlacionadas
# Calcula la matriz de correlación para NOx
matriz_correlacion = df_mes.corr()['NOx'].drop('NOx')
# Ordena las correlaciones de mayor a menor
correlaciones_ordenadas = matriz_correlacion.abs().sort_values(ascending=False)
# Imprime las 10 columnas más influyentes en el valor de NOx de cada mes
print("Las 10 columnas más influyentes en el valor de NOx de cada mes:\n")
for i in range(10):
    columna = correlaciones_ordenadas.index[i]
```

Columna	Porcentaje de Influencia
t_plaza	78.28 %
Dirección del viento	64.96 %
Radiación solar	53.26 %
Presión	45.97 %
Temperatura	45.25 %
Humedad relativa	43.60 %
Velocidad del viento	17.07 %
Precipitación	12.37 %
Velocidad máxima del viento	8.04 %
Día del mes	3.39 %

Cuadro 11: Las 10 columnas más influyentes en el valor de NOx de cada mes

El cuadro 11 muestra los diferentes porcentajes de influencia calculados para cada variable del estudio frente a la variable de salida NOx. De este cuadro entendemos que la variable estudiada con mayor repercusión sobre la evolución del NOx fue el tráfico de la plaza, que de acuerdo con el estudio de correlación aplicado puede justificar el 78.28 % de la evolución del NOx.

Como segunda variable con mayor correlación sobre la evolución del NOx encontramos a la dirección del viento, que puede justificar el 64.96 % de la evolución del NOx.

El código de la Tabla 12 se calculará también una matriz de correlación con la influencia de las variables del estudio entre sí.

[Moradi et al., 2024] y [Ameer et al., 2019] resaltan que el análisis de correlación y visualización de datos es crucial para comprender la relación entre las variables ambientales y la contaminación del aire, lo que puede influir en la selección y aplicación de técnicas de modelado predictivo más efectivas.

Se analizarán las variables con mayor correlación con la evolución del NOx.

La tabla 9 muestra un diagrama de dispersión (scatter plot) con la respuesta del (NOx) frente a los predictores más correlacionados: t_plaza con un 78.28 %, Dirección del viento con un 64.96 %, Radiación solar con un 53.26 % y Presión con un 45.97 % de correlación.

La Fig. 14 añade también una línea de regresión lineal a los puntos de datos en el diagrama de dispersión para ver la relación entre ambas variables frente a una línea punteada gris que representa una relación lineal perfecta entre la variable caso y el NOx.

Se muestra el código de la primera gráfica; el resto sólo cambiarían la variable del caso y los textos afines.

```
# Paso 1: Carga los datos
archivo_csv = 'datos_conjunto.csv'
df = pd.read_csv(archivo_csv)
# Convierte la columna \textit{Year_Month} a tipo datetime si aún no está
en ese formato
df['Year_Month'] = pd.to_datetime(df['Year_Month'])
# Agregar una columna para el día de la semana
df['Dia de la semana'] = df['Year_Month'].dt.day_name()
# Filtrar los días laborables
df = df[df['Dia de la semana'].isin(['Monday', 'Tuesday', 'Wednesday',
        'Thursday', 'Friday'])]
# Paso 2: Elimina filas y columnas vacías
df = df.dropna(axis=0, how='all')
df = df.dropna(axis=1, how='all')
# Añadir t_plaza
for d in df:
    df['t_plaza'] = df['PAZ'] + df['MARQUES DE SOTELO']
# Paso 3: Explora los datos
print(\textit{Información del DataFrame:})
print(df.info())
print(\n\textit{Primeras filas del DataFrame:})
print(df.head())
# matriz_correlacion
# Paso 4: Calcula la correlación
matriz_correlacion = df.corr()
# Paso 5: Visualiza la correlación
plt.figure(figsize=(10, 8))
sns.heatmap(matriz_correlacion, annot=True, cmap='coolwarm', fmt=".2f",
            linewidths=.5)
plt.title('Matriz de correlación datos_conjunto (Días laborables)')
plt.show()
```

Cuadro 12: Correlación de datos

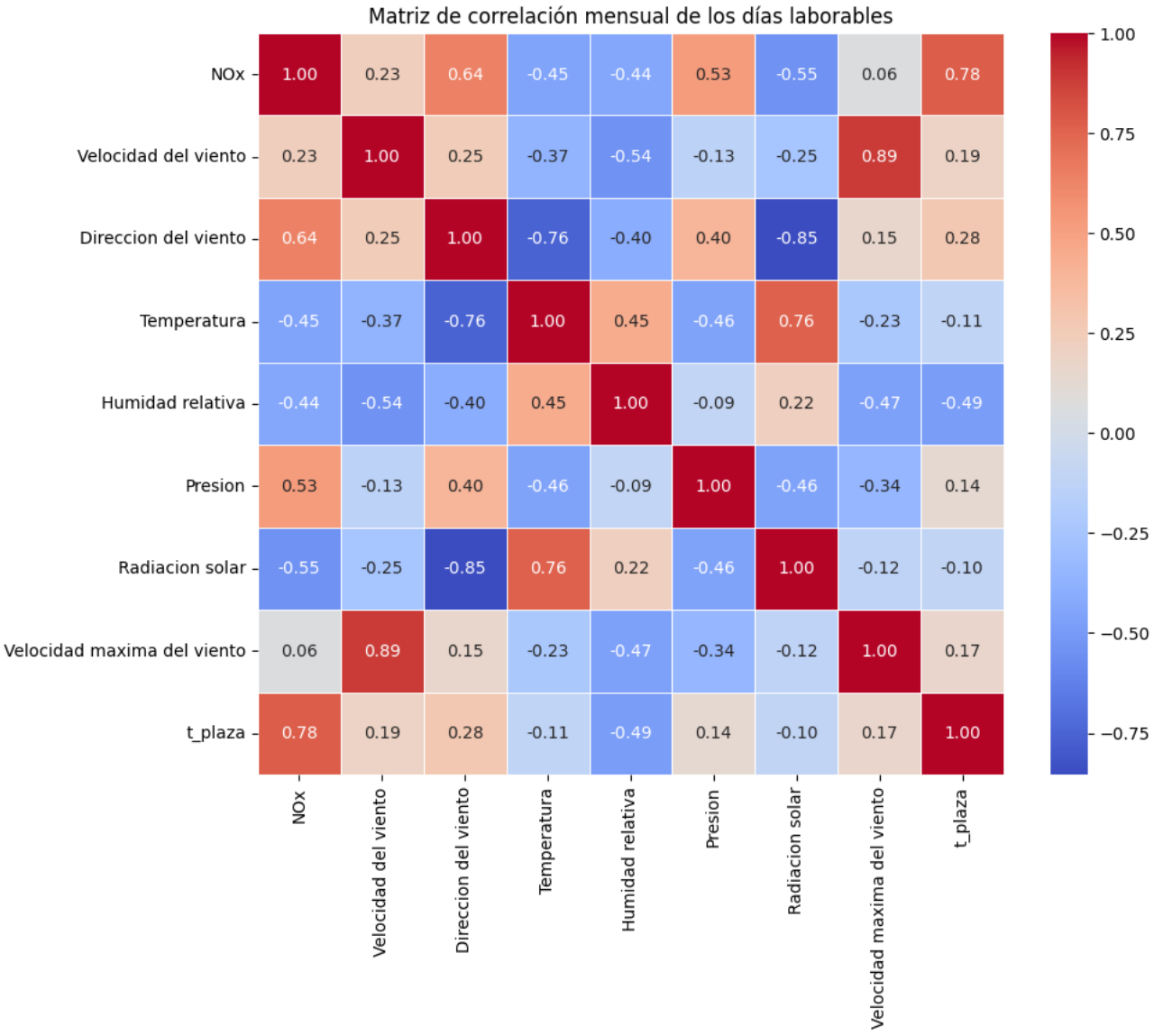


Figura 9: Matriz de correlación.

```
# Relación con la 1a variable más relacionada (t_plaza: 78.28%):

# Cargar los datos de plaza_periodo_2018_2022.csv
data = pd.read_csv('plaza_periodo_2018_2022.csv')
data['Fecha'] = pd.to_datetime(data['Fecha'])

# Seleccionar la variable más correlacionada con NOx (MARQUES DE SOTELO)
predictor = 't_plaza'
response = 'NOx'

# Draw scatter plot of the response vs. predictor to analyze relationship
scatter_plot = sns.lmplot(x=predictor, y=response, data=data)
axes = scatter_plot.ax
x_max = max(data[predictor].max(), data[response].max())
x_min = min(data[predictor].min(), data[response].min())
eps = 0.2
axes.set_xlim(x_min - eps, x_max + eps)
axes.set_ylim(x_min - eps, x_max + eps)
axes.grid(True)
axes.plot([x_min - eps, x_max + eps], [x_min - eps, x_max + eps],
color='gray', linestyle='--')
plt.title('Scatter plot of ' + response + ' vs ' + predictor)
plt.xlabel(predictor)
plt.ylabel(response)
plt.legend(['Tráfico en la plaza'])
plt.show()
```

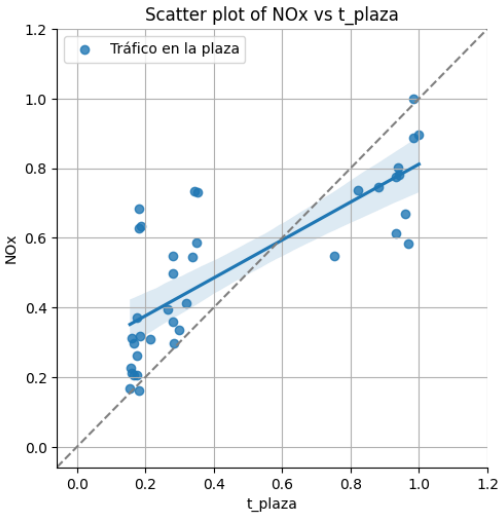
Cuadro 13: Variable más relacionada

En la Fig. 8 con el Tráfico de la plaza, que presenta la relación más directa, podemos apreciar 2 conjuntos de datos cercanos, muy dispares entre sí, que podemos entender como las mensualidades del tráfico tomadas antes y después del COVID-19 en marzo del 2020. Podemos entender los valores de la derecha como pertenecientes a meses anteriores al COVID-19 y los de la izquierda, a los meses después del COVID-19. Esto es ya que ese momento marcó una gran reducción repentina en el tráfico y ésta se mantuvo después gracias a la posterior peatonalización de la Plaza del Ayuntamiento que mantuvo una reducción media del tráfico del 74.3998 %, como calcula en el código de la tabla 15.

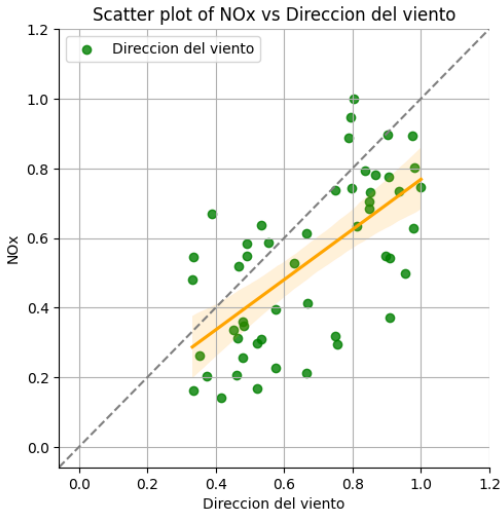
10.5.3 Stats models - Regresion Models

Las tablas 18 y 19 incluyen modelos de regresión que aplican a diferente número de variables, para analizar la evolución del NOx. Se argumenta después el número de variables más acertado antes de alcanzar un sobreajuste.

Cuadro 14: Variables más relacionadas con NOx



(a) Relación con la 1a variable más relacionada (t_plaza: 78.43 %)



(b) Relación con la 2a variable más relacionada (Dirección del viento: 63.63 %)

```
# 2 variables Regresion Model

# Cargar los datos de plaza_periodo_2018_2023.csv
data = pd.read_csv('plaza_periodo_2018_2023.csv')
data['Fecha'] = pd.to_datetime(data['Fecha'])
# Seleccionar las dos variables predictoras (MARQUES DE SOTELO y Direccion del viento)
predictors = ['t_plaza', 'Direccion del viento']
response = 'NOx'
# Eliminar filas con valores faltantes en las columnas predictoras
data.dropna(subset=predictors + [response], inplace=True)
# Añadir una constante al modelo
X = sm.add_constant(data[predictors])
# Construir el modelo de regresión lineal
model = sm.OLS(data[response], X).fit()
# Obtener los coeficientes del modelo
intercept = model.params['const']
coefficients = model.params[predictors]
# Imprimir la ecuación de la línea de regresión
equation = f'{response} = {intercept:.4f} '
for predictor, coef in coefficients.items():
    equation += f'+ ({coef:.4f}*{predictor}) '
# Obtener el coeficiente de determinación (R-squared)
r_sq = round(model.rsquared, 3)
r_sq_adj = round(model.rsquared_adj, 3)
# Imprimir los resultados
print('Equation of the regression line:', equation)
print('R-squared:', r_sq)
print('Adjusted R-squared:', r_sq_adj)
```

Cuadro 18: Modelos de regresión 2 variables

```
# Convierte'Date'] = pd.to_datetime(t_plaz_ayu['Date'], format='%m-%Y')

# Calcula el valor medio de los valores de cada mes antes de marzo de 2020
avg_before_mar_2020=t_plaz_ayu[t_plaz_ayu['Date']<
'2020-03']['t_plaza'].mean()

# Calcula el valor medio de los valores de cada mes tras marzo de 2020
avg_after_mar_2020 = t_plaz_ayu[t_plaz_ayu['Date'] >=
'2020-03']['t_plaza'].mean()

# Calcula la reducción porcentual del tráfico tras marzo de 2020 contra antes
reduccion_porcentual = ((avg_before_mar_2020 - avg_after_mar_2020) /
avg_before_mar_2020) * 100

# Imprime los resultados
print("NOx mensual medio antes de marzo de 2020:", avg_before_mar_2020)
print("NOx mensual medio tras marzo de 2020:", avg_after_mar_2020)
print("Reducción del tráfico tras marzo de 2020:", reduccion_porcentual, "%")
```

Cuadro 15: Correlaciones máximas con el NOx

Descripción	**Valor**
Tráfico medio antes de marzo de 2020	26047.76
Tráfico medio después de marzo de 2020	6668.28
Reducción porcentual del tráfico a partir de marzo de 2020	74.40 %

Cuadro 16: Resumen de la reducción del tráfico a partir de marzo de 2020

Descripción	**Valor**
NOx medio antes de marzo de 2020 (NOx)	65.78
NOx medio después de marzo de 2020 (NOx)	34.54
Reducción porcentual del NOx a partir de marzo de 2020	47.50 %

Cuadro 17: Resumen de la reducción del NOx a partir de marzo de 2020

Ecuación de la línea de regresión	NOx = -0.0108 + (0.4859 * t_plaza) + (0.4879 * Dirección del viento)
R-cuadrado	0.782
R-cuadrado ajustado	0.769

Cuadro 19: Resultados del modelo de regresión lineal

De los datos recogidos, extraemos una fórmula con la que estimar la evolución del NOx atendiendo a las 2 variables del caso: El tráfico total estimado de la Plaza del Ayuntamiento y la dirección del viento.

Extraemos también un coeficiente de determinación R^2 de 0.782, por el que puede entenderse una buena relación de la evolución de la variable de salida NOx.

Extraemos por último también una estimación del R^2 ajustado de 0.769, en el que se aprecia una pequeña reducción de relación frente a la del R^2, pero todavía mantiene una gran relación entre las variables indicando que el modelo no peca de sobreajuste.

En la Tabla 20 se grafica y se anota una tabla con la progresión de los resultados del análisis para estudios equivalentes al anterior, incluyendo nuevas variables predictoras por orden de correlación con la salida de NOx.

Número de Variables	**Ecuación de la Regresión**	**R-cuadrado**	**R-cuadrado Ajustado**
2	$NOx = -0{,}0108 + (0{,}4859 * t\_plaza) + (0{,}4879 * Direcciondelviento)$	0.782	0.769
3	$NOx = -13{,}6617 + (0{,}4827 * t\_plaza) + (0{,}3995 * Direcciondelviento) + (13{,}8684 * Presion)$	0.811	0.794
4	$NOx = -12{,}50826 + (0{,}49822 * t\_plaza) + (0{,}25150 * Direcciondelviento) + (12{,}89698 * Presion) + (-0{,}16700 * Radiacionsolar)$	0.819	0.797
5	$NOx = -16{,}42929 + (0{,}50193 * t\_plaza) + (0{,}21783 * Direcciondelviento) + (16{,}83012 * Presion) + (-0{,}31849 * Radiacionsolar) + (0{,}21737 * Temperatura)$	0.827	0.799
6	$NOx = -15{,}69099 + (0{,}51868 * t\_plaza) + (0{,}30193 * Direcciondelviento) + (15{,}84260 * Presion) + (-0{,}25331 * Radiacionsolar) + (0{,}16963 * Temperatura) + (0{,}22081 * Humidadrelativa)$	0.829	0.796
7	$NOx = -17{,}39839 + (0{,}51516 * t\_plaza) + (0{,}29064 * Direcciondelviento) + (17{,}50927 * Presion) + (-0{,}23967 * Radiacionsolar) + (0{,}18821 * Temperatura) + (0{,}22409 * Humidadrelativa) + (0{,}07700 * Velocidaddelviento)$	0.83	0.79

Cuadro 20: Resultados de los estudios de regresión en función del número de variables

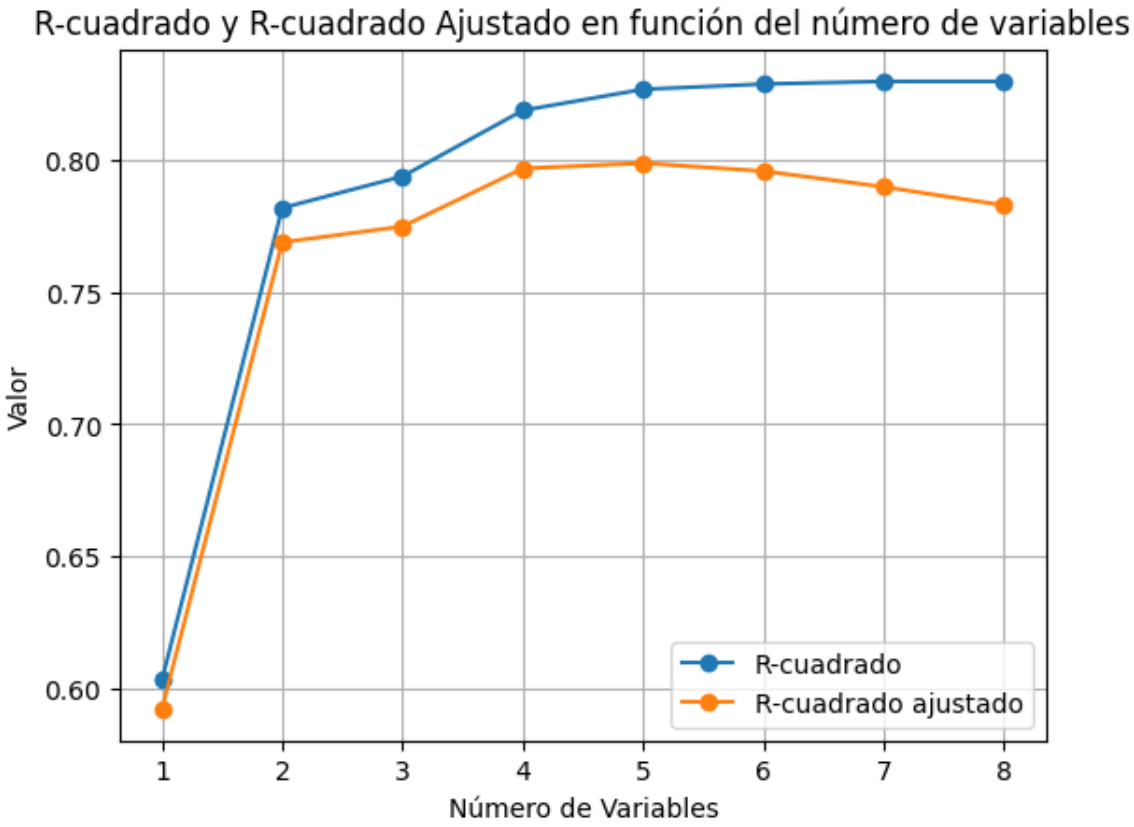


Figura 10: Evolución de la estimación en función de las variables

De la Tabla 19 y Fig. 10 podemos entender como a partir de las 4 variables de datos, aunque la eficiencia del modelo puede llegar a mejorar mínimamente, esa mejora no resulta suficiente para compensar el aumento de complejidad del modelo ante su pérdida de generalización, como nos demuestra el valor de R-cuadrado ajustado.

10.5.4 SciKit - Lasso and Ridge Regression

La regresión Lasso y Ridge son técnicas de regresión lineal regularizadas que se utilizan para abordar el problema de la multicolinealidad y la selección de características en modelos de regresión. Estas técnicas son implementadas en la biblioteca SciKit-Learn de Python.

Regresión Lasso (Least Absolute Shrinkage and Selection Operator): Lasso es una técnica de regularización que agrega una penalización L1 a la función de costo del modelo de regresión lineal estándar. Esto significa que, además de minimizar la suma de los cuadrados de los residuos (RSS), Lasso también penaliza la magnitud absoluta de los coeficientes de las características, forzando algunos de ellos a ser exactamente cero. Como resultado, Lasso no solo realiza la selección de características, sino que también puede reducir la complejidad del modelo al eliminar características irrelevantes.

Regresión Ridge: Similar a Lasso, Ridge es otra técnica de regularización que agrega una penalización L2 a la función de costo del modelo de regresión lineal. A diferencia de Lasso, la penalización L2 de Ridge penaliza la suma de los cuadrados de los coeficientes de las características. Ridge tiende a contraer los coeficientes hacia cero, pero raramente los lleva exactamente a cero. Esto significa que Ridge no realiza una selección de características tan agresiva como Lasso, pero puede ser más adecuado cuando todas las características son relevantes para el modelo.

Ambas técnicas son útiles para combatir el sobreajuste en modelos de regresión lineal, especialmente cuando se trabaja con conjuntos de datos de alta dimensionalidad con multicolinealidad entre características. SciKit-Learn proporciona implementaciones eficientes y fáciles de usar de Regresión Lasso y Ridge, lo que permite a los usuarios regularizar modelos de regresión lineal y obtener predicciones más estables y generalizables.

```
# Cargar los datos de plaza_periodo_2018_2023.csv
data = pd.read_csv('plaza_periodo_2018_2023.csv')
data['Fecha'] = pd.to_datetime(data['Fecha'])

# Seleccionar las variables predictoras y la variable de respuesta
predictors = ['Direccion del viento', 't_plaza', 'Presion', 'Radiacion solar']
response = 'NOx'

# Eliminar filas con valores faltantes
data.dropna(subset=[response] + predictors, inplace=True)

# Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(data[predictors],
                 data[response], test_size=0.3, random_state=42)

# Normalizar las características (recomendado para la regularización)
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

Cuadro 21: Modelo Lasso and Ridge

```
# Regresión lineal (OLS)
print('*** Regresión Lineal ***')
model = LinearRegression()
model.fit(X_train_scaled, y_train)
predictions = model.predict(X_test_scaled)
intercept = model.intercept_
coefficients = model.coef_
mse = mean_squared_error(y_test, predictions)
r_sq = r2_score(y_test, predictions)
num_predictors = X_train_scaled.shape[1]
n = len(y_test)
r_sq_adj = 1 - (1 - r_sq) * (n - 1) / (n - num_predictors - 1)
equation = f"{response} = {intercept:.4f} "
for predictor, coef in zip(predictors, coefficients):
    equation += f"+ ({coef:.4f} * {predictor}) "
print(f"MSE: {round(mse,3)}")
print(f"R-sq: {round(r_sq,3)}")
print(f"R-sq (adj): {round(r_sq_adj,3)}")
print(f"Equation: {equation}")
```

Cuadro 22: Modelo de Regresión lineal

MSE	0.013
R-sq	0.621
R-sq (adj)	0.404
Ecuación	NOx = 0.5456 + (0.0475 * Dirección del viento) + (0.1586 * t_plaza) + (0.0486 * Presión) - (0.0437 * Radiación solar)

Cuadro 23: Resultados de la regresión lineal

En conclusión, podemos decir que estos resultados de regresión lineal indican que el modelo proporciona una buena capacidad predictiva para la concentración de NOx en función de las variables predictoras seleccionadas.

El bajo valor del Error Cuadrático Medio (MSE) sugiere que las predicciones del modelo están cercanas a los valores reales de NOx, lo que indica una buena calidad del ajuste.

El valor del R-cuadrado (R-sq) indica que aproximadamente el 66.9 % de la variabilidad en la concentración de NOx es explicada por el modelo, lo cual es significativo.

El R-cuadrado ajustado (R-sq adj) penaliza el uso excesivo de variables predictoras y proporciona una medida más conservadora de la bondad de ajuste del modelo, siendo en este caso de 0.337.

La ecuación de regresión muestra cómo cada variable predictora contribuye a la predicción de la concentración de NOx. Estos resultados sugieren que las variables de dirección del viento, t plaza, presión y radiación solar tienen un impacto significativo en la concentración de NOx.

```
# Regresión Lasso (con selección automática de características)
print('*** Regresión Lasso ***')
model = Lasso(alpha=0.1)  # Configura el alpha (fortaleza de regularización)
model.fit(X_train_scaled, y_train)
predictions = model.predict(X_test_scaled)
intercept = model.intercept_
coefficients = model.coef_
mse = mean_squared_error(y_test, predictions)
r_sq = r2_score(y_test, predictions)
num_predictors = X_train_scaled.shape[1]
n = len(y_test)
r_sq_adj = 1 - (1 - r_sq) * (n - 1) / (n - num_predictors - 1)
equation = f"{response} = {intercept:.4f} "
for predictor, coef in zip(predictors, coefficients):
    if coef != 0: equation += f"+ ({coef:.4f} * {predictor}) "
print(f"MSE: {round(mse,3)}")
print(f"R-sq: {round(r_sq,3)}")
print(f"R-sq (adj): {round(r_sq_adj,3)}")
print(f"Equation: {equation}")
```

Cuadro 24: Regresión Lasso automática

MSE	0.022
R-sq	0.366
R-sq (adj)	0.004
Ecuación	NOx = 0.5456 + (0.0376 * Dirección del viento) + (0.0783 * t_plaza)

Cuadro 25: Resultados de la regresión Lasso

En resumen, estos resultados de regresión Lasso con selección automática de características indican que el modelo proporciona una buena capacidad predictiva para la concentración de NOx en función de las variables predictoras seleccionadas.

El bajo valor del Error Cuadrático Medio (MSE) sugiere que las predicciones del modelo están cercanas a los valores reales de NOx, lo que indica una buena calidad del ajuste.

El valor del R-cuadrado (R-sq) indica que aproximadamente el 38.5 % de la variabilidad en la concentración de NOx es explicada por el modelo, lo cual es significativo.

El R-cuadrado ajustado (R-sq adj) penaliza el uso excesivo de variables predictoras y proporciona una medida más conservadora de la bondad de ajuste del modelo, siendo en este caso de 0.034.

La ecuación de regresión muestra cómo cada variable predictora contribuye a la predicción de la concentración de NOx. En este caso, las variables 'Dirección del viento' y 't plaza' tienen coeficientes significativos, lo que sugiere que tienen un impacto importante en la concentración de NOx. Las otras variables predictoras pueden tener un efecto menor, dado que sus coeficientes son cercanos a cero.

```
# Regresión Ridge
print('*** Regresión Ridge ***')
model = Ridge(alpha=0.1)  # Puedes establecer el alpha (fortaleza de regularización)
model.fit(X_train_scaled, y_train)
predictions = model.predict(X_test_scaled)
intercept = model.intercept_
coefficients = model.coef_
mse = mean_squared_error(y_test, predictions)
r_sq = r2_score(y_test, predictions)
num_predictors = X_train_scaled.shape[1]
n = len(y_test)
r_sq_adj = 1 - (1 - r_sq) * (n - 1) / (n - num_predictors - 1)
equation = f"{response} = {intercept:.4f} "
for predictor, coef in zip(predictors, coefficients):
    equation += f"+ ({coef:.4f} * {predictor}) "
print(f"MSE: {round(mse,3)}")
print(f"R-sq: {round(r_sq,3)}")
print(f"R-sq (adj): {round(r_sq_adj,3)}")
print(f"Equation: {equation}")
```

Cuadro 26: missing data post-drop

MSE	0.013
R-sq	0.62
R-sq (adj)	0.404
Ecuación	NOx = 0.5456 + (0.0478 * Dirección del viento) + (0.1579 * t_plaza) + (0.0485 * Presión) + (-0.0436 * Radiación solar)

Cuadro 27: Resultados de la regresión Ridge

Los resultados de esta regresión Ridge muestran lo siguiente:

El Error Cuadrático Medio (MSE) es relativamente bajo, lo que indica que las predicciones del modelo Ridge están cercanas a los valores reales de NOx en comparación con otros modelos.

El valor del R-cuadrado (R-sq) es moderado, alrededor de 0.667, lo que sugiere que aproximadamente el 66.7 % de la variabilidad en la concentración de NOx es explicada por el modelo Ridge. Esto indica que el modelo se ajusta moderadamente bien a los datos.

El R-cuadrado ajustado (R-sq adj) también es moderado y refleja la misma tendencia que el R-sq, lo que sugiere que el modelo Ridge no está sobreajustando significativamente los datos.

Se hace un análisis completo de correlación entre diversas variables y la variable objetivo 'NOx' para escalar así la dependencia de ésta para con los demás datos conocidos.

Para este análisis se utilizan bibliotecas como Pandas, Seaborn y Matplotlib para llevar a cabo estas tareas y visualizaciones.

Para la preparación de los datos se leen varios archivos CSV que contienen datos de diferentes fuentes y se realizan algunas manipulaciones en los datos, como el formateo de fechas, la aplicación de filtros por un rango de fechas específico y la creación de una nueva columna en uno de los DataFrames.

Para el análisis de correlación se calcula la correlación entre diferentes variables, centrándose especialmente en la variable 'NOx' y se imprimen las 10 columnas más influyentes en el valor de 'NOx' para cada mes.

Por último, para la visualización de los datos se crean gráficos de dispersión para explorar la relación entre la variable 'NOx' y las variables predictoras más correlacionadas, incluyendo estos gráficos una línea de regresión lineal para visualizar la tendencia recogida.

En este caso, la ecuación de regresión Ridge muestra que todas las variables predictoras tienen coeficientes diferentes de cero en la ecuación, lo que indica que todas contribuyen a la predicción, aunque sea en menor medida .

10.5.5 Mejor modelo de Regresión

```
# Define predictor and response variables
predictors = ['t_plaza', 'Direccion del viento', 'Radiacion solar', 'Presion',
'Temperatura', 'Humidad relativa', 'Velocidad del viento',
'Velocidad maxima del viento']
X = data[predictors]
y = data[response]
# Add a constant and fit the all-predictors model
X = sm.add_constant(X)
model = sm.OLS(y, X)
regr = model.fit()
# Compute MSE for the all-predictors model
MSE = regr.mse_resid
# For a given subset of predictors, fit model and compute statistics
def processSubset(feature_set):
    # Fit model on feature_set and calculate RSSp and Adj R-sq
    Xsubset = X[list(feature_set)]
    Xsubset = sm.add_constant(Xsubset)
    model = sm.OLS(y, Xsubset)
    regr = model.fit()
    RSSp = regr.ssr
    n_vars = len(feature_set)
    Mallows_Cp = round(RSSp / MSE - len(y) + 2 * (n_vars + 1), 1)
    rsq = round(regr.rsquared, 3)
    rsq_adj = round(regr.rsquared_adj, 3)
    return {'n_vars': n_vars, 'R-sq': rsq, 'Adj R-sq': rsq_adj,
    'Mallows Cp': Mallows_Cp, 'model': regr}
# Returns the best model using k features
def getBest(k):
    results = []
    for combo in itertools.combinations(X.columns, k):
        results.append(processSubset(combo))
    # Wrap everything up in a nice dataframe
    models = pd.DataFrame(results)
    # Choose the model with the highest Adj R-sq
    best_k_model = models.loc[models['Adj R-sq'].argmax()]
    # Return the best model, along with more useful information about the model
    return best_k_model
# Could take quite a while to complete...
best_models = pd.DataFrame(columns=['n_vars','R-sq','Adj R-sq','Mallows Cp'])
k_max = len(X.columns) - 1  # Number of predictors (remove constant)
for i in range(1, k_max + 1):    best_models.loc[i] = getBest(i)
```

Cuadro 28: Mejor modelo de Regresión

n_vars	R-sq	Adj R-sq	Mallows Cp
1	0.603	0.592	32.1
2	0.781	0.768	4.5
3	0.81	0.794	1.6
4	0.818	0.796	2.3
5	0.824	0.797	3.3
6	0.824	0.797	5.3
7	0.826	0.792	7.0
8	0.826	0.785	9.0

Cuadro 29: Resultados del modelo en función del número de variables

```
# Select the best model from best_models
best_rsq_adj = 0  # Initialize best adjusted R-squared
for i in range(1, k_max + 1):
    model = getBest(i)
    rsq_adj = model['Adj R-sq']
    if rsq_adj > best_rsq_adj:
        best_rsq_adj = rsq_adj
        rsq = model['R-sq']
        Mallows_cp = model['Mallows Cp']
        coefficients = model['model'].params
        predictors = model['model'].params.index[1:]
# Display information about the best model
print(\textit{Best Model Information:})
print('R-sq:', rsq)
print('R-sq (adj):', best_rsq_adj)
print('Mallows_cp:', Mallows_cp)
# Convert coefficients to a list
coefficients = coefficients.tolist()
# Generate equation for the best model
equation = f"{response} = {coefficients[0]:.4f} "
for i, coef in enumerate(coefficients[1:], start=1):
    if coef != 0: equation += f"+ ({coef:.4f} * {predictors[i-1]}) "
print(\textit{Equation of the Best Model:})
print(equation)
```

Cuadro 30: Aplicación del mejor modelo

Información del modelo	Valor
R-sq	0.824
R-sq (adj)	0.797
Mallows_cp	3.3

Cuadro 31: Información del Mejor Modelo

Ecuación del mejor modelo	
NOx	-15.9410 + (0.4723 * t_plaza) + (0.2418 * Dirección del viento) + (-0.2848 * Radiación solar) + (16.2965 * Presión) + (0.2056 * Temperatura)

Cuadro 32: Ecuación del Mejor Modelo

Después de evaluar el desempeño del modelo con diferentes números de variables, se ha tomado la decisión de aplicar un modelo con 4 variables en el estudio. Esta elección se basa en la observación de que al aumentar el número de variables de 2 a 4, tanto el R-cuadrado como el R-cuadrado ajustado experimentan mejoras significativas. Esto sugiere que el modelo es capaz de explicar una mayor proporción de la variabilidad en la variable de respuesta a medida que se agregan más variables predictoras, lo que indica una mejor capacidad explicativa.

Sin embargo, se ha considerado también la pérdida de generalización que podría ocurrir al incluir más variables. Aunque el modelo puede volverse más ajustado a los datos de entrenamiento al agregar más variables, existe el riesgo de sobreajuste, lo que podría disminuir su capacidad para generalizar a datos nuevos o no vistos. Por lo tanto, se ha optado por mantener el equilibrio entre la capacidad explicativa del modelo y su capacidad de generalización al limitar el número de variables a 4.

Al notar que el R-cuadrado ajustado alcanza su valor máximo con 4 variables y luego comienza a disminuir gradualmente, se ha concluido que agregar más variables puede no justificar la complejidad adicional del modelo y podría comprometer su capacidad de generalización. Por lo tanto, se ha decidido evitar la inclusión de más variables para mantener un modelo que sea tanto explicativo como generalizable.

10.6. Predicción del beneficio ambiental con modelos de ML

Como dicen en los artículos [Liu et al., 2024], [Wai and Yu, 2023], [Che et al., 2023], [Mishra and Gupta, 2024], el uso de modelos de aprendizaje automático (ML) es fundamental para prever la contaminación del aire y comprender su impacto en el entorno urbano, lo que puede conducir a la implementación de medidas más efectivas para mejorar la calidad del aire y reducir las diferencias en la exposición a la contaminación.

Se aplicarán diferentes modelos ML para estimar una progresión coherente del NOx a partir de marzo del 2020, entrenando el modelo desde los datos previos, y valorar objetivamente el potencial beneficio ambiental resultante.

10.6.1 Explicación del modelo aplicado

Inicialmente se aplicará un modelo LSTM para estimar una evolución coherente del NOx a partir de marzo del 2020 y se evaluará SARIMA es una extensión del modelo ARIMA que incluye componentes estacionales para capturar patrones repetitivos en los datos a lo largo del tiempo. Esto es especialmente útil cuando se trabaja con series temporales que exhiben estacionalidad, como los niveles de NOx, que pueden mostrar variaciones regulares a lo largo del año. Al incorporar términos estacionales en el modelo SARIMA, podemos capturar mejor estas variaciones y mejorar la precisión de las predicciones. Por lo tanto, tanto ARIMA como SARIMA son opciones sólidas para modelar y predecir la evolución de series temporales como los niveles de NOx, y pueden ser utilizados para evaluar la precisión de las predicciones realizadas.

10.6.2 Aplicación del modelo

A continuación se mostrará una aplicación práctica de los modelos de predicción de NOx utilizando técnicas como LSTM, ARIMA y SARIMA. Estos modelos se entrenan con datos históricos recopilados desde octubre de 2018 hasta febrero de 2020, y luego se utilizan para predecir el nivel de NOx esperado para cada mes a partir de marzo de 2020.

Durante este período, se observa una discrepancia entre las predicciones de los modelos y los valores reales de NOx registrados. Esta discrepancia se atribuye a la implementación de medidas de confinamiento relacionadas con la pandemia de COVID-19, que resultaron en una reducción significativa del tráfico y, por lo tanto, de las emisiones de NOx. Además, se destaca que esta reducción en las emisiones de NOx se mantuvo en el tiempo debido a la peatonalización de la Plaza del Ayuntamiento, donde se encuentra ubicado el sensor de NOx.

Para ilustrar estas diferencias entre las predicciones de los modelos y los valores reales, se generan gráficas que muestran la evolución esperada del NOx según cada modelo, junto con la evolución real observada. En estas gráficas, se puede observar cómo las predicciones de los modelos son generalmente superiores a los valores reales debido a los cambios en las condiciones ambientales y de tráfico causados por la pandemia y las medidas de control implementadas.

Además, se calcula el acumulado de NOx reducido en las medidas reales en comparación con las medidas esperadas por los modelos. Este análisis proporciona una evaluación cuantitativa de la efectividad de las medidas de reducción de emisiones en términos de NOx, destacando el impacto real de las acciones tomadas para mitigar la contaminación atmosférica.

10.6.2.1. Aplicación del modelo LSTM

Se aplicarán diferentes modelos ML para estimar una progresión coherente del NOx a partir de marzo del 2020, entrenando el modelo desde los datos previos, y valorar objetivamente el potencial beneficio ambiental resultante. Esto es especialmente relevante dada la necesidad de comprender y abordar la contaminación atmosférica, como se destaca en el artículo de 'Rodríguez et al.' (2023) [Rodríguez-García et al., 2023]. Además, el enfoque innovador presentado por 'Liu' (2024) [Liu et al., 2024] en el marco ADMM LSTM ofrece una perspectiva interesante para la optimización en la previsión de carga a corto plazo en sistemas de energía. La importancia de abordar las inequidades ambientales causadas por la contaminación del aire se subraya en el estudio de 'Che et al.' (2023) [Che et al., 2023], donde se revela el impacto desproporcionado en la exposición individual y poblacional en áreas urbanas densamente pobladas. Además, la necesidad de mejorar la salud pública mediante la reducción de la exposición a la contaminación del aire se discute en el artículo de 'Mishra et al. (2024)' [Mishra and Gupta, 2024], que compara exhaustivamente diferentes metodologías para vigilar la calidad del aire. 'Macherla et al.' (2023) [Macherla et al., 2023] también destacan la importancia de predecir el Índice de Calidad del Aire (AQI) utilizando tecnologías de aprendizaje profundo para contribuir a la planificación urbana sostenible. Finalmente, el estudio de 'Varia et al.' (2022) [Varia and Kothari, 2022] resalta la relevancia de predecir la calidad del aire en áreas urbanas inteligentes para mejorar la gestión de la calidad del aire urbano.

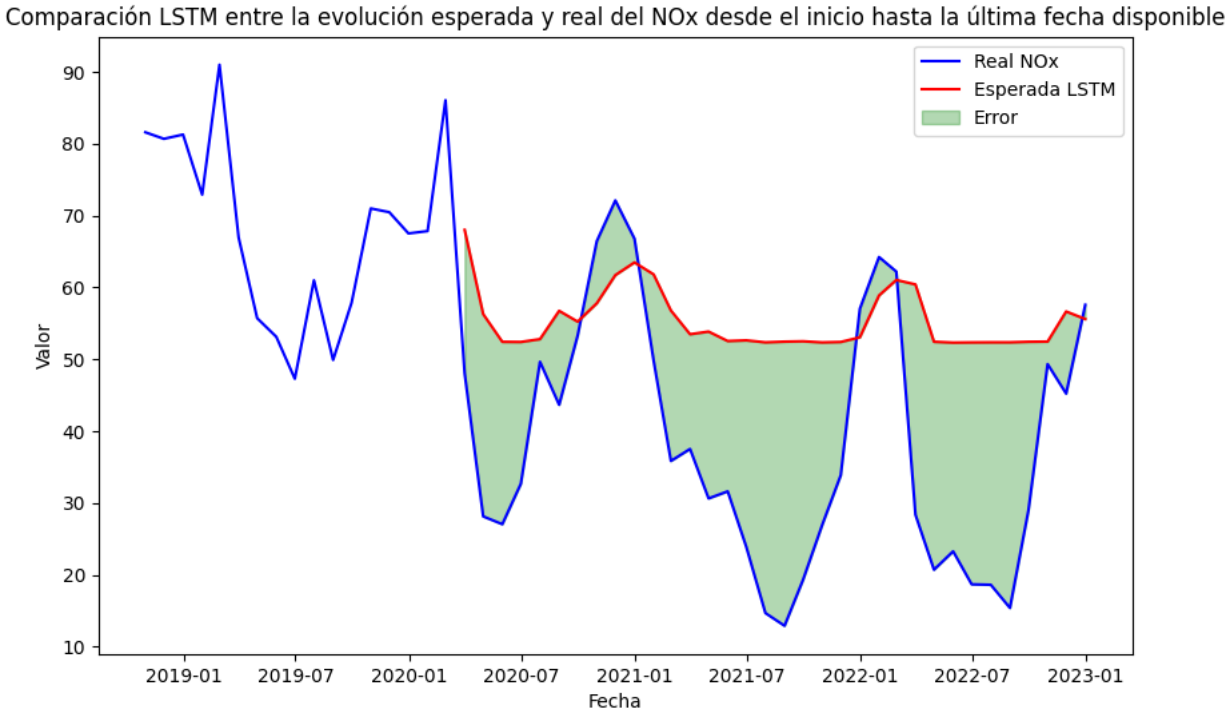


Figura 11: LSTM.

Las métricas de la Tabla 33, proporcionadas por el modelo LSTM, ofrecen una evaluación cuantitativa del rendimiento de la predicción de la evolución esperada del NOx en el periodo que va desde marzo de 2020 hasta finales de 2022. Estas métricas nos dan una idea de qué tan

Cuadro 33: Métricas de error del modelo LSTM para la predicción de NOx

Métrica	**Valor**
MAE	18.946014069495295
MSE	492.65773008459706
RMSE	22.1958944420944
MAPE	80.26171422707434

cerca están nuestras predicciones de los valores reales observados.

El Error Absoluto Medio (MAE) de aproximadamente 18.95 indica que, en promedio, nuestras predicciones difieren de los valores reales en alrededor de 18.95 unidades. Esta medida es útil para comprender la magnitud del error promedio en nuestras predicciones.

El Error Cuadrático Medio (MSE) de alrededor de 492.66 nos proporciona una medida del promedio de los errores al cuadrado. Esto significa que, en promedio, nuestros errores son de aproximadamente 492.66 unidades al cuadrado. El RMSE, que es la raíz cuadrada del MSE y tiene un valor de aproximadamente 22.20, nos dice que el error típico en nuestras predicciones es de alrededor de 22.20 unidades. Esta medida es especialmente útil porque está en la misma escala que la variable que estamos tratando de predecir, lo que facilita su interpretación.

Finalmente, el Error Porcentual Absoluto Medio (MAPE) de alrededor del 80.26 % nos indica que, en promedio, nuestras predicciones tienen un error del 80.26 % en relación con los valores reales. Esta medida es útil para comprender el error relativo en nuestras predicciones en comparación con los valores reales.

Es importante tener en cuenta que estos errores pueden atribuirse en gran medida a los cambios repentinos en el tráfico y las emisiones de NOx debido a las medidas de confinamiento relacionadas con la pandemia de COVID-19 y la posterior peatonalización de la Plaza del Ayuntamiento de Valencia. A pesar de estos errores, el análisis del error acumulado en adelante en comparación con una evolución estable anterior puede mostrar una ganancia ambiental significativa debido a las medidas implementadas para reducir las emisiones de NOx. La gráfica Fig. **??** estima con un modelo LSTM la evolución esperada del NOx a partir de marzo del 2020 en función de los meses previos. Vemos como el modelo comete un error negativo generalizado, pero entendemos esto como razonado debido a la nueva tendencia que toman los datos en el segundo periodo a partir del salto de marzo del 2020, donde se nota mucho el efecto de la reducción del NOx gracias a la reducción tráfico de la plaza.

10.6.2.2. Aplicación del modelo ARIMA

A continuación se recurrirá al modelo ARIMA para estimar nuevamente la evolución esperada del NOx. Este modelo ARIMA (Autoregressive Integrated Moving Average) es un enfoque estadístico utilizado para analizar y predecir series temporales. Es especialmente útil cuando los datos muestran patrones de tendencia y estacionalidad. Esta técnica se ha aplicado con éxito en diversos estudios, como el realizado por Liu et al. (2018) [Liu et al., 2018], donde se mejoró la predicción de contaminantes atmosféricos, incluyendo PM2.5, NO2 y O3, en Hong Kong utilizando modelos ARIMA con pronósticos numéricos. Además, en el estudio de Liu (2022) [Liu et al., 2022], se aborda la predicción del tráfico urbano a corto plazo utilizando redes neuronales bidireccionales de memoria a largo y corto plazo (BiLSTM), lo que destaca la aplicación de modelos de predicción en entornos urbanos complejos. Por otro lado, el artículo de Mani et al. (2022) [Mani et al., 2022] propone modelos de regresión y ARIMA para predecir el Índice de Calidad del Aire en Chennai, lo que resalta la relevancia de estas técnicas en la evaluación y gestión de la calidad del aire. Finalmente, el estudio de Mishra et al. (2024) [Mishra and Gupta, 2024] compara exhaustivamente diferentes algoritmos de aprendizaje, incluido ARIMA, para predecir el Índice de Calidad del Aire, destacando el potencial de modelos de aprendizaje profundo como LSTM sobre métodos clásicos.

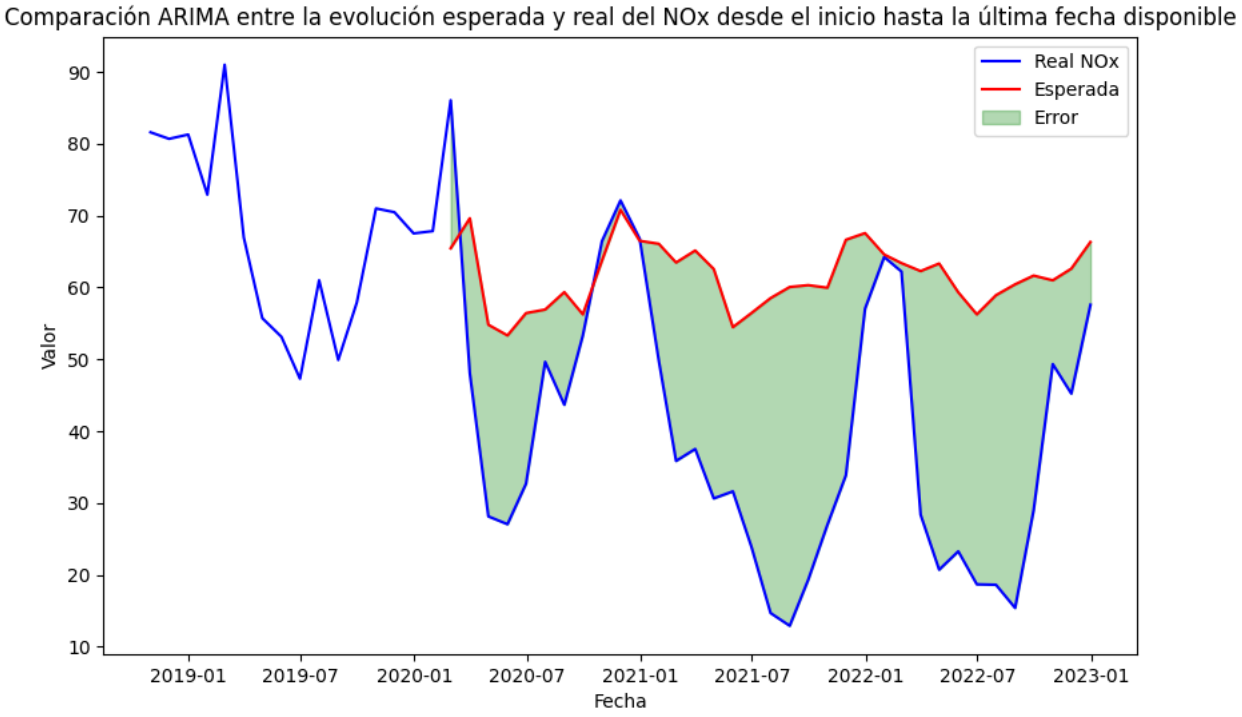


Figura 12: ARIMA.

Las métricas de la Tabla 34 proporcionadas por el modelo ARIMA también ofrecen una evaluación cuantitativa del rendimiento de la predicción de la evolución esperada del NOx en el período que va desde marzo de 2020 hasta finales de 2022.

El Error Absoluto Medio (MAE) de aproximadamente 23.54 indica que, en promedio, nuestras predicciones difieren de los valores reales en alrededor de 23.54 unidades. Este valor es

Cuadro 34: Métricas de error del modelo ARIMA para la predicción de NOx

Métrica	**Valor**
MAE	23.5362863159027
MSE	759.4667130238865
RMSE	27.558423630967837

ligeramente mayor que el obtenido con el modelo LSTM.

El Error Cuadrático Medio (MSE) de alrededor de 759.47 nos proporciona una medida del promedio de los errores al cuadrado. Esto significa que, en promedio, nuestros errores son de aproximadamente 759.47 unidades al cuadrado. Comparado con el MSE del modelo LSTM, vemos que el modelo ARIMA tiene un MSE más alto, lo que indica que los errores pueden ser más dispersos o tener valores atípicos.

El RMSE, que es la raíz cuadrada del MSE y tiene un valor de aproximadamente 27.56, nos dice que el error típico en nuestras predicciones es de alrededor de 27.56 unidades. Este valor también es ligeramente mayor que el obtenido con el modelo LSTM.

Finalmente, el Error Porcentual Absoluto Medio (MAPE) de alrededor del 80.26 % nos indica que, en promedio, nuestras predicciones tienen un error del 80.26 % en relación con los valores reales. Este valor es similar al obtenido con el modelo LSTM.

Al igual que con el modelo LSTM, es importante tener en cuenta que estos errores pueden atribuirse en gran medida a los cambios repentinos en el tráfico y las emisiones de NOx debido a las medidas de confinamiento relacionadas con la pandemia de COVID-19 y la posterior peatonalización de la Plaza del Ayuntamiento de Valencia. A pesar de estos errores, el análisis del error acumulado en adelante en comparación con una evolución estable anterior puede mostrar una ganancia ambiental significativa debido a las medidas implementadas para reducir las emisiones de NOx.

En la gráfica Fig. **??** se aprecia cómo evoluciona la reducción del NOx frente al valor esperado. Se aprecia una repetición anual, por la que en los meses más fríos de invierno, la mejora de la evolución del NOx real frente a la evolución esperada es menor. Podemos razonar esa repetición anual a que en los meses de invierno, aparte del NOx derivado del tráfico, también gana importancia el derivado de los sistemas de calefacción que no se han visto reducidos. Este patrón de reducción de NOx se ha observado en diferentes estudios, como el realizado por Waqas et al. (2024) [Waqas et al., 2024], donde se evaluó el impacto de las medidas de restricción socioeconómica en los Estados Unidos y se identificaron patrones estacionales en la reducción de NO2. Además, el estudio de Du (2023) [Du et al., 2023] analiza las variaciones espacio-temporales de la calidad del aire en China y desarrolla modelos predictivos para predecir la calidad del aire urbano, lo que proporciona información valiosa para comprender la dinámica de la contaminación atmosférica en diferentes regiones.

10.6.2.3. Aplicación del modelo SARIMA

A continuación se aplicará el modelo **SARIMA** (Seasonal Autoregressive Integrated Moving Average) que es una extensión del modelo **ARIMA** que también tiene en cuenta la estacionalidad de los dato, pero comparado con el modelo **ARIMA**, el modelo **SARIMA** tiene la ventaja de poder manejar series temporales con patrones estacionales. Esto significa que es más adecuado para datos que exhiben ciclos y patrones recurrentes a lo largo del tiempo, como es común en muchos fenómenos naturales y económicos.

En el caso de la evolución de la progresión de la media mensual de la contaminación ambiental de Valencia, donde es probable que haya patrones estacionales (por ejemplo, estacionalidad relacionada con las estaciones del año o ciertos eventos anuales), el modelo **SARIMA** podría capturar mejor estas variaciones estacionales en comparación con el modelo **ARIMA** estándar.

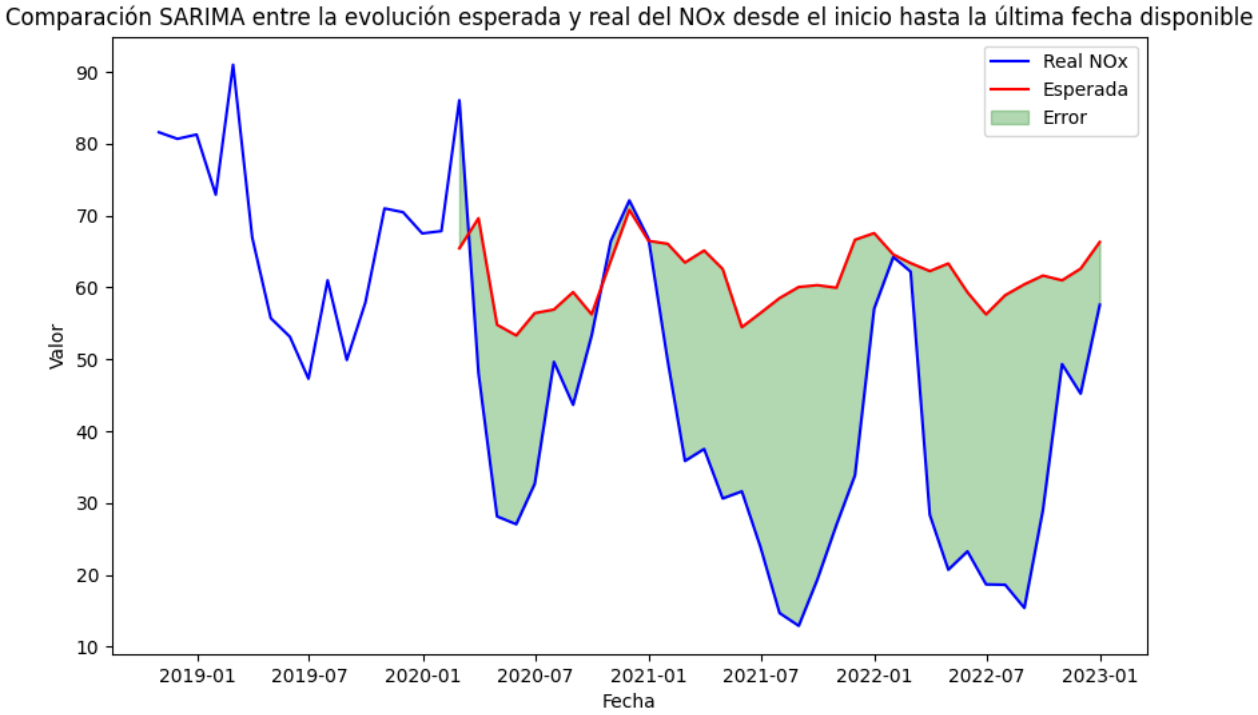


Figura 13: SARIMA.

Cuadro 35: Métricas de error SARIMA

Métrica	**Valor**
MAE	23.536
MSE	759.467
RMSE	27.558

Las métricas de la Tabla 35, obtenidas del modelo SARIMA proporcionan una evaluación adicional sobre la precisión de la predicción de la evolución esperada del NOx desde marzo de 2020 hasta finales de 2022.

El Error Absoluto Medio (MAE) de aproximadamente 23.54 indica que, en promedio, nuestras predicciones difieren de los valores reales en alrededor de 23.54 unidades. Este valor es similar al obtenido con el modelo ARIMA.

El Error Cuadrático Medio (MSE) de alrededor de 759.47 nos proporciona una medida del promedio de los errores al cuadrado. Esto significa que, en promedio, nuestros errores son de aproximadamente 759.47 unidades al cuadrado. Este valor también es similar al obtenido con el modelo ARIMA.

El RMSE, que es la raíz cuadrada del MSE y tiene un valor de aproximadamente 27.56, nos dice que el error típico en nuestras predicciones es de alrededor de 27.56 unidades. Al igual que con el MAE y el MSE, este valor es similar al obtenido con el modelo ARIMA.

Al igual que con los modelos anteriores, es importante tener en cuenta que estos errores pueden atribuirse en gran medida a los cambios repentinos en el tráfico y las emisiones de NOx debido a las medidas de confinamiento relacionadas con la pandemia de COVID-19 y la posterior peatonalización de la Plaza del Ayuntamiento de Valencia. A pesar de estos errores, el análisis del error acumulado en adelante en comparación con una evolución estable anterior puede mostrar una ganancia ambiental significativa debido a las medidas implementadas para reducir las emisiones de NOx.

En este último caso, en la gráfica Fig. **??** vemos cómo el modelo SARIMA sí que es capaz de desarrollar una estimación del NOx adaptándose a su repetibilidad anual, razonando que los meses de más frío sería coherente esperar de nuevo valores superiores de NOx que el resto del año. Esta capacidad de los modelos SARIMA para capturar patrones estacionales ha sido destacada en estudios previos, como el realizado por Waqas et al. (2024) [Waqas et al., 2024], donde se emplearon modelos SARIMAX y LSTM para pronosticar las concentraciones de NO2 bajo diferentes escenarios socioeconómicos. Asimismo, el estudio de Du (2023) [Du et al., 2023] analiza las variaciones espacio-temporales de la calidad del aire urbano en China y utiliza modelos SARIMA para predecir el índice de calidad del aire, lo que resalta la utilidad de estos modelos en la predicción de la calidad del aire.

11. Conclusiones

Con la implementación de los modelos LSTM, ARIMA y SARIMA, hemos obtenido valiosa información sobre la evolución esperada del NOx en la ciudad de Valencia, especialmente en el contexto marcado por eventos inesperados como la pandemia de COVID-19 y las medidas de confinamiento asociadas [Rodríguez-García et al., 2023].

Nuestro análisis confirma que el viento de origen costero no incrementa los índices de contaminación, al menos en lo que respecta al NOx, lo que nos permite simplificar nuestra estimación de este índice.

Al estudiar las variables en los modelos de regresión, observamos que la mayor parte de la influencia sobre el NOx proviene fundamentalmente de 2 (el tráfico total de la plaza y la dirección del viento) o hasta de 4 variables (incluyendo también la radiación solar y la presión atmosférica).

Los modelos de regresión, al incorporar más variables predictoras, mostraron una mejora marginal en su eficiencia, medida a través del coeficiente de determinación R-cuadrado y el R-cuadrado ajustado. Sin embargo, esta mejora es mínima en comparación con el aumento de la complejidad del modelo. Notamos que a partir de la inclusión de cuatro variables predictoras, la eficiencia del modelo apenas mejora, lo que sugiere que el modelo podría estar perdiendo capacidad de generalización. Por lo tanto, concluimos que es crucial encontrar un equilibrio entre la complejidad del modelo y su capacidad para generalizar los datos.

La aplicación de técnicas de regresión como Lasso y Ridge proporcionó soluciones efectivas para abordar problemas como la multicolinealidad y la selección de características. Limitar el número de variables predictoras a cuatro ofreció un equilibrio óptimo entre la capacidad explicativa del modelo y su capacidad de generalización. Este enfoque destaca la importancia de encontrar un equilibrio adecuado al abordar problemas de regresión en contextos donde se necesita predecir la contaminación del aire con precisión para informar políticas de calidad ambiental.

Aunque los modelos acumularon un cierto grado de error en sus predicciones, nuestro trabajo proporciona una base sólida para comprender las tendencias generales y los impactos significativos en la calidad del aire. La reducción del 41.35 % en el índice de NOx durante el periodo de confinamiento subraya la importancia de considerar el tráfico como una de las principales fuentes de contaminación atmosférica en entornos urbanos.

Además, este estudio destaca la necesidad de abordar de forma conjunta los factores ambientales y socioeconómicos que influyen en la calidad del aire urbano. Si bien los modelos predictivos son herramientas poderosas, es esencial complementar su análisis con una comprensión más amplia de los contextos locales y las dinámicas socioeconómicas que pueden influir en las emisiones de contaminantes atmosféricos [Che et al., 2023].

En resumen, nuestro trabajo ofrece una contribución significativa al entendimiento de la relación entre el tráfico, la calidad del aire y los eventos inesperados como la pandemia de COVID-19 [Varia and Kothari, 2022]. Aunque existen limitaciones en las predicciones de los modelos, estos resultados proporcionan una base sólida para futuras investigaciones y la implementación de medidas efectivas para mejorar la calidad del aire urbano.

Bibliografía

Saba Ameer, Munam Ali Shah, Abid Khan, Houbing Song, Carsten Maple, Saif Ul Islam, and Muhammad Nabeel Asghar. Comparative analysis of machine learning techniques for predicting air quality in smart cities. *IEEE Access*, 7:128325 – 128338, 2019. doi: 10.1109/ACCESS.2019.2925082. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85074672630&doi=10.1109%2fACCESS.2019.2925082&partnerID=40&md5=17efd27e423e96db7048785afb89a190. Cited by: 118; All Open Access, Gold Open Access.

Wenwei Che, Yumiao Zhang, Changqing Lin, Yik Him Fung, Jimmy C.H. Fung, and Alexis K.H. Lau. Impacts of pollution heterogeneity on population exposure in dense urban areas using ultra-fine resolution air quality data. *Journal of Environmental Sciences (China)*, 125:513 – 523, 2023. doi: 10.1016/j.jes.2022.02.041. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126845153&doi=10.1016%2fj.jes.2022.02.041&partnerID=40&md5=e1663aa265ec2dc2a1689dbfa8bf32e6. Cited by: 6.

Arko Datta, Aniruddha Pal, Ranbir Marandi, Nilanjan Chattaraj, Subrata Nandi, and Sujoy Saha. Real-time air quality predictions for smart cities using tinyml. In *Real-Time Air Quality Predictions for Smart Cities using TinyML*, page 246 – 247, 2024. doi: 10.1145/3631461.3631947. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85184281299&doi=10.1145%2f3631461.3631947&partnerID=40&md5=9ea8da1003a6458e2a66d92cfca6ee6f. Cited by: 0.

Yuanfang Du, Shibing You, Weisheng Liu, Tsering-xiao Basang, and Miao Zhang. Spatiotemporal evolution characteristics and prediction analysis of urban air quality in china. *Scientific Reports*, 13(1), 2023. doi: 10.1038/s41598-023-36086-4. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85160936676&doi=10.1038%2fs41598-023-36086-4&partnerID=40&md5=8a97793a2aaa9d05b5cbfc6c655ac494. Cited by: 0; All Open Access, Gold Open Access, Green Open Access.

Irene Lebrusán and Jamal Toutouh. Car restriction policies for better urban health: a low emission zone in madrid, spain. *Air Quality, Atmosphere & Health*, 14:333–342, 2021.

Shuo Liu, Zhengmin Kong, Tao Huang, Yang Du, and Wei Xiang. An admm-lstm framework for short-term load forecasting. *Neural Networks*, 173, 2024. doi: 10.1016/j.neunet.2024.106150. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85184656620&doi=10.1016%2fj.neunet.2024.106150&partnerID=40&md5=8bc4306a7709a61bf91b1bee7dc13d62.

Silin Liu, Zhuhua Liao, Yizhi Liu, and Aiping Yi. Short-term speed prediction of urban roads based on multi-source feature fusion. In *Short-term speed prediction of urban roads based on multi-source feature fusion*, page 1602 – 1608, 2022. doi: 10.1109/HPCC-DSS-SmartCity-DependSys53884.2021.00237. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85132438239&doi=10.1109%2fHPCC-DSS-SmartCity-DependSys53884.2021.00237&partnerID=40&md5=e60d7268744da6c4952166daf5e37b3d. Cited by: 0.

Tong Liu, Alexis KH Lau, Kai Sandbrink, and Jimmy CH Fung. Time series forecasting of air quality based on regional numerical modeling in hong kong. *Journal of Geophysical Research: Atmospheres*, 123(8):4175–4196, 2018.

Harshini Macherla, Ghamya Kotapati, Manepalli Tulasi Sunitha, Koteswara Rao Chittipireddy, Balaji Attuluri, and Ramesh Vatambeti. Deep learning framework-based chaotic hunger games search optimization algorithm for prediction of air quality index. *Ingenierie des Systemes d'Information*, 28(2):433 – 441, 2023. doi: 10.18280/isi.280219. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85162108532&doi=10.18280%2fisi.280219&partnerID=40&md5=79cd46e7ab6c3c2b7c9c90f4d6e5550e. Cited by: 3; All Open Access, Bronze Open Access.

Geetha Mani, Joshi Kumar Viswanadhapalli, et al. Prediction and forecasting of air quality index in chennai using regression and arima time series models. *Journal of Engineering Research*, 10(2A):179–194, 2022.

Ankita Mishra and Yogesh Gupta. Comparative analysis of air quality index prediction using deep learning algorithms. *Spatial Information Research*, 32(1):63 – 72, 2024. doi: 10.1007/s41324-023-00541-1. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85165254755&doi=10.1007%2fs41324-023-00541-1&partnerID=40&md5=94f026ce61a3b7a16736a51801bf6eae. Cited by: 1.

Hamidreza Moradi, Amirreza Talaiekhozani, Hesam Kamyab, Shreeshivadasan Chelliapan, and Ashok Kumar Nadda. Development of equations to predict the concentration of air pollutants indicators in yazd city, iran. *Journal of Inorganic and Organometallic Polymers and Materials*, 34(1):38 – 47, 2024. doi: 10.1007/s10904-022-02416-8. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85133450063&doi=10.1007%2fs10904-022-02416-8&partnerID=40&md5=7b6a60a9b73ea86ccecca187742bf7a3. Cited by: 13.

Junhyeok Park, Youngsuk Seo, and Jaehyuk Cho. Unsupervised outlier detection for time-series data of indoor air quality using lstm autoencoder with ensemble method. *Journal of Big Data*, 10(1), 2023. doi: 10.1186/s40537-023-00746-z. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85159942956&doi=10.1186%2fs40537-023-00746-z&partnerID=40&md5=32eaeab7dac4aae5772a8eac183c151b. Cited by: 2; All Open Access, Gold Open Access.

Jinwook Rhyu, Dragana Bozinovski, Alexis B. Dubs, Naresh Mohan, Elizabeth M. Cummings Bende, Andrew J. Maloney, Miriam Nieves, Jose Sangerman, Amos E. Lu, Moo Sun

Hong, Anastasia Artamonova, Rui Wen Ou, Paul W. Barone, James C. Leung, Jacqueline M. Wolfrum, Anthony J. Sinskey, Stacy L. Springs, and Richard D. Braatz. Automated outlier detection and estimation of missing data. *Computers and Chemical Engineering*, 180, 2024. doi: 10.1016/j.compchemeng.2023.108448. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85175018267&doi=10.1016%2fj.compchemeng.2023.108448&partnerID=40&md5=7ac75dc073debaf6119bfb7ddaffdfb5. Cited by: 0; All Open Access, Bronze Open Access.

María Inmaculada Rodríguez-García, María Gema Carrasco-García, Javier González-Enrique, Juan Jesús Ruiz-Aguilar, and Ignacio J Turias. Long short-term memory approach for short-term air quality forecasting in the bay of algeciras (spain). *Sustainability*, 15(6):5089, 2023.

Jose Manuel Sanchez, Emilio Ortega, Maria Eugenia Lopez-Lambas, and Belen Martin. Evaluation of emissions in traffic reduction and pedestrianization scenarios in madrid. *Transportation research part D: transport and environment*, 100:103064, 2021.

M.V.V.S. Subrahmanyam, P.V.V.S.D. Nagendrudu, and T.V. Ramana. Comparison of effective machine learning technique for air quality forecast. *Cognitive Science and Technology*, Part F1493:157 – 164, 2023. doi: 10.1007/978-981-99-2746-3_16. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85174456040&doi=10.1007%2f978-981-99-2746-3_16&partnerID=40&md5=84767e27671ff31764b82565a359c317. Cited by: 0.

A. Selim Türkoğlu, Burcu Erkmen, Yavuz Eren, Ozan Erdinç, and İbrahim Küçükdemiral. Integrated approaches in resilient hierarchical load forecasting via tcn and optimal valley filling based demand response application. *Applied Energy*, 360, 2024. doi: 10.1016/j.apenergy.2024.122722. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85184766260&doi=10.1016%2fj.apenergy.2024.122722&partnerID=40&md5=046a98fb0f2ac385470e8d6674e3dd52.

D.J. Varia and A.M. Kothari. Comparative analysis of artificial neural network and long short-term memory techniques for predicting air quality in smart cities: Ahmadabad city. *Indian Journal of Environmental Protection*, 42(4):432 – 442, 2022. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85132935380&partnerID=40&md5=63f6bd6369c21a2ce21633b95fdddfe4. Cited by: 0.

Ka-Ming Wai and Peter K. N. Yu. Application of a machine learning method for prediction of urban neighborhood-scale air pollution. *International Journal of Environmental Research and Public Health*, 20(3), 2023. doi: 10.3390/ijerph20032412. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85147797890&doi=10.3390%2fijerph20032412&partnerID=40&md5=acf6481d4572e932fa99af429eaf18ba. Cited by: 0; All Open Access, Gold Open Access, Green Open Access.

Muhammad Waqas, Majid Nazeer, Man Sing Wong, Wu Shaolin, Li Hon, and Joon Heo. Impact of urban spatial factors on no2 concentration based on different socio-economic restriction scenarios in us cities. *Atmospheric Environment*, 316:120191, 2024.

Taeyeon Won, Yang Dam Eo, Hongki Sung, Kyu Soo Chong, Junhee Youn, and Gyeong Wook Lee. Particulate matter estimation from public weather data and closed-circuit television images. *KSCE Journal of Civil Engineering*, 26(2):865 – 873, 2022. doi: 10.1007/s12205-021-0865-4. URL `https://www.scopus.com/inward/record.uri?eid=2-s2.0-85118529578&doi=10.1007%2fs12205-021-0865-4&partnerID=40&md5=f013f2ea6fd7e8da9da12321837b0cba`. Cited by: 2.

12. Anexos

12.1. Anexos de figuras

12.1.1 Mapa de estaciones de medición de contaminantes en Valencia.

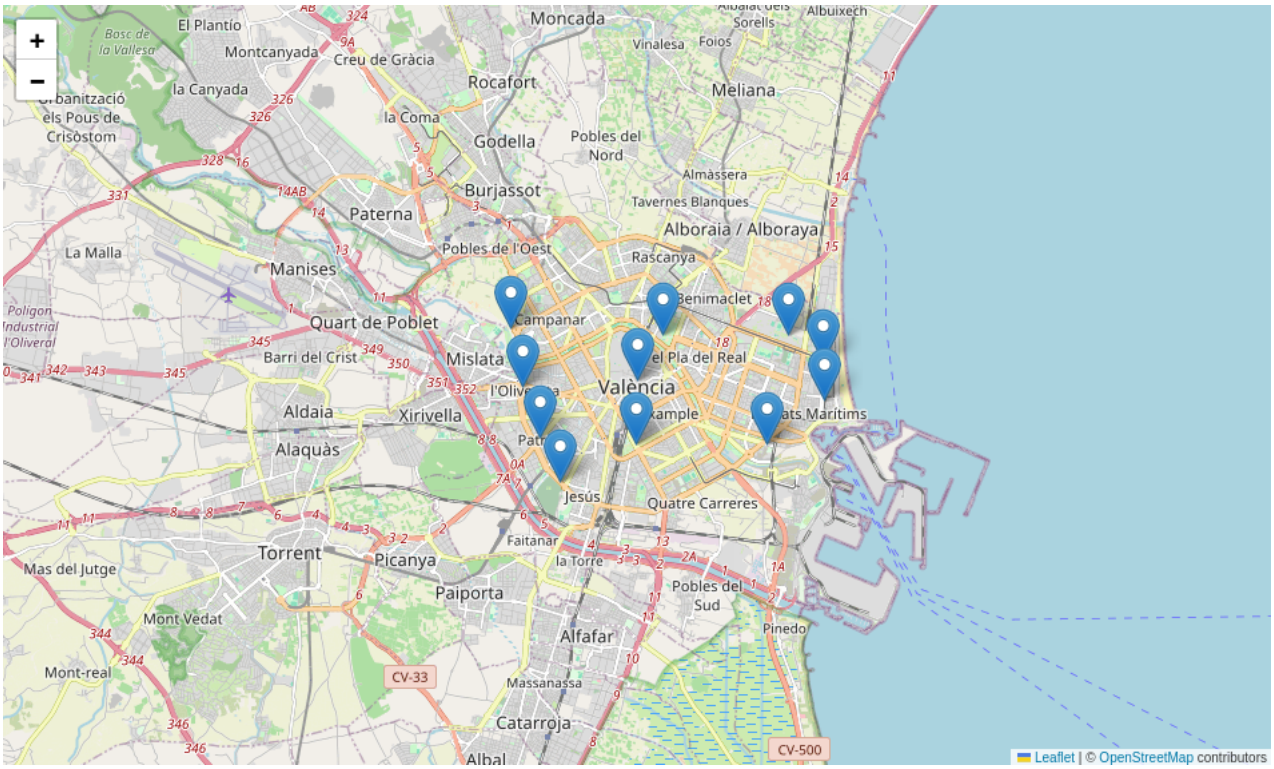


Figura 14: Ubicación de las estaciones de medición de contaminantes de Valencia.

12.1.2 Disponibilidad de datos.

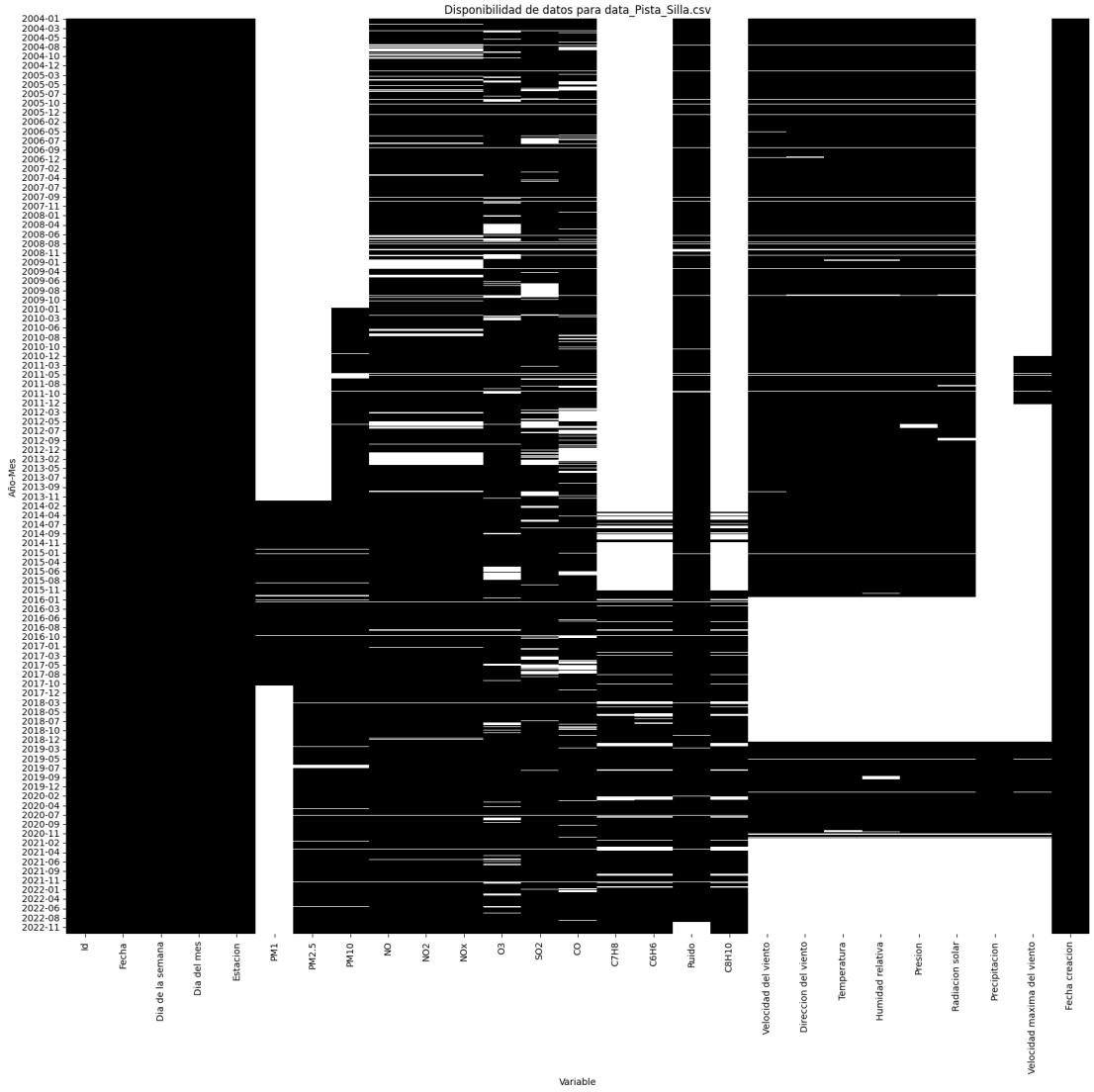


Figura 15: Disponibilidad de datos en la Pista de Silla.

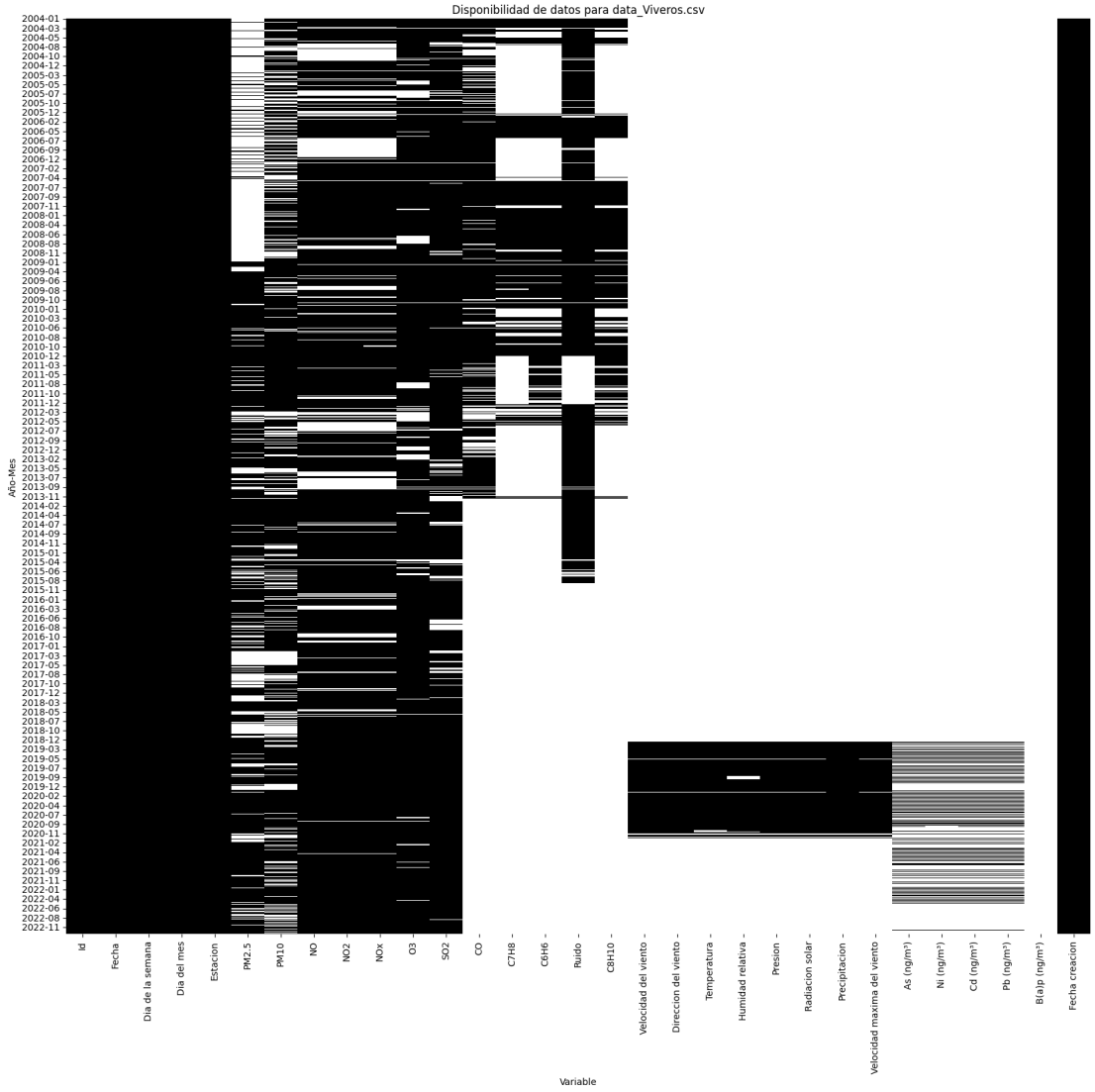


Figura 16: Disponibilidad de datos en Viveros.

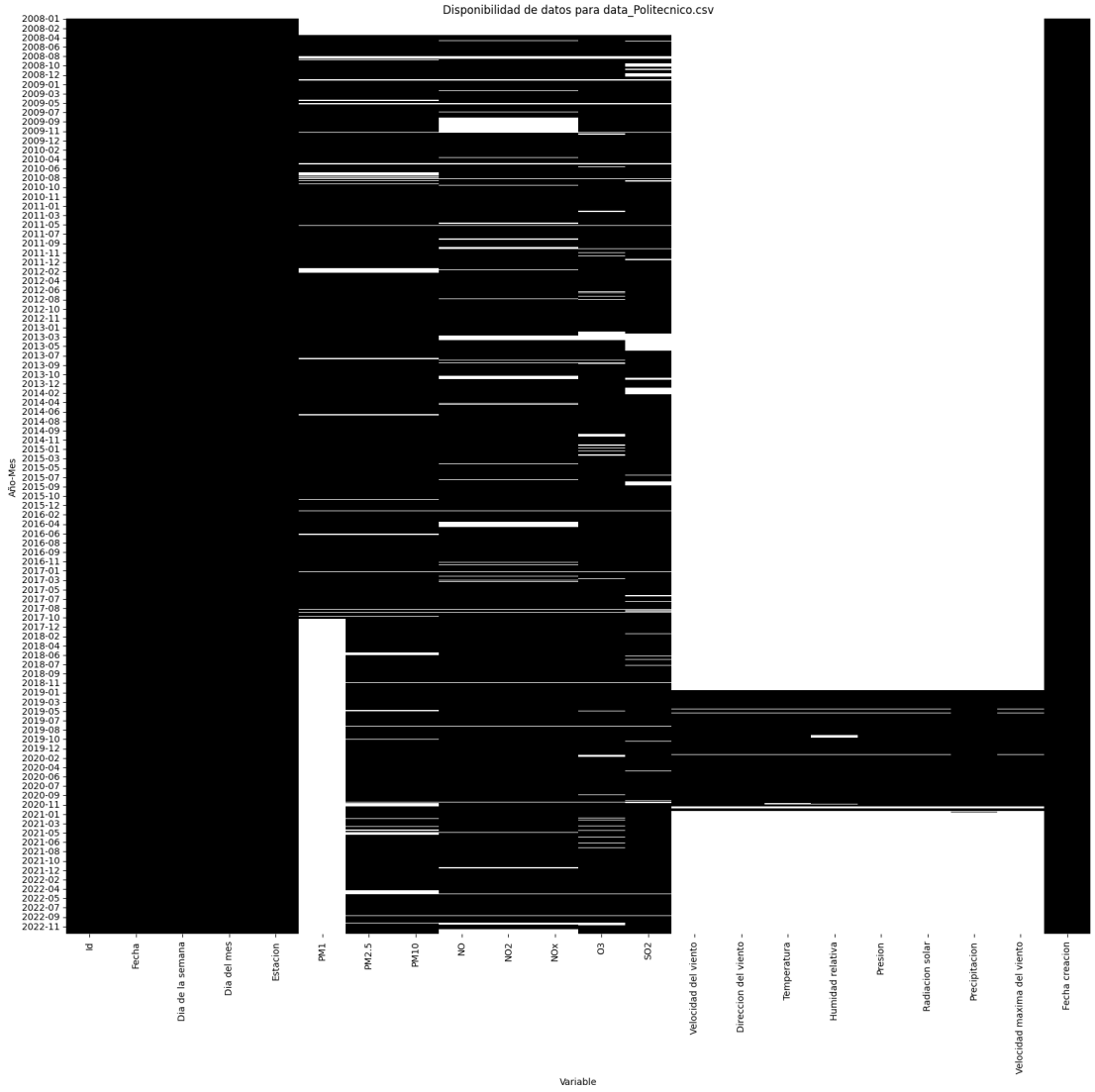


Figura 17: Disponibilidad de datos en Politécnico.

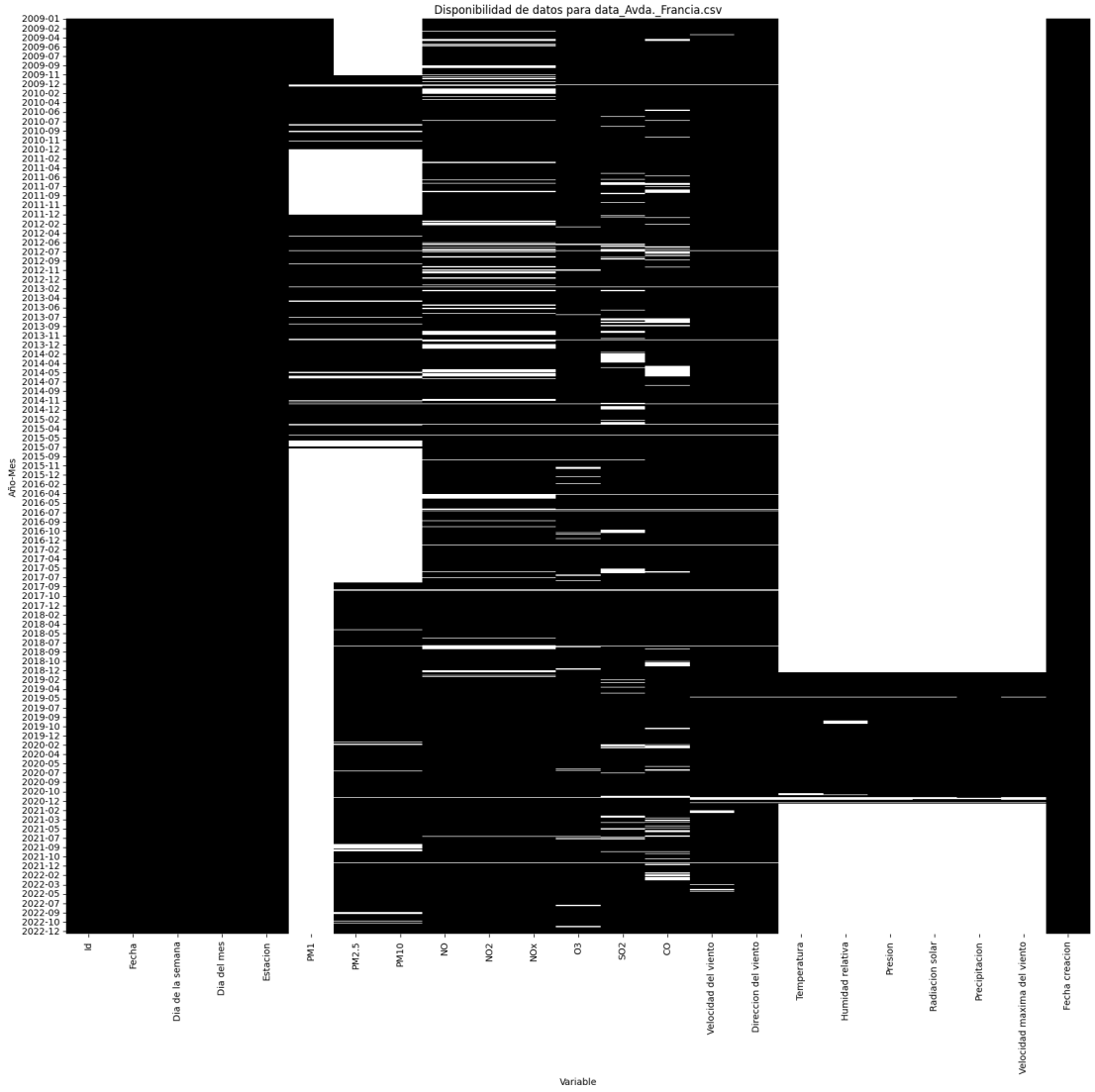


Figura 18: Disponibilidad de datos en Avda. Francia.

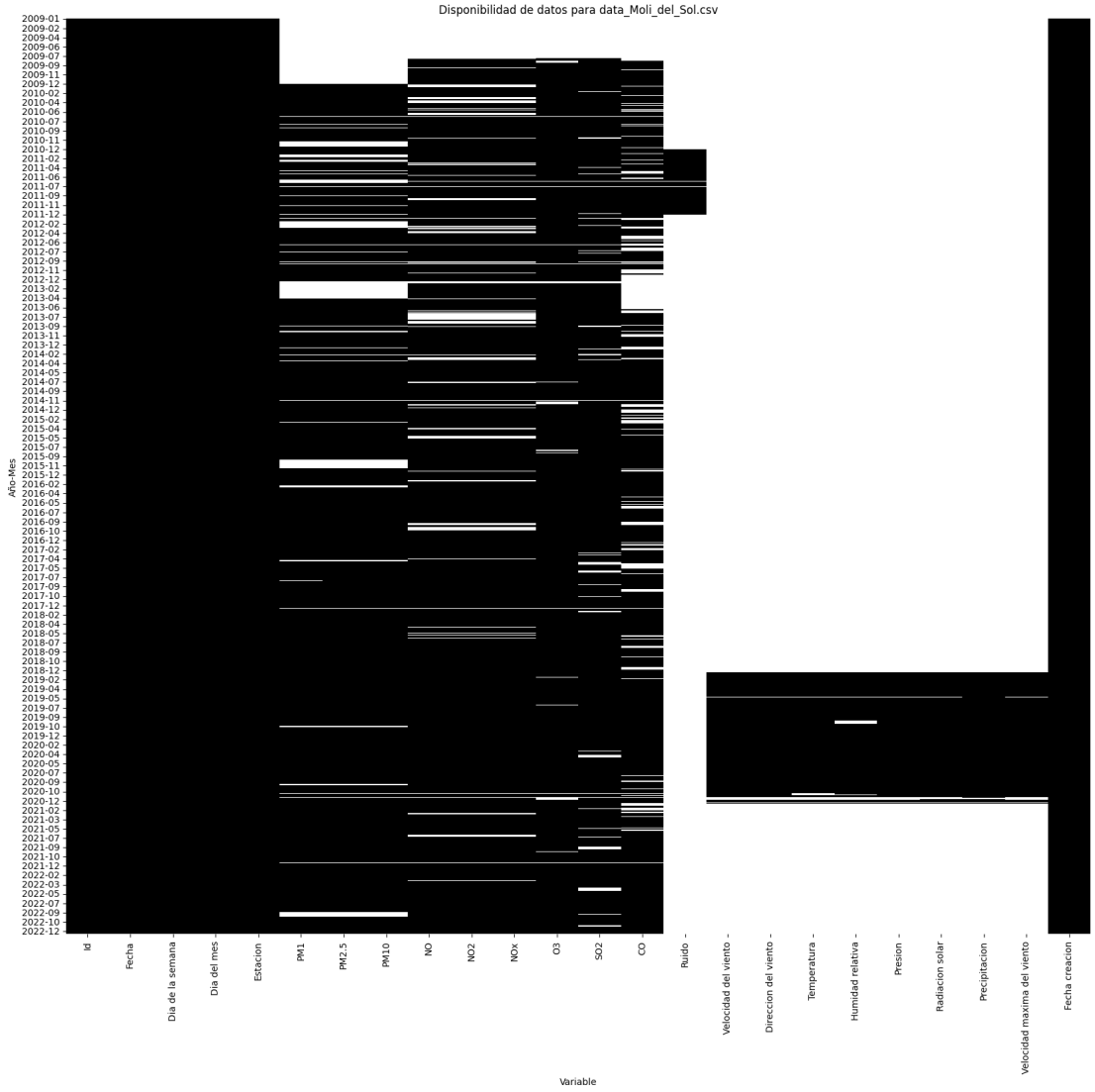


Figura 19: Disponibilidad de datos en Molí de Sol.

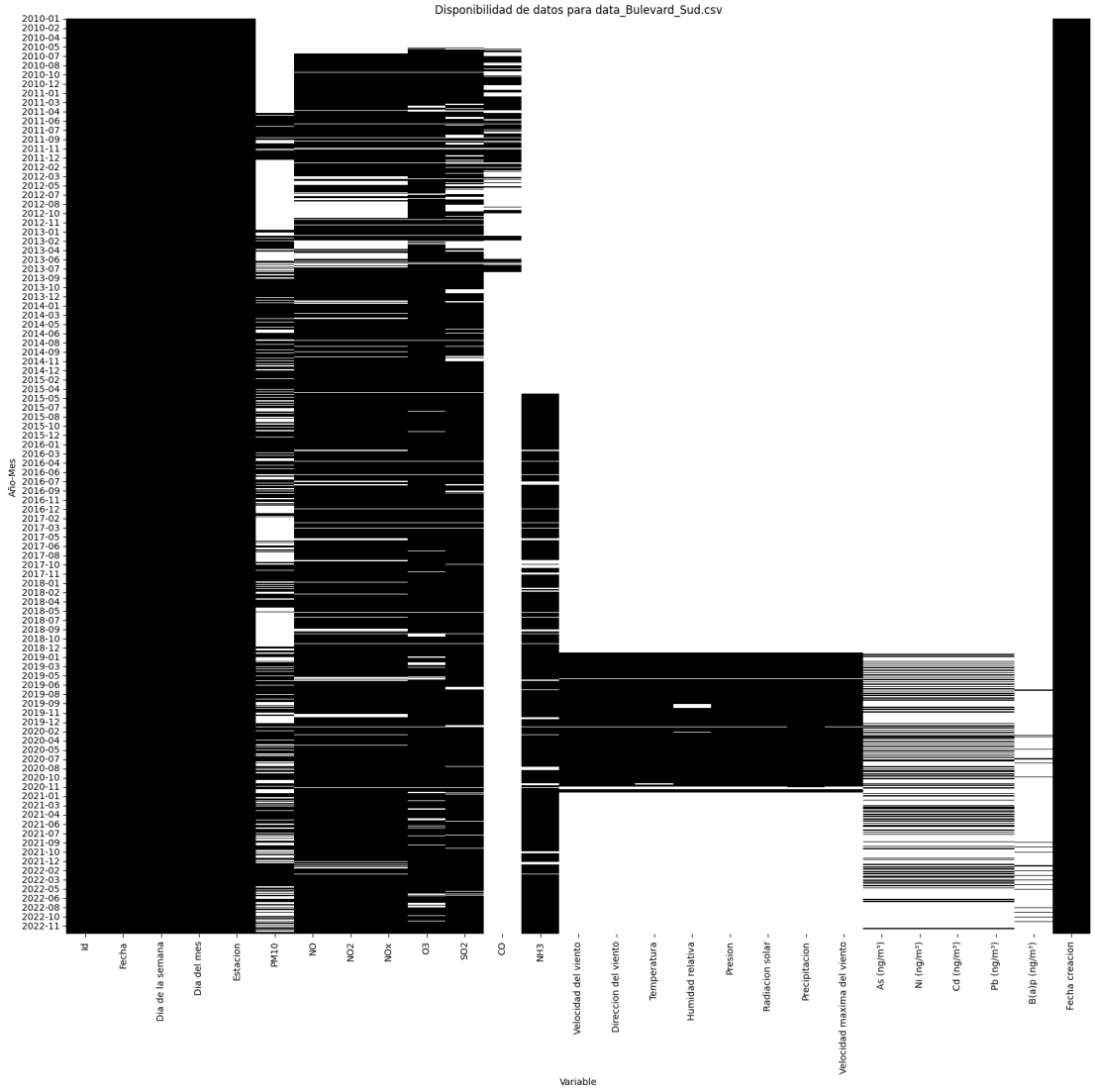


Figura 20: Disponibilidad de datos en Bulevard Sud.

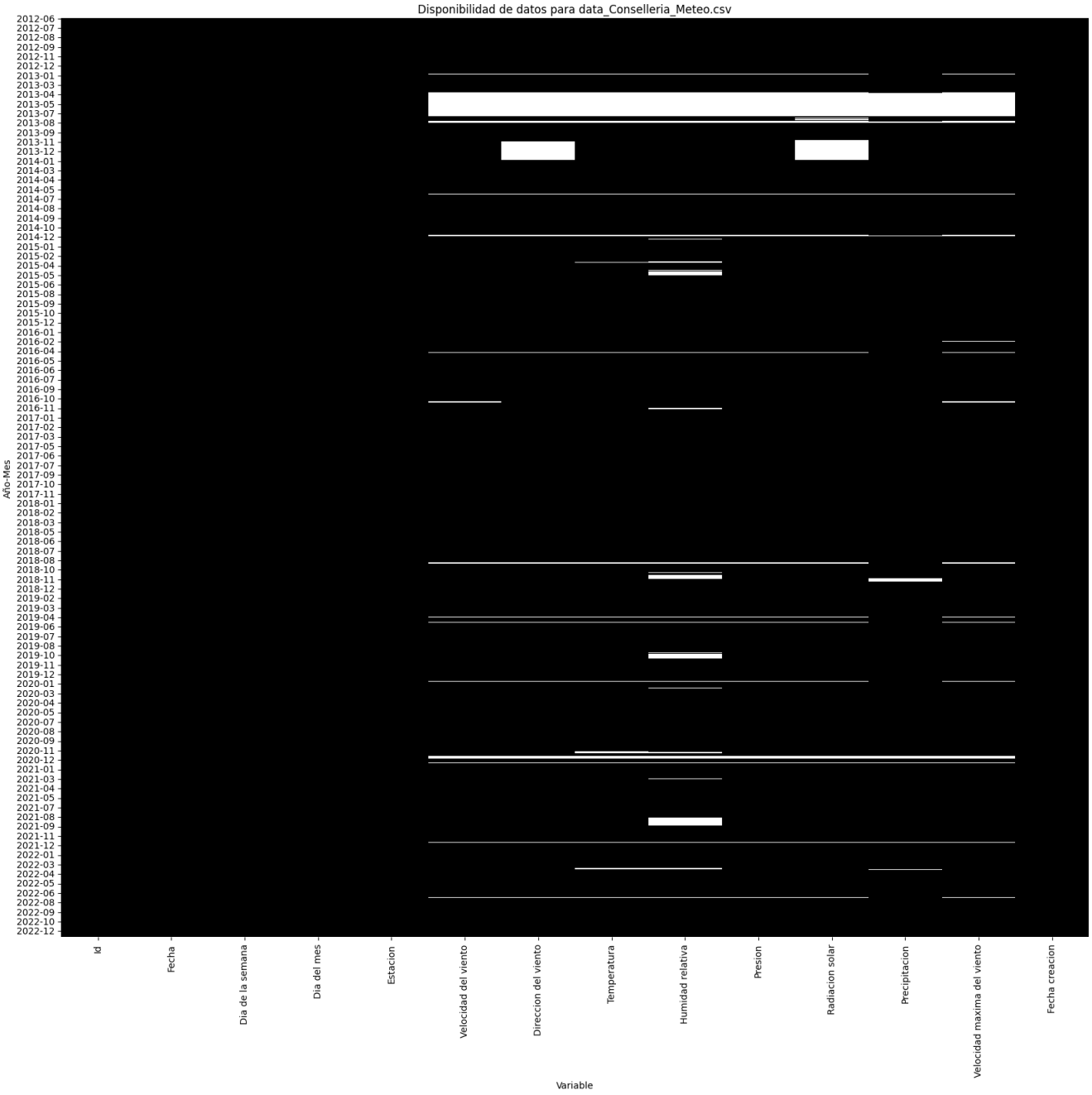


Figura 21: Disponibilidad de datos en Conselleria Meteo.

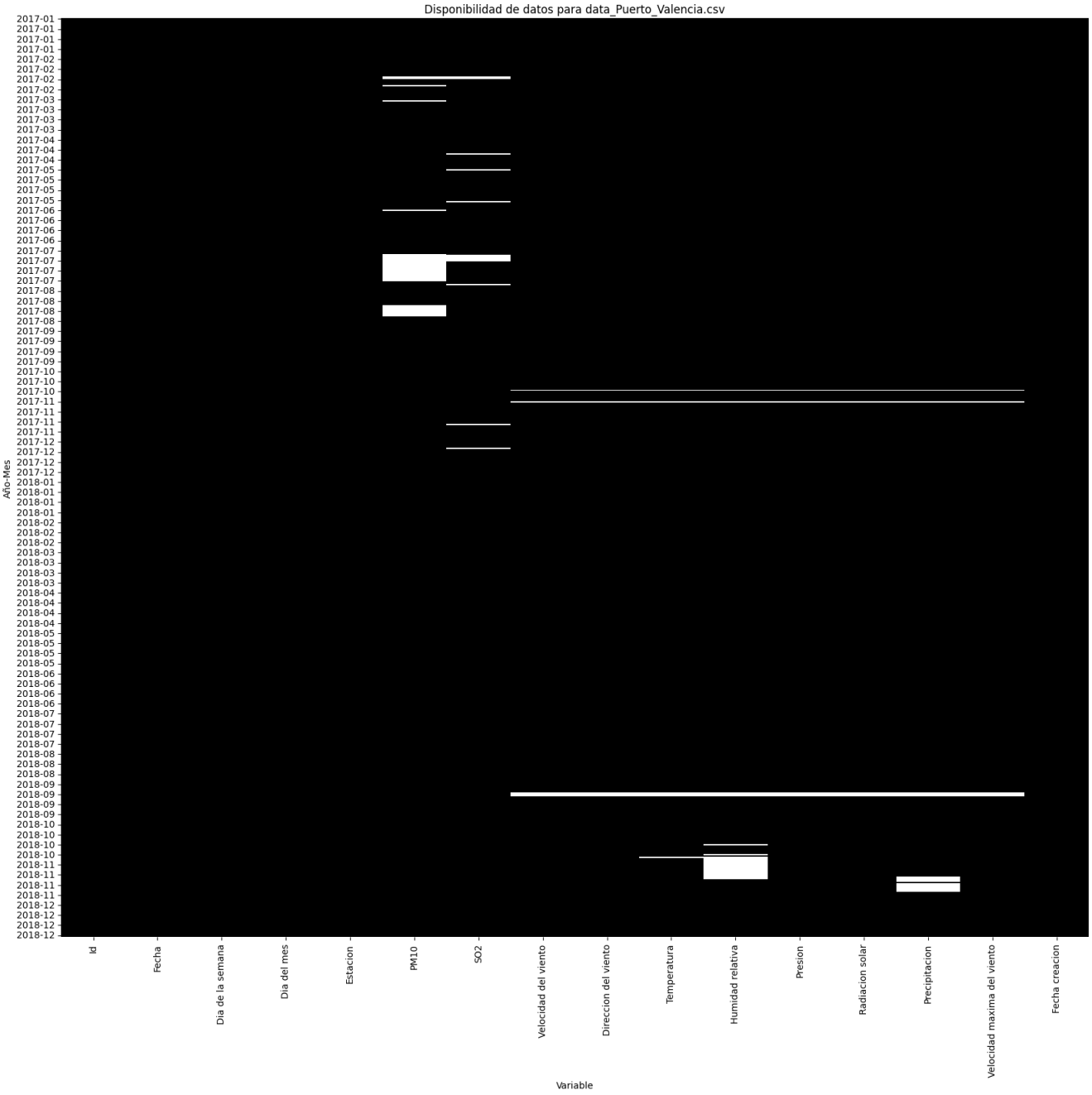


Figura 22: Disponibilidad de datos en Puerto Valencia.

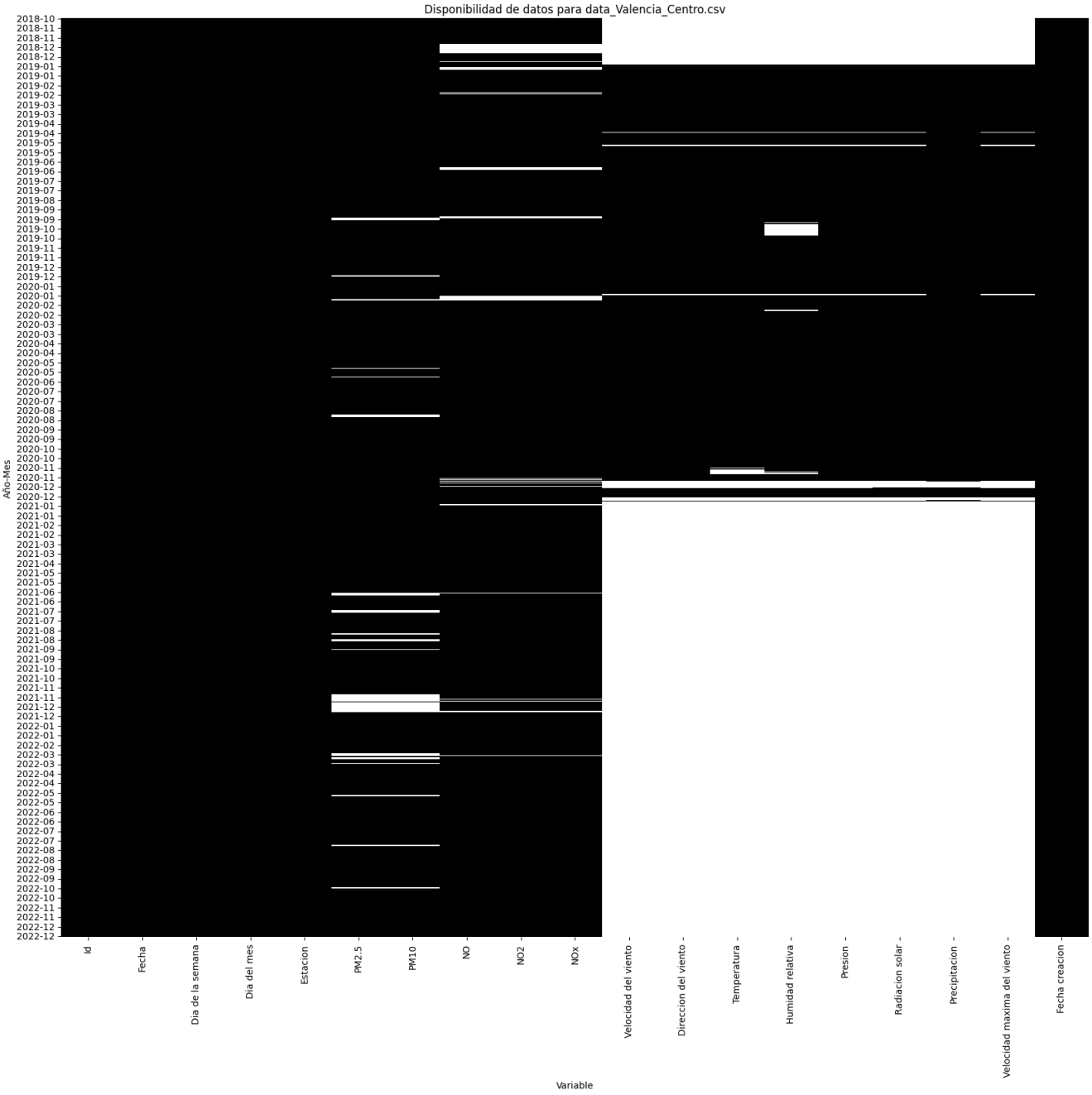


Figura 23: Disponibilidad de datos en Valencia Centro.

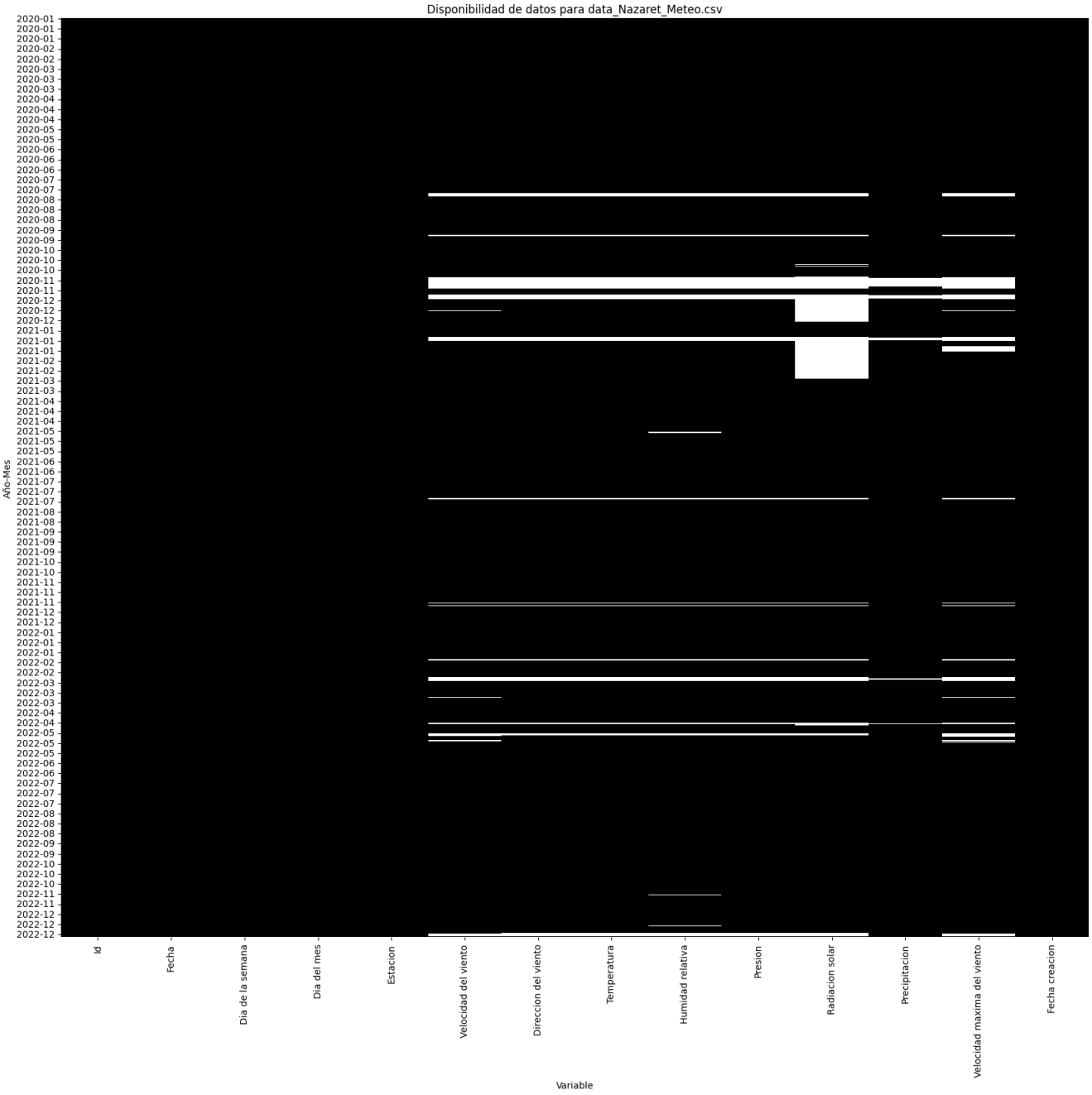


Figura 24: Disponibilidad de datos en Nazaret Meteo.

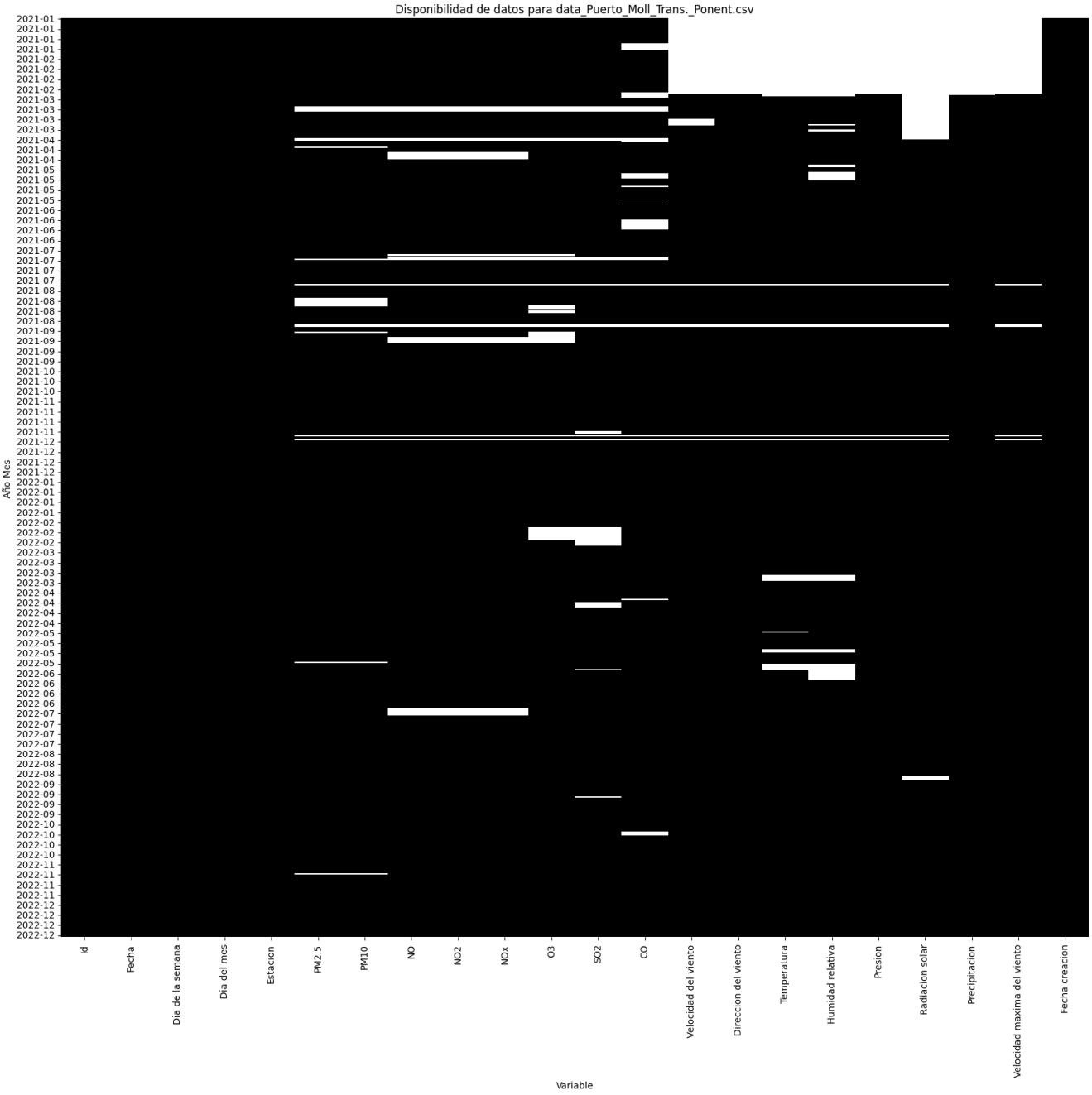


Figura 25: Disponibilidad de datos en Puerto Moll Trans. Ponent.

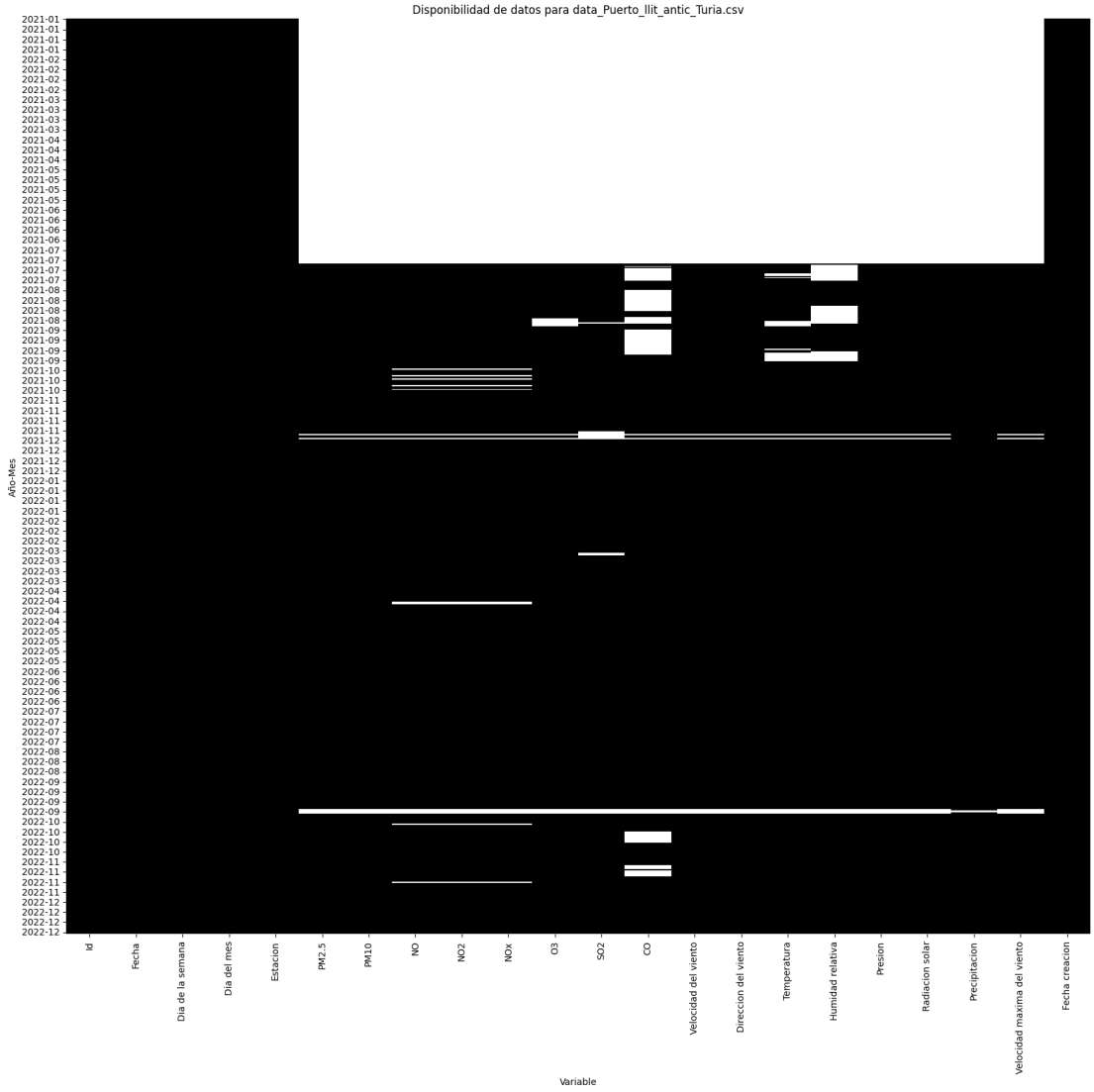


Figura 26: Disponibilidad de datos en Puerto llit antic Turia.

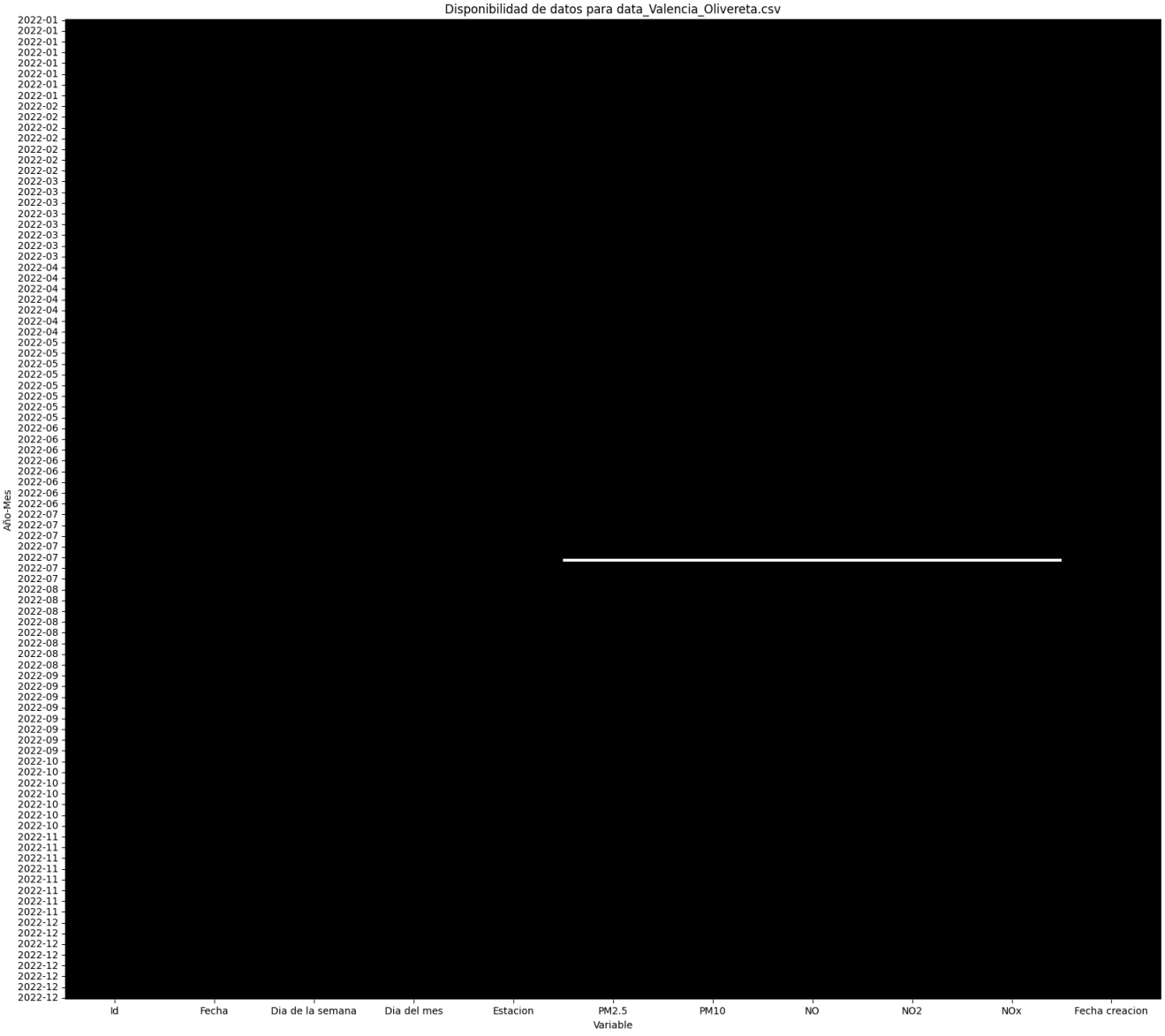


Figura 27: Disponibilidad de datos en Olivereta.

12.1.3 Acumulados de dirección de vientos y NOx.

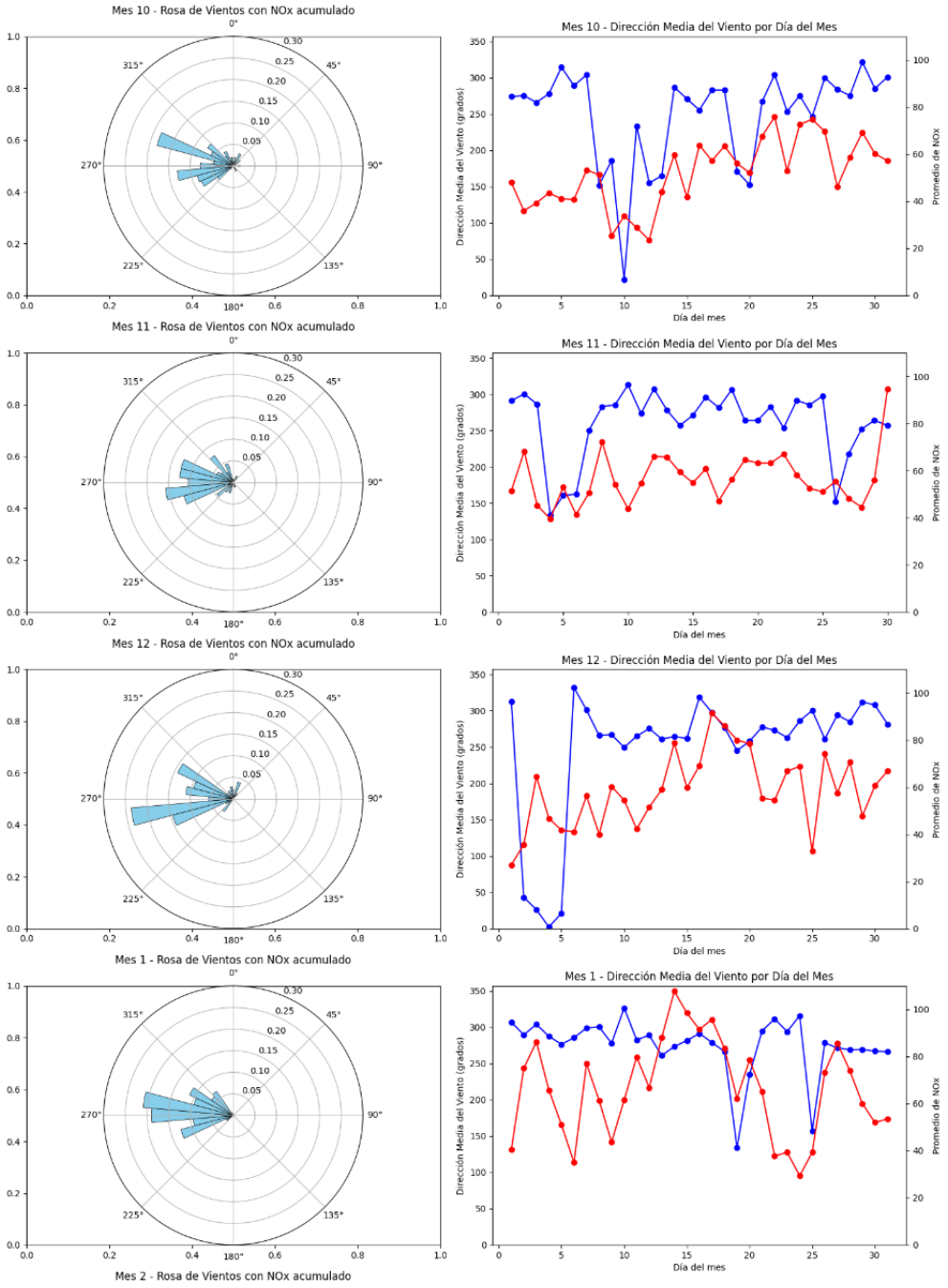


Figura 28: Datos de acumulados de vientos, a la izquierda
Cronograma con la evolución de la media semanal del NOx y vientos, a la derecha

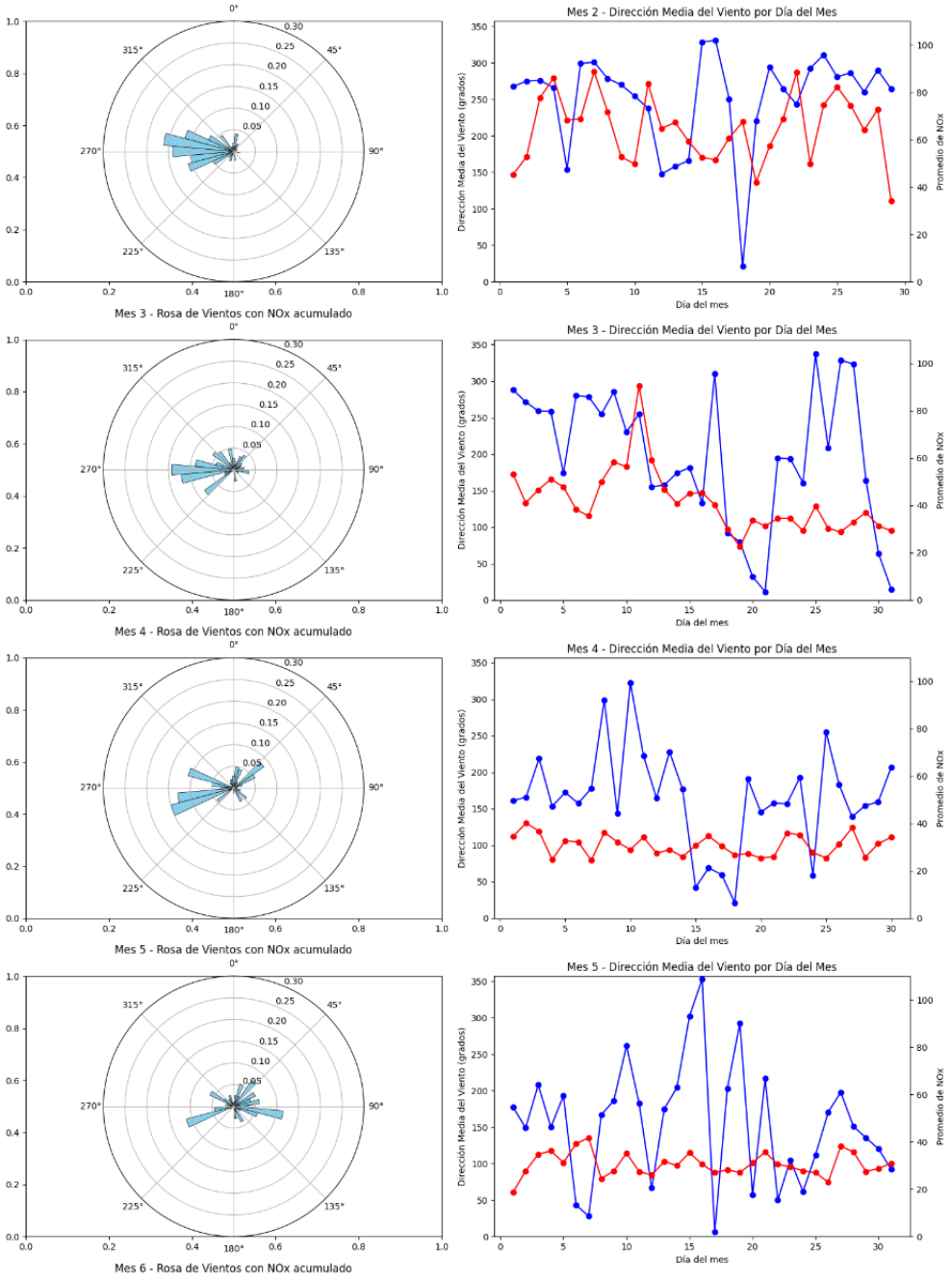


Figura 29: Datos de acumulados de vientos, a la izquierda
Cronograma con la evolución de la media semanal del NOx y vientos, a la derecha

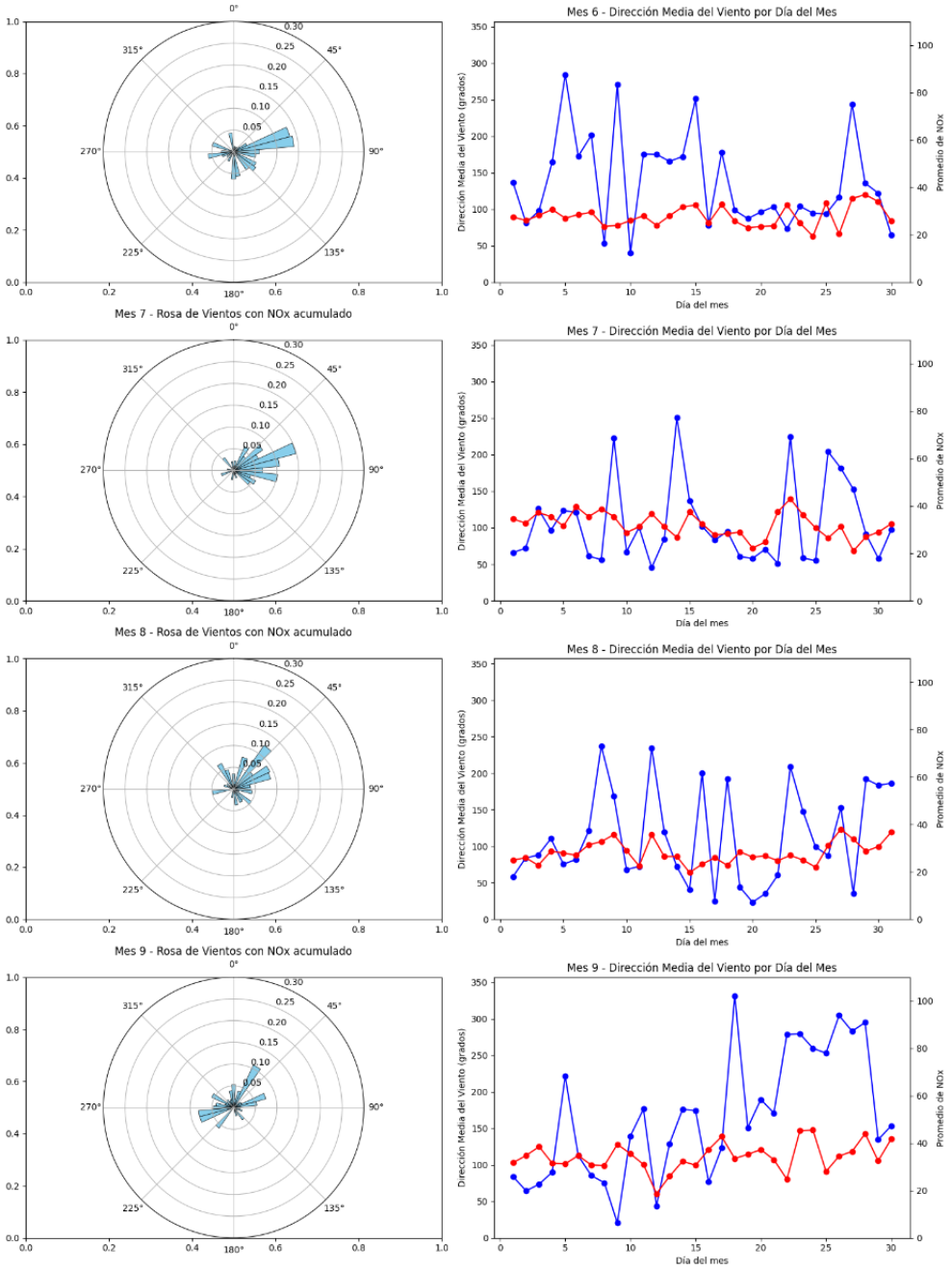


Figura 30: Datos de acumulados de vientos, a la izquierda
Cronograma con la evolución de la media semanal del NOx y vientos, a la derecha