



GRADO DE ESTADÍSTICA

— TRABAJO FIN DE ESTUDIOS —

*Modelos estadísticos
en análisis de
cambio climático*

Tutor: José Luis Pino Mejías

Alumno: Alberto Pérez García

Sevilla, Junio de 2023

Índice general

Prólogo	III
Resumen	V
Abstract	VI
Índice de Figuras	VII
Índice de Tablas	IX
1. Introducción	1
2. Cambio Climático	3
2.1. Debate	3
2.2. Actualidad	4
2.3. Política Europea	4
2.4. Cumbre del clima	5
3. Fuentes de datos	7
3.1. Global Carbon Project	7
3.2. Oficina Nacional de Administración Oceánica y Atmosférica	8
3.3. Organización Meteorológica Mundial	9
3.4. Agencia Estatal de Meteorología	9
3.5. Instituto Internacional de Derecho y Medio Ambiente	10
3.6. Sistema Integrado de Observación de Carbono	11
3.7. Copernicus	11
3.8. Grupo Intergubernamental de Expertos sobre el Cambio Climático	12
3.9. Centro Nacional de Datos de Hielo y Nieve	13
4. Ejemplo de modelos estadísticos	15
4.1. Meta-Análisis	15
4.2. Métodos basados en datos de panel	19
4.2.1. Enfoques de paneles transversales y lineales	21
4.2.1.1. Modelos de sección transversal	21
4.2.1.2. Modelos de panel lineal	21
4.2.2. Enfoques híbridos emergentes	22
4.2.2.1. Modelos de panel no lineal y métodos multietapa para estimar efectos marginales heterogéneos	22
4.2.2.1.1. Modelos de panel no lineal	22
4.2.2.1.2. Modelos multietapa	22
4.2.2.2. Modelos de diferencias largas	23
4.2.2.3. Modelos de variación de particionamiento	23
4.3. Modelos aditivos generalizados	23

5. Ejemplos de aplicaciones	27
5.1. Modelización de los gases de efecto invernadero	27
5.1.1. Descripción de los datos	27
5.1.2. Modelo de series temporales	29
5.2. Modelización de los elementos del clima	42
5.2.1. Descripción de los datos	42
5.2.2. Modelo de regresión lineal múltiple	45
5.3. Modelización de la concentración de ozono	51
5.3.1. Descripción de los datos	51
5.3.2. Modelo aditivo generalizado	52
5.3.3. Modelo de árbol de regresión	56
6. Conclusiones	61
A. Apéndice: Exploración y transformación de datos	63
A.1. Lectura y transformación de datos	63
A.2. Gráficos	66
B. Apéndice: Librerías y análisis	69
B.1. Librerías	69
B.2. Análisis	69
Bibliografía	73

Prólogo

Para empezar, me gustaría agradecer a mi familia, en especial a mis padres, a mi hermana y a mis abuelos por todo el apoyo y la paciencia durante todos estos cuatro años donde siempre han apostado por mí y aunque mi abuelo ya no esté entre nosotros, seguro que estaría muy orgulloso.

También, dar las gracias a mis profesores de Estadística que me han enseñado todo este campo y lo apasionante que es.

Por último, pero no menos importante, agradecer a mi tutor José Luis Pino Mejías por su dedicación, entrega e interés para poder realizar este trabajo.

Resumen

Nos hallamos en un momento crucial para afrontar con éxito el mayor desafío que nos toca vivir: el cambio climático.

Actualmente, es indispensable hacer para cualquier estudio sobre el cambio climático un análisis de datos para conocer y anticiparse a todos los cambios que se producen en el medio ambiente, y que tanto perjuicio ocasionan.

Este trabajo consiste en la recopilación de datos atmosféricos y climáticos, y la aplicación de ciertas técnicas estadísticas para dichos datos. Con tal objetivo, se pretende analizar la influencia de estos elementos en el calentamiento global y por ello en el cambio climático a partir de varios conjuntos de datos obtenidos de distintas fuentes de información.

Abstract

We are at a decisive moment to face successfully the biggest challenge that it is time to live: climate change.

Currently, it is essential to carry out a data analysis for any study on climate change to know and anticipate all the changes that occur in the environment, and that so much damage causes.

This project focuses on the research of atmospheric and climate data, and the application of certain statistical techniques for that data already said. With this objective, it is intended to analyze the influence of these elements in global warming and therefore, in climate change from some data got from different sources of information.

Índice de figuras

4.1. Fases de realización de un Meta-Análisis. Fuente: Introducción al análisis de datos en Meta-Análisis	16
5.1. Evolución de la concentración de GEI	29
5.2. Valores medios mensuales de CO2 (2001-2022)	30
5.3. FAS de datos CO2	32
5.4. Datos tras una diferencia regular	32
5.5. FAS tras la primera diferencia	33
5.6. FAS tras una diferencia regular y otra estacional	33
5.7. Datos tras una diferencia regular y otra estacional	34
5.8. Predicción de la concentración de CO2 en la atmósfera	42
5.9. Comparación de las temperaturas en Sevilla	44
5.10. Comparación de las temperaturas en Tromso	45
5.11. Gráfico de normalidad	49
5.12. Efectos parciales	53
5.13. Árbol de regresión	57
5.14. Árbol de regresión podado	59

Índice de tablas

4.1. Evaluación de la precisión de las estimaciones de Tol (2009) y Tol (2014)	18
--	----

Capítulo 1

Introducción

El cambio climático es un fenómeno que incluye alteraciones de los parámetros climáticos: temperaturas, precipitaciones, fenómenos climatológicos diversos, etc. Estos cambios pueden ser naturales, como las variaciones del ciclo solar. Normalmente este tipo de cambios se producen a lo largo de los siglos permitiendo a las especies animales y vegetales adaptarse de forma gradual a las nuevas condiciones climatológicas.

Sin embargo, cuando el clima se modifica de forma rápida, los ecosistemas pierden la oportunidad de adaptarse, con la consecuencia de producir migraciones de los seres vivos e incluso producirse la muerte de aquellos que no logran emigrar.

Desde el siglo XIX, la industrialización ha sido la principal causa de este cambio climático. Sobre todo, esto es debido a la quema de combustibles fósiles que producen gases de efecto invernadero que envuelven a la tierra, atrapando el calor del sol y por ello, elevando las temperaturas.

Como consecuencia del cambio climático, se producen sequías, escasez de agua, grandes incendios, deshielo de los polos, y por ende, aumento del nivel del mar, inundaciones, tormentas catastróficas y disminución de la biodiversidad.

Capítulo 2

Cambio Climático

2.1. Debate

Como afirma la química y divulgadora científica, Deborah García, ” *Hay cuestiones acerca del cambio climático sobre las que hay que debatir, por ejemplo, cómo mitigar o cómo adaptarse al calentamiento global. Sin embargo, no existe un debate en la comunidad científica acerca de la existencia, causas y consecuencias inmediatas del calentamiento global. El consenso científico es abrumador. Quienes niegan estos hechos no son galileos, sino ignorantes*”. [García, 2022]

A pesar del consenso científico sobre la existencia de esta crisis medioambiental, todavía hay personas la niegan. Pese a que cada vez representan un porcentaje menor, siguen entorpeciendo en decisiones para salvar el planeta. Las principales afirmaciones empleadas por los negacionistas para apoyar su teoría son las siguientes:

- Se han producido otros cambios climáticos a lo largo de la historia.
- Si hace frío, ¿cómo se va a estar produciendo un cambio climático?
- Los seres humanos no son responsables, sino que es un proceso natural.
- La temperatura no ha variado en la última década.
- No todas las personas pertenecientes a la comunidad científica confirman el cambio climático.
- Que suba la temperatura 2 o 3 grados no supone una gran alteración.
- Cuanto más CO_2 haya, mejor es para las plantas.
- Aparecerán zonas fértiles que antes no lo eran.

Por tanto, para no caer en los engaños de los negacionistas hay que tener claro varios términos:

- La confusión entre tiempo y clima. El tiempo meteorológico es el estado de la atmósfera en un momento y lugar determinado definido por variables como la presión, el viento, la radiación solar, la precipitación, etc. Sin embargo, el clima es el conjunto de condiciones atmosféricas de un lugar durante un periodo largo. La meteorología estudia el tiempo mientras que la climatología se encarga del clima.
- El dióxido de carbono es responsable del calentamiento global. El CO_2 es capaz de retener el calor y que no escape de la atmósfera. Esto es conocido como efecto invernadero. Desde hace medio siglo, ha aumentado la cantidad de moléculas de dióxido de carbono en la atmósfera y provoca mayor calentamiento global. Por otro

lado, el exceso de éste proviene de la actividad humana. La quema de combustibles fósiles provoca un claro desbalance entre el carbono-12 y el carbono-13.

- La ocurrencia de otros cambios climáticos en el tiempo no significa que éste sea igual. Ni mucho menos, nunca se había producido un calentamiento así en un periodo de tiempo tan breve.

2.2. Actualidad

El cambio climático nos guste o no, ya es una realidad. Según unos estudios, la temperatura media global ha aumentado alrededor de 1°C desde comienzos del siglo anterior a causa de los gases de efecto invernadero. El mayor crecimiento de la temperatura se acusa más en las zonas polares. Así pues, el nivel del mar ha subido considerablemente, unos 20 cm, por el derretimiento de los casquetes polares.

Por otro lado, la acidificación de los océanos se ha producido debido a la absorción de gran cantidad de dióxido de carbono, originando una disminución del pH del agua y aumentando así su acidez. Los ecosistemas marinos como las ostras, arrecifes y animales con caparazón tienen un serio problema ya que se les dificulta formar conchas y esqueletos por la reducción de la concentración de carbonato de calcio en el agua.

Asimismo han ocurrido cambios en los fenómenos naturales. Ha crecido la intensidad y frecuencia de huracanes, inundaciones, incendios, etc.

Por todo ello, muchas especies no han tenido más remedio que migrar y buscar otros lugares y hábitats a causa de estas dificultades.

A pesar de los pequeños avances que se están realizando, éstos todavía siguen sin estar a la altura de la importancia del problema que existe. La solución ya no está solamente en dejar de quemar combustibles fósiles sino en aplicar mecanismos para la adaptación a los peores impactos del cambio climático. Así, entre otras medidas, es imprescindible un cambio a un modelo energético basado en la descentralización y las energías renovables.

Como ya es sabido, los países industrializados son los principales causantes del cambio climático ya que emiten cerca del 75% de los gases de efecto invernadero (GEI) a la atmósfera. La ciencia concluye que si aumentamos la temperatura de la Tierra desde ahora hasta finales de siglo en 1.5°C, los ecosistemas tendrán un grave problema. Por desgracia, los principales países que sufrirán el calentamiento serán los más pobres y menos desarrollados aunque los industrializados también lo acusarán.

En este sentido, en la Conferencia Internacional acontecida el 9 de enero de 2023, el Secretario General de las Naciones Unidas, en referencia a las lluvias torrenciales e inundaciones sufridas en Pakistán, declaró " *Las palabras no son suficientes. Sin acción, la catástrofe climática se acerca para todos nosotros.*" [Guterres citado en ONU, 2022]

2.3. Política Europea

En la actualidad, muchos artículos, trabajos y conferencias han expuesto las consecuencias de las actividades humanas de producción y consumo que están teniendo sobre el medio ambiente y la calidad de vida. De ahí que la inquietud y la toma de conciencia social acerca de dichos problemas medioambientales, ha procedido a establecer unas reglas

para llevar en concordancia la factibilidad de los ecosistemas con el desarrollo económico. Cabe destacar el Protocolo de Kioto y las normas IPPC (de prevención y control integrado de la contaminación).

De la misma manera para preservar el planeta, la protección del medio ambiente es reconocida en los tratados de la Unión Europea. La política medioambiental de la UE, basada en el artículo 191 del Tratado de Funcionamiento de la Unión Europea (TFUE), tiene como objetivo mejorar la calidad del medio ambiente y la vida humana. Además se focaliza en el uso cuidadoso y responsable de los recursos naturales y promueve hacer frente a desastres como el cambio climático y la pérdida de la biodiversidad.

Así, estas políticas medioambientales están trazadas para valorar y ser consciente del medio ambiente y cuidar la salud humana. Se pretende mejorar la capacidad para adaptarse a situaciones adversas. En diciembre de 2019, la Comisión Europea anunció el plan para abordar el cambio climático y la degradación medioambiental llamado Pacto Verde Europeo. Esta hoja de ruta tiene como finalidad realizar una transición hacia la sostenibilidad mediante:

- el impulso del uso eficiente de los recursos.
- el paso a una economía moderna, eficiente y circular.
- la restauración de la biodiversidad.
- la reducción de la contaminación.

En el paquete de medidas del Pacto Verde Europeo, una de éstas fue la nueva Estrategia Europea de Adaptación al Cambio Climático publicada en 2021 para acentuar el deseo climático de la UE para 2030 y 2050. Esta estrategia consiste en preparar a todos los países pertenecientes a la UE ante los efectos irremediables y ofrecer una respuesta determinante frente a esta crisis. La estrategia incluyó también que se actuase en los sectores que pueden resultar más vulnerables. Incorporó el decreto de actuar en tres áreas diferenciadas: la acción en los ámbitos de la política agrícola común, política de cohesión y la política pesquera común.

En esa misma línea, se desarrolló una herramienta sobre adaptación que contuviera toda la información sobre esta materia. Así nació Climate-ADAPT, la Plataforma Europea de Adaptación al Cambio Climático, establecida en 2012 a través de la colaboración entre la Comisión Europea y la Agencia Europea de Medio Ambiente. De esta forma, dejaba de no existir una base de datos específica sobre adaptación a nivel europeo y se podía tomar una decisión única en los diferentes niveles competenciales.

2.4. Cumbre del clima

Con la necesidad de hacer frente al cambio climático, se han tenido que realizar acuerdos entre países como el Protocolo de Kioto en 1997 (que no entró en vigor hasta principios de 2005) fue elaborado en la Convención Marco de las Naciones Unidas, cuyo objetivo era reducir las emisiones de los seis principales gases de efecto invernadero aunque no incluyó a todos los países, sólo a los industrializados. Sin embargo, China y Estados Unidos quedaron fuera de este tratado.

Unos años después se celebró la Cumbre de Copenhague (2009) pero ésta fue un auténtico desastre. La misión era sustituir el Protocolo de Kioto y sirvió para aumentar las diferencias entre países ricos y pobres. Continuando con esta finalidad, se llevó a cabo

la Cumbre Climática de París en 2015 y se pactó el Acuerdo de París. *” El gran reto de la Cumbre de París es sustituir al Protocolo de Kyoto y llegar a un acuerdo que incluya a todas las naciones (cubriendo prácticamente el 100 % de los gases emitidos, frente al 11 % actual), independientemente de su nivel de riqueza.”*[Pilar, 2015]. Este acuerdo ha dado lugar a soluciones como la bajada de emisiones de carbono y a nuevos mercados.

En 2019 se celebró la Cumbre del Clima de Madrid pero no se obtuvieron grandes logros debido a la falta de acuerdos entre países. La activista medioambiental Greta Thunberg expuso en su discurso ante el plenario de la COP (Conferencia de las Naciones Unidas sobre el cambio climático) que *” la esperanza viene del hecho de que en todas partes la gente corea con verdadera pasión las consignas que reclaman justicia climática.”*[Thunberg citado en Liguori, 2019]. Es por lo que todos debemos comprometernos a renovar nuestro compromiso con la justicia climática.

Dos años después, a causa de la COVID-19, se concretó la Cumbre Climática en Glasgow. Ecologistas y científicos coinciden en que se podía haber conseguido más y apuntan que el objetivo de 1.5°C escasamente se mantiene. La decisión final incluye el objetivo de reducir las emisiones de gases de efecto invernadero en un 45 % en 2030 y lograr la descarbonización de la economía. En líneas generales, para el Secretario General de la Organización Mundial, Petteri Taalas; y el director ejecutivo de la Climate Analytics, Bill Hare, se dio un pequeño paso pero insuficiente en la lucha del clima. Por último, el economista Nicholas Stern declaró: *” Hasta la próxima Cumbre del Clima (COP27) en Sharm el-Sheij, Egipto, hay mucho trabajo por hacer porque cuanto más se tarde en desbloquear la financiación, más se tardará en aumentar la ambición”*. [Stern citado en SINC, 2021].

Continuando con la afirmación del economista británico y tras la celebración de la COP27 (2022), se adoptaron una serie de decisiones muy optimistas para continuar con el proyecto climático. Entre las medidas aprobadas, se ha firmado un acuerdo para proporcionar financiación por *” pérdidas y daños”* a los países más vulnerables. Se han producido avances significativos respecto al tema de adaptación que concluirá en la próxima COP28 y la reafirmación del compromiso de limitar la subida de la temperatura.

Por otra parte, el informe sobre la Situación de los Bosques del Mundo 2022 destaca la necesidad de intensificar la acción para liberar el potencial de los bosques en lucha contra el cambio climático y la pérdida de la biodiversidad. Este documento muestra las tres vías para afrontar estos problemas: detener la deforestación y conservar los bosques, restaurar las tierras degradadas y utilizar los bosques de manera sostenible.

Capítulo 3

Fuentes de datos

En esta sección se presentan algunos organismos que difunden datos y gráficos relevantes sobre temas relacionados con el cambio climático como el aumento de la temperatura global, las emisiones de gases de efecto invernadero, el aumento del nivel del mar, entre otros.

3.1. Global Carbon Project

El Proyecto Global de Carbono (GCP) (<https://www.globalcarbonproject.org/>) es un Proyecto de Investigación Global de Future Earth y un socio de investigación del Programa Mundial de Investigación Climática. Se creó para apoyar el debate sobre las políticas a adoptar y detener el aumento de los gases de efecto invernadero en la atmósfera. Establecido en 2001, sus proyectos incluyen presupuestos globales para tres gases de efecto invernadero importantes: el dióxido de carbono (CO_2), metano (CH_4) y óxido nitroso (N_2O).

El objetivo principal ha sido comprender el ciclo del carbono. La estrategia de implementación del GCP se organiza en torno a tres temas científicos: patrones y variabilidad, procesos e interacciones y gestión del carbono. El primero, determina las distribuciones geográficas y temporales actuales flujos en el ciclo global del carbono; el segundo, establece los mecanismos de control y retroalimentación que determinan la dinámica del ciclo del carbono; y el último, las dinámicas del sistema carbono-clima-humano en el futuro.

El GCP tiene como mandato desarrollar un marco de investigación para la incorporación de componentes biogeoquímicos y humanos del ciclo global del carbono; sintetizar la comprensión actual del ciclo; proporcionar una plataforma de coordinación global para programas de carbono; fortalecer los programas de investigación de carbono; desarrollar nuevas iniciativas a medio plazo sobre problemas; y, fomentar nuevas investigaciones sobre el carbono en regiones que proporcionen mejores restricciones de los presupuestos de carbono mundiales.

La investigación sobre el ciclo del carbono es una parte fundamental de muchas actividades que plantean la ciencia ambiental de todo el sistema terrestre y la agenda internacional de desarrollo sostenible.

Este proyecto trabaja con numerosos datos. Para ilustrarlo, estudia:

- La tendencia del dióxido de carbono atmosférico a nivel global en la superficie terrestre al nivel del mar y en un lugar en concreto, como Mauna Loa a una altitud de 3400 metros.
- Las concentraciones medias mensuales de gases de efecto invernadero tomadas en la estación de contaminación del aire de Kennaook, Tasmania. Por otro lado, las tasas de crecimiento anual y decenal de los GEI en la atmósfera.
- Las emisiones de dióxido de carbono por combustible y por lugar.
- La evolución de los sumideros terrestres y oceánicos en el mundo.
- Cuantificar el almacenamiento de carbono orgánico en el suelo de la región permafrost circumpolar del norte.

3.2. Oficina Nacional de Administración Oceánica y Atmosférica

La Oficina Nacional de Administración Oceánica y Atmosférica (NOAA) (<https://www.noaa.gov/>) es una agencia científica cuyas actividades se enfocan en observar los estados de los océanos y la atmósfera y emitir advertencias acerca de condiciones meteorológicas peligrosas. La misión consiste en entender y predecir los cambios en el ambiente del planeta, conservar y utilizar los recursos marinos y costeros para completar las necesidades en relación a lo económico, social y ambiental.

Además la NOAA tiene cuatro metas, cada una se corresponde con las actividades programadas:

- **Garantizar** el uso sostenible de los recursos de los ecosistemas costeros y marinos.
- **Comprender** el cambio climático para que se pueda responder de la mejor manera posible.
- **Suministrar** datos y pronósticos del tiempo, tormentas, sequías e inundaciones.
- **Abastecer** información sobre el tiempo, el clima y los ecosistemas para ratificar que el transporte comercial e individual sea seguro sin dañar el medio ambiente.

La misión de ciencia, servicio y administración está guiada a una visión del futuro en el que las sociedades y sus ecosistemas son resistentes, seguros y saludables frente a cambios inesperados o duraderos. Una visión de resiliencia ayudará a la NOAA a reducir la vulnerabilidad de los sistemas ecológicos a corto plazo, mientras que coopera con las personas para que puedan adaptarse a cambios económicos, ambientales... De tal manera, es necesario ceñirse a las situaciones actuales para poder trazar cambios futuros y tomar decisiones convenientes.

Los datos que emplea para realizar estudios están basados en el tiempo y el clima, satélites y el océano/costa. Respecto al primero de ellos, presenta:

- Datos sobre el tiempo severo (estadísticas de tornados).
- Datos meteorológicos históricos.
- Registros meteorológicos de EEUU.
- Pronóstico del tiempo de invierno.

- Datos climáticos mensuales.
- Muertes relacionadas con el clima.
- Evolución de las anomalías de la temperatura mundial en los meses de verano.
- Aumento de la temperatura en el mundo en los meses de junio.
- Grados de diferencia con el promedio de temperatura de 1951 a 1980 en agosto.
- Pronóstico del Índice de Calor.
- Estadísticas de daños ambientales y económicos.
- Informes de la capa de nieve de EEUU.
- Perspectivas de sequía y estacional en EEUU.

Por otro lado, gracias a los satélites se obtienen datos acerca de la atmósfera, criosfera, océano, entre otros.

3.3. Organización Meteorológica Mundial

La Organización Meteorológica Mundial (OMM) (<https://public.wmo.int/>) fue establecida en 1950 gracias a la confirmación del Convenio de la OMM y un año después se convirtió en el organismo especializado de las Naciones Unidas para la meteorología, la hidrología y otras ciencias. Por ello, lidera la cooperación internacional en el uso de servicios meteorológicos, climáticos, hidrológicos, en definitiva, medioambientales y concede conocimientos especializados en esos ámbitos.

Este organismo facilita y fomenta el establecimiento de una red de observación de la Tierra para proporcionar datos meteorológicos, climáticos y relacionados con el agua; el mantenimiento de centros de gestión de datos y sistemas de telecomunicaciones; la creación de estándares de seguimiento para comprobar la uniformidad de los procedimientos empleados y la homogeneidad de los datos; la prestación de servicios relacionados con el tiempo, el clima y el agua para disminuir los riesgos de desastres medioambientales; y por último, la coordinación de la investigación y formación en meteorología.

La OMM visualiza que para el año 2030, las naciones del mundo, sobre todo las más vulnerables, serán mucho más resistentes a las adversidades socioeconómicas provocadas por fenómenos extremos medioambientales. Así, tiene como misión facilitar la cooperación internacional, apoyar el intercambio de información meteorológica, impulsar la investigación y extender el uso de la meteorología a sectores como la aviación, la agricultura y la gestión del agua.

La Organización Meteorológica Mundial difunde datos sobre la temperatura mínima y máxima media por meses para ciudades seleccionadas de todo el mundo además de la precipitación media total mensual. Los promedios mensuales se basan en el periodo de 30 años 1981-2010.

3.4. Agencia Estatal de Meteorología

La Agencia Estatal de Meteorología (AEMET) (<https://www.aemet.es/es/>) es un organismo público español adscrito al Ministerio para la Transición Ecológica a través

de la Secretaría de Estado de Medio Ambiente. Tiene como finalidad, el desarrollo, implantación y prestación de los servicios meteorológicos y el apoyo a otras políticas colaborando con la seguridad de las personas y el desarrollo sostenible de la sociedad española.

La AEMET se dedica a la emisión de avisos y predicciones de fenómenos meteorológicos que puedan afectar al bienestar de las personas, servicios de apoyo a la aviación y navegación marítima, atención de las condiciones climáticas y de la atmósfera, actualización de datos meteorológicos, realización de estudios medioambientales y desarrollo de técnicas necesarias para ello, entre otras actividades.

Los propósitos sobre predicción meteorológica conforman la parte fundamental de las actividades I+D+i de la AEMET. El principal objetivo es garantizar que la prestación de la predicción esté basada en el progreso científico. Entre los proyectos de I+D+i destacan los estudios de fenómenos meteorológicos significativos, predicción numérica del tiempo, predicción inmediata, predicción probabilística y la meteorología marítima.

Asimismo, ha reforzado sus actividades de I+D+i para suministrar mejores servicios climáticos que son esenciales para la toma de decisiones. Estos servicios climáticos proporcionan información del clima pasado, presente y futuro y así poder evitar riesgos socioeconómicos muy delicados al clima.

Entre la información que proporciona, se puede encontrar datos sobre valores climatológicos diarios, mensuales/anuales registrados como, por ejemplo, la duración media de las olas de calor en España. Datos relativo al nivel de concentración de CO_2 en la atmósfera regulado por el Observatorio de Izaña, Tenerife. El índice normalizado de vegetación, la temperatura del agua del mar y predicciones acerca del cambio climático para el siglo XXI, marítimas, de radiación ultravioleta, etc.

3.5. Instituto Internacional de Derecho y Medio Ambiente

El Instituto Internacional de Derecho y Medio Ambiente (IIDMA) (<https://iidma.org/>) es organización fundada en 1997 con el objetivo de contribuir al desarrollo sostenible y a la protección del medio ambiente. Está especializada en la lucha contra el cambio climático, la transición energética, biodiversidad y océanos.

La organización pretende conseguir su misión a través de la investigación, difusión, estudio y aplicación correcta del Derecho. Para que así sea, esto requiere que las instituciones nacionales e internacionales sean conscientes de que respetando el Derecho se lograrán transformar las políticas ambientales en resultados.

Los objetivos prioritarios son cuatro:

- **Investigar**, desarrollar y fomentar la aplicación y cumplimiento del Derecho como instrumento de protección del medio ambiente y de desarrollo sostenible.
- **Promover** la cooperación al desarrollo en el ámbito de la protección ambiental y el desarrollo sostenible.
- **Sensibilizar** sobre los problemas ecológicos. Propagar la normativa ambiental para su ejecución.

- **Responder** a los retos jurídicos que representan los avances científicos y los efectos que éstos provocan en la sociedad y el medio ambiente.

Los ámbitos de especialidad del IIDMA son el cambio climático; la transición energética, fomentando un cambio estructural en los sistemas de energía con la descarbonización de este sector; la gobernanza y sostenibilidad; los recursos hídricos a través del impulso de la eficiencia de los programas dirigidos a la protección de este recurso; y, la biodiversidad y océanos.

Se pueden encontrar datos sobre las muertes prematuras debidas a emisiones de carbón por comunidades autónomas.

3.6. Sistema Integrado de Observación de Carbono

El Sistema Integrado de Observación de Carbono (ICOS) (<https://www.icos-cp.eu/>) es una infraestructura de investigación, financiada por 16 países europeos, que proporciona datos de más de 170 estaciones de medición. Estas estaciones observan las concentraciones de los gases de efecto invernadero (GEI) en la atmósfera, al igual que los flujos de carbono entre la atmósfera, la superficie terrestre y los océanos. El conocimiento basado en ICOS ayuda a la toma de decisiones y políticas para combatir el cambio climático y sus efectos.

El principal objetivo es elaborar observaciones estandarizadas y precisas para facilitar la investigación del ciclo del carbono y conceder información necesaria sobre los GEI. Para ello, se promueven desarrollos tecnológicos y demostraciones enlazando investigación, educación e innovación.

Las estaciones ICOS actúan en tres aspectos diferentes: atmósfera, ecosistema y océano. Dentro de cada dominio se coordinan las observaciones y proporcionan análisis de gases. La red de estaciones atmosféricas de ICOS observa los gases de efecto invernadero en la atmósfera. La red de estaciones de ecosistemas de ICOS observa los gases de efecto invernadero, además de los componentes vivos y no vivos, agua y energía entre los ecosistemas y la atmósfera. Por último, la red estaciones oceánicas de ICOS observa los niveles de carbono en la superficie del océano para facilitar los datos necesarios para comprobar el rol de los océanos en el ciclo del carbono.

Los datos de ICOS colaboran en dar cuenta del sistema de la Tierra y su respuesta al cambio climático. Generan conocimiento científico que avanza en el cumplimiento de los objetivos de Desarrollo Sostenible de las Naciones Unidas y de la Unión Europea. Por otro lado, ayuda a la CMNUCC (Convención Marco de las Naciones Unidas sobre el Cambio Climático) a conseguir su objetivo de mitigar y adaptarse al cambio climático.

En cuanto a la obtención de datos, ICOS ofrece colecciones de datos, así como objetos de datos individuales para buscarlos en el portal de datos. Los datos de observaciones se clasifican en 3 niveles: el nivel 0 (o datos sin procesar), el nivel 1 (datos observacionales intermedios en el procesamiento de esta clase de datos) y el nivel 2 (datos observacionales de calidad final controlada). Asimismo existe otro nivel en el que se encuentran productos ya elaborados por comunidades científicas.

3.7. Copernicus

Copernicus (<https://www.copernicus.eu/>) es el Programa de Observación de la Tierra

de la Unión Europea que vigila nuestro planeta y su medio ambiente para obtener el máximo beneficio de toda la ciudadanía europea. Además, ofrece servicios de información basados en datos de la Tierra por satélite e in situ. Por un lado, los satélites Sentinel están preparados con el propósito de satisfacer las necesidades de los servicios de Copernicus. Por otro lado, los sistemas in situ son estaciones meteorológicas terrestres, boyas oceánicas y redes de seguimiento de la calidad del aire, que suministra información fiable para validar los datos de los satélites.

Los servicios de Copernicus utilizan y analizan estos datos para transformarlos en información con valor añadido y poder identificar características y anomalías y como consecuencia, extraer información estadística.

Las actividades de las que se encarga en ofrecer servicios son la atmósfera, el cambio climático, entre otros.

El Servicio de Monitoreo de la Atmósfera permite muchas aplicaciones en una gran variedad de dominios. Este servicio se centra en cinco áreas principales:

- calidad del aire y composición atmosférica mediante el seguimiento y la previsión de los GEI;
- capa de ozono y radiación ultravioleta;
- emisiones y flujos superficiales;
- radiación solar, que es de gran relevancia para la salud y la agricultura;
- forzamiento climático.

El Servicio de Cambio Climático (C3S) dispone de informes de la situación climática de Europa y el resto del mundo. La meta del C3S es proporcionar información fiable relativo al cambio climático en favor de las políticas de adaptación y suavización de la UE.

Los conjuntos de datos se pueden filtrar respecto al tipo de producto (índices climáticos, observaciones in situ y satelitales, pronósticos estacionales. . .), al dominio de la variable (atmósfera, Tierra u océano), la cobertura espacial y temporal y el sector (agricultura, energía, salud, etc).

3.8. Grupo Intergubernamental de Expertos sobre el Cambio Climático

El Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC) (<https://www.ipcc.ch/>) fue establecido en 1988, por la Organización Meteorológica Mundial (OMM) y el Programa de las Naciones Unidas para el Medio Ambiente (PNUMA), y es el órgano internacional encargado de evaluar los conocimientos científicos relacionados con el cambio climático.

El IPCC es un órgano científico que examina y evalúa la bibliografía científica, técnica y socioeconómica más actual para la comprensión del cambio climático. Asimismo es un órgano intergubernamental del que forman parte todos los países miembros de las Naciones Unidas y de la OMM.

El objetivo es claramente proporcionar evaluaciones científicas sobre el cambio climático, sus implicaciones y riesgos futuros, además de fórmulas de adaptación y mitigación. Mediante dichas evaluaciones, identifica dónde hay acuerdo en la comunidad científica y dónde se necesita más investigación.

Este órgano ha comprendido cinco ciclos de evaluación y ha entregado también cinco Informes científicos sobre el cambio climático, que son los más completos en todo el mundo. Cada uno de ellos se ha incorporado a la formulación de políticas climáticas internacionales. Veamos sus puntos claves.

El Primer Informe de Evaluación del IPCC enfatizó la relevancia del cambio climático como un desafío global que requiere cooperación internacional. Fue decisivo en la creación de la CMNUCC. El Segundo Informe de Evaluación proporcionó elementos importantes para que los gobiernos se basaran en el periodo previo a la admisión del Protocolo de Kioto. El Tercer Informe de Evaluación se enfocó en los impactos del cambio climático y la obligación de adaptación. El Cuarto Informe de Evaluación consistió en establecer un acuerdo posterior al de Kioto en el que limitara el calentamiento a 2°C. Finalmente, el Quinto Informe de Evaluación concedió la base científica del Acuerdo de París. Como apunte, actualmente se encuentra elaborando el Sexto Informe de Evaluación.

El Centro de Distribución de Datos proporciona datos climáticos, socioeconómicos y ambientales, tanto del pasado como proyecciones futuras. Además se deparan técnicas para la selección y el uso de las diferentes clases de datos. Como observación, los datos proporcionados son de libre y fácil acceso.

3.9. Centro Nacional de Datos de Hielo y Nieve

El Centro Nacional de Datos de Hielo y Nieve (NSIDC) (<https://nsidc.org/>) es un centro de información de EEUU que apoya la investigación polar y crioesférica. NSIDC archiva y distribuye datos sobre avalanchas, glaciares, capas de hielo, permafrost y núcleos de hielo. Por otra parte, sirve como uno de los doce centros de archivo activo distribuido financiados por la Administración Nacional de Aeronáutica y del Espacio para archivar y distribuir datos de los programas de medición de campo y satélites pasados y actuales de la NASA.

NSIDC contribuye a las soluciones del cambio climático, afectando de manera positiva al futuro entre la relación del planeta con la humanidad. Desde este centro se avanza en la comprensión de las regiones congeladas de la Tierra y los cambios que tienen lugar para informar la toma de decisiones.

Respecto a la gestión de datos, los programas del NSIDC encargados para ello, se centran principalmente en preservar, documentar y facilitar acceso a los datos criosféricos y geofísicos relacionados. Asimismo almacena y administra datos científicos para organizaciones gubernamentales. Las recopilaciones de datos incluyen:

- Datos relacionados con la criosfera de misiones satelitales de observación de la Tierra, observaciones meteorológicas, etc.
- Datos producidos por el grupo de personas que componen el NSIDC y científicos externos.
- Conocimiento y observaciones de las regiones del Ártico para compartir, usar y preservar datos.

Capítulo 4

Ejemplo de modelos estadísticos

La modelización estadística es una herramienta matemática que pretende reconstruir la realidad de forma simplificada a través de métodos estadísticos. Se quiere describir la relación entre la variable respuesta u objetivo (dependiente) y las variables predictoras o explicativas (independientes). Los dos principales objetivos son:

- **Predecir** el valor de una variable de acuerdo con los valores conocidos de las variables explicativas. En este caso no nos interesa conocer la interacción entre las variables o de la manera en que interactúan. Basta con que las predicciones sean adecuadas.
- **Explicar** la relación entre las variables independientes y la dependiente y comprobar si es significativa o no.

Existen muchos métodos y técnicas para modelar datos. A continuación veremos varios ejemplos de artículos e investigaciones que han puesto en práctica modelos estadísticos para determinar el impacto económico del cambio climático y, en menor medida, cómo afecta a la salud de las personas.

4.1. Meta-Análisis

Este término fue acuñado por Glass y lo definió como *” el análisis de análisis... [es decir,] el análisis estadístico de una gran colección de resultados de análisis de estudios individuales con el propósito de integrar los hallazgos”* (Glass, 1976). Estas técnicas comenzaron a utilizarse en el ámbito de las ciencias sociales y más adelante se incluyeron en los ensayos clínicos para evaluar la eficacia de los tratamientos. El Meta-Análisis es un método de agregación de información cuantitativo que, en parte, suprime la aproximación subjetiva de las revisiones de un experto sobre un tema basadas en habilidades cualitativas.

El procedimiento de un Meta-Análisis consta de dos fases:

1. Recogida de información fundamental: dado los miles de artículos disponibles relacionados con el propósito del estudio, es necesario buscar un número adecuado de trabajos fiables que proporcionen certeza sobre el objetivo.
2. Combinación de la información disponible: se realizan análisis estadísticos que permiten la combinación de resultados y así obtener efectos globales, intervalos de

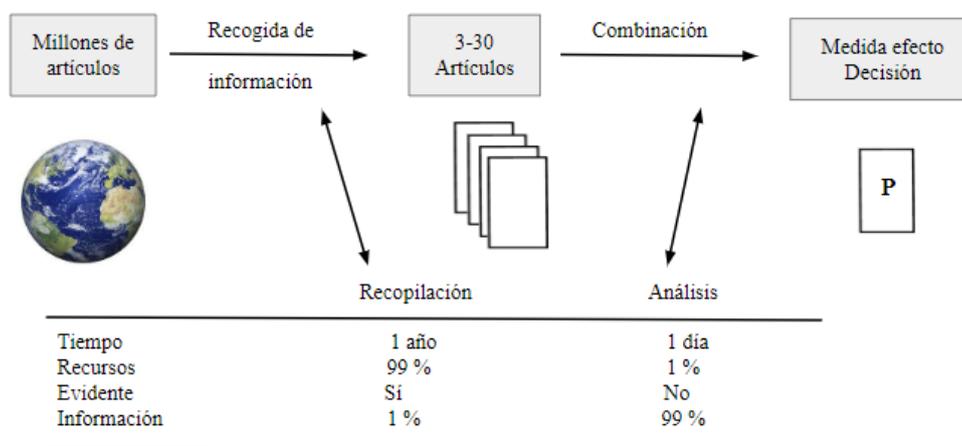


Figura 4.1: Fases de realización de un Meta-Análisis. Fuente: Introducción al análisis de datos en Meta-Análisis

confianza y la evidencia significativa. En consecuencia, estas herramientas pueden ayudar a la toma de decisiones.

Con esta breve explicación, pasamos a comentar el artículo *A survey of global impacts of climate change: replication, survey methods and a statistical analysis* [Nordhaus and Moffat, 2017]. Los autores William D. Nordhaus y Moffat declaran no tener conflicto de intereses económicos con este estudio. Los documentos de trabajo de NBER tienen como compromiso su distribución con fines de discusión y explicaciones.

Este documento se ha organizado de la siguiente manera: una breve introducción sobre el estudio en cuestión junto a la metodología que se va emplear, un resumen de la investigación sistemática sobre los impactos del cambio climático, el análisis estadístico y los principales resultados con la esperada conclusión.

Para comenzar, la pregunta de si el cambio climático tendrá efectos significativos en la actividad económica tiene una clara respuesta. Los modelos económicos son los encargados de predecirlo. Los sectores claves son la agricultura, la salud, los ecosistemas y los sistemas energéticos. Por ello, para investigar los efectos significativos, el estudio consta de dos partes.

En la primera se comprueban las estimaciones de los impactos globales del cambio climático a través de encuestas realizadas por investigadores. A raíz de este análisis es probable que apreciemos algunos errores y puede que éstos hayan servido para la toma de decisiones de científicos y de políticas sobre dicho tema. En particular, Tol en 2009 realizó un estudio estadístico que contuvo muchos errores. Debido a eso, hizo correcciones (2014) pero seguía manteniendo errores. En la segunda aportación se estudiaron estimaciones alternativas para usarlas en funciones de daño empíricas en modelos de evaluación integrados (IAM).

En cuanto a la metodología se presentan dos métodos fundamentales para la síntesis de la investigación: la narrativa y la cuantitativa. Por un lado, la narrativa es el recurso más universal debido a su simplicidad pero presenta una desventaja como la subjetividad en la

perspectiva del investigador. Por otro lado, el enfoque cuantitativo suele ser agrupado en tres clases: el metanálisis (con muestra de población bien definida), la síntesis de investigación con un muestreo claro (síntesis de investigación sistemática) y aquella sin un muestreo evidente (síntesis de investigación no sistemática).

Por consiguiente, para realizar un correcto metanálisis hay que tener en cuenta unos elementos claves:

- Comprobar que al recopilar las estimaciones de estudios separados, éstas sean independientes.
- Los tamaños de muestra de distintos estudios conceden una información sustancial en el cálculo de la estadística combinada. El tamaño de muestra de un estudio es un indicador de la confiabilidad de sus resultados.
- Se pueden obtener conclusiones más relevantes que en estudios individuales.
- Conseguir una interpretación estadística de una síntesis de estudios diferentes es complicado y es posible que no se logre en muchos casos.

Con relación a la muestra, es importante pensar en la población de la que se extraen los datos. En el metanálisis clásico, si las muestras son de la misma población y son independientes, tiene sentido juntar las muestras para hacer una metamuestra y así obtener una mayor precisión que en los estudios individuales. Si bien cada vez más se utiliza este método, existen casos donde se emplea mal el término siendo los estudios resúmenes de investigación. Por ello, debido a la naturaleza de los metanálisis no se suelen aplicar a temas económicos ni medioambientales.

Continuando con el enfoque cuantitativo, pasamos a comentar la investigación sistemática y no sistemática aplicadas a los impactos del cambio climático.

Con el objetivo de realizar una SRS (síntesis de investigación sistemática) lo primero que hay que determinar es el área de investigación y el lugar donde buscar. Se valoraron varias plataformas como Google Scholar, EconLit y JSTOR. Para determinar cuál de ellas elegir, se le aplicaron dos criterios: primero, si una búsqueda aparecía en la mayoría de los estudios de Tol (2014) y segundo, nuestra capacidad para usar dicha plataforma. Por tanto, se escogió EconLit a pesar de que Google Scholar fuera más completa (pero contenía un volumen de resultados impresionante). Lo siguiente por hacer fue establecer el área de investigación. Se requirió definir una cadena de búsqueda de palabras claves para limitar el número de estudios. Y en última instancia, se tomaron 11 que fueron los que ofrecían información suficiente sobre el tema a tratar.

En pocas palabras, el enfoque de una SRS tenía un valor limitado para resumir los resultados. Más aún, *los impactos del cambio climático* es un tema muy amplio que abarca muchos sectores y al ir seleccionando una plataforma u otra, se están despreciando algunos artículos importantes.

Puesto que la SRS tiene una capacidad reducida, se combinará con la síntesis no sistemática. Por ende, incluirán en el resumen los estudios que utilizó Tol (2014) y aquellos destacados del SRS pero no incorporados en Tol (2014). De esta forma, la mayoría proporcionaron estimaciones sobre los daños globales, no obstante, hubo resultados que no coincidían con los de Tol. La ejecución de la replicación por parte de los autores fue muy complicada y no se pudieron determinar los resultados de muchos estudios. De esta

forma, aunque se hayan realizado con mucha dedicación, es posible que éstos contengan inexactitudes.

Una vez recogidos los resultados, se crearon varias tablas para organizar la información. Los resultados por parte de Tol y de las síntesis se agruparon en una tabla en la que cada fila era la estimación de daño global de cada estudio y las columnas indicaban el aumento de la temperatura media mundial por encima de la preindustrial y estimaciones de Tol (2009), Tol (2014) y del actual análisis.

Asimismo en una nueva tabla se incluyeron estudios posteriores a la síntesis de investigación y en otra, la precisión de los estudios de Tol tomando como referencia las estimaciones utilizadas en el presente análisis. Sin ir más lejos, el siguiente cuadro corresponde a ésta última tabla descrita.

Tabla 4.1: Evaluación de la precisión de las estimaciones de Tol (2009) y Tol (2014)

Categoría	2009	2014
Errores	6	6
No se puede replicar	1	4
Omitido de la encuesta	17	10
Confirmar estimación	7	11

La mayoría de los casos de error y de ausencia de replicabilidad aparecieron como consecuencia del hecho de que Tol no proporcionó los métodos que usó para obtener estimaciones.

Ahora bien, para elaborar una estimación central se han utilizado mecanismos para mostrar la dependencia entre los estudios. Se comenzó con 36 estimaciones aprovechables de 27 estudios y se ponderaron con unos pesos según criterios. Posteriormente, se mostró los pesos de cada una de las estimaciones en una gráfica y se concluye que la mayoría de estudios estiman un aumento de temperatura en torno a los 2-4°C y coinciden en una pérdida del 0 al 4% de la producción global.

Como segundo objetivo, se estimaron distintas funciones de daño. Se combinaron regresiones ponderadas y no ponderadas; mínimos cuadrados ordinarios; regresiones por cuantiles; en nuestro caso, la mediana; especificaciones cuadráticas, lineales cuadráticas y polinómicas.

Por un lado, si bien se obtuvieron estimaciones diferentes, la regresión preferida por los autores fue la regresión ponderada cuadrática mediana. Además, con un ajuste del 25% para cubrir los sectores no cuantificados, se estimó un impacto de -2,04% de ingresos con un calentamiento de 3°C y -8,16% de ingresos con un calentamiento de 6°C. Por otro lado, fue fundamental conocer el grado de convexidad de la función de daño. Así, un resultado propone una convexidad algo menor que la cuadrática. No obstante, da la impresión que se pueden aproximar de manera adecuada los resultados de la forma cuadrática o lineal-cuadrática.

Para finalizar, como conclusión final y global, resaltar el carácter limitado del trabajo sobre impactos. Hay pocas organizaciones que sean capaces de integrar diferentes estudios para el uso posterior de las estimaciones en el modelado económico-climático. A diferencia de otros estudios respecto al cambio climático, los estudios integrales de impacto han

comenzado muy recientemente aunque sean fundamentales para la toma de decisiones políticas. Es por ello que se necesitan más trabajos sobre los impactos.

4.2. Métodos basados en datos de panel

Se define el término de datos de panel como aquellos que combinan una dimensión temporal con otra transversal. Por un lado, la dimensión temporal recoge las observaciones en el tiempo para cada individuo que forman la muestra. Por otro lado, el conjunto transversal de datos contiene las observaciones de todos los individuos para cada instante de tiempo. Un conjunto de datos de panel reúne observaciones acerca de muchos fenómenos a lo largo de periodos concretos. La dimensión temporal potencia la estructura de datos y puede añadir información que no consta en un único corte.

El motivo de uso de la técnica basada en datos de panel es aprovechar la información que proporciona la variabilidad transversal y estimación de los parámetros de una función respuesta valiéndose de la variación que presentan las variables independientes en el modelo. Si hubiera más variabilidad transversal que temporal, el estudio con datos de panel añadiría un mejor resultado para esa estimación. Se representa de la forma siguiente:

$$Y_{it} = \alpha + \beta X_{it} + \epsilon_{it} \quad \text{donde} \quad \begin{array}{l} i = 1, \dots, N \quad \text{grupos o individuos} \\ t = 1, \dots, T \quad \text{periodos de tiempo} \end{array}$$

Esta formulación permite la combinación de múltiples parámetros individuales y temporales. Entre las muchas técnicas de los modelos de datos de panel, los más recurrentes son: el modelo de *efectos fijos* y el modelo de *efectos aleatorios*.

La diferencia entre ambos es la siguiente: $\alpha_i = \alpha + u_i$. O sea, en lugar de considerar α como fija, suponemos que es una variable aleatoria con valor medio α y una desviación aleatoria u_i de este valor medio. A su vez $u_i \sim N(0, \sigma_u^2)$.

Sustituyendo α_i en $Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$ obtenemos:

$$Y_{it} = \alpha + \beta X_{it} + u_i + \epsilon_{it}$$

Por tanto, el nuevo término de perturbación sería $\omega_{it} = u_i + \epsilon_{it}$. Bajo el supuesto de que no están correlacionados, la varianza de $\omega = \sigma_u^2 + \sigma_\epsilon^2$.

Si en este caso se estimara por mínimos cuadrados ordinarios, los estimadores serían consistentes, pero no eficientes al no considerar σ_u^2 . Por ello, el método de estimación eficiente en este caso y dado la composición del término de perturbación aleatoria es el de los mínimos cuadrados generalizados.

Entonces, ¿cuándo deberíamos usar efectos fijos o efectos aleatorios?

Utilizamos estos modelos cuando N es grande y T pequeño. Un modelo de efectos fijos es mejor cuando se disponen datos de todos los miembros de la población. En cambio, si la población es bastante grande y tenemos una muestra, un modelo de efectos aleatorios parece que sería mejor.

Para determinar si es preferible estimar el modelo con efectos fijos o aleatorios, se opta por el **Test de Hausman**. Es un contraste clásico de robustez frente a eficiencia en los

estimadores. Se plantea escoger entre dos estimadores; uno robusto $\hat{\theta}_R$ consistente tanto en la hipótesis nula H_0 como en la alternativa H_1 , y otro eficiente $\hat{\theta}_E$ en la hipótesis nula sólo.

La prueba se basa en la medida H , que es la distancia entre los efectos fijos y los estimadores de los efectos aleatorios que bajo la hipótesis nula sigue una distribución χ^2 con grados de libertad número de variables incluidas. Si H es un valor grande, significa que el estimador de efectos aleatorios no es consistente y es preferible emplear los estimadores del modelo de efectos fijos. Analíticamente queda:

$$H = (\beta_c - \beta_e)'(V_c - V_e)^{-1}(\beta_c - \beta_e), H \sim \chi_n^2$$

donde β_c es el vector de estimaciones del estimador consistente $\hat{\theta}_R$
 β_e es el vector de estimaciones del estimador eficiente $\hat{\theta}_E$
 V_c es la matriz de covarianzas del estimador consistente
 V_e es la matriz de covarianzas del estimador eficiente
 n son los grados de libertad (número de variables)

Por tanto, si se rechaza H_0 se deben asumir las estimaciones de efectos fijos. En caso contrario, el estimador más eficiente, el del efectos aleatorios, debe ser seleccionados.

Los enfoques de panel tienen ventajas significativas sobre las regresiones transversales, no sólo por el tipo de datos, sino porque controlan variables omitidas no observadas.

Una preocupación que ha existido desde hace bastante tiempo es que el efecto de la variación del tiempo sobre los impactos económicos no puede ser utilizado para ver el efecto del cambio climático ya que las oscilaciones del tiempo atmosférico a corto plazo pueden ser diferentes al cambio permanente del clima.

El artículo *Estimating the Economic Impacts of Climate Change Using Weather Observations* [Kolstad and Moore, 2020] estudia diferentes métodos para estimar los impactos climáticos desde diferentes perspectivas, tales como el enfoque clásico transversal, los enfoques de panel más recientes que usan la variación interanual del tiempo y los enfoques híbridos.

Para los autores, el tiempo atmosférico tiene una aleatoriedad inherente, es decir, se puede considerar como una distribución de probabilidad. De tal manera, el clima se definirá como la distribución de probabilidad completa de las estadísticas meteorológicas oportunas. Por tanto, el cambio climático es un cambio en la distribución de probabilidad de la que se consigue la estadística meteorológica en cada periodo de tiempo.

Entonces, ¿qué ocurrirá si el clima cambia? Pues que cambiará el tiempo que enfrenta el agente. Si éste conoce el cambio en el clima, podrá responder de manera eficiente compensando los costos del cambio. Otro punto para analizar es el grado de cambio en el clima. Un cambio instantáneo en el clima provoca la incapacidad de anticipación y la reducción del bienestar aumentando los costos de ajuste. Dado un cambio gradual en el clima, los agentes pueden preveer dicho cambio y por ende, los costos de ajuste serán menores.

Para cuantificar los daños causados por el cambio climático es necesario contabilizar los costos económicos netos que incluyen los costos de equilibrio (abarcando los daños residuales como los costos de adaptación) y los costos de ajuste y poder así determinar empíricamente

estos daños. Para estimar los efectos del tiempo y del clima estudiaremos varios métodos empíricos.

4.2.1. Enfoques de paneles transversales y lineales

A continuación se describirán los enfoques transversales y lineales para estimar los efectos del tiempo y del clima. Destacar la capacidad del enfoque transversal para notificar los efectos a largo plazo y la capacidad del enfoque de panel lineal para establecer con precisión los impactos a corto plazo del tiempo.

4.2.1.1. Modelos de sección transversal

El enfoque transversal proporciona una estimación de la respuesta a largo plazo al cambio climático y, en consecuencia, estimar los costos de equilibrio de un cambio en el clima.

En cuanto a la especificación econométrica, contiene una variable dependiente que varía en el espacio en un determinado instante y se modela en función del clima y otras variables observables; y por último, el término de error.

La principal ventaja para usar este enfoque es que estima los efectos de equilibrio a largo plazo al incluir los beneficios netos de todas las adaptaciones posibles. En cuanto a las desventajas, existen dos principalmente. La primera es si el conjunto de variables transversales observables establecidas como controles son capaces de eliminar todas las variables de confusión que sesgarían las estimaciones del efecto del clima si se prescindieran de la ecuación de estimación. O sea, si los factores no observables están correlacionados con el clima y también influyen en el resultado, en tal caso, el efecto estimado del clima estará sesgado. El segundo inconveniente es definir las variables climáticas como variables explicativas ya que, por ejemplo, el clima no es directamente observable. Por un lado, si el clima es estacionario, la distribución del tiempo puede usarse como aproximación imperfecta del clima. En cambio, si el clima cambia, su estimación es algo más complicada.

4.2.1.2. Modelos de panel lineal

Debido a la inquietud de los enfoques transversales, se han empezado a emplear los datos de panel para la estimación de estos impactos. Puesto que los datos de panel incorporan observaciones de muchas unidades durante muchos periodos de tiempo, es probable controlar las diferencias invariantes en el tiempo entre las unidades y los datos.

Respecto a la especificación econométrica, la variable dependiente, como la producción o el ingreso, varía con el tiempo y el espacio y se modela como una función lineal de una variable meteorológica (que varía también con el tiempo y el espacio), un efecto fijo (que varía en el espacio), un efecto fijo (que varía en el tiempo) y un término de error. Los efectos fijos controlan las diferencias no observables entre unidades y entre periodos de tiempo, lo que aumenta la confianza de que la relación estimada entre el resultado de interés y la variable meteorológica no esté sesgada por las variables omitidas.

El coeficiente del tiempo se calcula haciendo el promedio al efecto de la variación de la serie temporal meteorológica en cada ubicación. Para las diferencias en el clima entre distintas ubicaciones se encarga el efecto fijo y no afecta a los resultados económicos. La variación utilizada para estimar el impacto del tiempo suele proceder de choques

meteorológicos temporales e inesperados. Ya que no se esperarían inversiones a corto plazo como respuesta a estos cambios repentinos, este enfoque sería suficiente para estimar el equilibrio a largo plazo si los costes de adaptación y ajuste son pequeños. En los casos en los que la adaptación sea esencial, la respuesta a corto plazo puede provocar sesgos en la estimación de los daños. Otro dilema, es que si bien las variables omitidas no son un problema, las variables omitidas en el tiempo sí lo podrían ser.

4.2.2. Enfoques híbridos emergentes

Existen enfoques que emplean datos de panel de manera innovadora. Combinan la variación transversal, interanual y decenal atmosférica para estimar los efectos del cambio climático. Estos son los enfoques híbridos. Se dividen en tres categorías: modelos de panel que permiten efectos marginales heterogéneos, modelos de diferencias largas y modelos que dividen la variación disponible en los datos de panel.

4.2.2.1. Modelos de panel no lineal y métodos multietapa para estimar efectos marginales heterogéneos

Los datos de panel contienen información sobre el efecto de la variabilidad meteorológica interanual en numerosas ubicaciones y por eso se usan para modelar la heterogeneidad.

4.2.2.1.1. Modelos de panel no lineal

Los modelos de panel no lineal representan resultados económicos como una función no lineal del tiempo. El uso de ésta quiere decir que el efecto marginal de un cambio meteorológico es en sí mismo una función del tiempo. Si bien el efecto del tiempo se elimina gracias al efecto fijo de ubicación, el efecto marginal esperado de un cambio meteorológico en lugar determinado depende de la distribución del tiempo en dicho lugar. Como resultado, la no linealidad en el tiempo permite que el efecto marginal del tiempo varíe con el clima en todas las ubicaciones.

Las estimaciones de panel no lineal no manifiestan sólo la respuesta a corto plazo al cambio climático. Esta respuesta junto a la de a largo plazo dependerá de la variación transversal y de series temporales en el conjunto de datos.

4.2.2.1.2. Modelos multietapa

Sin desviarnos de los modelos de panel no lineales, éstos modelan el efecto del tiempo y del clima en dos pasos. En primer lugar, El efecto de la variación meteorológica se estima para cada unidad del conjunto de datos, lo que provoca que la respuesta varíe según el espacio. En segundo lugar, el coeficiente del tiempo se modela en función del clima y de otras variables de control.

En cuanto a las ventajas de emplear efectos marginales heterogéneos, la principal es que se pueden seguir utilizando para evaluar el efecto a largo plazo manteniendo los efectos fijos para controlar variables inobservables invariantes en el tiempo que, en otros términos, podrían sesgar la estimación de una regresión transversal estándar, por ejemplo.

No obstante, podría haber problemas con las variables omitidas ya que como la variación en el efecto marginal de las fluctuaciones meteorológicas se evalúa con la variación transversal, es posible insertar un sesgo de variable omitida si estas variables invariantes en el tiempo no se incorporan en la regresión.

4.2.2.2. Modelos de diferencias largas

En este caso, estos modelos se basan en que el tiempo atmosférico varía aleatoriamente en un rango de escalas de tiempo. Debido a que la ecuación de estimación normalmente incluye un término de intersección (que elimina las tendencias promedio en el tiempo), el poder estadístico del modelo emana de las variaciones a mediano y largo plazo.

Por lo tanto, este enfoque es sustancialmente una regresión transversal que emplea tendencias a largo plazo para suprimir el efecto de las variables omitidas invariantes en el tiempo que son un problema en la regresión transversal estándar. No obstante, una variable que esté correlacionada con las tendencias meteorológicas y las tendencias de la variable resultado provocará un sesgo si no se controla. Por ello, es fundamental que las tendencias a mediano plazo sean aleatorias.

4.2.2.3. Modelos de variación de particionamiento

Este último enfoque se refiere a que los datos de panel engloban variaciones tanto en el tiempo como en el clima para estimar conjuntamente los efectos de la variación a corto y largo plazo. A esto se le denomina “variación de partición” porque descompone la variación en la variable resultado en la parte asociada con el clima y la parte relacionada con las fluctuaciones interanuales alrededor del promedio climatológico, indicando la primera el efecto a largo plazo y la segunda el impacto a corto plazo.

Pese a que este enfoque utiliza datos de panel, la función a largo plazo se estima con la variación transversal y se obtiene los mismos inconvenientes que para el modelo anterior. Para combatirlos, ciertos autores han propuesto una pequeña modificación que se basa en que el clima de una ubicación varíe con el tiempo para estimar los efectos a corto y largo plazo en un panel que comprende efectos fijos de periodo y ubicación. Así, nos permite el control de las variables omitidas invariantes en el tiempo a través del efecto fijo de ubicación. Eso sí, observa el efecto a mediano plazo, no a largo plazo.

Una desventaja de dividir la variación es conocer cómo medir el clima a largo plazo. Si el clima no es estacionario, esto inducirá a un error de medición y sesgará la estimación de la respuesta a corto plazo.

Para finalizar, hemos visto diferentes enfoques para estimar los efectos a corto, mediano y largo plazo. Sin embargo, para identificar los costos totales asociados al cambio climático habría que calcular la tasa y el costo de la adaptación a un cambio en el clima.

4.3. Modelos aditivos generalizados

El cambio climático, la contaminación del aire y los riesgos para la salud son sectores en los que existe una gran preocupación a nivel mundial. Numerosos estudios han analizado cómo los factores meteorológicos junto con la contaminación del aire pueden influir en la salud y bienestar de la sociedad.

Los métodos empleados por los científicos constan de análisis de series temporales con modelos lineales generalizados (GLM) o modelos aditivos generalizados (GAM) utilizando funciones paramétricas y no paramétricas para estimar los efectos en la salud en periodos largos y cortos de estos factores. En particular, el modelo aditivo generalizado es útil

para modelar variables no lineales. Para desarrollar el GAM, debemos antes introducir los modelos lineales generales y los modelos lineales generalizados.

El modelo lineal general se usa como modelo principal para muchas aplicaciones de análisis estadístico. La generalización matemática de las regresiones múltiples puede transformarse en modelos lineales generales. La ecuación de regresión lineal más popular con m regresores, para predecir una variable de regresión es

$$Y = b_0 + b_1X_1 + \cdots + b_mX_m$$

donde Y es la variable dependiente, $X_1 \dots X_m$ son los valores para las variables predictoras y los $b_0 \dots b_m$ son los coeficientes de regresión.

En los GLM, la variable dependiente puede ser no normal y no continua. Una combinación lineal de las variables predictoras predice los valores de la variable dependiente, que se pueden relacionar a dicha variable gracias a una función de enlace. Veámoslo matemáticamente:

$$Y = g(b_0 + b_1X_1 + \cdots + b_mX_m) \quad \text{donde } g \text{ es una función}$$

La función de enlace es

$$g(\mu Y) = b_0 + b_1X_1 + \cdots + b_mX_m$$

donde $g(\mu Y)$ indica el valor esperado de la variable respuesta Y

Los GAM proporcionan una estructura para generalizar un modelo lineal general al permitir la adición de funciones no lineales de las variables. La ventaja de los GAM es limitar el error en la predicción de una variable dependiente de varias distribuciones mediante la evaluación de funciones no específicas que están relacionadas con la variable dependiente por la función de enlace. Este modelo permite tomar una gran variedad de distribuciones para la variable respuesta y enlazar las funciones para medir los efectos de las variables predictoras. Las distribuciones más utilizadas en GAM son Normal, Gamma y Poisson.

La principal aplicación de GAM se concentra en la contaminación del aire. Muchos modelos de series temporales que utilizan GAM y diferentes distribuciones muestran la relación entre la calidad del aire y los riesgos para la salud.

Respecto al ajuste de los modelos aditivos generalizados, el GLM se orientó para la relación no lineal de predicción para los predictores y la variable. En el proceso de iteración de GLM están involucrados dos algoritmos: bucle externo y bucle interno. Éstos ayudan a estimar los valores de los estimadores usando ecuaciones integrales de ajuste posterior no lineales, que no se pueden resolver explícitamente.

La base del modelo se basa en “suavizados”. Se pueden agrupar en 3 tipos: splines de regresión, regresión local y splines de suavizado. Los splines de regresión se usan por su facilidad y se pueden escribir como combinación lineal de la función base. La regresión local se basa en los tipos de suavizadores del vecino más cercano. Finalmente, los splines de suavizado estiman la función suave minimizando la suma de cuadrados penalizada. El principal inconveniente es el número efectivos de grados de libertad utilizados que son menores que la cantidad de nudos.

Haciendo referencia a los grados de libertad, éstos son una preocupación importante para ajustar estos modelos. Los grados de libertad dependen de los parámetros de suavizado. Si todos los parámetros de suavizado son muy altos entonces el modelo será bastante inflexible, con muy pocos grados de libertad. En particular, GAM requiere más grados de libertad que un modelo de regresión lineal regular.

Otro aspecto clave de los GAM es el “retraso”. Se denomina como tiempo de retraso al tiempo entre las dos series temporales que están correlacionando. En estos modelos relacionados con la calidad del aire, este valor se toma en función del objetivo del efecto a corto o largo plazo.

La estimación más importante con un GAM es predecir siempre la mejor estimación de la variable dependiente y las variables dependientes se asocian con predictores. El impacto de los predictores se puede evaluar en el modelo ajustado final.

Como cualquier modelo, el GAM posee ventajas y desventajas. En cuanto a los puntos positivos, GAM permite ajustar relaciones no lineales sin tener que realizar transformaciones manuales, realiza predicciones más precisas para la variable respuesta en un sistema de datos no lineal, proporciona el efecto para una variable X_i sobre Y manteniendo constante otra variable, los grados de libertad se pueden emplear para aclarar la suavidad de la función y también se puede usar con datos cualitativos. Respecto a los aspectos desfavorables, el mayor es la aditividad ya que restringe el impacto de varias interacciones que se implementarían en el modelo y los términos de adición e interacción debe ser añadidos manualmente por el usuario.

En otras palabras, los modelos aditivos generalizados proporcionan el mejor ajuste para relaciones no lineales entre el predictor y la variable independiente. No obstante, se debe prestar atención al aplicarlos ya que un sobreajuste del modelo suele causar un error al aplicar un modelo complejo. Asimismo, hay que tener precaución a la hora de interpretar los resultados de GAM y modelos lineales generales. De tal manera se aconseja emplear modelos simples para predecir casos futuros y usar GAM para aquellos casos no lineales.

Como conclusión, para analizar los efectos a corto plazo se ha propuesto GAM con spline penalizado como mejor opción para estudiar la relación entre el medioambiente, clima y salud.

Capítulo 5

Ejemplos de aplicaciones

En este capítulo se lleva a cabo la aplicación práctica de técnicas estadísticas para datos acerca de la temperatura, precipitaciones, concentración de los gases de efecto invernadero y aumento del nivel del mar, con el fin de contrastar la existencia del cambio climático. Para ello vamos a obtener varios conjuntos de datos.

5.1. Modelización de los gases de efecto invernadero

5.1.1. Descripción de los datos

La base de datos de la que se dispone contiene la serie temporal con los valores medios mensuales globales al nivel del mar, centrados en la mitad de cada mes, de los gases de efecto invernadero más representativos que provocan un aumento del calentamiento global. Estos datos se han extraído de la NOAA por separado. La División de Monitoreo Global de NOAA ha medido el dióxido de carbono, el óxido nitroso y el metano durante varias décadas en una red distribuida globalmente de sitios de muestreo de aire. Los datos perteneciente al dióxido de carbono se encuentran en (<https://doi.org/10.15138/9N0H-ZH07>) y los correspondientes a los otros dos en (<https://doi.org/10.15138/P8XG-AA10>).

Además, se tienen recogidas la tendencia (después de la corrección del ciclo estacional promedio) y la incertidumbre de cada media mensual basada en la variabilidad observada de cada GEI. Los datos se anuncian como una fracción molar (ppm) de aire seco definida como el número de moléculas del GEI en específico dividido por el número de todas las moléculas en el aire, eliminado el vapor de agua.

En primer lugar, realizamos un breve análisis descriptivo de los datos almacenados en el fichero “gei.RData”. Seleccionamos las variables que muestran los datos medios mensuales de dióxido de carbono (CO_2), óxido nitroso (N_2O) y metano (CH_4).

```
load("datos/GEI.RData")  
  
# Selección de las variables que nos interesan  
library(dplyr)  
GEI<-select(GEI, Año, Mes, Decimal, MediaCO2, MediaN2O, MediaCH4)
```

Debido a que los valores de cada gas han comenzado a tomarse en un año distinto, aparecen valores perdidos al haberlos agrupado en una sola base de datos. De tal manera,

para realizar el resumen descriptivo se van a utilizar los años comunes, es decir, el periodo que abarca desde enero de 2001 hasta enero de 2023.

```
gei<-filter(GEI, (Año>=2001 & Decimal!=2023.125))
```

Se realiza una exploración de los datos para conocer la estructura y un resumen de éstos

```
str(gei) # Estructura de los datos
```

```
## 'data.frame': 265 obs. of 6 variables:
## $ Año : int 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 2001 ...
## $ Mes : Factor w/ 12 levels "Enero","Febrero",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Decimal : num 2001 2001 2001 2001 2001 ...
## $ MediaCO2: num 371 371 372 372 372 ...
## $ MediaN2O: num 0.316 0.316 0.316 0.316 0.316 ...
## $ MediaCH4: num 1.77 1.77 1.77 1.77 1.77 ...
```

```
summary(gei) # Resumen de las variables
```

```
##      Año           Mes           Decimal           MediaCO2           MediaN2O
## Min.   :2001   Enero   : 23   Min.   :2001   Min.   :368.3   Min.   :0.3161
## 1st Qu.:2006   Febrero: 22   1st Qu.:2007   1st Qu.:381.3   1st Qu.:0.3198
## Median :2012   Marzo   : 22   Median :2012   Median :392.0   Median :0.3249
## Mean   :2012   Abril   : 22   Mean   :2012   Mean   :393.0   Mean   :0.3251
## 3rd Qu.:2017   Mayo    : 22   3rd Qu.:2018   3rd Qu.:405.3   3rd Qu.:0.3297
## Max.   :2023   Junio   : 22   Max.   :2023   Max.   :419.4   Max.   :0.3365
##                (Other):132
##      MediaCH4
## Min.   :1.763
## 1st Qu.:1.780
## Median :1.806
## Mean   :1.818
## 3rd Qu.:1.850
## Max.   :1.925
##
```

Seguidamente, se muestra una representación gráfica de la abundancia de los principales gases de efecto invernadero.

Gases de efecto invernadero 2001–2023

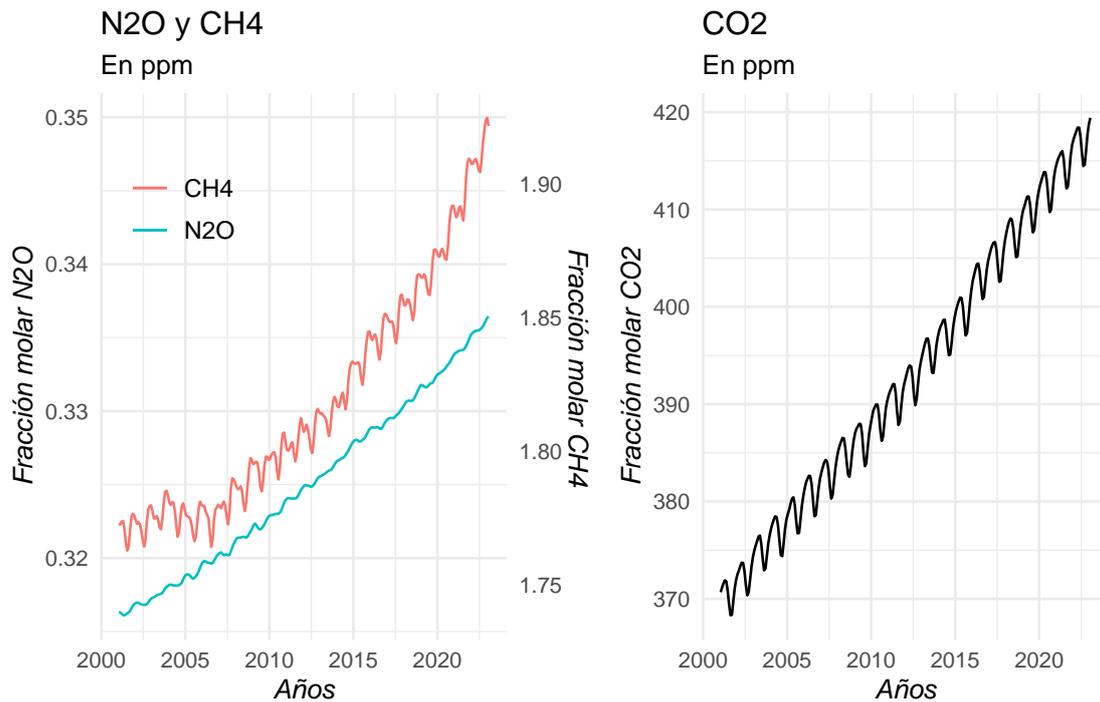


Figura 5.1: Evolución de la concentración de GEI

A simple vista se observa que la abundancia de los gases se ha ido incrementando con el paso de los años.

Con relación a dicho incremento, el óxido nitroso parece que su tendencia es más suave, no hay mucha variación entre los meses de cada año. Por su parte, el metano y el dióxido de carbono presentan una conducta similar ya que anualmente existe un comportamiento parecido con más cantidad en unos meses y menos en otros. En cuanto al metano, en la franja de 2005 a 2007 aproximadamente, ha habido un retroceso en abundancia.

Por otro lado, con ayuda de la escala de los ejes, observamos que el CO_2 es el gas de efecto invernadero más abundante en la atmósfera.

Como objeto de estudio, se va aplicar un modelo de series temporales a la variable correspondiente a los valores medios de CO_2 .

5.1.2. Modelo de series temporales

Una serie temporal se define como una sucesión de observaciones correspondientes a una variable en distintos momentos de tiempo. Los objetivos principales de estos modelos son describir la evolución que la serie ha tenido en el pasado y predecir sus valores respecto a un futuro cercano.

Para realizar este modelo, se aplicará la **metodología de Box-Jenkins** con el fin de encontrar el mejor ajuste de la serie temporal y así generar pronósticos más acertados.

Se pretende llevar a la práctica este estudio a fin de observar la evolución de la concentración de CO_2 en la atmósfera desde enero de 2001 a diciembre de 2022.

Por ello, establecemos los nuevos datos:

```
datos_CO2<-select(gei,Año,Mes,Decimal,MediaCO2)[gei$Año<2023,]
```

Podemos ya crear la serie temporal con la función `ts` y representarlos.

```
datos_st<-ts(datos_CO2$MediaCO2,frequency = 12,start = c(2001,1))
```

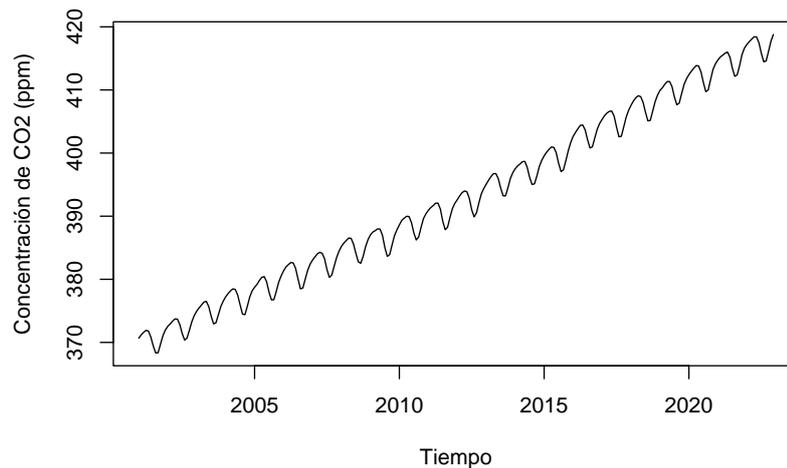


Figura 5.2: Valores medios mensuales de CO2 (2001-2022)

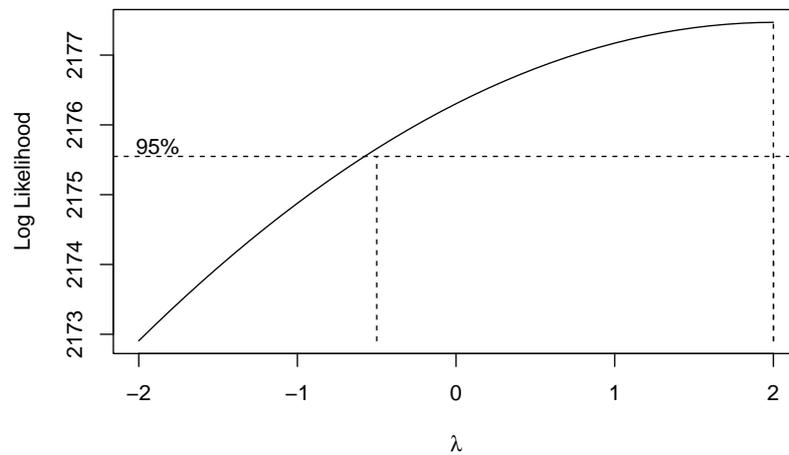
A continuación, pasamos a comprobar y a relizar las transformaciones necesarias para que la varianza y la media sean estables en el tiempo.

En primer lugar, mediante el gráfico de la serie, no parece que haya que aplicar alguna transformación para la varianza ya que se aprecia que es constante. Aun así, vamos a comprobarlo con la familia de **transformaciones de Box-Cox**:

$$Y_t^{(\lambda)} = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(Y_t) & \lambda = 0 \end{cases}$$

Para ello empleamos la función `BoxCox.ar` del paquete `TSA`.

```
#install.packages("TSA")
library(TSA)
bc=BoxCox.ar(y=datos_st, method = "ols"); bc
```



```
## $lambda
## [1] -2.0 -1.9 -1.8 -1.7 -1.6 -1.5 -1.4 -1.3 -1.2 -1.1 -1.0 -0.9 -0.8 -0.7 -0.6
## [16] -0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
## [31] 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0
##
## $loglike
## [1] 2172.906 2173.127 2173.343 2173.554 2173.759 2173.959 2174.153 2174.342
## [9] 2174.526 2174.704 2174.877 2175.045 2175.207 2175.363 2175.514 2175.659
## [17] 2175.799 2175.934 2176.063 2176.186 2176.304 2176.416 2176.522 2176.623
## [25] 2176.718 2176.808 2176.892 2176.970 2177.043 2177.110 2177.171 2177.227
## [33] 2177.277 2177.321 2177.360 2177.392 2177.419 2177.440 2177.456 2177.466
## [41] 2177.469
##
## $mle
## [1] 2
##
## $ci
## [1] -0.5 2.0
```

Como el $1 \in IC$ para λ , no transformamos los datos.

En segundo lugar, en cuanto a la estabilidad de la media en el tiempo, hacemos uso de la función de autocorrelación muestral del proceso *FAS* con el comando `acf`.

```
acf(datos_st, lag=50, main="FAS de CO2")
```

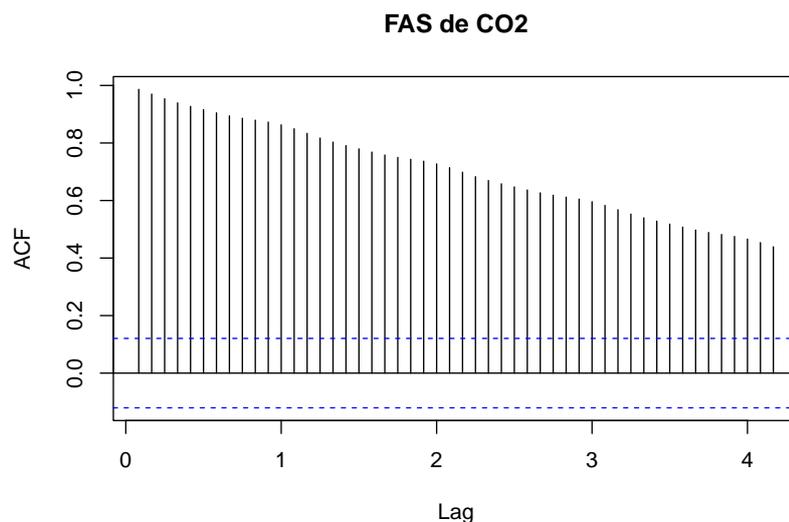


Figura 5.3: FAS de datos CO2

Desciende muy lentamente, lo que indica que se trata de un proceso integrado. Aplicamos una diferencia regular:

```
datos_st_d1<-diff(datos_st, lag = 1, diff=1)
```

Veamos qué aspecto tiene la serie tras una diferencia regular:

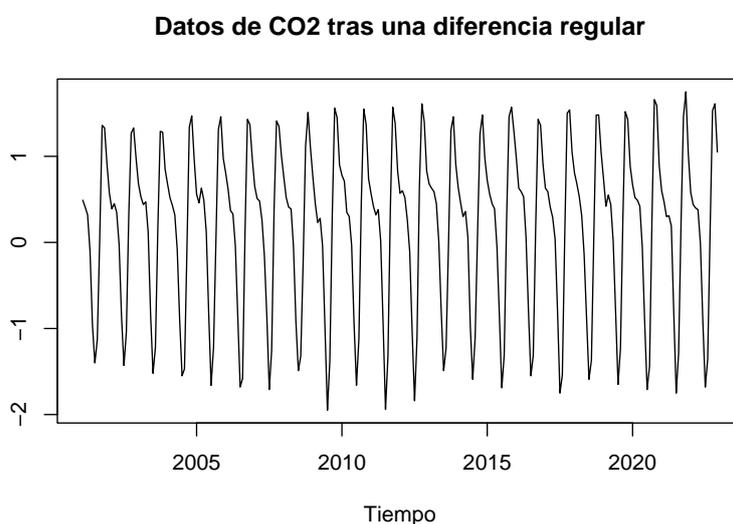


Figura 5.4: Datos tras una diferencia regular

Ahora miramos la FAS:

```
acf(datos_st_d1,lag=50, main="FAS de primera diferencia")
```

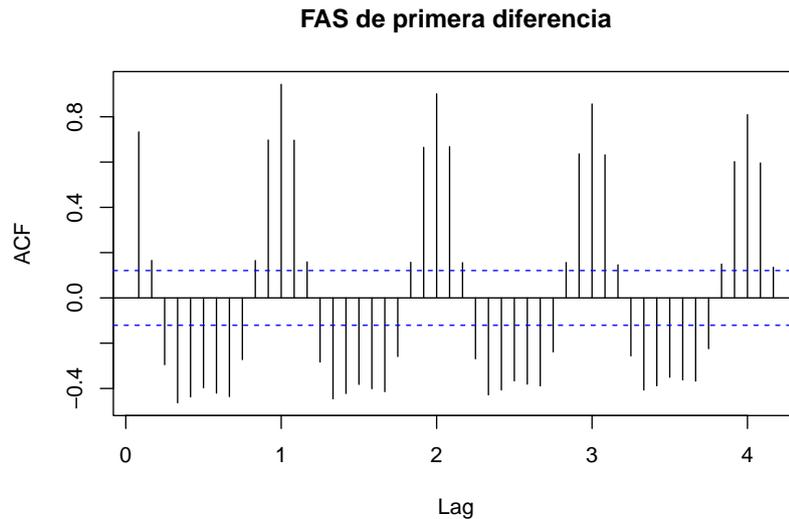


Figura 5.5: FAS tras la primera diferencia

Las autocorrelaciones estacionales descienden también lentamente, así que aplicamos una diferencia estacional:

```
datos_st_d<-diff(datos_st_d1,lag = 12,diff=1)
acf(datos_st_d,lag=50, main="FAS de tras una diferencia regular
y otra estacional")
```

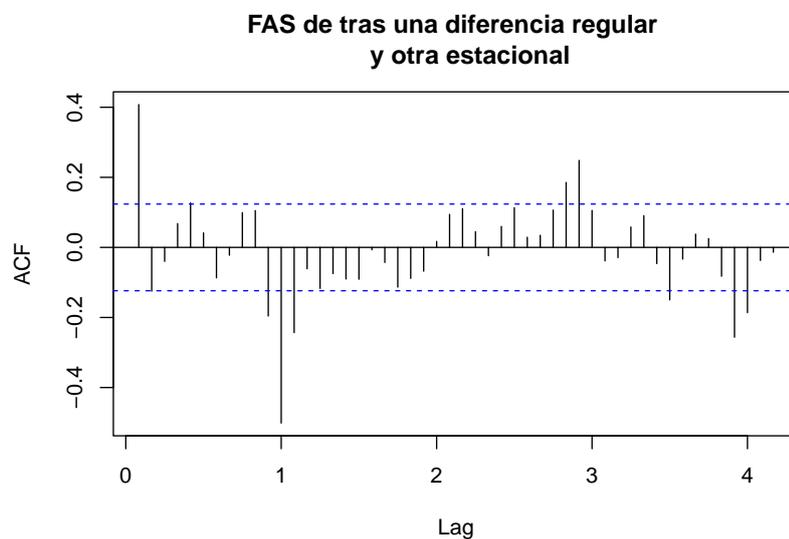


Figura 5.6: FAS tras una diferencia regular y otra estacional

Ahora el gráfico de la FAS tiene el aspecto de un proceso estacionario (no integrado). La serie tras una diferencia regular y otra estacional queda así:

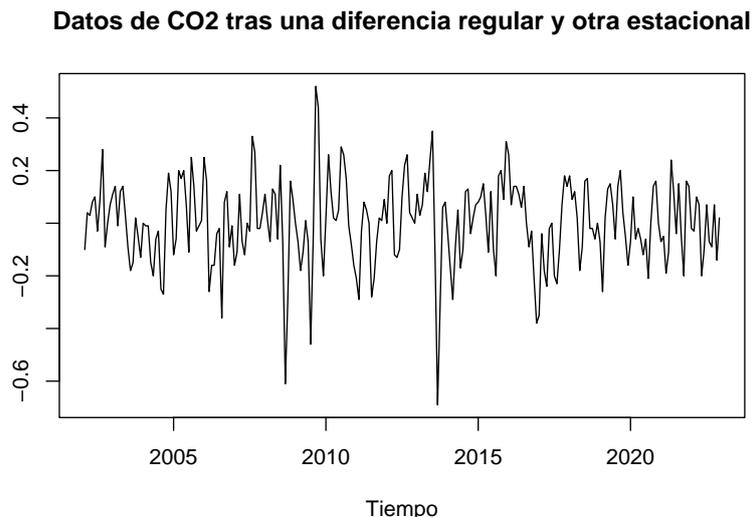


Figura 5.7: Datos tras una diferencia regular y otra estacional

Seguidamente contrastamos la estacionariedad a través del **test de raíz unitaria de Dickey-Fuller** con las siguientes hipótesis:

$$\begin{cases} H_0 : \text{El polinomio autorregresivo tiene 1 raíz unitaria} \\ H_1 : \text{Las raíces del polinomio autorregresivo son estacionarias} \end{cases}$$

Para ello, debemos cargar la librería `tseries`.

```
library(tseries)
adf.test(datos_st_d)

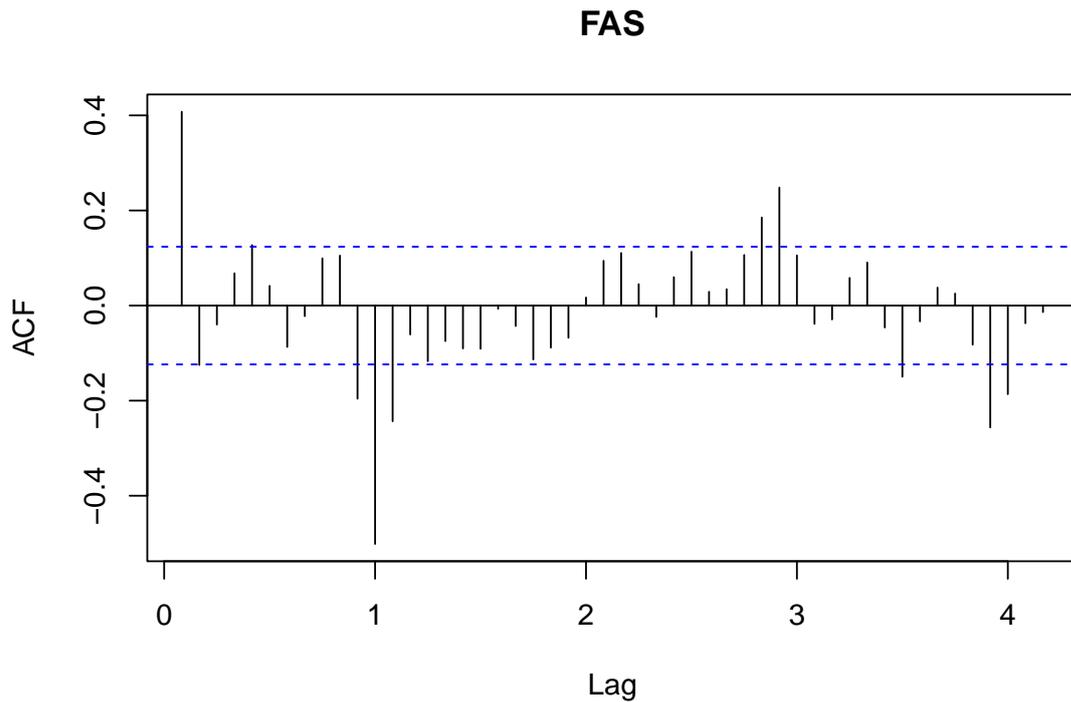
##
## Augmented Dickey-Fuller Test
##
## data:  datos_st_d
## Dickey-Fuller = -5.4067, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

Como $p\text{-value} = 0.01$ rechazamos la hipótesis nula, o sea, las raíces del polinomio autorregresivo son estacionaria, la serie transformada se puede considerar estacionaria.

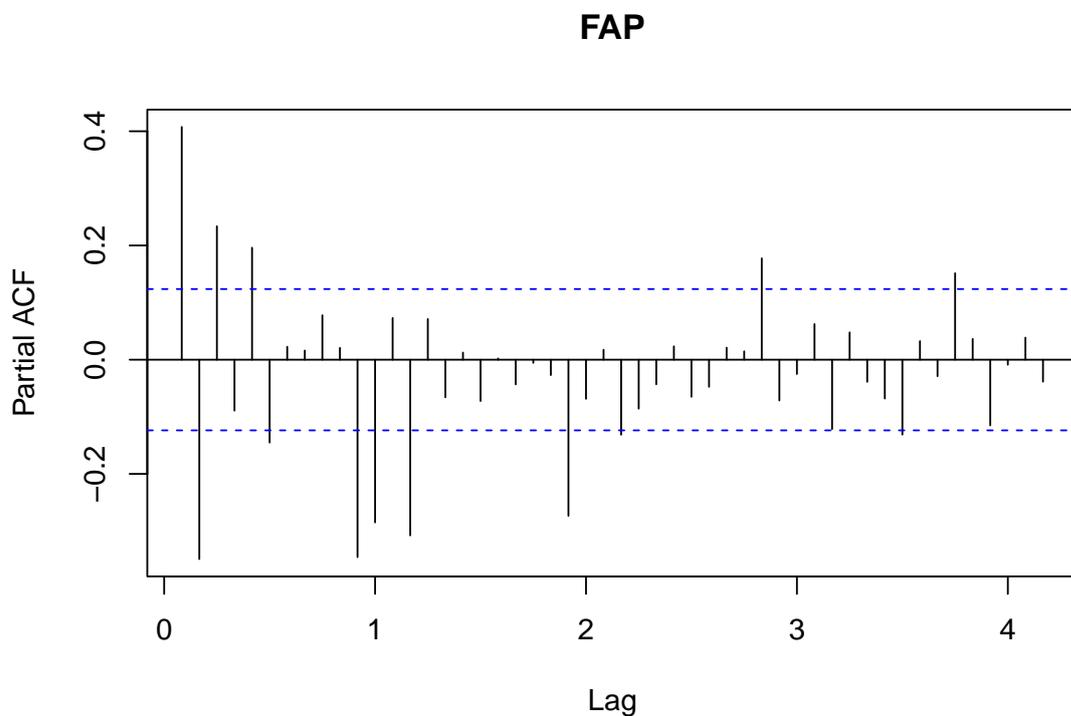
Seguidamente establecemos la estructura *ARIMA*. Se puede proceder de dos formas:

En el **primer caso**, se identifica manualmente con las funciones de autocorrelación simple (*FAS*) y parcial (*FAP*).

```
acf(datos_st_d, lag=50, main="FAS")
```



```
pacf(datos_st_d, lag=50, main="FAP")
```



Respecto a la FAS, vemos una autocorrelación no nula (o quizás dos) en los retardos bajos y otra en el retardo 12. Con relación a la FAP, en los retardos bajos no se anula y en la parte estacional, sí. Por ello, es conveniente considerar los siguientes modelos: $MA(1) \times AR(1)_{12}$, $MA(1) \times AR(2)_{12}$, $MA(2) \times AR(1)_{12}$ y $MA(2) \times AR(2)_{12}$.

Vamos a comprobarlos:

Modelo 1: $MA(1) \times AR(1)_{12}$:

```
ajuste1=arima(datos_st_d,order = c(0,0,1),
              seasonal = list(order=c(1,0,0),period=12));ajuste1

##
## Call:
## arima(x = datos_st_d, order = c(0, 0, 1), seasonal = list(order = c(1, 0, 0),
##   period = 12))
##
## Coefficients:
##          ma1          sar1  intercept
##          0.8259  -0.5220    0.0012
## s.e.    0.0377   0.0535    0.0084
##
## sigma^2 estimated as 0.01207:  log likelihood = 195.7,  aic = -385.4
```

Los coeficientes estimados son significativamente no nulos.

```
confint(ajuste1)

##              2.5 %      97.5 %
## ma1          0.75203491  0.89981190
## sar1         -0.62689434 -0.41712335
## intercept -0.01536102  0.01773087
```

Como la media del modelo incluye al 0, la eliminamos.

```
ajuste1=arima(datos_st_d,order = c(0,0,1),
              seasonal = list(order=c(1,0,0),period=12),
              include.mean = FALSE);ajuste1

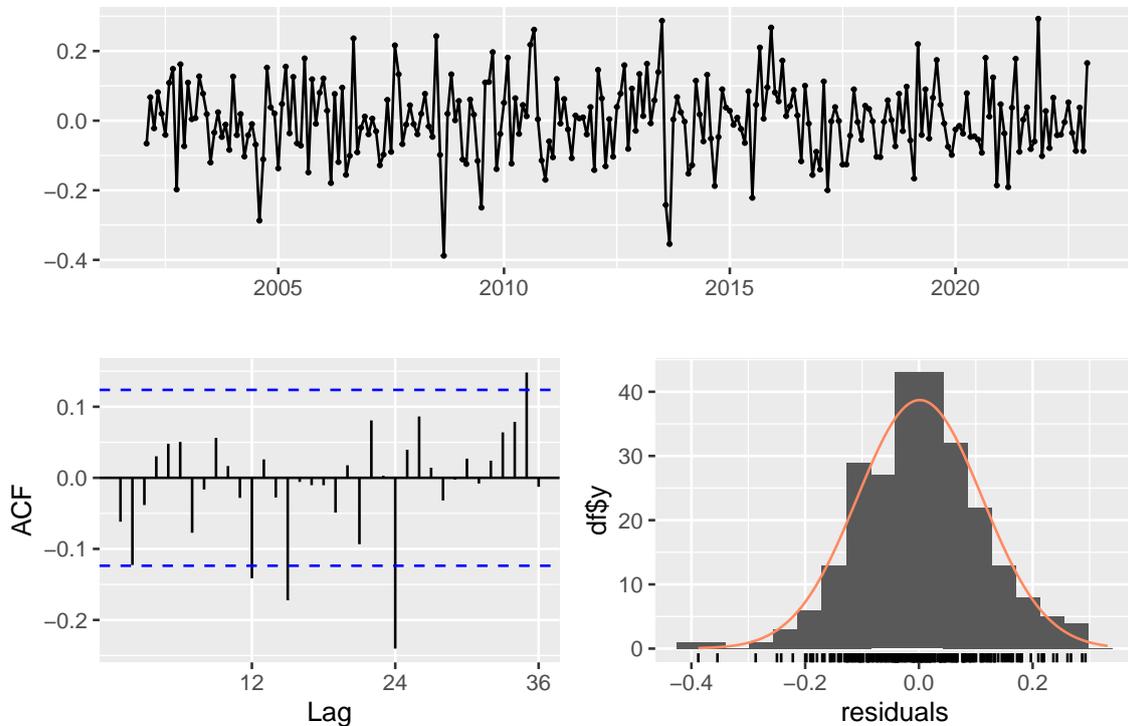
##
## Call:
## arima(x = datos_st_d, order = c(0, 0, 1), seasonal = list(order = c(1, 0, 0),
##   period = 12), include.mean = FALSE)
##
## Coefficients:
##          ma1          sar1
##          0.8259  -0.5220
## s.e.    0.0377   0.0535
##
## sigma^2 estimated as 0.01207:  log likelihood = 195.69,  aic = -387.38
```

Una vez estimado el modelo es necesario examinar de manera cuidadosa si el modelo es considerado adecuado. Si es así, entonces la serie se debería comportar como un ruido blanco. Se utiliza el **test de Ljung-Box**.

En el software **R**, es necesario instalar el paquete `forecast` con la función `checkresiduals`.

```
#install.packages("forecast")
library(forecast)
checkresiduals(ajuste1)
```

Residuals from ARIMA(0,0,1)(1,0,0)[12] with zero mean



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(0,0,1)(1,0,0)[12] with zero mean
## Q* = 44.276, df = 22, p-value = 0.003275
##
## Model df: 2. Total lags used: 24
```

Desechamos el primer modelo porque no pasa este test. Se rechaza que las innovaciones están incorreladas.

Modelo 2: $MA(1) \times AR(2)_{12}$:

```
ajuste2=arima(datos_st_d,order = c(0,0,1),
              seasonal = list(order=c(2,0,0),period=12));ajuste2
```

```
##
## Call:
## arima(x = datos_st_d, order = c(0, 0, 1), seasonal = list(order = c(2, 0, 0),
##   period = 12))
##
## Coefficients:
##          ma1          sar1          sar2  intercept
##          0.7981    -0.6732    -0.2874         0.0010
## s.e.        0.0391     0.0608     0.0613         0.0063
```

```
##
## sigma^2 estimated as 0.01103: log likelihood = 205.95, aic = -403.91
```

Los coeficientes estimados son significativamente no nulos.

```
confint(ajuste2)
```

```
##              2.5 %      97.5 %
## ma1          0.72158279  0.87466175
## sar1         -0.79232245 -0.55408267
## sar2         -0.40757926 -0.16718401
## intercept -0.01126239  0.01328871
```

Como la media del modelo incluye al 0, la eliminamos.

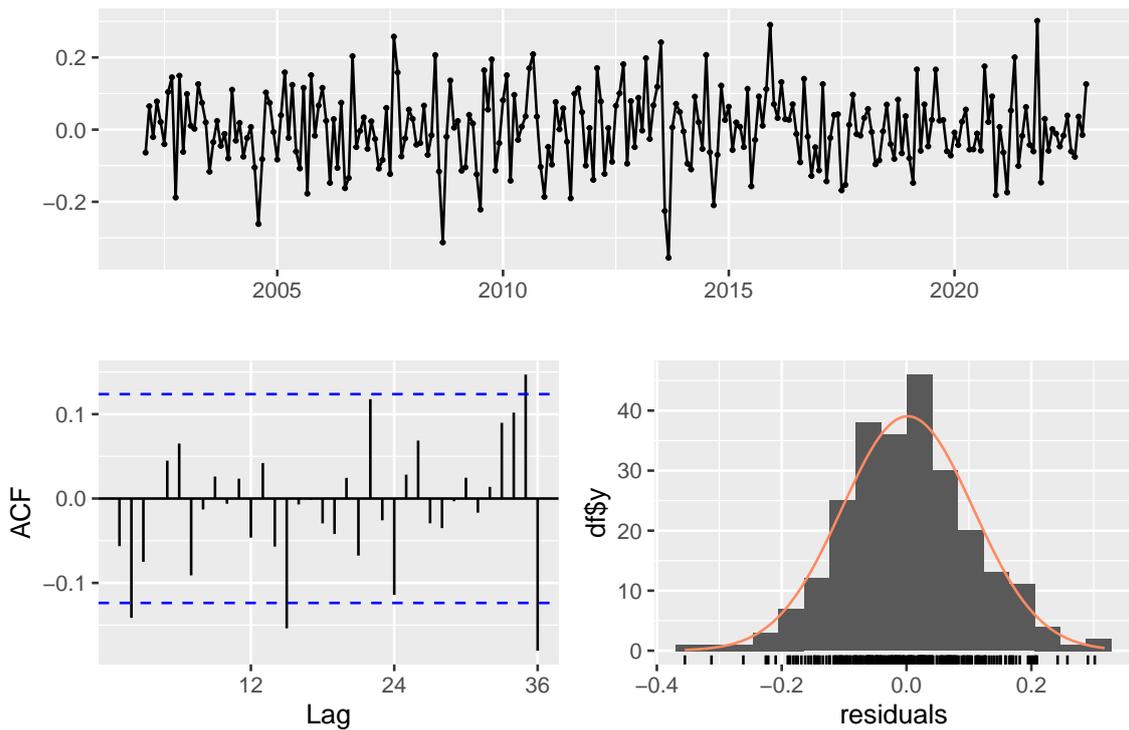
```
ajuste2=arima(datos_st_d,order = c(0,0,1),
              seasonal = list(order=c(2,0,0),period=12),
              include.mean = FALSE);ajuste2
```

```
##
## Call:
## arima(x = datos_st_d, order = c(0, 0, 1), seasonal = list(order = c(2, 0, 0),
##   period = 12), include.mean = FALSE)
##
## Coefficients:
##          ma1      sar1      sar2
##          0.7982 -0.6732 -0.2873
## s.e.  0.0391  0.0608  0.0613
##
## sigma^2 estimated as 0.01103: log likelihood = 205.94, aic = -405.88
```

Comprobamos la diagnosis del modelo.

```
checkresiduals(ajuste2)
```

Residuals from ARIMA(0,0,1)(2,0,0)[12] with zero mean



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(0,0,1)(2,0,0)[12] with zero mean
## Q* = 29.618, df = 21, p-value = 0.09994
##
## Model df: 3. Total lags used: 24
```

Nos quedamos con este modelo porque sí pasa este test para $\alpha = 0.05$. El valor de *AIC* es de -405.88.

Modelo 3: $MA(2) \times AR(1)_{12}$:

```
ajuste3=arima(datos_st_d,order = c(0,0,2),
              seasonal = list(order=c(1,0,0),period=12));ajuste3
```

```
##
## Call:
## arima(x = datos_st_d, order = c(0, 0, 2), seasonal = list(order = c(1, 0, 0),
##   period = 12))
##
## Coefficients:
##      ma1      ma2      sar1  intercept
##    0.7386 -0.1021 -0.5159    0.0011
## s.e. 0.0710  0.0724  0.0538    0.0076
##
## sigma^2 estimated as 0.01197: log likelihood = 196.67, aic = -385.35
```

Los coeficientes estimados son significativamente no nulos.

```
confint(ajuste3)
```

```
##              2.5 %      97.5 %
## ma1          0.5994910  0.87778641
## ma2          -0.2440020  0.03984488
## sar1         -0.6213442 -0.41039972
## intercept   -0.0137189  0.01596058
```

Si eliminamos los términos no significativos, entonces se reduce al primer modelo, que fue descartado.

Modelo 4: $MA(2) \times AR(2)_{12}$

```
ajuste4=arima(datos_st_d,order = c(0,0,2),
              seasonal = list(order=c(2,0,0),period=12));ajuste4
```

```
##
## Call:
## arima(x = datos_st_d, order = c(0, 0, 2), seasonal = list(order = c(2, 0, 0),
##   period = 12))
##
## Coefficients:
##          ma1          ma2          sar1          sar2  intercept
##          0.7093  -0.1036  -0.6677  -0.2865          0.0010
## s.e.    0.0744   0.0756   0.0609   0.0612          0.0056
##
## sigma^2 estimated as 0.01095:  log likelihood = 206.89,  aic = -403.77
```

Los coeficientes estimados son significativamente no nulos.

```
confint(ajuste4)
```

```
##              2.5 %      97.5 %
## ma1          0.563400775  0.85511499
## ma2          -0.251789448  0.04456006
## sar1         -0.787182722 -0.54828312
## sar2         -0.406500959 -0.16646566
## intercept   -0.009998342  0.01192692
```

Si eliminamos los términos no significativos, entonces se reduce al modelo 2, que fue seleccionado.

En el **segundo caso** se recurre a la función `auto.arima`.

```
ajuste5=auto.arima(datos_st_d,step=F); ajuste5
```

```
## Series: datos_st_d
## ARIMA(2,0,1)(1,0,1)[12] with zero mean
##
## Coefficients:
##          ar1          ar2          ma1          sar1          sma1
##          0.0536  -0.1695  0.7116  0.0010  -0.8718
## s.e.    0.0918   0.0793  0.0744  0.0824   0.0729
```

```
##
## sigma^2 = 0.009581: log likelihood = 220.64
## AIC=-429.29 AICc=-428.94 BIC=-408.14
```

Este modelo presenta un *AIC* menor que el ajuste del segundo modelo. No obstante, hay que tener precaución al compararlos, pues han sido calculados con funciones distintas.

```
auto.arima(datos_st_d, stepwise = F)
```

```
## Series: datos_st_d
## ARIMA(2,0,1)(1,0,1)[12] with zero mean
##
## Coefficients:
##      ar1      ar2      ma1      sar1      sma1
##      0.0536 -0.1695  0.7116  0.0010 -0.8718
## s.e.  0.0918  0.0793  0.0744  0.0824  0.0729
##
## sigma^2 = 0.009581: log likelihood = 220.64
## AIC=-429.29 AICc=-428.94 BIC=-408.14
```

Ajustamos la serie diferenciada. Quedarían los modelos de tal manera:

Modelo 2: $ARIMA(0, 1, 1) \times ARIMA(2, 1, 0)_{12}$

```
ajuste_bueno1=arima(datos_st, order = c(0,1,1),
                    seasonal = list(order=c(2,1,0),
                                    period=12)); ajuste_bueno1
```

```
##
## Call:
## arima(x = datos_st, order = c(0, 1, 1), seasonal = list(order = c(2, 1, 0),
##      period = 12))
##
## Coefficients:
##      ma1      sar1      sar2
##      0.7983 -0.6733 -0.2871
## s.e.  0.0390  0.0608  0.0613
##
## sigma^2 estimated as 0.01103: log likelihood = 206.03, aic = -406.06
```

Modelo 5: $ARIMA(2, 1, 1) \times ARIMA(1, 1, 1)_{12}$

```
ajuste_bueno2=arima(datos_st, order = c(2,1,1),
                    seasonal = list(order=c(1,1,1),
                                    period=12)); ajuste_bueno2
```

```
##
## Call:
## arima(x = datos_st, order = c(2, 1, 1), seasonal = list(order = c(1, 1, 1),
##      period = 12))
##
## Coefficients:
##      ar1      ar2      ma1      sar1      sma1
```

```
##          0.0537  -0.1696  0.7115  0.0017  -0.8732
## s.e.    0.0918   0.0793  0.0744  0.0824   0.0734
##
## sigma^2 estimated as 0.009379:  log likelihood = 220.73,  aic = -431.47
```

Una vez comparados ambos, el que tiene menor AIC es el calculado con `auto.arima`, o sea, el modelo $SARIMA(2, 1, 1) X (1, 1, 1)_{12}$.

Ya establecido el modelo, el propósito es predecir los resultados. Predecimos para los próximos 24 meses.

```
pred=predict(ajuste_bueno2,n.ahead = 24)
```

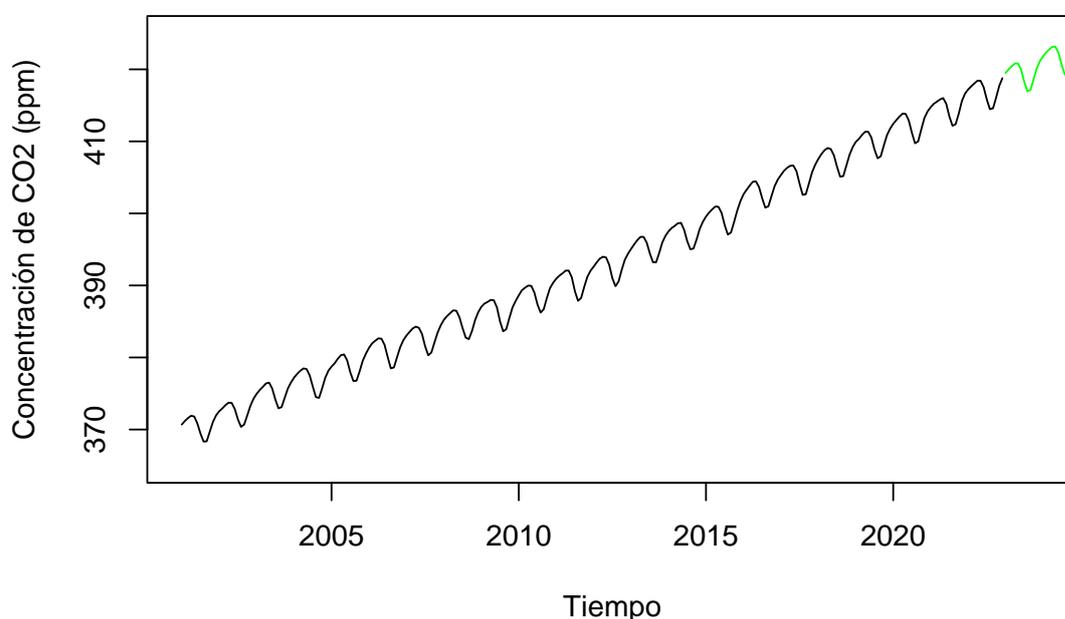


Figura 5.8: Predicción de la concentración de CO2 en la atmósfera

5.2. Modelización de los elementos del clima

5.2.1. Descripción de los datos

En este segundo apartado, se consigue obtener una base de datos referente a los elementos del clima. En meteorología, se define como elementos del clima al conjunto de componentes que caracterizan el tiempo atmosférico y que interactúan entre sí en la troposfera. Los principales son: la temperatura, la precipitación, la humedad, el viento, la presión atmosférica y la nubosidad.

Estos datos se han extraído de la **Ogimet** (<https://www.ogimet.com/gsynres.phtml>) donde se pueden recoger de distintos estados, estaciones y periodos de tiempo. En nuestro caso, el conjunto de datos incluye estos elementos en durante dos años (2003 y 2022), en los meses de enero y julio; y, en las ciudades de Sevilla (España) y Tromso (Noruega).

Se han seleccionado estos dos lugares para evaluar si el cambio de temperatura es muy distinto en cada ciudad.

Como se realizó en el apartado anterior, se comenzará realizando un análisis descriptivo de los datos para ver la estructura y un resumen de los mismos.

```
load("datos/ogi.RData")
str(ogi) # Estructura de los datos

## 'data.frame':    248 obs. of  14 variables:
## $ Año   : chr  "2022" "2022" "2022" "2022" ...
## $ Mes   : Factor w/ 2 levels "Enero","Julio": 1 1 1 1 1 1 1 1 1 1 ...
## $ Dia   : int   1 2 3 4 5 6 7 8 9 10 ...
## $ Lugar: chr   "Sevilla" "Sevilla" "Sevilla" "Sevilla" ...
## $ TMax  : num   19.5 19.5 21.3 16.1 16.2 15.3 15.1 16.3 15.1 16.1 ...
## $ TMin  : num    7.7  6.4  7.8  9.8  5.8  6.6  4.1  5.1  1.3  5.2 ...
## $ TMed  : num   13.4 11.3 12.4 12 11.1 11.3 8.7 9.4 7.1 9.6 ...
## $ HRMed : num   74.7 85.2 82 91.8 73.3 74.4 79.5 71.7 78.9 85.6 ...
## $ VelV  : num    8.4  6.5  6.3  6.8 12 10.9 9 11.7 4.8 4.9 ...
## $ PresM : num  1027 1029 1027 1021 1023 ...
## $ Prec  : num    0 0 0 0.1 0 0.1 0 0 0 0 ...
## $ NubT  : num    1.3 2.1 1.1 6 2.5 1.4 0.4 1.2 1.1 1.5 ...
## $ Vis   : num   30 19.2 22.9 12.3 28.8 30 30 30 29.4 27.5 ...
## $ EspN  : num   NA ...
```

```
summary(ogi) # Resumen de las variables climáticas
```

```
##      Año           Mes           Dia           Lugar
## Length:248      Enero:124      Min.    : 1      Length:248
## Class :character      Julio:124      1st Qu.: 8      Class :character
## Mode  :character           Median :16      Mode  :character
##                               Mean   :16
##                               3rd Qu.:24
##                               Max.   :31
##
##      TMax           TMin           TMed           HRMed
## Min.    :-11.100      Min.    :-13.400      Min.    :-11.70      Min.    :17.20
## 1st Qu.:  9.225      1st Qu.:  0.850      1st Qu.:  4.40      1st Qu.:57.15
## Median : 16.700      Median :  8.600      Median : 12.00      Median :73.00
## Mean   : 17.354      Mean   :  7.631      Mean   : 12.22      Mean   :67.56
## 3rd Qu.: 29.800      3rd Qu.: 15.700      3rd Qu.: 21.30      3rd Qu.:81.92
## Max.   : 44.800      Max.   : 28.900      Max.   : 36.20      Max.   :98.50
##                               NA's    :1           NA's    :1
##      VelV           PresM           Prec           NubT
## Min.    : 2.600      Min.    : 96.8      Min.    : 0.000      Min.    :0.000
## 1st Qu.:  7.675      1st Qu.:1007.2      1st Qu.:  0.000      1st Qu.:0.700
## Median :10.100      Median :1014.0      Median :  0.000      Median :3.250
## Mean   :11.204      Mean   :1008.8      Mean   :  2.422      Mean   :3.603
## 3rd Qu.:13.825      3rd Qu.:1019.6      3rd Qu.:  1.100      3rd Qu.:6.500
## Max.   :31.400      Max.   :1032.4      Max.   :42.000      Max.   :8.000
```

```

##                                     NA's   :3           NA's   :2
##      Vis                          EspN
##  Min.   : 4.70   Min.    :25.00
##  1st Qu.:18.30   1st Qu.:48.00
##  Median :29.40   Median :65.00
##  Mean   :35.26   Mean   :63.15
##  3rd Qu.:49.10   3rd Qu.:81.00
##  Max.   :75.00   Max.   :94.00
##                                     NA's   :187

```

A continuación, se muestra una representación gráfica de las temperaturas máximas y mínimas de las dos estaciones meteorológicas recogidas en cada año para cada lugar.

Temperaturas máximas y mínimas en Sevilla

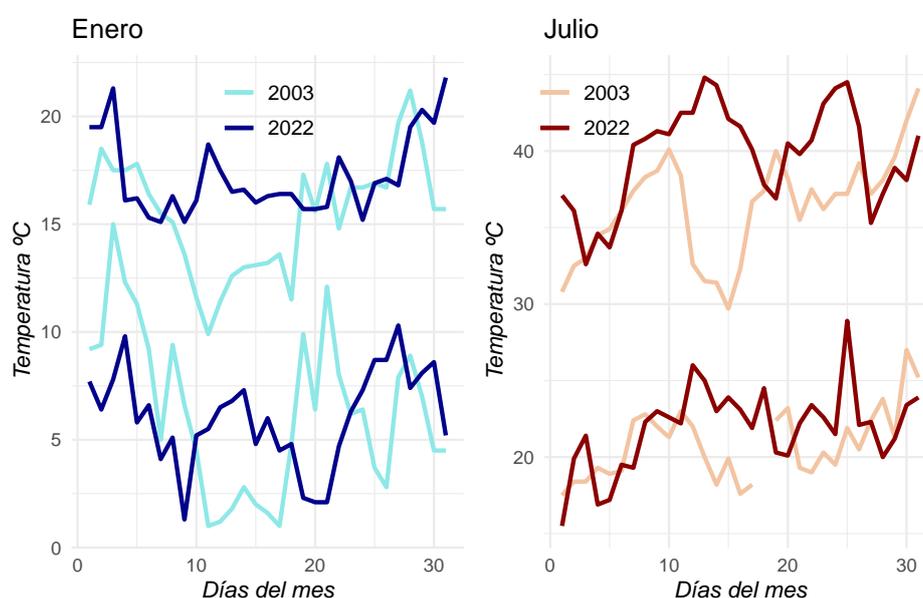


Figura 5.9: Comparación de las temperaturas en Sevilla

Se observan las temperaturas mínimas y máximas en enero y julio para los años 2003 y 2022 en Sevilla.

Respecto al primer gráfico, las mínimas se puede considerar que no han variado demasiado, sin embargo, donde sí hay una clara diferencia es en las máximas, son más cálidas las actuales.

En cuanto al segundo, se observa que el 18 de julio de 2003 no tiene representación, pues es debido a la ausencia de éste en el conjunto de datos (*NA*). Al igual que en enero, las máximas se han incrementado en estas dos décadas.

En resumen, se comprueba el aumento de la temperatura global.

Temperaturas máximas y mínimas en Tromso

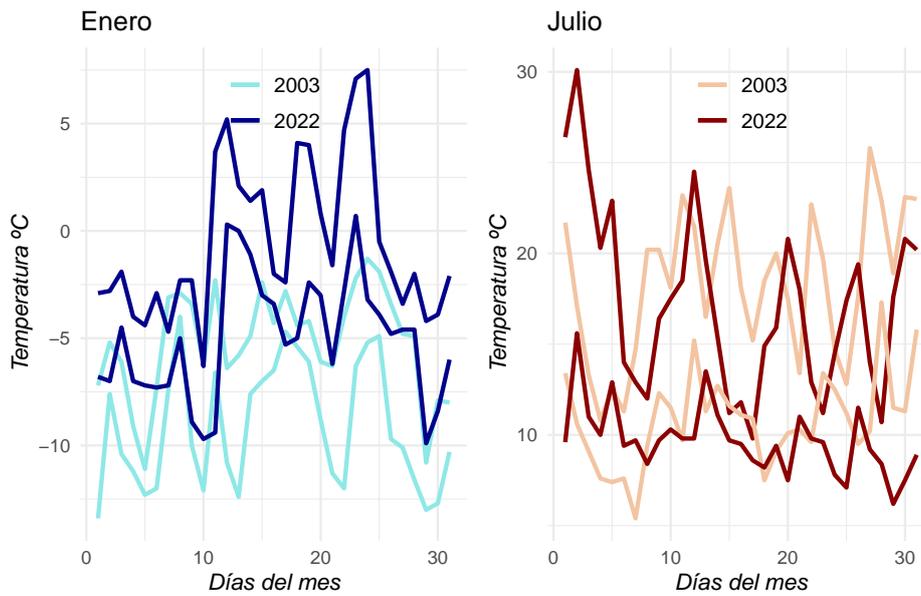


Figura 5.10: Comparación de las temperaturas en Tromsø

En este caso, las temperaturas son de la ciudad noruega de Tromsø.

Como dato preocupante, en el gráfico de la izquierda, se puede observar el hecho de que la temperatura mínima en 2022 en el mes de enero sea levemente superior a la temperatura máxima en el año 2003.

Estas dos ciudades sirven como corroboración del calentamiento global. Este aumento de temperatura es suficiente para que el clima cambie de forma rápida y profunda. También es de analizar que las zonas más frías de la Tierra son las que más sufren estas consecuencias.

Vamos a considerar un modelo de regresión en el que la variable respuesta es la temperatura media.

5.2.2. Modelo de regresión lineal múltiple

Se va a utilizar el modelo de regresión lineal múltiple. Las variables explicativas que se van a seleccionar son *HRMed*, *PresM*, *Prec*, *NubT*, *Año*, *Mes* y *Lugar*.

En primer lugar, debemos transformar las variables categóricas en tipo factor.

```
ogi$Año<-factor(ogi$Año)
ogi$Mes<-factor(ogi$Mes)
ogi$Lugar<-factor(ogi$Lugar)
```

Generamos el modelo:

```
mod1<-lm(TMed~HRMed+PresM+Prec+NubT+Año+Mes+Lugar, data = ogi)
summary(mod1)
```

```
##
## Call:
## lm(formula = TMed ~ HRMed + PresM + Prec + NubT + Año + Mes +
```

```

##      Lugar, data = ogi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.6443 -1.6876 -0.1522  1.5846  9.4833
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.871464   3.030185   4.578 7.63e-06 ***
## HRMed        -0.124093   0.016373  -7.579 8.09e-13 ***
## PresM         0.003264   0.002906   1.123  0.26250
## Prec         0.052000   0.033001   1.576  0.11644
## NubT         0.556414   0.103986   5.351 2.08e-07 ***
## Año2022      1.065509   0.341132   3.123  0.00201 **
## MesJulio     16.862164   0.441842  38.163 < 2e-16 ***
## LugarTromso -15.221644   0.534397 -28.484 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.632 on 234 degrees of freedom
## (6 observations deleted due to missingness)
## Multiple R-squared:  0.9576, Adjusted R-squared:  0.9563
## F-statistic: 754.5 on 7 and 234 DF,  p-value: < 2.2e-16

```

Podemos comentar que el modelo es capaz de explicar el 95.76 % de la varianza observada en la temperatura media. ($R_{ajustado}^2 = 0.9563$).

Realizando el contraste fundamental del modelo se obtiene que:

$$\begin{cases} H_0 : \beta_1 = \dots = \beta_{11} = 0 \\ H_1 : \exists i \text{ tal que } \beta_i \neq 0 \end{cases}$$

El *p-valor* es significativo, por tanto el modelo es útil y existe relación lineal entre los predictores y la variable respuesta (al menos un coeficiente es distinto de 0).

En el caso de contrastar la relación por separado, entre las variables explicativas con la variable respuesta, se verifica que los predictores que tienen relación significativa son: *HRMed*, *NubT*, *Año*, *Mes* y *Lugar*.

Para la estimación de los coeficientes, se deben tener en cuenta las categorías de referencia para las variables cualitativas: Año:2003, Mes: Enero, Lugar: Sevilla.

A nivel de interpretación, por cada porcentaje de *HRMed* disminuye la temperatura (-0.12) manteniéndose constante el resto de variables. De la misma manera el resto de predictores.

A continuación, con el método *stepwise* mixto y el *AIC* seleccionamos los predictores y determinaremos la calidad del modelo.

```
step(mod1, direction = "both", trace = 1)
```

```
## Start:  AIC=476.22
```

```

## TMed ~ HRMed + PresM + Prec + NubT + Año + Mes + Lugar
##
##           Df Sum of Sq      RSS      AIC
## - PresM  1          8.7  1629.5  475.52
## <none>                                1620.8  476.22
## - Prec   1         17.2  1638.0  476.77
## - Año    1         67.6  1688.4  484.10
## - NubT   1        198.3  1819.1  502.15
## - HRMed  1        397.9  2018.7  527.34
## - Lugar  1       5619.6  7240.4  836.44
## - Mes    1      10088.0 11708.8  952.76
##
## Step:  AIC=475.52
## TMed ~ HRMed + Prec + NubT + Año + Mes + Lugar
##
##           Df Sum of Sq      RSS      AIC
## <none>                                1629.5  475.52
## - Prec   1         17.2  1646.7  476.06
## + PresM  1          8.7  1620.8  476.22
## - Año    1         70.8  1700.4  483.82
## - NubT   1        193.3  1822.9  500.65
## - HRMed  1        389.9  2019.4  525.44
## - Lugar  1       5795.3  7424.8  840.52
## - Mes    1      10302.6 11932.1  955.33
##
## Call:
## lm(formula = TMed ~ HRMed + Prec + NubT + Año + Mes + Lugar,
##     data = ogi)
##
## Coefficients:
## (Intercept)          HRMed          Prec          NubT          Año2022          MesJulio
##  17.0655        -0.1222         0.0520         0.5479         1.0887         16.9209
## LugarTromso
## -15.3056

```

La variable *PresM* ha sido la única que ha sido excluida en el proceso de selección. Reajustamos el nuevo modelo sin esa variable:

```

mod2 <- update(mod1, formula = ~ . -PresM)
summary(mod2)

```

```

##
## Call:
## lm(formula = TMed ~ HRMed + Prec + NubT + Año + Mes + Lugar,
##     data = ogi)
##
## Residuals:
##    Min       1Q   Median       3Q      Max
## -5.562 -1.716 -0.159  1.609  9.493

```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.06546    1.04729  16.295 < 2e-16 ***
## HRMed       -0.12216    0.01629  -7.499 1.31e-12 ***
## Prec        0.05200    0.03302   1.575 0.11667
## NubT        0.54793    0.10377   5.280 2.93e-07 ***
## Año2022     1.08875    0.34069   3.196 0.00159 **
## MesJulio    16.92087    0.43898  38.546 < 2e-16 ***
## LugarTromso -15.30563    0.52943 -28.909 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.633 on 235 degrees of freedom
## (6 observations deleted due to missingness)
## Multiple R-squared:  0.9573, Adjusted R-squared:  0.9563
## F-statistic: 879.1 on 6 and 235 DF,  p-value: < 2.2e-16
```

El modelo queda:

$$\text{Temperatura media} = 17.065 - 0.122 \text{ HRMed} + 0.052 \text{ Prec} + 0.548 \text{ NubT} + 1.089 \text{ Año} + 16.921 \text{ Mes} - 15.306 \text{ Lugar}$$

Como se puede observar, el R^2 casi no ha variado, esto refleja que la variable excluida no ayudaba a representar la variabilidad de la variable respuesta. Los intervalos de confianza para cada coeficiente son los siguientes:

```
confint(mod2)

##           2.5 %           97.5 %
## (Intercept) 15.00218513  19.12873580
## HRMed       -0.15426140 -0.09006785
## Prec        -0.01305562  0.11704678
## NubT        0.34349482  0.75236825
## Año2022     0.41754106  1.75995036
## MesJulio    16.05602689  17.78571815
## LugarTromso -16.34867146 -14.26258424
```

Comprobamos la validación de condiciones del modelo. En primer lugar, la prueba de normalidad de los residuos gráficamente.

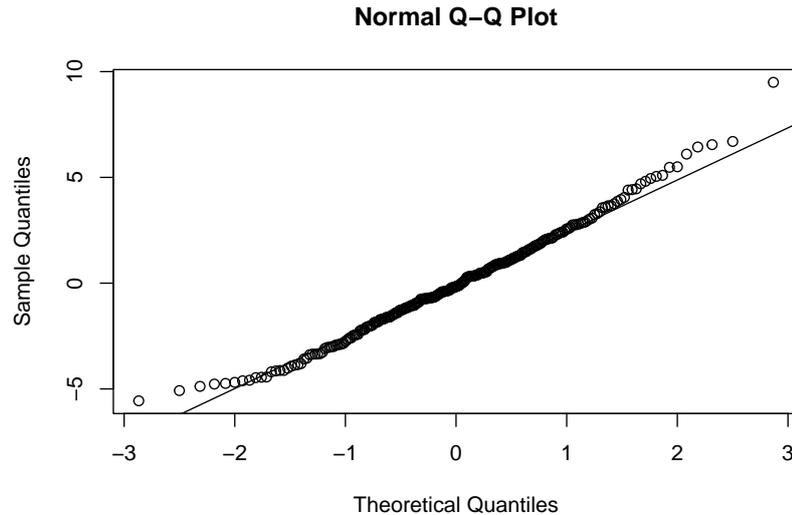


Figura 5.11: Gráfico de normalidad

Los puntos caen a lo largo de la línea diagonal, por tanto es probable que el conjunto de datos siga una distribución normal. Analíticamente con el **test de Shapiro-Wilk**.

$$\begin{cases} H_0 : X \sim N(\mu, \sigma^2) \\ H_1 : X \not\sim N(\mu, \sigma^2) \end{cases}$$

```
shapiro.test(mod2$residuals)
```

```
##
## Shapiro-Wilk normality test
##
## data: mod2$residuals
## W = 0.9891, p-value = 0.06473
```

El p-valor = 0.06473, se acepta la hipótesis nula. Ambas técnicas confirman la normalidad. En segundo lugar, contrastamos la homocedasticidad a través del **test de hipótesis de Breusch-Pagan**.

$$\begin{cases} H_0 : \text{Homocedasticidad} \\ H_1 : \text{Heterocedasticidad} \end{cases}$$

```
library(lmtest)
bptest(mod2)
```

```
##
## studentized Breusch-Pagan test
##
## data: mod2
## BP = 16.478, df = 6, p-value = 0.0114
```

Se observa la ausencia de homocedasticidad. Existen evidencias de alta colinealidad entre variables. Con la función `vif` se puede calcular el factor de inflación de la varianza y detectar las variables con mayor colinealidad.

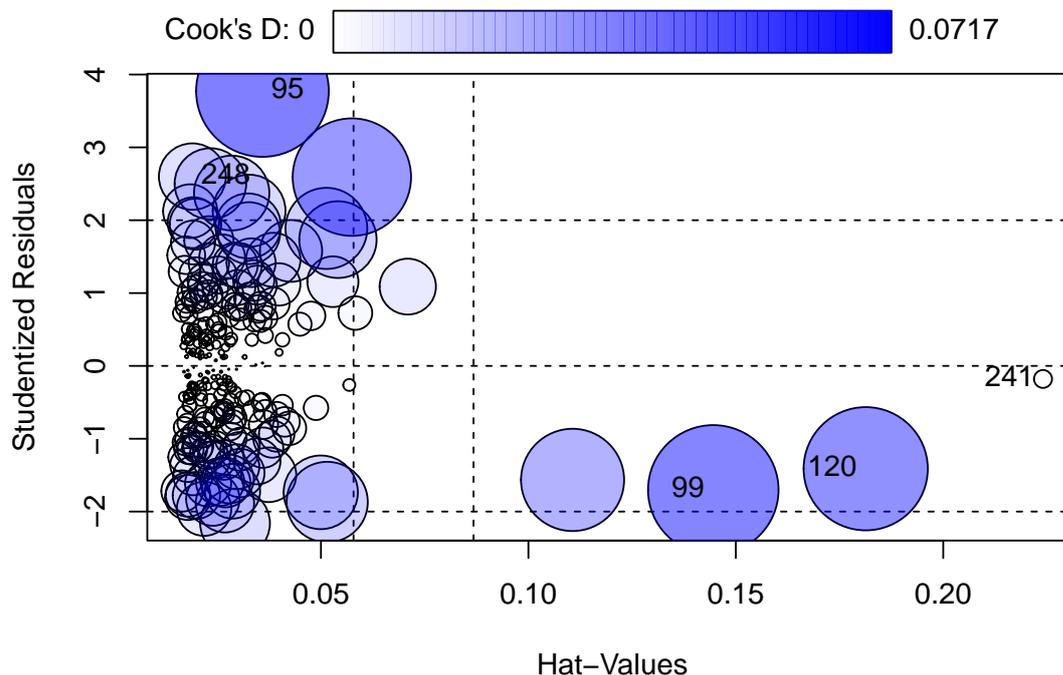
```
library(car)
vif(mod2)
```

```
##      HRMed      Prec      NubT      Año      Mes      Lugar
## 3.503348 1.347330 3.074505 1.012441 1.681340 2.444919
```

Como $1 < VIF < 5$, la regresión puede verse afectada por cierta multicolinealidad.

Por último, podemos identificar posibles valores atípicos o influyentes.

```
require(car)
influencePlot(mod2)
```



```
##      StudRes      Hat      CookD
## 95  3.7733383 0.03589755 0.071695936
## 99 -1.7018565 0.14456805 0.069365611
## 120 -1.4118990 0.18131306 0.062804151
## 241 -0.1791231 0.22404626 0.001328922
## 248  2.5991505 0.01900091 0.018245765
```

Detección de los residuos studentizados considerados como outliers.

```
which(rstudent(mod2) > 3)
```

```
## 95
## 94
```

```
outlierTest(mod2)
```

```
##      rstudent unadjusted p-value Bonferroni p
## 95 3.773338      0.00020416      0.049408
```

Detección de los *hat values* y *Distancia de Cook*.

```
which(hatvalues(mod2)>2*(6+1)/248) # criterio leverage
```

```
## 43 56 86 99 108 120 127 241
## 43 56 85 98 107 119 126 236
```

```
which(cooks.distance(mod2)>4/(248-6-1)) # criterio distancia cook
```

```
## 85 94 95 99 108 110 120 127 128 129 191 192 244 245 248
## 84 93 94 98 107 109 119 126 127 128 189 190 238 239 242
```

Los análisis muestran posibles observaciones influyentes (99, 120, 241 y 248) que superan los límites de preocupación para los valores *Leverages* o *Distancia de Cook*.

5.3. Modelización de la concentración de ozono

5.3.1. Descripción de los datos

Por último, se emplea un conjunto de datos procedente de una librería del software **R**. Este conjunto de datos describe la concentración de ozono (O_3) en Los Ángeles a través de 10 variables. Dichos datos se denominan *ozone1* y se encuentran en la librería **earth**. Entre las variables disponibles, éstas miden la altura de presión de 500 milibares, la velocidad del viento en mph, el porcentaje de humedad, la temperatura en grados Fahrenheit, la visibilidad en millas, y todo, durante un periodo de tiempo.

El ozono es un gas conocido por filtrar la radiación ultravioleta en las capas de la atmósfera, sobre todo, en las más altas. Sin embargo, en las capas más bajas de la atmósfera, no ha tenido la misma atención entre los medios.

Por un lado, afecta perjudicialmente a nuestra salud ya que la mayor fuente de ozono de la Troposfera procede de motores de combustión y emisiones industriales. Por otro lado, disminuye la capacidad de absorción de CO_2 de los ecosistemas, además de ser un gas de efecto invernadero. Es decir, un aumento de O_3 colabora en el calentamiento global. Es por eso que se ha escogido esta información.

Comenzamos cargando dicho conjunto, veremos su estructura y un resumen de éstos.

```
library(earth)
data("ozone1")
str(ozone1) # Estructura de los datos
```

```
## 'data.frame': 330 obs. of 10 variables:
## $ O3 : num 3 5 5 6 4 4 6 7 4 6 ...
## $ vh : num 5710 5700 5760 5720 5790 5790 5700 5700 5770 5720 ...
## $ wind : num 4 3 3 4 6 3 3 3 8 3 ...
## $ humidity: num 28 37 51 69 19 25 73 59 27 44 ...
## $ temp : num 40 45 54 35 45 55 41 44 54 51 ...
## $ ibh : num 2693 590 1450 1568 2631 ...
## $ dpq : num -25 -24 25 15 -33 -28 23 -2 -19 9 ...
## $ ibt : num 87 128 139 121 123 182 114 91 92 173 ...
## $ vis : num 250 100 60 60 100 250 120 120 120 150 ...
## $ doy : num 33 34 35 36 37 38 39 40 41 42 ...
```

```
summary(ozone1) # Resumen de las variables
```

```
##           O3                vh                wind                humidity
## Min.      : 1.00      Min.      :5320      Min.      : 0.000      Min.      :19.00
## 1st Qu.:  5.00      1st Qu.:5690      1st Qu.:  3.000      1st Qu.:47.00
## Median   :10.00      Median   :5760      Median   :  5.000      Median   :64.00
## Mean     :11.78      Mean     :5750      Mean     :  4.848      Mean     :58.13
## 3rd Qu.:17.00      3rd Qu.:5830      3rd Qu.:  6.000      3rd Qu.:73.00
## Max.     :38.00      Max.     :5950      Max.     :11.000      Max.     :93.00
##           temp                ibh                dpg                ibt
## Min.     :25.00      Min.     : 111.0      Min.     : -69.00      Min.     : -25.0
## 1st Qu.:51.00      1st Qu.: 877.5      1st Qu.:  -9.00      1st Qu.:107.0
## Median   :62.00      Median   :2112.5      Median   : 24.00      Median   :167.5
## Mean     :61.75      Mean     :2572.9      Mean     : 17.37      Mean     :161.2
## 3rd Qu.:72.00      3rd Qu.:5000.0      3rd Qu.: 44.75      3rd Qu.:214.0
## Max.     :93.00      Max.     :5000.0      Max.     :107.00      Max.     :332.0
##           vis                doy
## Min.     :  0.0      Min.     : 33.0
## 1st Qu.: 70.0      1st Qu.:120.2
## Median   :120.0      Median   :205.5
## Mean     :124.5      Mean     :209.4
## 3rd Qu.:150.0      3rd Qu.:301.8
## Max.     :350.0      Max.     :390.0
```

Tras esta breve descripción, aplicaremos varios modelos estadísticos a estos datos.

5.3.2. Modelo aditivo generalizado

Como se comentó anteriormente, el modelo aditivo generalizado es un modelo lineal generalizado en el que la variable respuesta depende linealmente de funciones suaves de algunas variables predictoras. Las variables que se van a utilizar son *wind*, *humidity* y *temp*.

En primer lugar, emplearemos un modelo aditivo no paramétrico con componentes “suaves”.

```
library(mgcv)
m1<-gam(O3~s(wind)+s(humidity)+s(temp),data = ozone1)
summary(m1)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## O3 ~ s(wind) + s(humidity) + s(temp)
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.7758      0.2461  47.85  <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df      F  p-value
## s(wind)    1.620  2.042   3.39   0.035 *
## s(humidity) 1.655  2.050  15.55 9.47e-07 ***
## s(temp)    3.718  4.647 105.53 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.689   Deviance explained = 69.5%
## GCV = 20.481   Scale est. = 19.985     n = 330
```

Todos los predictores son significativos. El coeficiente de determinación ajustado es 0.689. La desviación explicada es del 69.5%.

Representamos los efectos parciales:

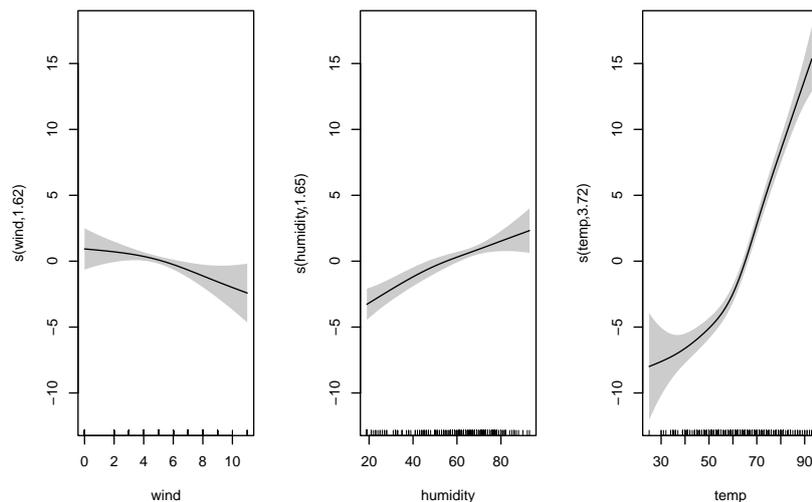


Figura 5.12: Efectos parciales

Observando los efectos se podría considerar que el efecto de los predictores *wind* y *humidity* podría ser lineal.

```
m2<-gam(O3~wind+humidity+s(temp),data = ozone1)
summary(m2)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## O3 ~ wind + humidity + s(temp)
##
```

```
## Parametric coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.64196    0.93768   9.216 < 2e-16 ***
## wind        -0.29751    0.12323  -2.414  0.0163 *
## humidity     0.07872    0.01394   5.647 3.58e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df      F p-value
## s(temp)  3.81  4.761 104.8 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.687   Deviance explained = 69.2%
## GCV = 20.545   Scale est. = 20.121     n = 330
```

En este caso, el coeficiente de determinación ajustado es menor. Comparamos ambos modelos.

```
anova(m2,m1,test = "F")
```

```
## Analysis of Deviance Table
##
## Model 1: O3 ~ wind + humidity + s(temp)
## Model 2: O3 ~ s(wind) + s(humidity) + s(temp)
##   Resid. Df Resid. Dev    Df Deviance      F Pr(>F)
## 1     322.24     6502.8
## 2     320.26     6435.3 1.9779   67.413 1.7054 0.1837
```

Como el *p-valor* es mayor que 0.05, aceptamos que los dos modelos son iguales.

Si consideramos interacción entre los predictores, quedaría:

```
m3<-gam(O3~s(wind,humidity,temp),data = ozone1)
summary(m3)
```

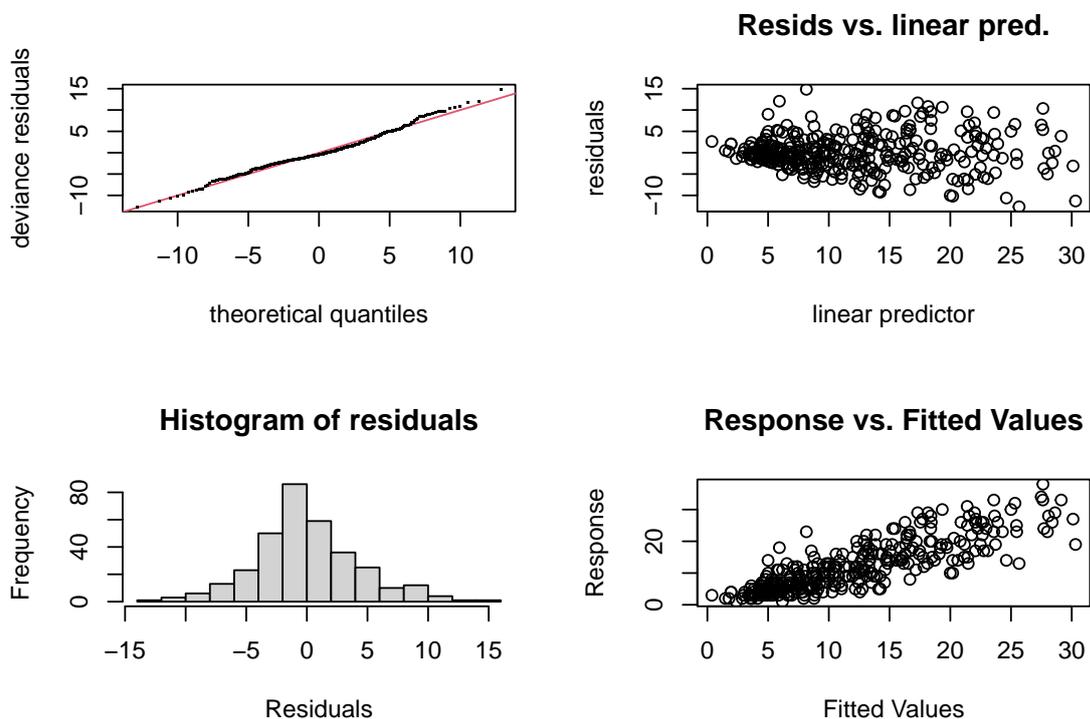
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## O3 ~ s(wind, humidity, temp)
##
## Parametric coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.7758    0.2388   49.3 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##           edf Ref.df      F p-value
```

```
## s(wind,humidity,temp) 9.458  9.889 80.99  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.707   Deviance explained = 71.5%
## GCV = 19.441   Scale est. = 18.825     n = 330
```

Al contrario que el segundo modelo, éste contiene un coeficiente de determinación mayor (0.707).

Aplicamos la diagnosis al tercer modelo.

```
gam.check(m3)
```



```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 7 iterations.
## The RMS GCV score gradient at convergence was 6.308476e-05 .
## The Hessian was positive definite.
## Model rank =  110 / 110
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##           k'    edf k-index p-value
## s(wind,humidity,temp) 109.00  9.46  0.98  0.23
```

Lo perfecto sería apreciar normalidad en los dos gráficos de la izquierda, ausencia de patrón en el superior derecho y ajuste a una recta el inferior derecho. En este ejemplo, los

dos primeros se cumple, pero en el tercero, no hay falta de patrón. En general, se puede considerar que se comporta de manera aceptable.

5.3.3. Modelo de árbol de regresión

Los árboles de regresión dividen el conjunto de datos en grupos más pequeños y luego ajustan un modelo simple para cada subgrupo. Consiste en una serie de reglas de “reparto” del espacio definido por las variables predictoras. Nuestro objetivo es analizar y predecir la concentración de ozono promedio frente a las variables explicativas. Utilizamos la librería `rpart`.

```
set.seed(123)
library(rpart)
```

Con objeto de validar el árbol de regresión, aplicamos validación cruzada con 10 pliegues.

```
(arb<-rpart(O3 ~ vh+wind+humidity+temp+ibh+dpg+ibt+vis, ozone1,
            control=rpart.control(xval=10)))
```

```
## n= 330
##
## node), split, n, deviance, yval
##      * denotes terminal node
##
## 1) root 330 21115.4100 11.775760
## 2) temp< 67.5 214 4114.3040 7.425234
## 4) ibh>=3573.5 108 689.6296 5.148148 *
## 5) ibh< 3573.5 106 2294.1230 9.745283
## 10) dpg< -9.5 35 362.6857 6.457143 *
## 11) dpg>=-9.5 71 1366.4790 11.366200
## 22) ibt< 159 40 287.9000 9.050000 *
## 23) ibt>=159 31 587.0968 14.354840 *
## 3) temp>=67.5 116 5478.4400 19.801720
## 6) ibt< 226.5 55 1276.8360 15.945450
## 12) humidity< 59.5 10 167.6000 10.800000 *
## 13) humidity>=59.5 45 785.6444 17.088890 *
## 7) ibt>=226.5 61 2646.2620 23.278690
## 14) dpg< -10.5 7 305.4286 15.714290 *
## 15) dpg>=-10.5 54 1888.3700 24.259260
## 30) ibt< 294 43 1149.1630 23.139530 *
## 31) ibt>=294 11 474.5455 28.636360 *
```

Para el nodo raíz se incluyen los 330 casos analizados, la suma de cuadrados de los residuos $SC E_{res}^{[root]} = 21115.4100$ y valor medio $\bar{y}_{[root]} = 11.775760$.

Este nodo raíz se divide en dos nodos hijos:

- Para $temp < 67.5$, se tienen 214 casos, que se ajustan con valor medio 7.425234 y suma de cuadrado de residuos de 4114.3040.
- Para $temp \geq 67.5$, se tienen 116 casos, que se ajustan con valor medio 19.801720 y suma de cuadrado de residuos de 5478.4400.

Gráficamente, el árbol quedaría de la siguiente forma:

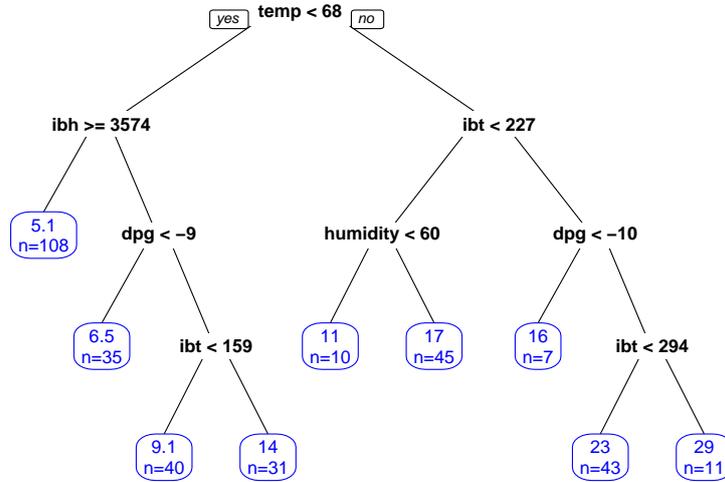


Figura 5.13: Árbol de regresión

Los nodos finales presentan la predicción y el número de casos en el nodo. A continuación, un resumen sobre el algoritmo.

La relevancia de las variables predictoras en valor absoluto:

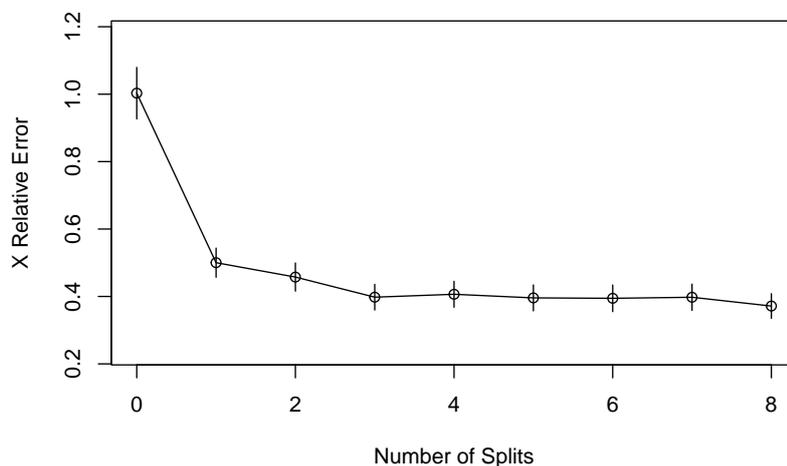
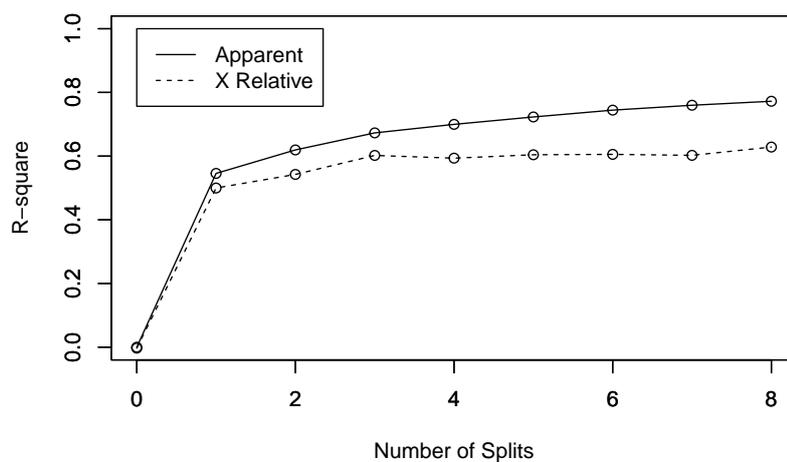
```
##      temp      ibt      vh      ibh      humidity      dpg      vis
## 13246.4006 10717.8679 8597.0017 4109.1323 2394.5995 2167.6568 1832.7884
##      wind
##    112.9916
```

Con `rsq.rpart` se obtiene:

- La tabla CP (tabla del parámetro de complejidad). Si se trata de minimizar el error cuadrático medio, dicho parámetro es el valor que penaliza la complejidad del árbol.
- El coeficiente R-cuadrado (aparente y X-relativo) frente al número de divisiones.
- Error relativo por validación cruzada frente al número de divisiones.

```
##
## Regression tree:
## rpart(formula = O3 ~ vh + wind + humidity + temp + ibh + dpg +
##       ibt + vis, data = ozone1, control = rpart.control(xval = 10))
##
## Variables actually used in tree construction:
## [1] dpg      humidity ibh      ibt      temp
##
## Root node error: 21115/330 = 63.986
##
## n= 330
##
##          CP nsplit rel error  xerror      xstd
```

```
## 1 0.545699      0  1.00000 1.00296 0.076477
## 2 0.073659      1  0.45430 0.50019 0.043274
## 3 0.053542      2  0.38064 0.45752 0.041780
## 4 0.026756      3  0.32710 0.39790 0.038133
## 5 0.023276      4  0.30034 0.40647 0.038570
## 6 0.021428      5  0.27707 0.39579 0.038492
## 7 0.015325      6  0.25564 0.39453 0.039488
## 8 0.012534      7  0.24032 0.39762 0.039075
## 9 0.010000      8  0.22778 0.37158 0.036779
```



Una vez determinado el árbol, se procede a podarlo cortando las divisiones menos importantes con ayuda del parámetro de complejidad.

Se obtiene el valor de error relativo por validación cruzada máximo admisible.

```
(Valor1SE= arb$cptable[which.min(arb$cptable[, "xerror"]), "xerror"]
+ arb$cptable[which.min(arb$cptable[, "xerror"]), "xstd"])
```

```
## [1] 0.4083617
```

A partir de ese valor, establecemos un CP ligeramente superior a 0.023276, por ejemplo $CP = 0.025$ y podamos el árbol.

```
(arb.pod<-prune(arb, cp=0.025))
```

```
## n= 330
##
## node), split, n, deviance, yval
##      * denotes terminal node
##
## 1) root 330 21115.4100 11.775760
##    2) temp< 67.5 214 4114.3040 7.425234
##      4) ibh>=3573.5 108 689.6296 5.148148 *
##      5) ibh< 3573.5 106 2294.1230 9.745283
##        10) dpg< -9.5 35 362.6857 6.457143 *
##        11) dpg>=-9.5 71 1366.4790 11.366200 *
##    3) temp>=67.5 116 5478.4400 19.801720
##      6) ibt< 226.5 55 1276.8360 15.945450 *
##      7) ibt>=226.5 61 2646.2620 23.278690 *
```

Gráficamente queda:

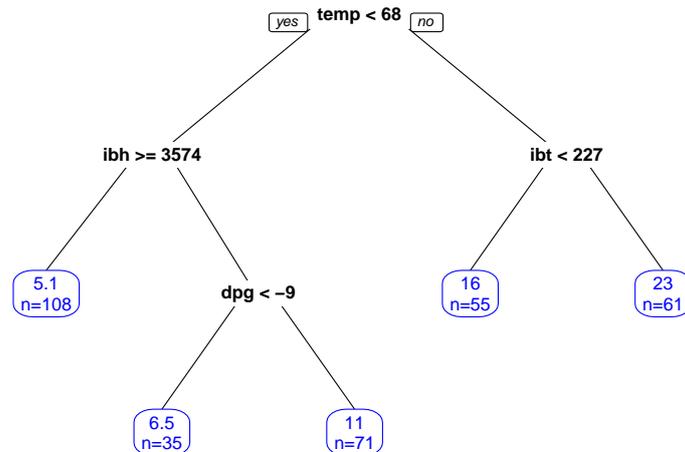


Figura 5.14: Árbol de regresión podado

En este nuevo árbol, la relevancia de las variables predictoras se ven modificadas:

```
arb.pod$variable.importance
```

```
##      temp      ibt      vh      ibh  humidity      vis      dpg
## 12814.7792 9832.2869 8134.8116 3636.1090 1828.9743 1592.3718 1443.0684
##      wind
##    112.9916
```

Resumiendo, el árbol podado ajusta la variable objetivo según las respuestas a las cuestiones.

- Si $(temp < 67.5)$, $(ibh \geq 3573.5)$ entonces ajustar $O_3 = 5.148148$.
- Si $(temp < 67.5)$, $(ibh < 3573.5)$, $(dpg < -9.5)$ entonces ajustar $O_3 = 6.457143$.
- Si $(temp < 67.5)$, $(ibh < 3573.5)$, $(dpg \geq -9.5)$ entonces ajustar $O_3 = 11.366200$.
- Si $(temp \geq 67.5)$, $(ibt < 226.5)$ entonces ajustar $O_3 = 15.945450$.
- Si $(temp \geq 67.5)$, $(ibt \geq 226.5)$ entonces ajustar $O_3 = 23.278690$.

La suma de cuadrados de los residuos del árbol queda $SC_{resid} = 6341.8923$, el coeficiente de determinación alcanzado es $R^2 = 0.6996557$ y por tanto la variabilidad no explicada por el árbol es $1 - R^2 = 0.300344$.

Capítulo 6

Conclusiones

El cambio climático es ya una realidad que se muestra en toda la superficie de la Tierra a través del aumento de las temperaturas, la subida del nivel del mar, la extinción de especies, etc. En España, se ha observado el alargamiento de los veranos, un descenso en las precipitaciones y caudales de los ríos, disminución en la nubosidad, periodos de sequía más frecuentes y olas de calor más intensas, entre otros.

Es por ello que se realizan estudios sobre los impactos, aunque no los suficientes, para poder actuar ante estos cambios y limitar o evitar los riesgos derivados del cambio climático estableciendo unas mejores condiciones climáticas futuras.

Respecto al modelado, existen numerosos métodos empíricos para cuantificar los daños causados por el cambio climático contabilizando los costos de equilibrio y de ajuste. Para estimar dichos efectos se emplean aquellos basados en datos de panel.

En este trabajo se emplean varios conjuntos de datos para observar dicho cambio.

En primer lugar, se ha estudiado la serie temporal mensual para la concentración de dióxido de carbono global obteniendo el mejor modelo ARIMA para luego su predicción. Los tres principales gases de efecto invernadero tienen una tendencia clara y ascendente como se vio en la [Figura 5.1].

En segundo lugar, a través de un modelo de regresión múltiple se quiso obtener la relación entre los elementos del clima. En nuestro caso, la temperatura media respecto a la precipitación, humedad relativa, etc. Se obtuvo un modelo con un alto R^2 aunque se apreció cierta multicolinealidad entre las variables explicativas. Algunas observaciones podían ser influyentes.

Por último, para la concentración de O_3 en los Ángeles, se aplicó por un lado, un modelo aditivo generalizado con varias posibilidades (no paramétrico, semiparamétrico y con interacción). Se obtuvo que el modelo con interacción presenta mejor bondad de ajuste. Por otro lado, un modelo de árbol de regresión con el que se consiguió la relevancia de las variables predictoras más influyentes para ajustar la variable objetivo.

Apéndice A

Apéndice: Exploración y transformación de datos

A.1. Lectura y transformación de datos

```
# Datos de CO2 (https://doi.org/10.15138/9NOH-ZH07)
CO2<-readLines("datos/co2_mm_gl.csv")
I<-grepl("^#",CO2)
list_co2<-strsplit(CO2[!I],split = ",")
M1 <- matrix(unlist(list_co2[2:length(list_co2)]),
             nrow=length(list_co2[2:length(list_co2)]), byrow=TRUE)
colnames(M1) <- c("Año","Mes","Decimal","MediaCO2","MediaCO2_inc",
                 "TendenciaCO2","TendenciaCO2_inc")
tabCO2<-as.data.frame(M1)

# Renombrar las variables
tabCO2$Año=as.integer(tabCO2$Año)
tabCO2$Mes=factor(tabCO2$Mes, levels = 1:12,
                  labels = c("Enero","Febrero","Marzo","Abril","Mayo",
                             "Junio","Julio","Agosto","Septiembre",
                             "Octubre","Noviembre","Diciembre"))
tabCO2$Decimal=as.numeric(tabCO2$Decimal)
tabCO2$MediaCO2=as.numeric(tabCO2$MediaCO2)
tabCO2$MediaCO2_inc=as.numeric(tabCO2$MediaCO2_inc)
tabCO2$TendenciaCO2=as.numeric(tabCO2$TendenciaCO2)
tabCO2$TendenciaCO2_inc=as.numeric(tabCO2$TendenciaCO2_inc)
summary(tabCO2)
```

##	Año	Mes	Decimal	MediaCO2	MediaCO2_inc
##	Min. :1979	Enero : 45	Min. :1979	Min. :334.4	Min. :0.03000
##	1st Qu.:1990	Febrero: 45	1st Qu.:1990	1st Qu.:353.3	1st Qu.:0.08000
##	Median :2001	Marzo : 44	Median :2001	Median :370.1	Median :0.10000
##	Mean :2001	Abril : 44	Mean :2001	Mean :372.9	Mean :0.09894
##	3rd Qu.:2012	Mayo : 44	3rd Qu.:2012	3rd Qu.:392.0	3rd Qu.:0.11000
##	Max. :2023	Junio : 44	Max. :2023	Max. :419.8	Max. :0.29000

```
## (Other):264
## TendenciaCO2 TendenciaCO2_inc
## Min. :335.9 Min. :0.03000
## 1st Qu.:353.7 1st Qu.:0.05000
## Median :369.9 Median :0.06000
## Mean :372.9 Mean :0.06021
## 3rd Qu.:391.7 3rd Qu.:0.07000
## Max. :418.4 Max. :0.10000
##
```

```
# Datos de N2O (https://doi.org/10.15138/P8XG-AA10)
N2O<-readLines("datos/n2o_mm_gl.csv")
II<-grepl("^#",N2O)
list_n2o<-strsplit(N2O[!II],split = ",")
M2 <- matrix(unlist(list_n2o[c(3:length(list_n2o))]),
             nrow=length(list_n2o[c(3:length(list_n2o))]), byrow=TRUE)
colnames(M2) <- c("Año","Mes","Decimal","MediaN2O","MediaN2O_inc",
                 "TendenciaN2O","TendenciaN2O_inc")
tabN2O<-as.data.frame(M2)

# Renombrar las variables
tabN2O$Año=as.integer(tabN2O$Año)
tabN2O$Mes=factor(tabN2O$Mes,levels = 1:12,
                 labels = c("Enero","Febrero","Marzo","Abril","Mayo",
                           "Junio","Julio","Agosto","Septiembre",
                           "Octubre","Noviembre","Diciembre"))
tabN2O$Decimal=as.numeric(tabN2O$Decimal)
tabN2O$MediaN2O=as.numeric(tabN2O$MediaN2O)
tabN2O$MediaN2O_inc=as.numeric(tabN2O$MediaN2O_inc)
tabN2O$TendenciaN2O=as.numeric(tabN2O$TendenciaN2O)
tabN2O$TendenciaN2O_inc=as.numeric(tabN2O$TendenciaN2O_inc)
summary(tabN2O)
```

```
## Año Mes Decimal MediaN2O MediaN2O_inc
## Min. :2001 Enero : 23 Min. :2001 Min. :316.1 Min. : -9.9900
## 1st Qu.:2006 Febrero: 22 1st Qu.:2007 1st Qu.:319.8 1st Qu.: 0.1300
## Median :2012 Marzo : 22 Median :2012 Median :324.9 Median : 0.1400
## Mean :2012 Abril : 22 Mean :2012 Mean :325.1 Mean : -0.3214
## 3rd Qu.:2017 Mayo : 22 3rd Qu.:2018 3rd Qu.:329.7 3rd Qu.: 0.1400
## Max. :2023 Junio : 22 Max. :2023 Max. :336.4 Max. : 0.1800
## (Other):132
## TendenciaN2O TendenciaN2O_inc
## Min. :316.0 Min. : -9.9900
## 1st Qu.:319.8 1st Qu.: 0.1200
## Median :324.7 Median : 0.1300
## Mean :325.1 Mean : -0.3314
## 3rd Qu.:329.8 3rd Qu.: 0.1300
## Max. :336.3 Max. : 0.1600
##
```

```

# Datos de CH4 (https://doi.org/10.15138/P8XG-AA10)
CH4<-readLines("datos/ch4_mm_gl.csv")
III<-grepl("^#",CH4)
list_ch4<-strsplit(CH4[!III],split = ",")
M3 <- matrix(unlist(list_ch4[c(3:length(list_ch4))]),
             nrow=length(list_ch4[c(3:length(list_ch4))]), byrow=TRUE)
colnames(M3) <- c("Año","Mes","Decimal","MediaCH4","MediaCH4_inc",
                 "TendenciaCH4","TendenciaCH4_inc")
tabCH4<-as.data.frame(M3)

# Renombrar las variables
tabCH4$Año=as.integer(tabCH4$Año)
tabCH4$Mes=factor(tabCH4$Mes,levels = 1:12,
                  labels = c("Enero","Febrero","Marzo","Abril","Mayo",
                              "Junio","Julio","Agosto","Septiembre",
                              "Octubre","Noviembre","Diciembre"))
tabCH4$Decimal=as.numeric(tabCH4$Decimal)
tabCH4$MediaCH4=as.numeric(tabCH4$MediaCH4)
tabCH4$MediaCH4_inc=as.numeric(tabCH4$MediaCH4_inc)
tabCH4$TendenciaCH4=as.numeric(tabCH4$TendenciaCH4)
tabCH4$TendenciaCH4_inc=as.numeric(tabCH4$TendenciaCH4_inc)
summary(tabCH4)

```

```

##      Año           Mes           Decimal           MediaCH4
## Min.   :1983   Enero       : 40   Min.   :1984   Min.   :1626
## 1st Qu.:1993   Julio         : 40   1st Qu.:1993   1st Qu.:1737
## Median :2003   Agosto        : 40   Median :2003   Median :1775
## Mean   :2003   Septiembre    : 40   Mean   :2003   Mean   :1775
## 3rd Qu.:2013   Octubre       : 40   3rd Qu.:2013   3rd Qu.:1813
## Max.   :2023   Noviembre     : 40   Max.   :2023   Max.   :1925
##                (Other)    :235
##  MediaCH4_inc   TendenciaCH4  TendenciaCH4_inc
## Min.   : -9.9900  Min.   :1635   Min.   : -9.9900
## 1st Qu.:  0.8500  1st Qu.:1736   1st Qu.:  0.5300
## Median :  0.9800  Median :1774   Median :  0.5800
## Mean   :  0.7528  Mean   :1775   Mean   :  0.3333
## 3rd Qu.:  1.1500  3rd Qu.:1811   3rd Qu.:  0.6400
## Max.   :  2.6900  Max.   :1919   Max.   :  1.3900
##

```

```

#Juntar las tablas
GEI<-merge(tabCO2,merge(tabN2O,tabCH4,all = T),all = T)

# Reordenar los meses
GEI<-GEI[order(GEI$Decimal,decreasing = F),]

# Pasar las unidades a de ppb a ppm
GEI[,8:15]<-apply(GEI[, -(1:7)], 2, function(x) x/1000)

```

```

#Guardar los datos en un RData
save(GEI,file = "datos/GEI.RData")

# Datos de elementos del clima (https://www.ogimet.com/gsynres.phtml)
ogi<-read.table("datos/datos_ogimet.txt",header = T,sep = ";")
ogi[ogi == "---"] <- NA

# Renombrar las variables
ogi$Año<-as.character(ogi$Año)
ogi$Mes=factor(ogi$Mes,levels = c(1,7),
              labels = c("Enero","Julio"))
ogi$TMin<-as.numeric(ogi$TMin, na.omit=T)
ogi$TMed<-as.numeric(ogi$TMed, na.omit=T)
ogi$Prec[ogi$Prec == "Ip"] <- 0.1 # Ip significa Precipitación inapreciable
ogi$Prec<-as.numeric(ogi$Prec, na.omit=T)
ogi$NubT<-as.numeric(ogi$NubT, na.omit=T)
ogi$EspN<-as.numeric(ogi$EspN, na.omit=T)

# Cargar los datos
save(ogi,file = "datos/ogi.RData")

```

A.2. Gráficos

```

ggplot(gei,aes(x=Decimal))+
  geom_line(aes(y=MediaN20,colour="N20"))+
  theme_minimal()+
  geom_line(aes(y=MediaCH4/5.5,colour="CH4"))+
  scale_y_continuous(sec.axis = sec_axis(trans = ~.*5.5,
                                         name = "Fracción molar CH4"))+
  labs(title = "N20 y CH4",
       subtitle = "En ppm",x="Años",y="Fracción molar N20")+
  theme(axis.title = element_text(face = "italic"),
        legend.text = element_text(size = 10,colour = "black"),
        legend.title = element_blank(),legend.position = c(0.2,0.8))

ggplot(gei,aes(x=Decimal,y=MediaCO2))+
  geom_line(stat = "identity")+
  theme_minimal()+
  labs(title = "CO2",
       subtitle = "En ppm",x="Años",y="Fracción molar CO2")+
  theme(axis.title = element_text(face = "italic"))

ggplot(en_sev,aes(x=Dia))+
  geom_line(aes(y=TMax,color=Año),size=1)+
  theme_minimal()+
  geom_line(aes(y=TMin,color=Año),size=1)+

```

```

scale_color_manual(values=c("#8EE8E7", "darkblue"))+
labs(title = "Enero", x="Días del mes", y="Temperatura °C")+
theme(axis.title = element_text(face = "italic"),
      legend.text = element_text(size = 10, colour = "black"),
      legend.title = element_blank(), legend.position = c(0.5, 0.9))

ggplot(ju_sev, aes(x=Dia))+
  geom_line(aes(y=TMax, color=Año), size=1)+
  theme_minimal()+
  geom_line(aes(y=TMin, color=Año), size=1)+
  scale_color_manual(values=c("#F3C4A2", "darkred"))+
  labs(title = "Julio", x="Días del mes", y="Temperatura °C")+
  theme(axis.title = element_text(face = "italic"),
        legend.text = element_text(size = 10, colour = "black"),
        legend.title = element_blank(), legend.position = c(0.1, 0.9))

ggplot(en_tro, aes(x=Dia))+
  geom_line(aes(y=TMax, color=Año), size=1)+
  theme_minimal()+
  geom_line(aes(y=TMin, color=Año), size=1)+
  scale_color_manual(values=c("#8EE8E7", "darkblue"))+
  labs(title = "Enero", x="Días del mes", y="Temperatura °C")+
  theme(axis.title = element_text(face = "italic"),
        legend.text = element_text(size = 10, colour = "black"),
        legend.title = element_blank(), legend.position = c(0.5, 0.9))

ggplot(ju_tro, aes(x=Dia))+
  geom_line(aes(y=TMax, color=Año), size=1)+
  theme_minimal()+
  geom_line(aes(y=TMin, color=Año), size=1)+
  scale_color_manual(values=c("#F3C4A2", "darkred"))+
  labs(title = "Julio", x="Días del mes", y="Temperatura °C")+
  theme(axis.title = element_text(face = "italic"),
        legend.text = element_text(size = 10, colour = "black"),
        legend.title = element_blank(), legend.position = c(0.5, 0.9))

```


Apéndice B

Apéndice: Librerías y análisis

B.1. Librerías

```
# Instalación de paquetes y librerías

# install.packages("dplyr")
library(dplyr)
# install.packages("ggplot2")
library(ggplot2)
# install.packages("patchwork")
library(patchwork)
# install.packages("TSA")
library(TSA)
# install.packages("tseries")
library(tseries)
#install.packages("forecast")
library(forecast)
#install.packages("lmtest")
library(lmtest)
#install.packages("car")
library(car)
#install.packages("earth")
library(earth)
#install.packages("mgcv")
library(mgcv)
#install.packages("rpart")
library(rpart)
#install.packages("rpart.plot")
library(rpart.plot)
```

B.2. Análisis

```
ajuste1=arima(datos_st_d,order = c(0,0,1),
              seasonal = list(order=c(1,0,0),period=12),
              include.mean = FALSE);ajuste1
ajuste2=arima(datos_st_d,order = c(0,0,1),
              seasonal = list(order=c(2,0,0),period=12),
              include.mean = FALSE);ajuste2
ajuste3=arima(datos_st_d,order = c(0,0,2),
              seasonal = list(order=c(1,0,0),period=12));ajuste3
ajuste4=arima(datos_st_d,order = c(0,0,2),
              seasonal = list(order=c(2,0,0),period=12));ajuste4
ajuste5=auto.arima(datos_st_d,step=F); ajuste5
ajuste_bueno1=arima(datos_st,order = c(0,1,1),
                   seasonal = list(order=c(2,1,0),
                                   period=12)); ajuste_bueno1
ajuste_bueno2=arima(datos_st,order = c(2,1,1),
                   seasonal = list(order=c(1,1,1),
                                   period=12)); ajuste_bueno2
pred=predict(ajuste_bueno2,n.ahead = 24)
mod1<-lm(TMed~HRMed+PresM+Prec+NubT+Año+Mes+Lugar, data = ogi)
mod2 <- update(mod1, formula = ~ . -PresM)
m1<-gam(O3~s(wind)+s(humidity)+s(temp),data = ozone1)
m2<-gam(O3~wind+humidity+s(temp),data = ozone1)
m3<-gam(O3~s(wind,humidity,temp),data = ozone1)
(arb<-rpart(O3 ~ vh+wind+humidity+temp+ibh+dpq+ibt+vis, ozone1,
            control=rpart.control(xval=10)))
(arb.pod<-prune(arb,cp=0.025))
```

Bibliografía

- AEMET. "agencia estatal de meteorología". Disponible en <https://www.aemet.es/es/>.
- Copernicus. "copernicus". Disponible en <https://www.copernicus.eu/>.
- Amigos de la Tierra. "justicia climática". Disponible en <https://www.tierra.org/02clima-y-energia/04situacion-actual-cambio-climatico/>.
- Oficina de Publicaciones de la Unión Europea. *Diario Oficial de la Unión Europea*, C83, 2010.
- EPDATA. "el cambio climático, en datos y gráficos". Disponible en <https://www.epdata.es/datos/cambio-climatico-datos-graficos/447>, 2023.
- FAO. "el estado de los bosques del mundo 2022: vías forestales hacia la recuperación verde y la creación de economías inclusivas, resilientes y sostenibles". Disponible en <https://www.fao.org/3/cb9360es/online/cb9360es.html>, 2022.
- ferrovial. "¿qué es el protocolo de kioto?". Disponible en [https://www.ferrovial.com/es/recursos/protocolo-de-kioto/#:~:text=El%20Protocolo%20de%20Kioto%20es,\)%%2C%20hidrofluorocarbonos%20\(HFC\)%2C](https://www.ferrovial.com/es/recursos/protocolo-de-kioto/#:~:text=El%20Protocolo%20de%20Kioto%20es,)%%2C%20hidrofluorocarbonos%20(HFC)%2C).
- Gilberto A. Gamboa Bernal. Cumbre del cambio climático 2021: más escepticismo que compromisos. *Persona y Bioética*, 26(1):1–9, 2022.
- M. Dolores Jiménez Gamero. *Apuntes de la asignatura Series Temporales*, 2022.
- Déborah García. "no hay debate sobre el cambio climático". Disponible en https://www.lasexta.com/el-muro/deborah-garcia/hay-debate-cambio-climatico_2022090763186c381d9fe00001c86b3d.html, 2022.
- GCP. "proyecto global de carbono". Disponible en <https://www.globalcarbonproject.org/>.
- Rosa María Giles Carnero. La política de adaptación al cambio climático de la unión europea: desafíos y oportunidades para el proceso de integración europea. *Sostenibilidad de la Europa del s.XXI: económica, ambiental y social*, 4(4):273–299, 2020.
- Juan M. Gutiérrez. *El modelo de Datos de Panel*, 2020.
- ICOS. "sistema integrado de observación de carbono". Disponible en <https://www.icos-cp.eu/>.
- IIDMA. "instituto internacional de derecho y medio ambiente". Disponible en <https://iidma.org/>.

- IPCC. "grupo intergubernamental de expertos sobre el cambio climático". Disponible en <https://www.ipcc.ch/>.
- Rafael Pino Mejías Juan M. Pichardo. *Apuntes de la asignatura de Estadística Computacional II*, 2022.
- Suman Mor Khaiwal Ravindra, Preety Rattan and Ashutosh Nath Aggarwal. "generalized additive models: Building evidence of air pollution, climate change and human health". *Environment International*, 132, 2019.
- Charles D Kolstad and Frances C Moore. "estimating the economic impacts of climate change using weather observations". *Review of Environmental Economics and Policy*", 2020.
- X. Lan. "tendencias en el promedio global de co2 determinadas a partir de las mediciones del laboratorio de monitoreo global de la noaa". Disponible en <https://doi.org/10.15138/9N0H-ZH07>, 2023a.
- X. Lan. "tendencias en el promedio mundial de ch4, n2o y sf6 determinadas a partir de las mediciones del laboratorio de monitoreo global de la noaa". Disponible en <https://doi.org/10.15138/P8XG-AA10>, 2023b.
- Chiara Liguori. "la cumbre del clima de madrid ha fracasado. ¿y ahora qué?". Disponible en <https://www.es.amnesty.org/en-que-estamos/blog/historia/articulo/la-cumbre-del-clima-de-madrid-ha-fracasado-y-ahora-que/>, 2019.
- Emilio D. Lozano Aguilera. *Estadísticas económicas*. Universidad de Jaén. Servicio de Publicaciones, mayo de 2007 edition, 2007.
- Pedro L. Luque-Calvo. *Escribir un Trabajo Fin de Estudios con R Markdown*, 2017.
- Letón Molina and Pedromingo Marino. *Introducción al análisis de datos en Meta-Análisis*. Diaz de Santos, 2001.
- NOAA. "oficina nacional de administración oceánica y atmosférica". Disponible en <https://www.noaa.gov/>.
- William D. Nordhaus and Andrew Moffat. "a survey of global impacts of climate change: replication, survey methods and a statistical analysis". *Nber working paper series*", 2017.
- NSIDC. "centro nacional de datos de hielo y nieve". Disponible en <https://nsidc.org/>.
- OGIMET. "selección de resúmenes sinópticos diarios". Disponible en <https://www.ogimet.com/gsynres.phtml>.
- OMM. "organización meteorológica mundial". Disponible en <https://public.wmo.int/>.
- ONU. "la cop27 llega a un acuerdo decisivo sobre un nuevo fondo de pérdidas y daños para los países vulnerables". Disponible en <https://unfccc.int/es/news/la-cop27-llega-a-un-acuerdo-decisivo-sobre-un-nuevo-fondo-de-perdidas-y-danos-para-los-paises>, 2022.

- Samuel A. Pilar. "cumbre del clima de parís 2015, en 12 claves". Disponible en <https://www.rtve.es/noticias/20151130/cumbre-del-clima-paris-12-claves/1263241.shtml>, 2015.
- Montero. R. *Test de Hausman. Documentos de Trabajo en Economía Aplicada*. Universidad de Granada. España, 2005.
- Tim Rayner. Climate policy integration into eu energy policy: Progress and prospects, by claire dupont. *Carbon and Climate Law Review: CCLR*, 11(1):80–83, 2017.
- Héctor Rodríguez. "8 argumentos frente a los negacionistas del cambio climático". Disponible en https://www.nationalgeographic.com.es/ciencia/8-argumentos-ante-negacionistas-cambio-climatico_14979, 2022.
- SINC. "la cumbre del clima de glasgow da un paso pequeño pero insuficiente en la lucha climática". Disponible en <https://www.agenciasinc.es/Noticias/La-Cumbre-del-Clima-de-Glasgow-da-un-paso-pequeno-pero-insuficiente-en-la-lucha-climatica>, 2021.
- Plataforma sobre Adaptación al Cambio Climático en España. "unión europea". Disponible en <https://adaptecca.es/union-europea>.
- Naciones Unidas. "¿qué es el cambio climático?". Disponible en <https://www.un.org/es/climatechange/what-is-climate-change>.