



UNIVERSITAT DE VALÈNCIA (0-)

Facultad de Física

Máster oficial en Teledetección

Proyecto final de máster

**PRODUCCIÓN DE PRODUCTOS CLIMÁTICOS A
PARTIR DE DATOS DEL SENSOR IRS**

Septiembre, 2018

Autor: Antón Fort, Antonio

Tutores: Laparra Pérez-Muelas, Valero

Muñoz Marí, Jordi

ÍNDICE

ÍNDICE.....	1
ÍNDICE DE FIGURAS.....	3
ÍNDICE DE TABLAS.....	5
RESUMEN	6
PARTE I: INTRODUCCIÓN.....	7
1. INTRODUCCIÓN.....	7
2. COMPETENCIAS	7
3. JUSTIFICACIÓN Y OBJETIVOS	8
PARTE II: MARCO TEÓRICO	9
4. TELEDETECCIÓN APLICADA AL CLIMA.....	9
4.1. DEFINICIÓN E HISTORIA	9
4.1.1. INTRODUCCIÓN.....	9
4.1.2. DEFINICIÓN	9
4.1.3. HISTORIA.....	10
5. MAGNITUDES FÍSICAS	11
5.1. RADIANCIA	11
5.2. TEMPERATURA	11
6. MODELOS MATEMÁTICOS ESTADÍSTICOS.....	14
6.1. ANÁLISIS DE COMPONENTES PRINCIPALES	14
6.1.1. COVARIANZA Y MATRIZ DE COVARIANZA	15
6.1.2. VECTORES Y VALORES PROPIOS.....	16
6.1.3. ALGORITMO ACP.....	17
6.2. MÉTODOS DE REGRESIÓN Y ANÁLISIS DE PATRONES.....	18
6.2.1. REGRESIÓN LINEAL REGULARIZADA.....	18
6.2.2. REGRESIÓN KERNEL REGULARIZADA.....	19
PARTE III: DESARROLLO Y METODOLOGÍA.....	21
METODOLOGÍA EMPLEADA.....	21
7. MATERIAL EMPLEADO	21
8. SATÉLITE Y SENSOR	21

9.	DATOS	24
10.	PRE-PROCESADO	24
11.	APLICACIÓN DE LOS MODELOS Y OBTENCIÓN DE RESULTADOS	25
PARTE IV: RESULTADOS Y CONCLUSIONES		27
12.	RESULTADOS.....	27
12.1.	TEMPERATURA	27
12.2.	OZONO	32
12.3.	HUMEDAD	38
13.	MAPAS.....	43
14.	CONCLUSIONES.....	44
BIBLIOGRAFIA		48
PARTE V: ANEXOS		49
ANEXO I I: CÓDIGO MATLAB.....		49
Script PRINCIPAL: PCA		49
SCRIPT: MODELOS DE REGRESIÓN.....		52
SCRIPT: MAPEADO		54

ÍNDICE DE FIGURAS

<i>Figura I: Modelo climático simplificado (Fuente: https://scienceofdoom.com/2010/06/)</i>	12
<i>Figura II: Esquema de la variación térmica con la altura en la atmósfera. (Fuente: http://teleformacion.edu.aytolacoruna.es/AYC/document/atmosfera_y_clima/temperatura/temp3.htm)</i>	13
<i>Figura III: Modelo estructural del IRS, ofrecido por la ESA. (Fuente: https://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/MTGDesign/index.html)</i>	22
<i>Figura IV: Cobertura del sensor IRS. (Fuente: https://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/MTGDesign/index.html)</i>	23
<i>Figura V: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de superficie, para la temperatura. (Fuente: elaboración propia).</i>	27
<i>Figura VI: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de superficie, para la temperatura. (Fuente: elaboración propia).</i>	28
<i>Figura VII: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de media atmósfera, para la temperatura. (Fuente: elaboración propia).</i>	29
<i>Figura VIII: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de media atmósfera, para la temperatura. (Fuente: elaboración propia).</i>	30
<i>Figura IX: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF en lo alto de la atmósfera, para la temperatura. (Fuente: elaboración propia).</i>	31
<i>Figura X: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF en lo alto de la atmósfera, para la temperatura. (Fuente: elaboración propia).</i>	32
<i>Figura XI: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de superficie, para la presencia de ozono. (Fuente: elaboración propia).</i>	33
<i>Figura XII: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de superficie, para la presencia de ozono. (Fuente: elaboración propia).</i>	34
<i>Figura XIII: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de media atmósfera, para la presencia de ozono. (Fuente: elaboración propia).</i>	35
<i>Figura XIV: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de media atmósfera, para la presencia de ozono. (Fuente: elaboración propia).</i>	36
<i>Figura XV: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura final de la atmósfera, para la presencia de ozono. (Fuente: elaboración propia).</i>	37
<i>Figura XVI: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de final de atmósfera, para la presencia de ozono. (Fuente: elaboración propia).</i>	38
<i>Figura XVII: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de superficie, para la humedad. (Fuente: elaboración propia).</i>	39

<i>Figura XVIII: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de superficie, para la humedad. (Fuente: elaboración propia).</i>	39
<i>Figura XIX: Diagrama de dispersión entre la Regresión lineal y los datos ECMWF a altura de media atmósfera, para la humedad. (Fuente: elaboración propia).</i>	40
<i>Figura XX: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de media atmósfera, para la humedad. (Fuente: elaboración propia).</i>	41
<i>Figura XXI: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de final de atmósfera, para la humedad. (Fuente: elaboración propia).</i>	42
<i>Figura XXII: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de final de atmósfera, para la humedad. (Fuente: elaboración propia).</i>	42
<i>Figura XXIII: Mapa de temperaturas predicho por ECMWFNWP. (Fuente: elaboración propia)</i>	43
<i>Figura XXIV: Mapa de temperaturas predicho por el modelo de regresión lineal. (Fuente: elaboración propia)</i>	43
<i>Figura XXV: Mapa de error en temperaturas como resultado de restar el ideal y el predicho. (Fuente: elaboración propia)</i>	44
<i>Figura XXVI: Mapa de error humedad. (Fuente: elaboración propia)</i>	44
<i>Figura XXVII: Variación del error en temperatura predicha con la altura. (Fuente: elaboración propia)</i>	45
<i>Figura XXVIII: Variación del error en la presencia de ozono predicha con la altura. (Fuente: elaboración propia)</i>	46
<i>Figura XXIX: Variación del error en humedad predicha con la altura. (Fuente: elaboración propia)</i>	47

ÍNDICE DE TABLAS

Tabla I: Errores medios para la altura de suelo, para la temperatura en ambos modelos de regresión. (Fuente: elaboración propia).28

Tabla II: Errores medios para altura a mitad atmósfera, para la temperatura en ambos modelos de regresión, para la temperatura.(Fuente: elaboración propia).30

Tabla III: Errores medios para lo alto de la atmósfera, para la temperatura en ambos modelos de regresión. (Fuente: elaboración propia).32

Tabla IV: Errores medios para la altura de suelo, para la presencia de ozono atmosférico en ambos modelos de regresión. (Fuente: elaboración propia).34

Tabla V: Errores medios para altura a mitad atmósfera, para la presencia de ozono atmosférico en ambos modelos de regresión.(Fuente: elaboración propia).36

Tabla VI: Errores medios para lo alto de la atmósfera, para la presencia de ozono atmosférico en ambos modelos de regresión. (Fuente: elaboración propia).38

Tabla VII: Errores medios para la altura de suelo, para la humedad en ambos modelos de regresión. (Fuente: elaboración propia).40

Tabla VIII: Errores medios para altura a mitad atmósfera, para la humedad en ambos modelos de regresión.(Fuente: elaboración propia).41

Tabla IX: Errores medios para lo alto de la atmósfera, para la humedad en ambos modelos de regresión. (Fuente: elaboración propia).43

Tabla X: Errores para cada método de regresión, altura, y variable. (Fuente: elaboración propia)45

RESUMEN

La generación de productos climáticos globales nos permite conocer mejor los fenómenos meteorológicos, con continuidad espacial en todo el territorio.

Para la generación de estos productos, se utilizarán datos de radiancias adaptados al sensor IRS a escala global y datos climáticos para esas mismas fechas. Con ello podrán entrenarse distintos modelos de regresión que permitirán obtener estos datos de forma rápida y sencilla, en diferentes alturas de la atmosfera, para obtener datos de temperatura, ozono y humedad directamente desde las radiancias.

La finalidad en de este proyecto es obtener datos climáticos de forma rápida, sencilla y que puedan ser una alternativa viable, más rápida y suficientemente precisa en comparación con los habituales métodos, analizando los resultados obtenidos con los distintos métodos de regresión.

PARTE I: INTRODUCCIÓN

1. INTRODUCCIÓN

El presente documento recoge la memoria técnica del trabajo final del máster oficial en teledetección del alumno, y autor, Antonio Antón Fort.

Esta memoria se divide en distintas partes. La primera parte (PARTE I: INTRODUCCIÓN) tiene la intención de poner al lector en contexto.

En la segunda parte (PARTE II: MARCO TEÓRICO) se explican los distintos componentes que intervienen para la generación de los productos climáticos que se realiza en este trabajo, desde su historia, definición hasta sus bases físicas y matemáticas.

La tercera parte (PARTE III: DESARROLLO Y METODOLOGÍA) muestra el proceso de generación de estos productos y la metodología utilizada.

En la cuarta parte (PARTE IV: RESULTADOS Y CONCLUSIONES) se explican los resultados y las conclusiones que ha extraído el autor tras la realización del trabajo y se comentan posibles mejoras o desarrollos que podrían realizarse.

La quinta y última parte (PARTE V: ANEXOS) son los anexos, que también cobran gran importancia ya que en ellos se presenta el código utilizado.

2. COMPETENCIAS

En este trabajo se pretende mostrar las siguientes competencias:

Teledetección:

- Recoger, interpretar y tratar imágenes y datos satelitales *online*.

Programación:

- Implementar *scripts* MATLAB para realizar tareas de tratamiento de imágenes digitales y aplicación de modelos estadísticos para la extracción de información.

Competencias transversales:

- Desarrollar y diseñar proyectos en el campo de la teledetección.
- Planificar y desarrollar un proyecto de forma eficiente.
- Expresarse y comunicarse de forma oral en público.
- Resolver problemas técnicos.
- Utilizar correctamente el sistema de citas y referencias bibliográficas según APA.

3. JUSTIFICACIÓN Y OBJETIVOS

El objetivo de este proyecto es la obtención de productos climáticos mediante el uso de imágenes multiespectrales procedentes del sensor IRS y evaluar la calidad de los resultados obtenidos, para determinar los posibles usos.

Para ello se implementan *scripts* en MATLAB que realizan una compresión de las imágenes en su conjunto mediante un Análisis de Componentes Principales (ACP, PCA en inglés) y realizan una regresión con datos del sensor IRS como datos de entrada, y datos del European Center for Medium-Range Weather Forecasts (ECMWFNWP) como datos de salida. El ECMWFNWP proporciona estimaciones de varias variables a distintas alturas de la atmósfera, como la temperatura, humedad y ozono, utilizando modelos físicos. Estos modelos físicos son muy precisos, pero conllevan un importante coste computacional y son lentos. El objetivo del presente proyecto es obtener modelos matemáticos estadísticos que proporcionen las mismas estimaciones que los modelos físicos, es decir, que los emulen. La ventaja de utilizar modelos estadísticos es que estos, una vez entrenados, obtienen las mismas estimaciones que los modelos físicos (o muy parecidas), pero con un coste computacional mucho menor, lo que conlleva mayor rapidez en la obtención de resultados

PARTE II: MARCO TEÓRICO

4. TELEDETECCIÓN APLICADA AL CLIMA

4.1. DEFINICIÓN E HISTORIA

4.1.1. INTRODUCCIÓN

La superficie terrestre refleja la radiación electromagnética (REM) que le llega del sol, con distintas intensidades de acuerdo a la región del espectro (firma espectral) y la superficie observada, y estas son recogidas con distintos niveles digitales (ND) en el sensor. Este fenómeno es el eje fundamental de la teledetección.

Hoy en día, los datos son adquiridos en soporte digital y en formato numérico (ND), abriendo un enorme campo para la aplicación de la estadística a las ciencias de observación terrestre. Cualquier imagen puede pensarse como una matriz tridimensional, en la que cada una de las intersecciones de una fila y una columna corresponde a una posición geográfica discreta, y por lo tanto a un píxel, y la tercera dimensión está dada por la banda a la cual pertenece cada píxel.

Teniendo en cuenta este carácter matricial, sobre cualquier imagen pueden realizarse transformaciones y operaciones estadísticas, pudiendo realizar por ejemplo, descripciones alternativas de los datos, destacar ciertas características no perceptibles en un inicio, o preservar la información esencial y reducir el número de dimensiones.¹

4.1.2. DEFINICIÓN

La Agencia Espacial Europea nos aporta una definición de teledetección “Es un modo de obtener información acerca de objetos tomando y analizando datos sin que los instrumentos empleados para adquirir los datos estén en contacto directo con el objeto.”²

Otra definición posible aportada por la Universidad de Murcia es “Teledetección es la técnica que permite obtener información a distancia de objetos sin que exista un contacto material, en nuestro caso se trata de objetos situados sobre la superficie terrestre. Para que esta observación sea posible es necesario que, aunque sin contacto material, exista algún tipo de interacción entre los objetos y el

¹ Revista de Teledetección (2002). S.B.Ferrero, M.G.Palacio y O.R.Campanella Análisis de componentes principales en teledetección. Obtenido en: https://s3.amazonaws.com/academia.edu.documents/39895531/Anlisis_de_componentes_principales_en_te20151111-820-2sca50.pdf?

² ESA. ¿Qué es la teledetección?. Obtenido en: http://www.esa.int/SPECIALS/Eduspace_ES/SEMO1U3FEXF_0.html

sensor. En este caso la interacción va a ser un flujo de radiación que parte de los objetos y se dirige hacia el sensor.”³

Por otro lado el Centro de Astrobiología nos aporta una definición de meteorología “La ciencia del estudio de la atmósfera, de sus propiedades y de los fenómenos que en ella tienen lugar, los llamados meteoros. El estudio de la atmósfera se basa en el conocimiento de una serie de magnitudes, o variaciones meteorológicas, como la temperatura, la presión atmosférica o la humedad, las cuales varían tanto en el espacio como el tiempo”⁴

Como se puede deducir, el estudio del clima está ligado estrechamente a la teledetección hoy en día, ya que nos permite conocer tanto de forma local como global los distintos fenómenos meteorológicos que ocurren, así como su magnitud y variación.

4.1.3. HISTORIA

“Desde la más remota antigüedad los fenómenos atmosféricos han tenido una profunda influencia en la vida del hombre afectando a sus costumbres, a su modo de vida e incluso a su propio sustento...

...Estas fuerzas de la naturaleza llegaron a ser deificadas y más tarde, estudiadas. La forma de estudio de estos fenómenos siempre era desde la superficie, hasta finales del siglo XIX, cuando empezaron a utilizarse globos y poco más tarde el uso de cohetes y aviones permitió la observación más cercana al fenómeno, aunque la observación seguía siendo fundamentalmente desde la superficie...

...Sin embargo, cuando en el año 1958 el hombre fue capaz de colocar un nuevo satélite, fabricado por él, alrededor de la Tierra, comenzó una nueva era en las observaciones meteorológicas. Rápidamente se desarrolló la tecnología espacial necesaria para colocar instrumentos de observación a bordo de los satélites, iniciando de esta forma la observación de los fenómenos meteorológicos en toda su extensión y detalle, desde un nuevo punto de vista...”⁵

Con ello podemos concluir que el ser humano ha tenido la necesidad de conocer los fenómenos meteorológicos desde el inicio, sin poder obtener grandes conclusiones hasta el lanzamiento de los primeros globos que permitían conocer bien la atmósfera pero de forma puntual, y culminando más tarde, con el lanzamiento de satélites y el tratamiento de las imágenes que estos aportaban, es decir, la teledetección para la obtención de productos climáticos surge como respuesta a la necesidad de obtener datos meteorológicos de forma rápida y global, evitando crear gran cantidad de

³ Universidad de Murcia. Fundamentos físicos de la teledetección. Obtenido en: <http://www.um.es/geograf/sig/teledet/fundamento.html>

⁴ Centro de Astrobiología. Meteorología y Climatología. Obtenido en: <https://cab.inta-csic.es/uploads/culturacientifica/adjuntos/20130121115236.pdf>

⁵ Casanova, J.L. (1992). La teledetección en la enseñanza de la meteorología.

infraestructuras de medición y el lanzamiento globos y otros objetos con el fin de recoger muestras, siendo ahora una herramienta indispensable para la obtención de productos climáticos.

5. MAGNITUDES FÍSICAS

5.1. RADIANCIA

Tras el proceso de obtención de datos de satélite, normalmente se obtienen imágenes, las cuales hay que corregir geométricamente en primer lugar, atendiendo a la forma de toma del sensor y a la órbita del satélite. Estas imágenes representan radiancias alteradas por el propio sensor, a las cuales hay que aplicarles unos factores de corrección específicos de cada sensor (*offset* y *gain*), hasta entonces, son simples cuentas digitales que no representan ninguna magnitud física.

Tras el calibrado, obtenemos ya radiancias en el techo de la atmósfera, es decir, una magnitud física que no representa la superficie terrestre directamente, ya que esta alterada por el efecto de la atmósfera, pero ya tiene valor físico y nos permite obtener información.

Si se aplica una corrección atmosférica, ya sea mediante la aplicación del método de Chávez, o bien mediante la aplicación de la ecuación de transferencia radiativa o con métodos bicanales o biangulares, podemos extraer una magnitud física representativa de la superficie terrestre, la radiancia a nivel de superficie. Con esto ya es posible comparar distintas imágenes entre sí, distinguir objetos o materiales según su firma espectral o angular, etc.

Una buena definición de radiancia, según la Universidad de Murcia, es “el flujo radiante (ϕ) que abandona una unidad de área en una dirección particular siguiendo un ángulo sólido particular. Las unidades en que se mide son $Wm^{-2}sr^{-1}$. Su importancia estriba en que es la magnitud que detecta el sensor. Cuando la radiancia se refiere a una porción concreta del espectro electromagnético se le denomina radiancia espectral (L_λ).”⁶

Otra definición posible aportada por Zancajo Jimeno sería “El total de energía radiada por una superficie en una determinada dirección por unidad de área y por ángulo sólido de medida. Se mide en vatios por metro cuadrado y estereorradián ($Wm^{-2}sr^{-1}$).

Matemáticamente su expresión es:

$$L = \frac{\delta I}{\delta A * \cos \theta}$$

Y si la aplicamos a la radiancia espectral, se obtiene:

$$L = \frac{\delta L}{\delta \lambda} \quad " 7$$

Atendiendo a estas definiciones, se puede resumir que la radiancia es la cantidad de energía radiada por una superficie, ya sea energía reflejada, emitida o transmitida, en una dirección concreta.

5.2. TEMPERATURA

La temperatura es una magnitud escalar que mide la cantidad de energía térmica que tiene un cuerpo. En el caso de los gases su valor es proporcional a la energía cinética de las moléculas, según la expresión:

$$T = k * \langle E_c \rangle$$

Donde:

⁶ Universidad de Murcia. Fundamentos físicos de la teledetección. Obtenido en: <http://www.um.es/geograf/sig/teledet/fundamento.html>

⁷ Zancajo Jimeno, J. (2006). Teledetección: nociones y aplicaciones

- T es la temperatura: Su unidad de medida en el Sistema Internacional, en adelante S.I. , es el Kelvin (K)
- Constante universal k : Se trata de una constante igual para todos los gases. Sus unidades en el S.I. es el *Kelvin partido por Julio (K/J)*.
- Energía cinética promedio de las moléculas de gas $\langle E_c \rangle$: Su unidad de medida en el S.I. es el *Julio (J)*.⁸

La Tierra es un sistema termodinámico en equilibrio regido por la energía solar. El sistema climático en su conjunto distribuye la energía por el globo de forma que el clima actual responde a un balance energético entre la energía solar incidente -con su distribución geográfica y transporte global- y la energía emitida al exterior.⁹

La energía procedente de las radiaciones electromagnéticas del Sol que llegan a la Tierra y no han sido devueltas al exterior son absorbidas por la Tierra en distintos niveles y, posteriormente, transmitidas, absorbidas y/o emitidas por todo el sistema termodinámico. La absorción directa de esta energía por cualquier material se traduce en un incremento de su temperatura, lo cual es a su vez otra forma de emisión de energía (Figura I).

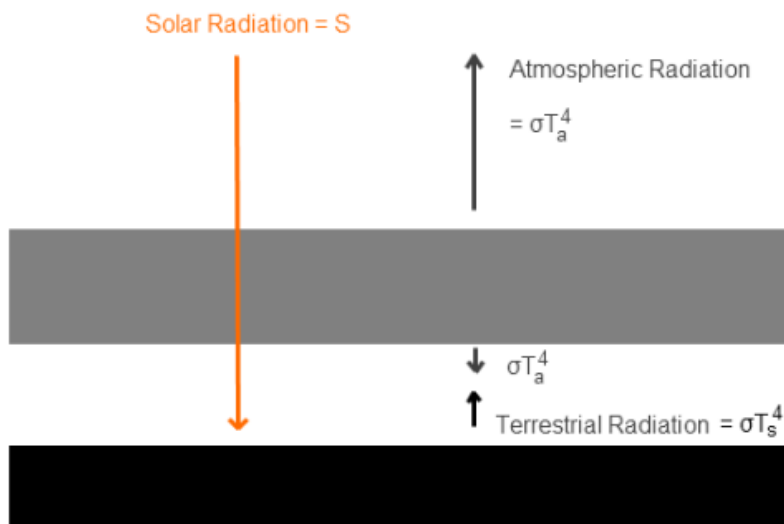


Figura I: Modelo climático simplificado (Fuente: <https://scienceofdoom.com/2010/06/>)

Generalmente, se absorbe más radiación en la superficie terrestre que en la atmósfera, ya que lógicamente, hay más cantidad de material y lo convierte en la fuente de emisión de calor principal de la Tierra. No

⁸ FisicaLab. Temperatura. Obtenido en:

<https://www.fiscalab.com/apartado/temperatura#contenidos>

⁹ L.G.Barrón. Universidad de Andalucía. (2009). Cambios climáticos y efectos ambientales.

Obtenido en:

https://www.researchgate.net/profile/Pablo_Garcia_Murillo/publication/260417035_Elementos_vegetales_del_paisaje_referentes_de_situaciones_climaticas_concretas/links/02e7e5310d233119bf000000.pdf#page=10

obstante, también se produce absorción y emisión a distintos niveles de la atmósfera, esto conlleva que la variación de temperatura a medida que aumenta la altura respecto a la superficie (gradiente térmico) no sea siempre constante, ya que incluso puede invertirse esta variación (Figura II).

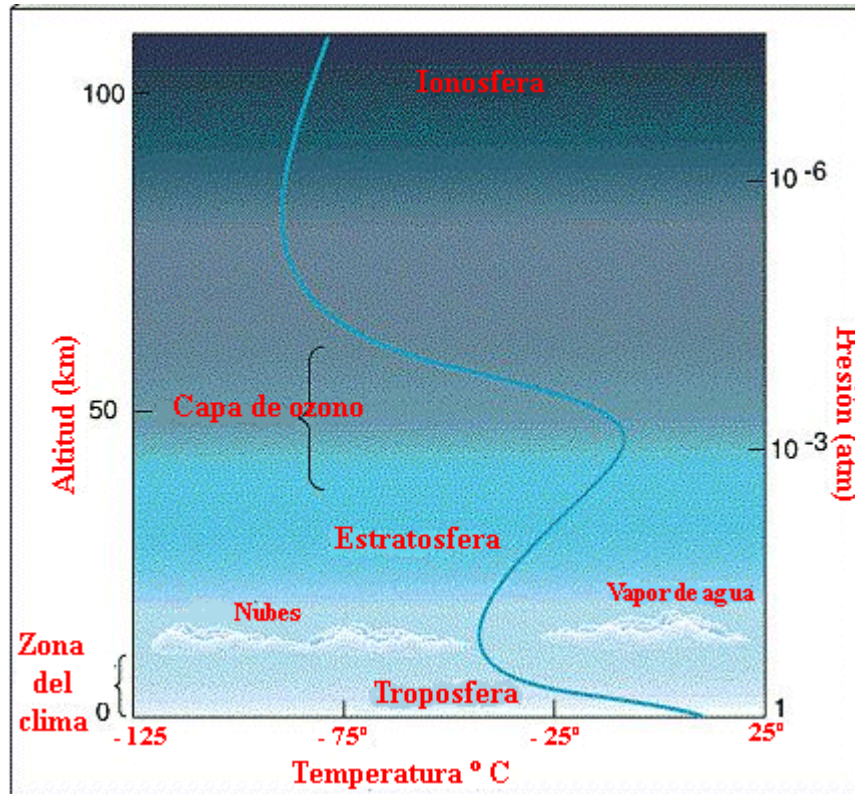


Figura II: Esquema de la variación térmica con la altura en la atmósfera. (Fuente: http://teleformacion.edu.aytolacoruna.es/AYC/document/atmosfera_y_clima/temperatura/temp3.htm)

6. MODELOS MATEMÁTICOS ESTADÍSTICOS

6.1. ANÁLISIS DE COMPONENTES PRINCIPALES

Como las imágenes digitales tienen carácter digital, tal y como se comentaba anteriormente (4.1.1), podemos realizar diversas transformaciones.

El Análisis de Componentes Principales, en adelante ACP, tiene por objeto transformar un conjunto de variables, a las que se denomina originales (en este caso son las bandas de cada imagen), en un nuevo conjunto de variables denominadas componentes principales. Esta transformación reduce la dimensionalidad de las imágenes, y con ello su tamaño, conservando los elementos más esenciales, es decir, resume un grupo amplio de variables en un nuevo grupo más pequeño y además hace que la capacidad para distinguir objetos o coberturas sea máxima.

Esta transformación puede aplicarse como realce previo a la interpretación visual o como procesamiento anterior a la clasificación. En general, esta técnica incrementa la eficiencia computacional de la clasificación y también es muy útil para detectar cambios en aplicaciones multitemporales, aunque en cualquier caso, se suele efectuar sobre variables tipificadas para evitar problemas derivados de la escala.

Desde el punto de vista estadístico, facilita la interpretación sobre los ejes de la variabilidad de la imagen, permitiendo identificar aquellos rasgos que aparecen en la mayoría de bandas u aquellos otros que solo son específicos de algunas de ellas. Las componentes principales obtenidas se caracterizan por estar incorrelacionadas entre sí y, además, pueden ordenarse de acuerdo con la información que llevan incorporada. Como medida de la cantidad de información que llevan incorporadas las distintas componentes principales, se utiliza su variancia. A mayor variancia, mayor cantidad de información aporta dicha componente principal. Por esta razón se selecciona como primera componente aquella que tenga mayor variancia y la última, la de menor variancia.

Algebraicamente, genera nuevas variables (componentes), mediante una combinación lineal de las ρ variables originales (bandas). Aunque se requieren las ρ componentes principales para reproducir la variabilidad total, normalmente la mayor parte de ella está contenido en un número menor de componentes m . En ese caso, reemplazando las ρ bandas por los m componentes, se reduce la dimensionalidad del problema conservando casi la totalidad de la información.

Esta característica es fundamental, ya que en Teledetección, la adquisición de imágenes en bandas adyacentes del espectro implica, con frecuencia, detectar información redundante (en apariencia, las bandas espectralmente próximas entre sí, se visualizan de forma similar). Por ello, los ND de los píxeles de una banda pueden presentar una importante relación con los de otra, resultando una o más de ellas irrelevantes, es decir, cuando las variables originales están muy correlacionadas entre sí, la

mayor parte de su variabilidad se puede explicar con muy pocas componentes, y si estuvieran muy incorreladas entre sí, el ACP carecería de interés ya que no causarían ningún efecto significativo.

La primera componente principal, al igual que las restantes, se expresa como una combinación lineal de las variables originales:

$$Z_{1i} = U_{11}X_{1i} + U_{12}X_{2i} + \dots + U_{1p}X_{pi}$$

Aunque la imagen suele estar presentada en una matriz tridimensional, para la realización del ACP se utiliza una matriz bidimensional. Formalmente los n píxeles de una imagen en p bandas pueden arreglarse en una matriz $X_{p \times n}$, donde cada fila representa los niveles digitales de una banda en concreto, obteniendo el sistema:

$$Z_1 = XU_1$$

$$Z_1 = \begin{bmatrix} Z_{11} \\ Z_{12} \\ \vdots \\ Z_{1n} \end{bmatrix} \quad X = \begin{bmatrix} X_{11} & \dots & X_{1n} \\ \vdots & \ddots & \vdots \\ X_{p1} & \dots & X_{pn} \end{bmatrix} \quad U_1 = \begin{bmatrix} U_{11} \\ U_{12} \\ \vdots \\ U_{1p} \end{bmatrix}$$

El estudio de la relación entre bandas, que es la base del ACP, puede realizarse de dos formas:

- **Con la matriz varianza-covarianza:** Trata de obtener como componentes principales, entre todas las combinaciones lineales posibles, aquellas que no están correlacionadas y la varianza es máxima.
- **Con la matriz de correlación:** Otorga a todas las variables la misma importancia, a partir del estandarizando de las mismas.

En este trabajo se realiza la primera forma, con lo que se profundizará en los fundamentos de esta (6.1.1 COVARIANZA Y MATRIZ DE COVARIANZA).^{10 y 12}

6.1.1. COVARIANZA Y MATRIZ DE COVARIANZA

Tenemos la matriz de variancia-covariancia S :

$$S = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{bmatrix}$$

En la que los elementos de la diagonal son las variancias de los ND en cada banda, habiendo obtenido la variancia de cada elemento Z_1 :

¹⁰ Revista de Teledetección, (2002). S.B.Ferrero, M.G.Palacio y O.R.Campanella. Análisis de componentes principales en teledetección: Consideraciones estadísticas para optimizar su interpretación. Obtenido en: https://s3.amazonaws.com/academia.edu.documents/39895531/Anlisis_de_componentes_principales_en_te20151111-820-2sca50.pdf

$$V(Z_1) = \frac{1}{n} \sum_{i=1}^n Z_{1i}^2 = \frac{1}{n} Z_1' Z_1 = \frac{1}{n} u_1' X' X u_1 = u_1' \left[\frac{1}{n} X' X \right] u_1 = u_1' S u_1$$

Los elementos fuera de la diagonal son las covariancias entre los ND de dos bandas. La covariancia es la media aritmética de los productos de las desviaciones de cada una de las variables respecto a sus medias respectivas. Matemáticamente, sean X e Y dos variables con n elementos, \bar{X} e \bar{Y} las medias de los elementos de X e Y respectivamente. Se define la covarianza entre ellas, $cov(X,Y)$ como:

$$cov(X,Y) = \sum_i^n \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

Como la covariancia entre la banda i y la j es la misma que entre la j y la i ($\sigma_{ij} = \sigma_{ji}$), la matriz Σ_X es simétrica. Cuando hay relación lineal entre los ND de dos bandas las covariancias son grandes en comparación con las variancias, por eso es que esta matriz sirve para estudiar la relación entre las bandas.

Como ya se ha comentado, los componentes principales son nuevas variables obtenidas como combinaciones lineales de las variables anteriores (bandas).

De todas las combinaciones lineales posibles, los componentes principales son las que no están correlacionadas y tienen máxima variancia.¹¹

6.1.2. VECTORES Y VALORES PROPIOS

Anteriormente teníamos que la variancia es:

$$V(Z_1) = u_1' S u_1$$

Donde S es la matriz variancia-covariancia de las observaciones. Sería posible maximizar la varianza sin límite aumentando el módulo del vector u_1 . Para que esta maximización tenga solución debemos imponer una restricción al módulo del vector u_1 , e imponer que $u_1' u_1 = 1$. Para ello se utilizan el multiplicador de Lagrange tal que:

$$M = u_1' S u_1 - \lambda(u_1' u_1 - 1)$$

Y para maximizar esta expresión derivamos a los componentes de u_1 e igualando a cero, entonces:

$$\frac{\partial M}{\partial u_1} = 2S u_1 - 2\lambda u_1 = 0$$

¹¹ De la Fuente Crespo, Laura. Componentes Principales. Obtenido en: [http://www.fuenterrebollo.com/Master-Econometria/Componentes Principales.pdf](http://www.fuenterrebollo.com/Master-Econometria/Componentes_Principales.pdf)

Cuya solución es:

$$Su_1 = \lambda u_1$$

Esto implica que **u** es un vector propio de la matriz **S**, y **λ** su correspondiente valor propio, para determinar qué valor propio de **S** es solución, hay que multiplicar por u_1' esta ecuación,

$$u_1' Su_1 = \lambda u_1' u_1 = \lambda$$

Con lo que se puede concluir que λ es la varianza de los valores del primer componente (Z_1), y como esta es la cantidad que queremos maximizar, **λ será el mayor valor propio de la matriz S**. Su vector asociado u_1 , define los coeficientes de cada variable en el primer componente principal.¹²

6.1.3. ALGORITMO ACP

Sea un conjunto de datos **X** dispuestos de forma matricial, a los que se desea aplicar el algoritmo ACP se procede del siguiente modo teniendo en cuenta lo visto anteriormente:

- Para cada elemento \widehat{X}_n se puede extraer su media \overline{X}_n y se resta a dicho elemento. Con ello se obtiene un nuevo conjunto de datos con media cero.

$$\widehat{X}_n = X_n - \overline{X}_n$$
- Obtener la matriz de covarianza (**C**) de todo el conjunto.
- Obtener los vectores (**V**) y valores propios (**D**) de la matriz de covarianza ordenados de mayor a menor.
- Escoger la cantidad de componentes principales.
 - Para ello hay que atender a los vectores propios que posean mayor valor propio, ya que son los más significativos, y podemos despreciar los que tengan menor valor propio pues son los que menos información aportan.
- Aplicar la ACP a nuestros datos aplicando:

$$ACP(X) = X * V$$

Tras esto, ya tenemos nuestro nuevo conjunto de datos, que se puede revertir siguiendo el proceso inverso:

$$X = ACP * V^{-1}$$

Recordando que **X** es el conjunto de datos con media cero, habría que sumarle la media obtenida en el primer paso para obtener el conjunto de datos original.

¹² Daniel Peña. Análisis de datos multivariantes (Capítulo 5: Componentes Principales).
Obtenido en: https://www.mhe.es/universidad/ciencias_matematicas/pena/home/CAPITULO.PDF

6.2. MÉTODOS DE REGRESIÓN Y ANÁLISIS DE PATRONES

6.2.1. REGRESIÓN LINEAL REGULARIZADA

La regresión lineal regularizada es una versión de la regresión lineal de mínimos cuadrados a la que se le añade un término de regularización para evitar el sobre-entrenamiento y condicionar mejor el problema de regresión. Matemáticamente se define como: sea $x_i \in \mathbb{R}^d$ el conjunto de datos de entrada e y_i la salida que se va a estimar, donde $i = 1, 2, \dots, n$ indican el índice de las n muestras de entrenamiento. Se quiere realizar la regresión lineal de mínimos cuadrados donde la ecuación es $\hat{y}_i = y_i + \varepsilon$, con ruido Gaussiano $\varepsilon \sim N(0, \sigma^2)$ de media cero y desviación estándar σ . A partir de este punto, descartamos el término bias b ya que puede ser incluido en los pesos del modelo agregando una columna de unos al conjunto de datos de entrada.

En el ajuste de la regresión lineal queremos minimizar la función de pérdida cuadrática L con respecto a los pesos del modelo w :

$$L = \frac{1}{n} \sum_i^n (x_i^\top w - y_i)^2 + \lambda \|w\|^2$$

Donde $e_i = y_i - x_i^\top w$ es el error de la regresión que queremos minimizar y λ es el parámetro de regularización que controla la norma de los pesos del modelo, $\|w\|^2$, para evitar el sobre-entrenamiento y forzar que haya más suavidad en la solución y así evitar valores de w muy desiguales.

Esta definición también puede expresarse en notación matricial pasa así considerar múltiples entradas y salidas. El modelo vendrá dado por $Y = XW$, donde X es la matriz de muestras de entrenamiento, $X = [x_1 \ x_2 \ \dots \ x_n]^\top$, de tamaño $n * d$, y $W \in \mathbb{R}^{d * p}$ es la matriz de pesos que definen la relación lineal entre las entradas $X \in \mathbb{R}^{n * p}$ y las salidas $Y \in \mathbb{R}^{n * p}$ que serán predichas. Derivando la función cuadrática de pérdida, $L = \|XW - Y\|^2 + \lambda \|W\|^2$, con respecto a los pesos del modelo W , e igualándolos a cero, se obtiene la solución de ecuación normal

$$W = (X^\top X + \lambda I)^{-1} X^\top Y$$

El único parámetro por ajustar (por ejemplo, por validación cruzada) es el parámetro de regularización λ , que es un valor (típicamente pequeño) que se suma a la diagonal de la matriz de covarianza estimada $\frac{1}{n-1} X^\top X$. Por lo tanto, en una regresión regularizada, se debe estimar la inversa de la matriz de covarianza (regularizada) para obtener los parámetros de recuperación (un peso por una característica de entrada y de salida estimada).

Finalmente, lo que se quiere en la práctica, es el valor de salida para nuevas muestras de entrada, X_* , que se calcula proyectándolo sobre la solución $\hat{Y}_* = X_* W$,

6.2.2. REGRESIÓN KERNEL REGULARIZADA

A menudo el espacio de datos de entrada no es apropiado para obtener un buen modelo puesto que las características demuestran tener relaciones no lineales. En estos casos se desea transformar los datos para plantear las distribuciones de manera que muestren dependencias lineales, que son más fáciles de analizar. Sin embargo, diseñar estas transformaciones no es tarea fácil. Una alternativa consiste en mapear los datos a un espacio de mayor dimensión, donde es más probable encontrar una solución lineal.

La regresión Kernel Ridge (KRR por sus siglas en inglés, Kernel Ridge Regression) es la versión kernel de la regresión lineal regularizada. La formulación es la siguiente:

Se quiere realizar una regresión lineal de mínimos cuadrados en un espacio de Hilbert, H , de gran dimensionalidad (posiblemente infinita) d_H , donde las muestras se mapean a través de una función no lineal $\phi(x_i)$. En notación matricial, el modelo viene dado por $Y = \phi W$ y, bajo la premisa de un modelo de ruido aditivo con variables independientes e idénticamente distribuidas, $\hat{Y} = Y + E$ con ruido Gaussiano $E \sim N(0, \sigma_n^2 I)$ de media cero y desviación estándar σ_n .

Luego, en el ajuste de la regresión lineal regularizada, se quiere minimizar la función de coste cuadrática regularizada, $L = \|\phi W - Y\|^2 + \lambda \|W\|^2$ con respecto a los pesos del modelo W . Por lo tanto, derivando con respecto a W e igualando los pesos a cero, se obtiene $W = (\phi^T \phi + \lambda I)^{-1} \phi^T Y$, donde ϕ es la matriz de las muestras mapeadas, $[\phi(x_1) \phi(x_2) \dots \phi(x_n)]^T$, con tamaño $n \times d_H$. Este problema no se puede resolver directamente ya que la inversa se ejecuta en la matriz $\phi \phi^T$ que es de tamaño $d_H \times d_H$, y $\phi()$ es en principio desconocido. Aquí se aplica el teorema del representante, con el que podemos expresar la solución como una combinación lineal de muestras mapeadas, $W = \phi^T A$, y entonces, la solución es derivada y expresada en función de los pesos duales A (una fila por muestra), $A = A = (\phi \phi^T + \lambda I)^{-1} Y$. Ahora el problema se puede resolver, ya que solo se necesita calcular la inversa de la matriz de Gram (regularizada) $K = \phi \phi^T$ de tamaño $n \times n$. A pesar de que el mapeo es desconocido, se puede reemplazar este producto de matrices con esta matriz de similitud entre las muestras, que se conoce como *matriz kernel*, K .

Para obtener nuevas predicciones X_* se proyecta sobre la solución W :

$$\hat{Y}_* = \phi_* W = \phi_* \phi^T A = K_* A = K_* (K + \lambda I)^{-1} Y$$

Donde la matriz K_* contiene las similitudes entre todas las muestras de prueba y todas las muestras de entrenamiento, cuyos elementos son, $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. Lo importante aquí es que solo se necesita acceso a la función kernel K . Ejemplos de típicas funciones kernel son las lineales $K(x_i, x_j) = x_i^T x_j$, las polinómicas, $K(x_i, x_j) = (x_i^T x_j + 1)^d$, o la que se utiliza en este trabajo, la función kernel base radial Gaussiana (RBF por sus siglas en inglés, Gaussian radial basis function):

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / (2\sigma^2))$$

Por lo tanto en KRR, se deben ajustar dos parámetros, el parámetro de regularización λ y el parámetro del ancho de kernel RBF σ .

PARTE III: DESARROLLO Y METODOLOGÍA

METODOLOGÍA EMPLEADA

7. MATERIAL EMPLEADO

Dado el carácter computacional de este proyecto, para llevarlo a cabo el único material necesario ha constado únicamente de cierto hardware y software, así como los datos proporcionados.

Respecto al hardware, se ha utilizado principalmente un ordenador portátil ACER Aspire F15, capaz de sustentar el software necesario y un disco duro portátil Western Digital Elements con capacidad para dos Terabytes para poder albergar los 202 Gigabytes de datos empleados en el proyecto.

Por otro lado, se ha empleado como sistema operativo Windows 10 con el software de pago MATLAB2015. Este software permite la realización de diversas rutinas sobre gran cantidad distinta de datos de forma “automática”, ya que habilita el fácil tratamiento y transformación de los datos en forma matricial (imágenes) mediante el uso del lenguaje de programación específico de dicho software.

8. SATÉLITE Y SENSOR

El sensor IRS (Infrared Souder) irá instalado en la constelación de satélites Meteosat, en concreto, en la serie de satélites Meteosat Tercera Generación, en adelante MTG, cuyo lanzamiento está previsto para 2021.

Esta serie de satélites, comprenderá cuatro satélites de formación de imágenes y dos de sondeo, los cuales ya no se estabilizarán mediante el giro de la plataforma, como ocurrió con las dos primeras generaciones de satélites Metosat, sino mediante plataformas estabilizadas de tres ejes.

Las plataformas satelitales albergarán los instrumentos *Flexible Combiner Imager* (FCI), *Lightning Imager* (LI) e *Infrared Souder* (IRS) que escanearán y sondearán la superficie terrestre y la atmósfera con canales adicionales con una mejor resolución espacial, temporal y radiométrica.

Constarán por un lado de los satélites de imágenes, MTG-I, utilizaran los sensores *Flexible Combined Imager* y el *Lightning Imager* (LI), y por otro, de los satélites de sondeo, MTG-S, incluirán un interferómetro, el **Infrared Souder (IRS)** y el instrumento Sentinel-4, el espectrómetro de alta resolución *Ultraviolet Visible Near-infrared* (UVN).

El sensor IRS será capaz de proporcionar información sobre estructuras de vapor de agua y temperatura de la atmósfera resueltas horizontal, vertical y temporalmente (4-dimensionales).

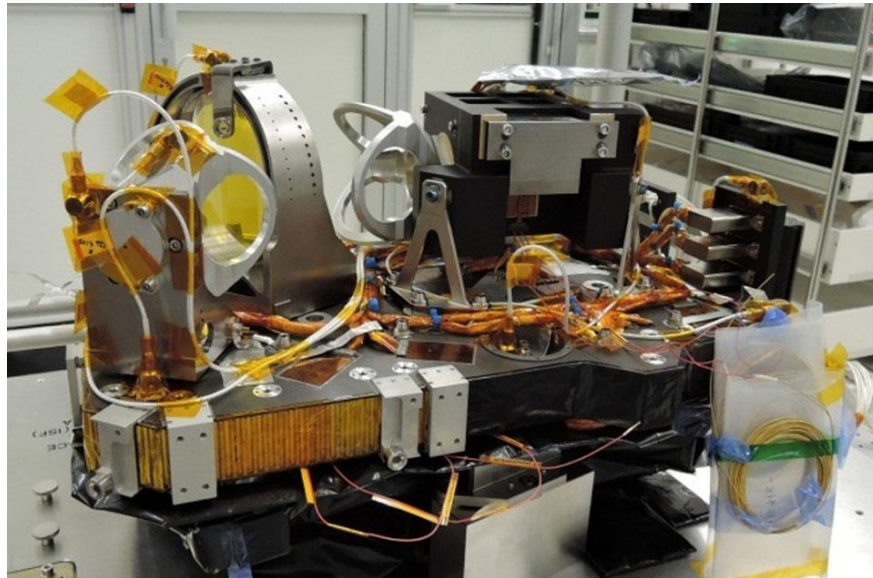


Figura III: Modelo estructural del IRS, ofrecido por la ESA. (Fuente: <https://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/MTGDesign/index.html>)

El IRS adquiere un número de sondeos espectrales simultáneamente sobre una zona usando un *array* de detector bidimensional. Utiliza un mecanismo de "*step-and-stare*" (paso y mirada) para sondear la atmósfera de la Tierra.

La cobertura de cada zona se escalona en dirección este/oeste para formar una línea de sondeos espectrales de zona, antes de desplazarse hacia el norte para formar la línea siguiente, cubriendo la cobertura de área local (LAC) dentro de la duración del ciclo de repetición (principio de "*step-and-stare*" para sondear la atmósfera).

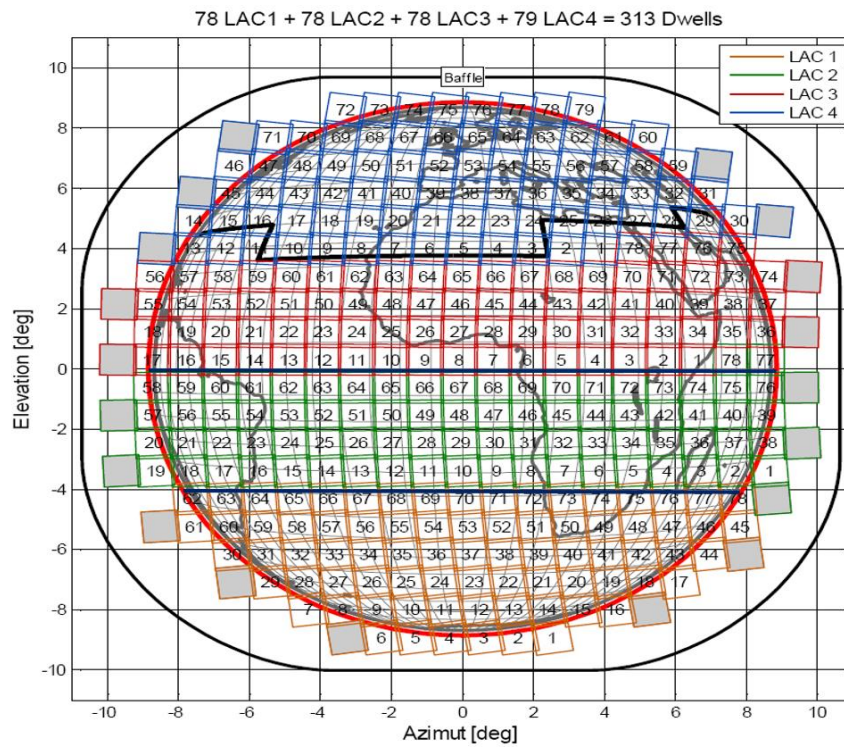


Figura IV: Cobertura del sensor IRS. (Fuente:

<https://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/MTGDesign/index.html>)

Se pueden definir hasta cuatro zonas de LAC separadas y las zonas de LAC se pueden explorar en cualquier orden con una longitud de secuencia máxima de 96 LAC antes de repetir la secuencia.

Las sondas espectrales se transmiten al suelo en forma de interferogramas y se transforman en canales espectrales como parte del procesamiento en tierra, antes de ser difundidas a los usuarios finales como conjuntos de datos de nivel 1.

El IRS se basa en un interferómetro de imágenes de Fourier con una resolución hiperespectral de $0,625 \text{ cm}^{-1}$ de número de onda, tomando medidas en dos bandas, el infrarrojo de onda larga (LWIR) y el infrarrojo de onda media (MWIR), con una resolución espacial de 4 km.

Resumen de especificaciones técnicas:

- Dos bandas espectrales: MWIR: 1600 a 2250 cm^{-1} (4,44-6,25 μm) y LWIR: 680 a 1210 cm^{-1} (8,26-14,70 μm).
- Cobertura total de disco en 60 min, región Europa (cobertura de área local) 30 min.
- Resolución terrestre de 4 km x 4 km.
- La medición radiométrica oscila entre 180 K y 313 K (temperatura equivalente del cuerpo negro).

- Ruido radiométrico espectral (excluida la calibración espectral) a 280 K en cuerpo negro: entre 170 y 900 mK dependiendo del número de onda considerado dentro de la banda de interés.¹³

9. DATOS

Para este trabajo, se ha utilizado datos facilitados por el *Image and Signal Proprocessing group* de la Universidad de Valencia, en adelante ISP.

Estos datos son realmente simulaciones para IRS ofrecidas en primera instancia por *European Organisation for Meteorological Satellites*, en adelante EUMETSAT, obtenidas a partir de los datos de IASI. Es decir, conociendo las características de ambos sensores y satélites correspondientes, se han transformado datos del sensor que ya está en funcionamiento (IASI) de forma que se asemejen lo máximo posible a cómo serán los ofrecidos por el sensor que aún no está en funcionamiento (IRS). Esto habilita la posibilidad de crear de antemano scripts y programas para tratar los datos de IRS directamente, reduciendo el periodo de “prueba”.

Estos datos consisten en 298 archivos digitales en formato de MATLAB (extensión .mat), los cuales están divididos en 2 grupos: Los datos de satélite y los datos climáticos.

Los datos de satélite están compuestos por las propias imágenes del sensor, representando cada archivo el conjunto de imágenes correspondientes a un “barrido” de este sensor entre determinadas fechas y horas. Estos archivos han sido cargados y guardados en formato de MATLAB, y constituirán las observaciones en sí de este proyecto.

Los datos climáticos constan de un archivo en formato MATLAB asociado a cada imagen de satélite, donde se aporta información atmosférica a distintos niveles de la atmósfera, como la temperatura, la cantidad de ozono, humedad etc. Estos serán las variables que nos permitirán entrenar nuestros modelos más adelante. Hay que destacar que estos datos no provienen de observaciones directas, sino que vienen dadas como resultados del análisis realizados del *European Centre for Medium-Range Weather Forecasts* (ECMWF). Estos datos constan de estimaciones para 137 niveles de presión diferentes (alturas) de entre 10^{-2} a 10^3 hPa (Hectopascuales), con una precisión de 0.5 grados.

10. PRE-PROCESADO

Antes de abordar el análisis y computación de los datos, los datos se dividen en dos conjuntos de forma aleatoria: el conjunto de entrenamiento (Training) y el conjunto de validación (Test), entrenar los modelos y validarlos con distintos datos, ya que de utilizar el mismo conjunto, los modelos

¹³ European Organisation for Meteorological Satellites (EUMETSAT). Meteosat Third Generation. Obtenido en: <https://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/MTGDesign/index.html>

obtendrían en todo caso excelentes resultados con dichos datos, pero no podría conocerse con cierta fiabilidad su comportamiento con un conjunto de datos distinto, lo cual es su finalidad.

Para ello, se cargan en MATLAB el directorio de datos, obteniendo un *array* o vector cuyos índices se corresponden con cada dato. Luego, se crea una “semilla” aleatoria pero “concreta” del tamaño de los datos, es decir, tendremos guardados unos nuevos índices que representaran a cada fichero pero barajados al azar en una variable, y aunque ejecutemos el programa en distintas ocasiones, los ficheros serán mezclados exactamente de la misma manera. Con esto ya podemos separar el conjunto de entrenamiento del conjunto de validación de forma aleatoria, evitando por ejemplo, que el conjunto de entrenamiento represente una zona muy concreta y distinta del conjunto de validación, y además, utilizando toda la extensión total tanto para el entrenamiento como para la validación.

Más tarde, para homogeneizar ambos grupos de datos, se calcula la media, se guarda su valor para un posterior uso, y se sustrae de cada conjunto de datos, obteniendo 2 nuevos conjuntos de datos iguales que los anteriores pero con media cero, es decir, centrados. Además, se desea desechar las imágenes que contienen demasiadas nubes utilizando los datos climáticos aportados inicialmente y obtener los datos necesarios, procedentes de las observaciones como los datos climáticos asociados, para poder realizar el ACP.

Para ello, se realizan dos bucles, uno para conjunto de datos, en los cuales se extrae la media, se compara las imágenes una a una con los datos climáticos para saber dónde hay o no nube y se obtienen las imágenes aptas (lo cual ayuda a reducir la cantidad de procesado al desechar imágenes con nubes). Tras acabar los bucles se guarda la media total de cada conjunto y se sustrae, obteniendo los nuevos conjuntos centrados.

Por otro lado, debido a la alta dimensionalidad y cantidad de los datos, se realiza un Análisis de Componentes Principales para cada grupo de datos (entrenamiento y validación), escogiendo un número de 100 componentes principales, es decir, se utilizaran las 100 componentes más significativas, ya que son suficientes para cumplir con nuestra finalidad y va reducir la dimensionalidad de los datos y permitir una computación más eficiente, tal y como se comentaba anteriormente (6.1 ANÁLISIS DE COMPONENTES PRINCIPALES).

Tras la aplicación del ACP, el siguiente paso será realizar ya los modelos de regresión.

11. APLICACIÓN DE LOS MODELOS Y OBTENCIÓN DE RESULTADOS

Para este proyecto, dadas las limitaciones procuradas por las características técnicas del hardware utilizado, se han empleado dos modelos de regresión: la regresión lineal regularizada y la regresión kernel regularizada. Estos modelos, por lo general, suelen ser suficientemente precisos con un coste computacional inferior al de otros modelos de regresión que también podrían haberse aplicado, como el SVM, redes neuronales o árboles de decisión.

Ambos modelos podrían haberse programado desde cero mediante MATLAB, no obstante, el ISP ofrece funciones que realizan todos estos modelos, descargables en su página web, estando estos ya preparados para su uso y reduciendo la cantidad de código a programar por cualquier usuario, y evitando posibles errores en la programación de los modelos y su consecuente pérdida de tiempo, ya que estos ya han sido probados. Cada modelo aportado por el ISP consta de dos funciones, una que entrena el modelo y se puede guardar en una variable, y otra que carga dicha variable y ejecuta la validación, es decir, esta última es la que aplica el modelo en sí.

Con esto, se aplican los modelos a los datos, tomando 10000 datos para la **regresión lineal regularizada**, y 5000 para la **regresión kernel regularizada**, ya que la toma de mayor cantidad de datos, se traduce en un tiempo de computación excesivamente alto debido a las características técnicas del hardware utilizado.

Aplicados los modelos, hay que tener en cuenta que los datos resultantes hay que sumarles de nuevo la media para obtener el valor correcto.

PARTE IV: RESULTADOS Y CONCLUSIONES

12. RESULTADOS

12.1. TEMPERATURA

Como resultados, se ha obtenido en primer lugar el diagrama de dispersión de cada método de regresión, y su error medio, a diferentes alturas para cada variable de estudio (Temperatura, ozono y humedad).

- A altura de la superficie terrestre:

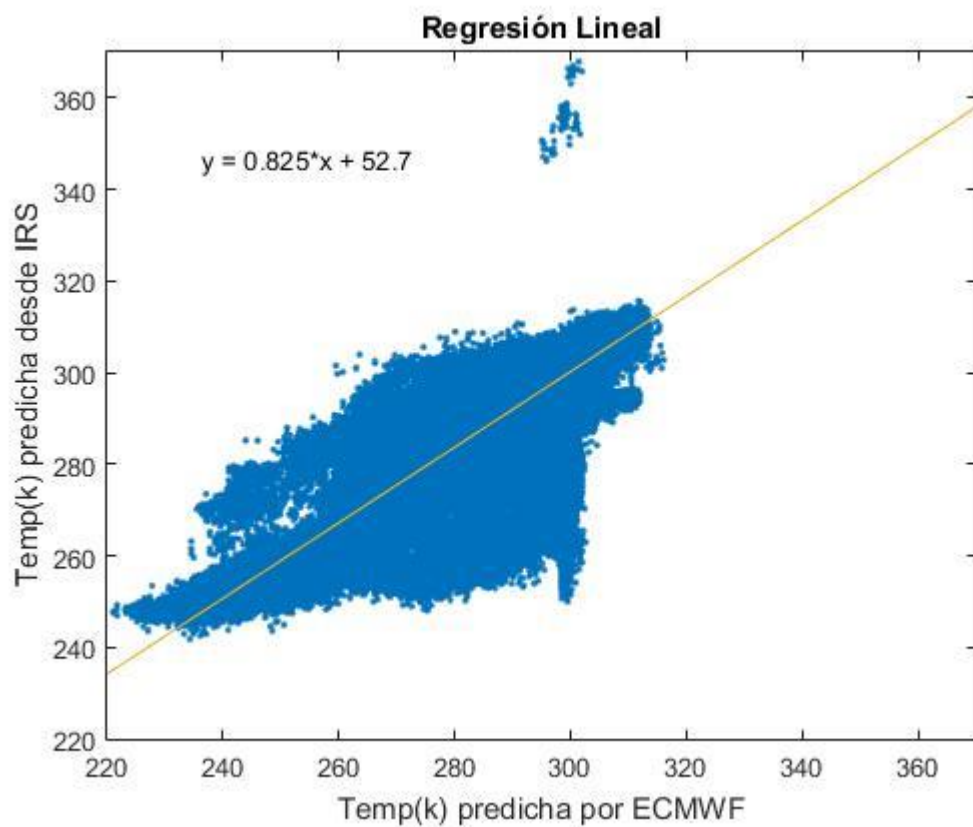


Figura V: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de superficie, para la temperatura. (Fuente: elaboración propia).

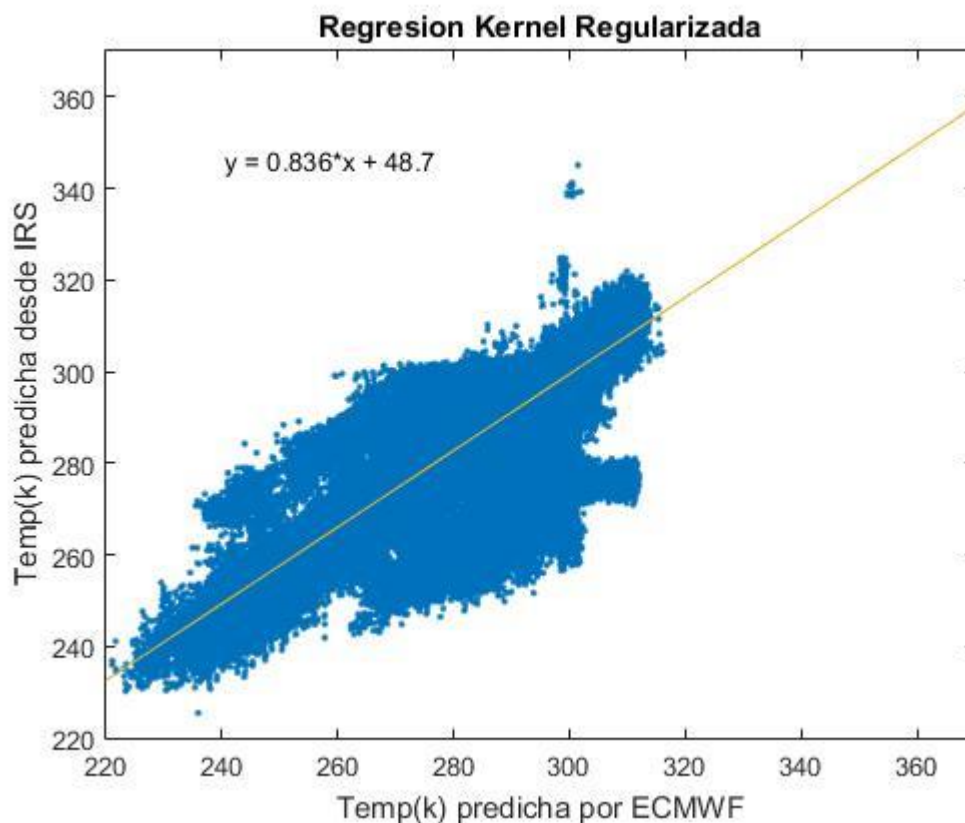


Figura VI: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de superficie, para la temperatura. (Fuente: elaboración propia).

Como podemos observar gráficamente (Figura V y Figura VI), la regresión lineal tiene algunas predicciones anómalas que se alejan mucho de la ideal, llegando a predecir con un error de hasta más 60ºKelvin, mientras que en la regresión kernel regularizada esto ocurre con un número ínfimo de predicciones y el error máximo apenas alcanza los 40ºKelvin en esos casos. En cuanto el balance global entre ambas, la regresión kernel regularizada muestra resultados algo mejores que la regresión lineal ya que se ajusta más a la línea de tendencia ideal, tal y como muestran los resultados analíticos, donde podemos ver que la pendiente de la línea de tendencia de la regresión kernel es algo menos pronunciada, teniendo un valor más cercano a la unidad ($1 > 0.836 > 0.825$).

	Temperatura (Kelvin)
Regresión Lineal	3.6453
Regresión Kernel	3.4844
Diferencia	0.1609

Tabla I: Errores medios para la altura de suelo, para la temperatura en ambos modelos de regresión. (Fuente: elaboración propia).

Tal y como se podía predecir tras el análisis visual y analítico de los diagramas de dispersión, al calcular el error medio de cada variable para cada método de regresión se observa (Tabla I) que la

regresión kernel tiene una precisión algo mayor, aunque la diferencia entre ambos métodos no es muy grande.

- A mitad de atmósfera aproximadamente:

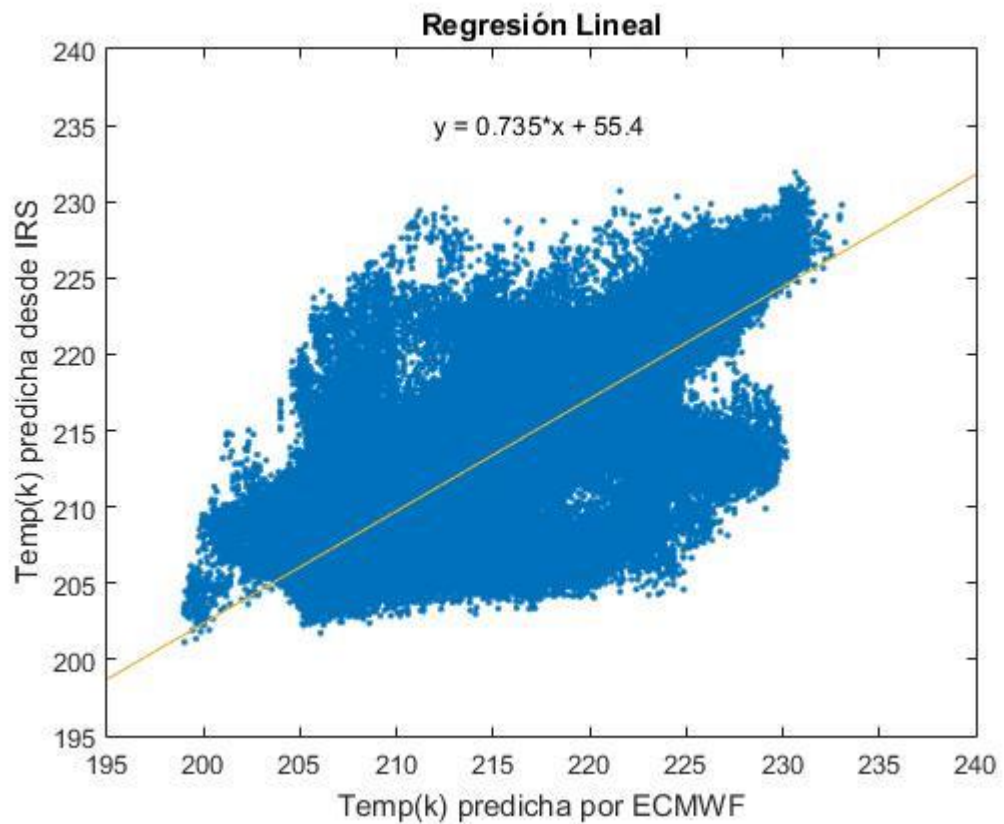


Figura VII: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de media atmósfera, para la temperatura. (Fuente: elaboración propia).

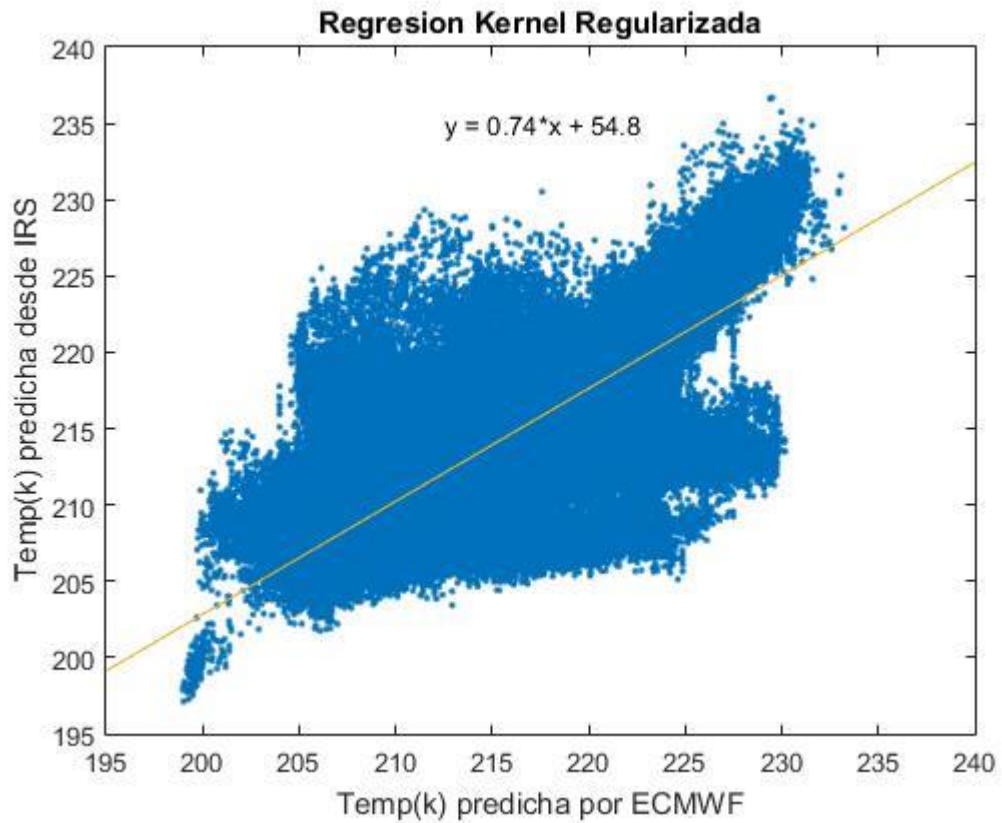


Figura VIII: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de media atmósfera, para la temperatura. (Fuente: elaboración propia).

Tras observar los gráficos (Figura VII y Figura VIII), se comprueba que apenas existen valores muy anómalos en ambos modelos en este caso, y además presentan una distribución muy parecida entre sí. No obstante, al observar analíticamente ambos resultados, se puede comprobar que la regresión kernel ha mejorado por muy poco los resultados de la regresión lineal, pero ambos modelos presentan casi los mismos resultados, con líneas de tendencia casi idénticas.

	Temperatura (Kelvin)
Regresión Lineal	2.0134
Regresión Kernel	1.9873
Diferencia	0.0261

Tabla II: Errores medios para altura a mitad atmósfera, para la temperatura en ambos modelos de regresión, para la temperatura. (Fuente: elaboración propia).

Coincidiendo con el análisis visual y analítico de los diagramas de dispersión, se comprueba que los errores medios extraídos para cada modelo de regresión (Tabla II) apenas se diferencian por un par de centésimas de grado, aunque es la regresión kernel regularizada la que obtiene el error más bajo.

- En el final de la atmósfera:

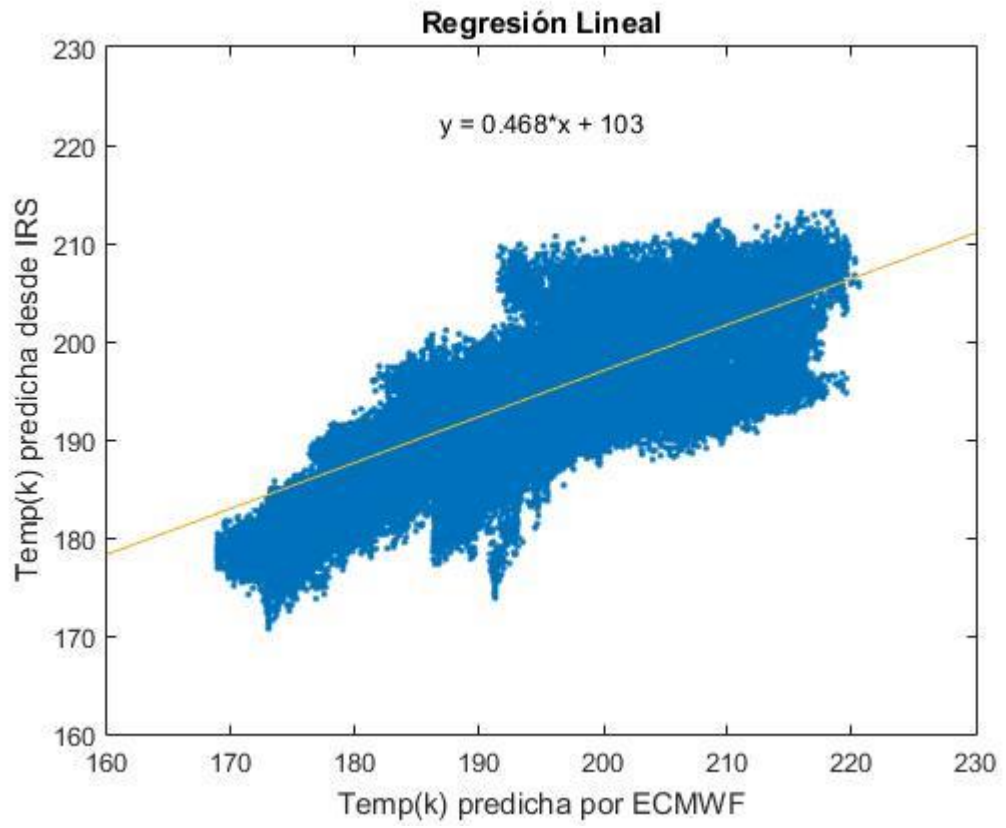


Figura IX: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF en lo alto de la atmósfera, para la temperatura. (Fuente: elaboración propia).

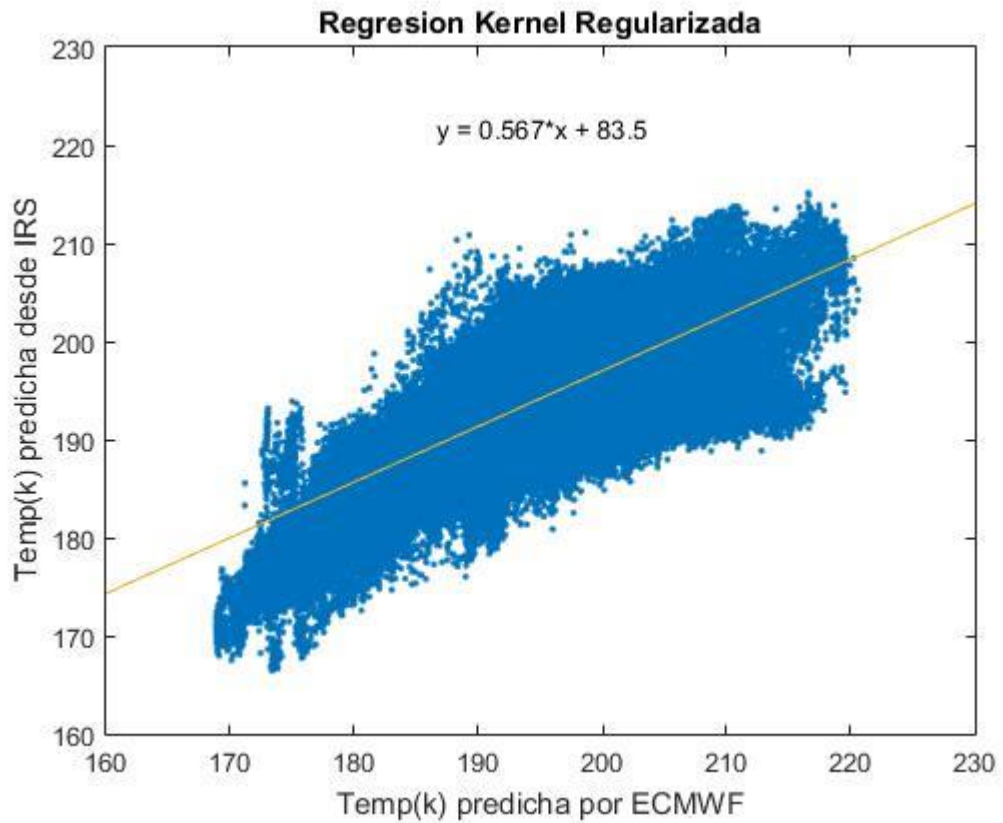


Figura X: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF en lo alto de la atmósfera, para la temperatura. (Fuente: elaboración propia).

Al analizar visualmente estos gráficos (Figura IX y Figura X), se observa que apenas aparece algún valor anómalo en ambos modelos, siendo el de regresión lineal el que menos disperso parece estar. En este caso, se observa una fuerte pendiente en ambos modelos, alejándose mucho de la línea de tendencia ideal y siendo muy similares entre sí.

	Temperatura (Kelvin)
Regresión Lineal	3.0512
Regresión Kernel	2.9590
Diferencia	0.0922

Tabla III: Errores medios para lo alto de la atmósfera, para la temperatura en ambos modelos de regresión. (Fuente: elaboración propia).

Al estudiar el error medio (Tabla III) destaca la similitud entre ambos modelos tal y como ocurría en el anterior caso, no obstante el error es mayor, sobre unos 3 grados de diferencia en ambos modelos.

12.2. OZONO

- A altura de la superficie terrestre:

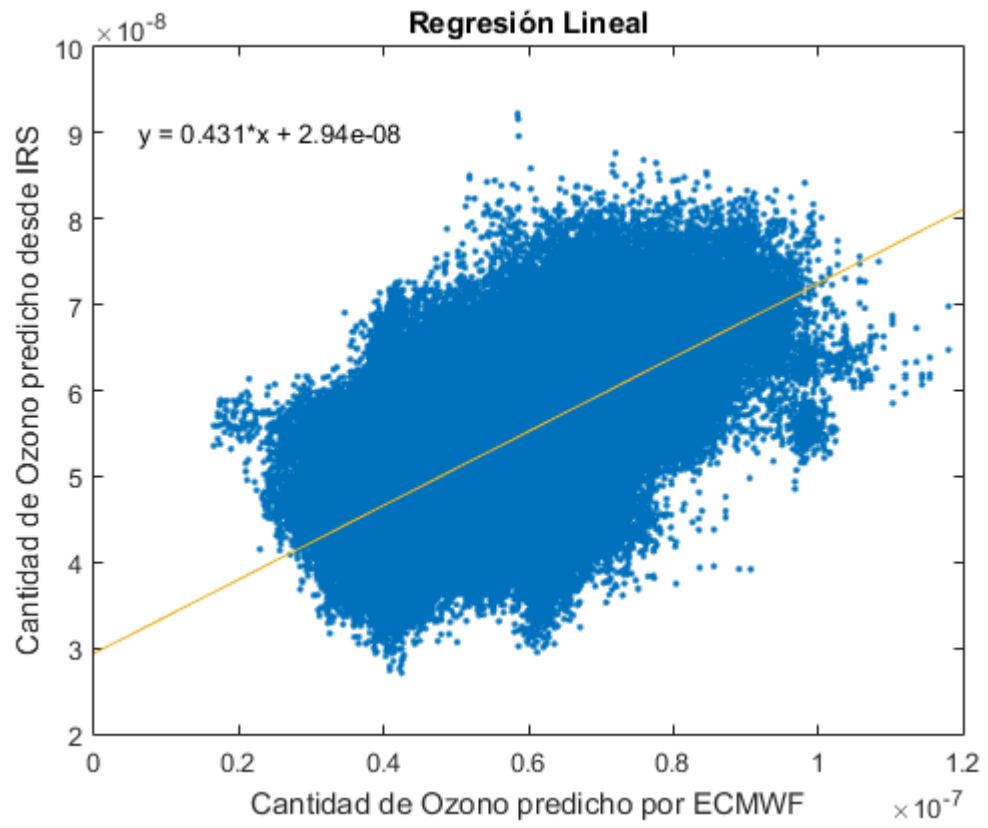


Figura XI: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de superficie, para la presencia de ozono. (Fuente: elaboración propia).

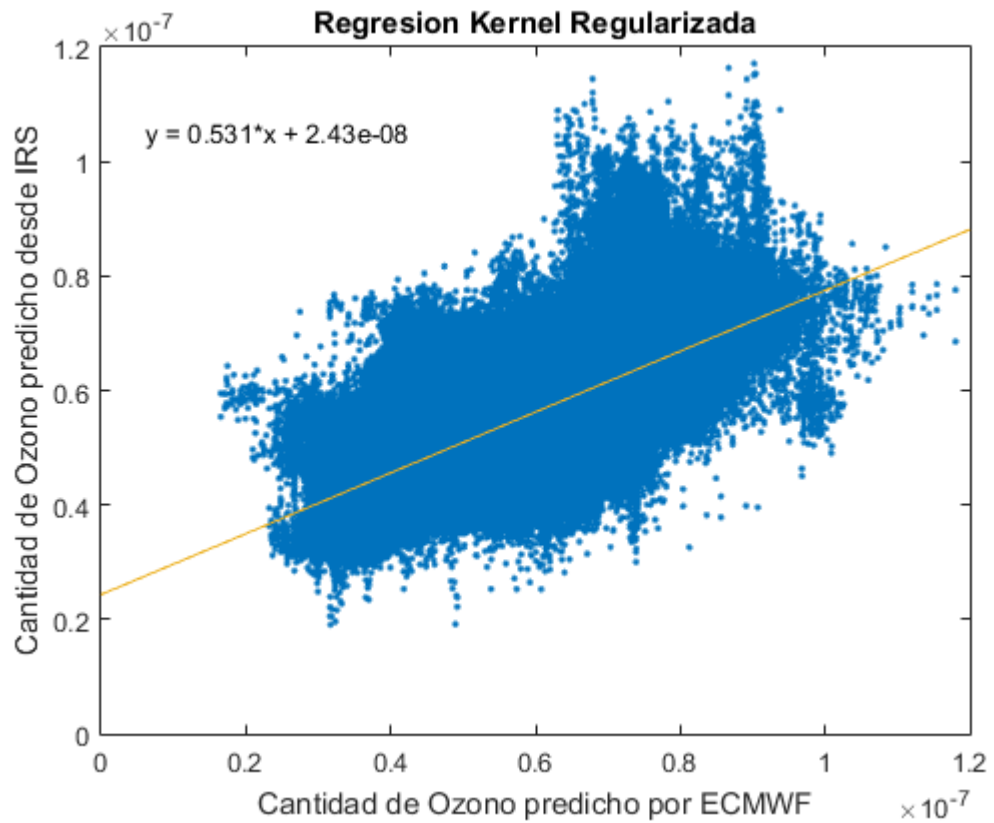


Figura XII: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de superficie, para la presencia de ozono. (Fuente: elaboración propia).

En este caso, las predicciones de la regresión kernel regularizada parecen contener más valores anómalos, no obstante, la tendencia general de estas predicciones es algo mejor que la obtenida con la regresión lineal, tal y como indica su línea de tendencia (Figura XI y Figura XII).

	Ozono
Regresión Lineal	$0.7332 \cdot 10^{-8}$
Regresión Kernel	$0.6827 \cdot 10^{-8}$
Diferencia	$0.0505 \cdot 10^{-8}$

Tabla IV: Errores medios para la altura de suelo, para la presencia de ozono atmosférico en ambos modelos de regresión. (Fuente: elaboración propia).

Como podemos ver (Tabla IV), el error medio obtenido para los diferentes modelos es muy similar, aunque, coincidiendo con los resultados obtenidos anteriormente, la regresión lineal presenta un error algo mayor.

- A mitad de atmósfera aproximadamente:

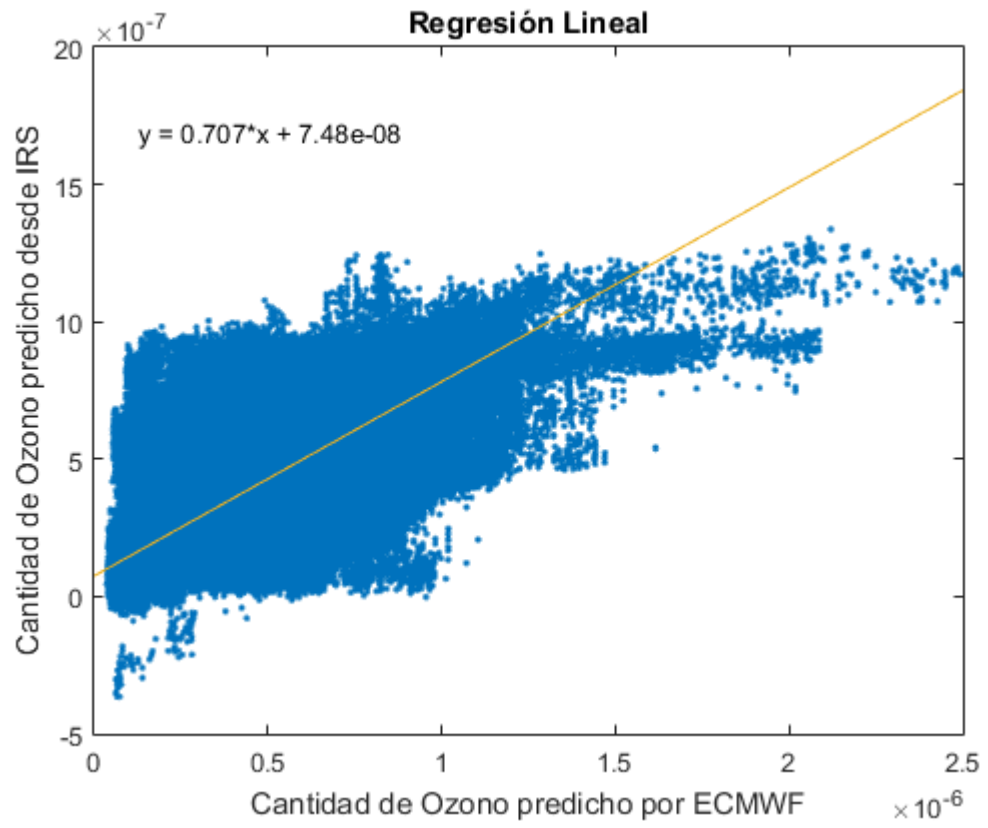


Figura XIII: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de media atmósfera, para la presencia de ozono. (Fuente: elaboración propia).

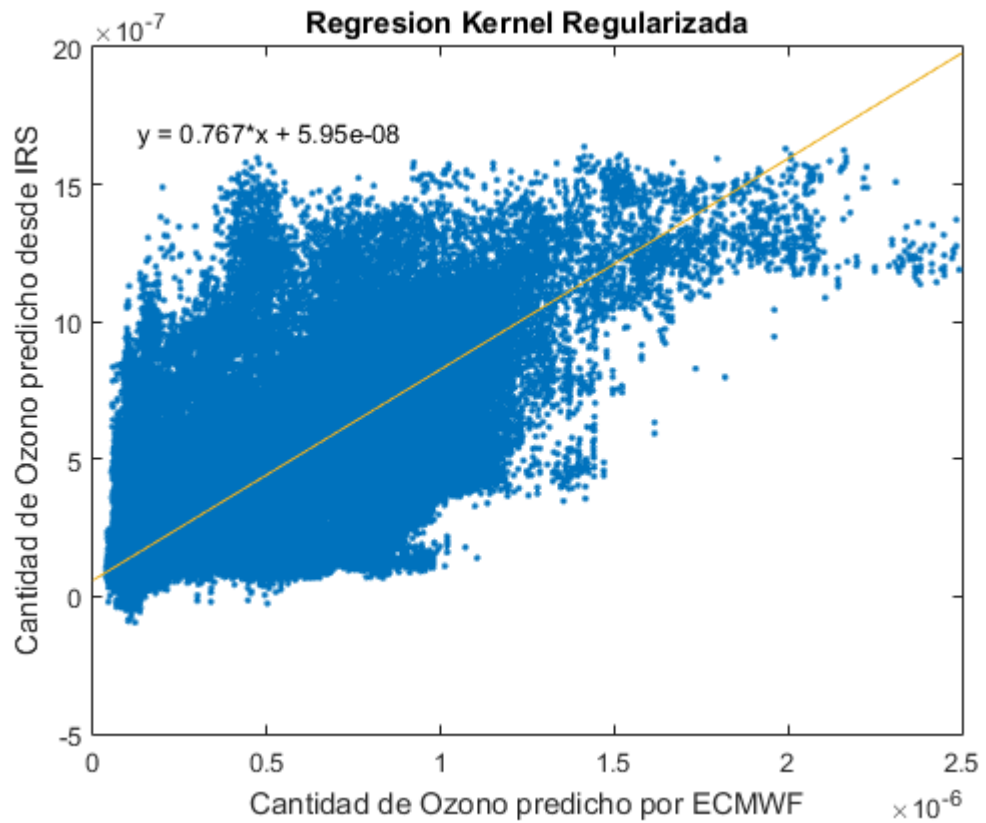


Figura XIV: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de media atmósfera, para la presencia de ozono. (Fuente: elaboración propia).

En esta zona de la atmosfera ocurre algo similar a la anterior, los resultados ofrecidos por la regresión kernel parecen tener valores más anómalos que la lineal, pero la tendencia general de los datos se acerca más al comportamiento ideal.

	Ozono
Regresión Lineal	$0.1485 \cdot 10^{-5}$
Regresión Kernel	$0.1440 \cdot 10^{-5}$
Diferencia	$0.0045 \cdot 10^{-5}$

Tabla V: Errores medios para altura a mitad atmósfera, para la presencia de ozono atmosférico en ambos modelos de regresión. (Fuente: elaboración propia).

- En el final de la atmósfera:

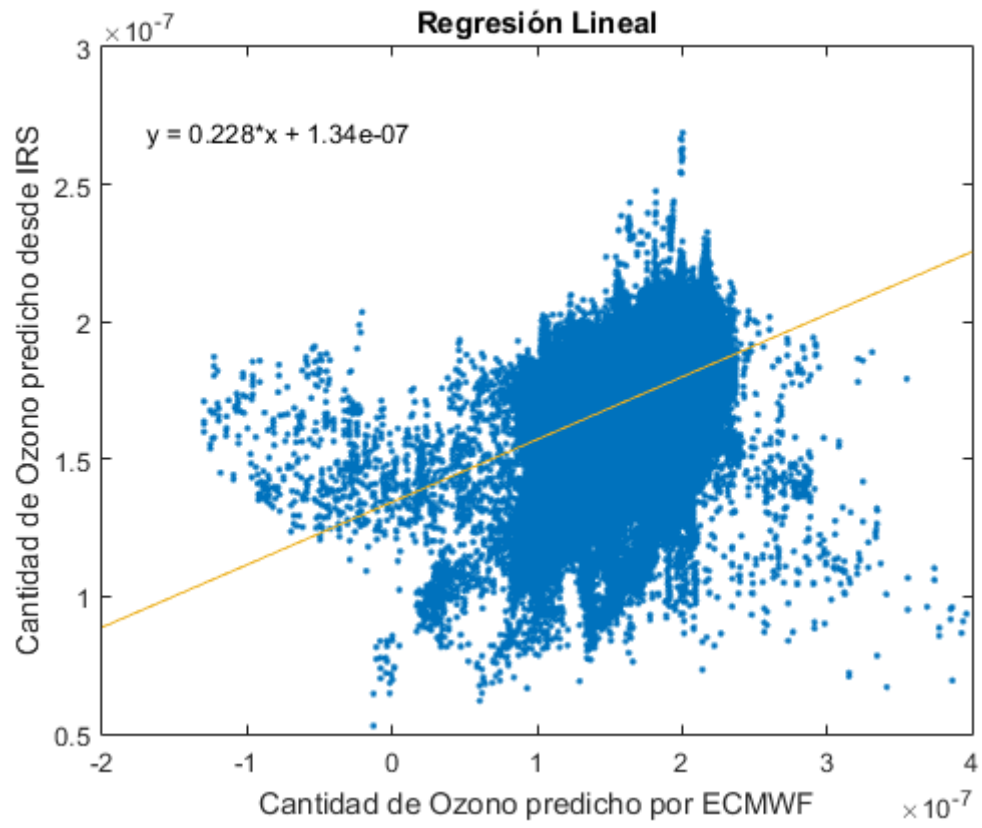


Figura XV: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura final de la atmósfera, para la presencia de ozono. (Fuente: elaboración propia).

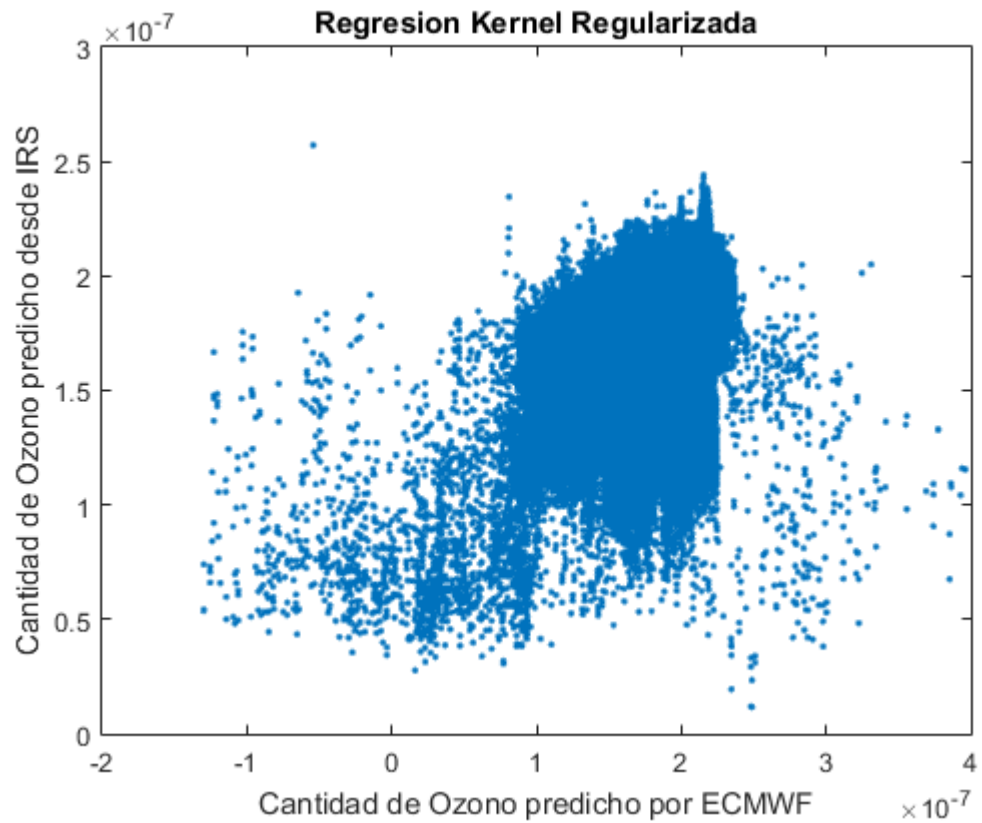


Figura XVI: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de final de atmósfera, para la presencia de ozono. (Fuente: elaboración propia).

En estas alturas de la atmosfera, las predicciones realizadas con ambos modelos sufren una gran dispersión, aunque el error medio cometido es muy pequeño.

	Ozono
Regresión Lineal	$0.2736 \cdot 10^{-6}$
Regresión Kernel	$0.2405 \cdot 10^{-6}$
Diferencia	0.0331

Tabla VI: Errores medios para lo alto de la atmósfera, para la presencia de ozono atmosférico en ambos modelos de regresión. (Fuente: elaboración propia).

12.3. HUMEDAD

- A altura de la superficie terrestre:

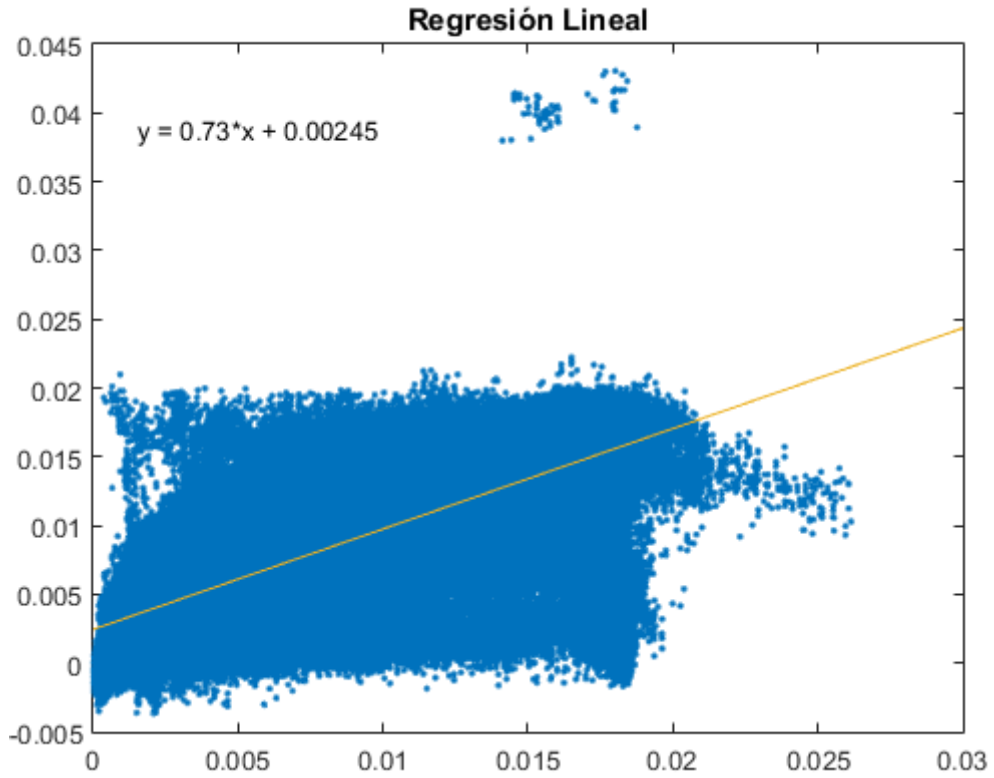


Figura XVII: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de superficie, para la humedad. (Fuente: elaboración propia).

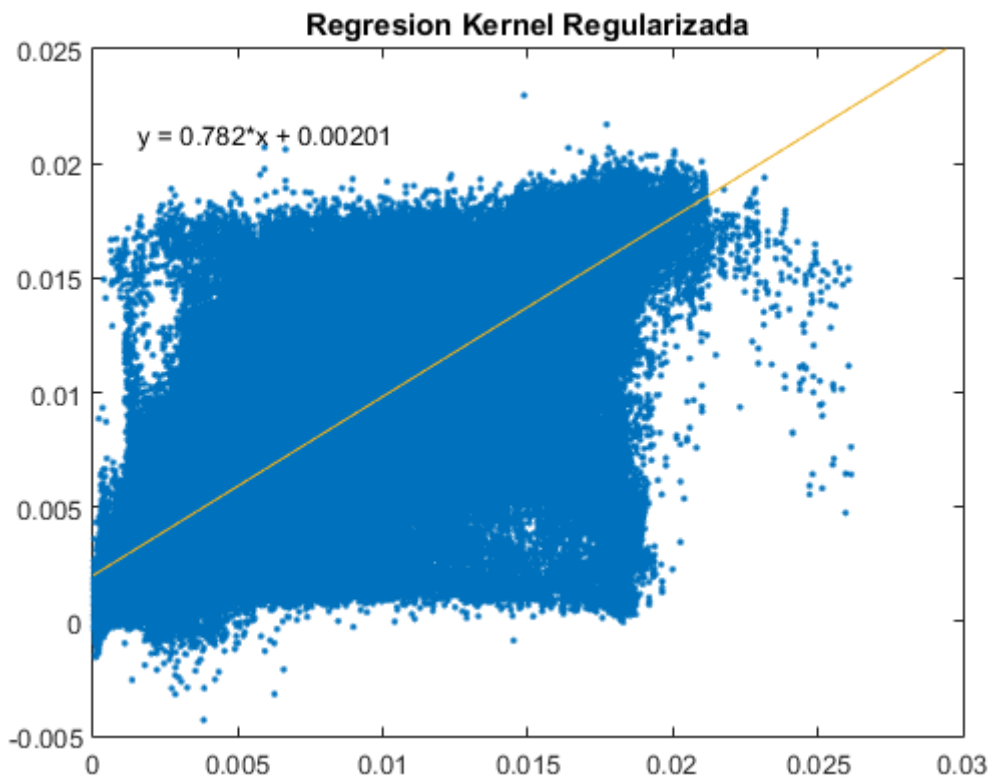


Figura XVIII: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de superficie, para la humedad. (Fuente: elaboración propia).

Al observar la dispersión de los datos (Figura XVII y Figura XVIII) se puede percibir un grupo de anomalías en la regresión lineal, que afectara negativamente a la precisión obtenida por este método. Excluyendo esta excepción, los resultados gráficos obtenidos por ambos métodos son extremadamente parecidos en este caso, de hecho, los 2 tienen una línea de tendencia muy similar que se acerca bastante a la ideal, no obstante, la regresión kernel consigue mejorar los resultados obtenidos por la regresión lineal.

	Humedad
Regresión Lineal	0.0021
Regresión Kernel	0.0018
Diferencia	0.0003

Tabla VII: Errores medios para la altura de suelo, para la humedad en ambos modelos de regresión. (Fuente: elaboración propia).

Los errores medios obtenidos tienen gran similitud, corroborando de nuevo los resultados gráficos comentados anteriormente.

- A mitad de atmósfera aproximadamente:

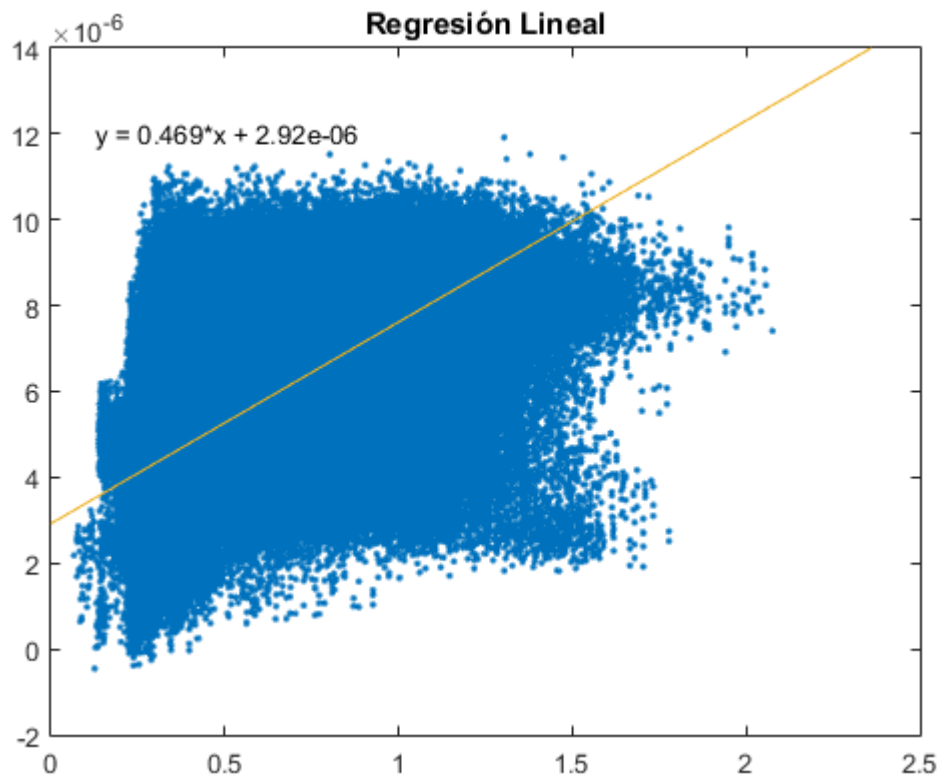


Figura XIX: Diagrama de dispersión entre la Regresión lineal y los datos ECMWF a altura de media atmósfera, para la humedad. (Fuente: elaboración propia).

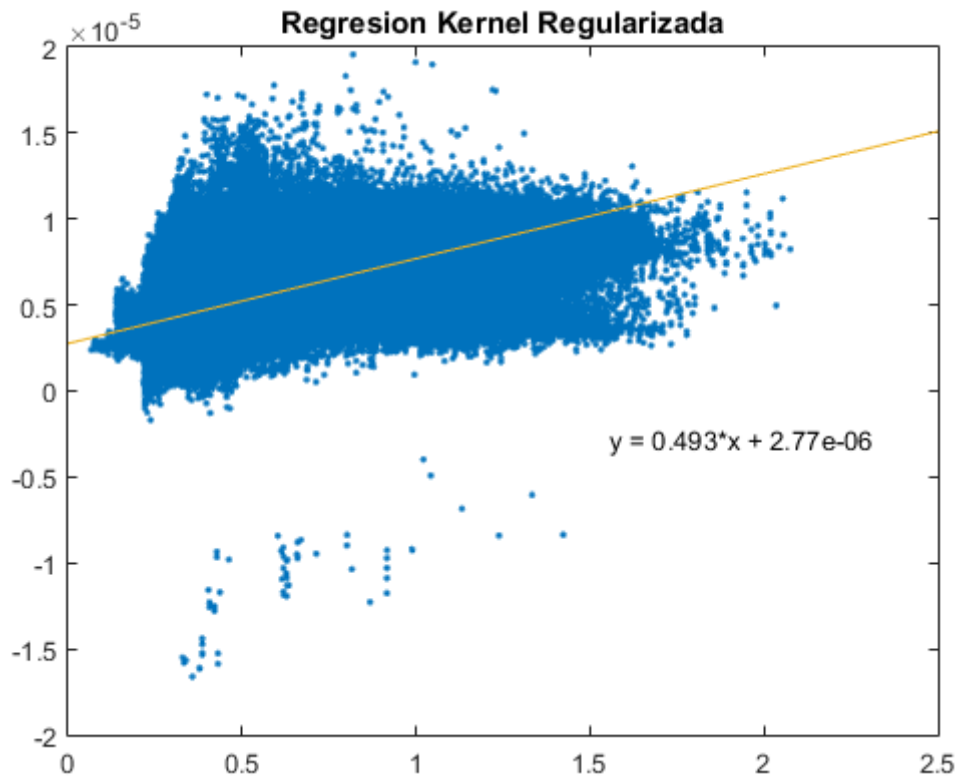


Figura XX: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de media atmósfera, para la humedad. (Fuente: elaboración propia).

Para este caso se producen diversas anomalías individuales en la regresión kernel, pese a ello, los resultados ofrecidos por esta regresión siguen siendo algo mejores que los de la regresión lineal, aunque muy parecidos, tal como podemos observar (Figura XIX y Figura XX) al analizar su línea de tendencia o el error medio obtenido para cada uno.

	Humedad
Regresión Lineal	$0.1492 \cdot 10^{-5}$
Regresión Kernel	$0.1437 \cdot 10^{-5}$
Diferencia	$0.0055 \cdot 10^{-5}$

Tabla VIII: Errores medios para altura a mitad atmósfera, para la humedad en ambos modelos de regresión. (Fuente: elaboración propia).

- En el final de la atmósfera:

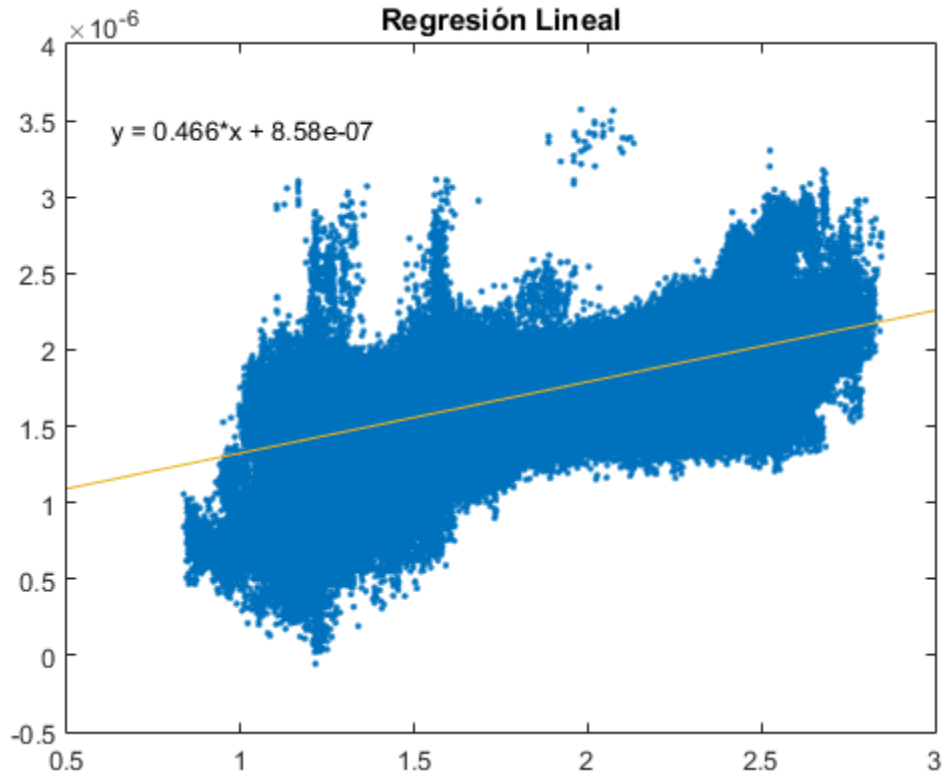


Figura XXI: Diagrama de dispersión entre la Regresión Lineal y los datos ECMWF a altura de final de atmósfera, para la humedad. (Fuente: elaboración propia).

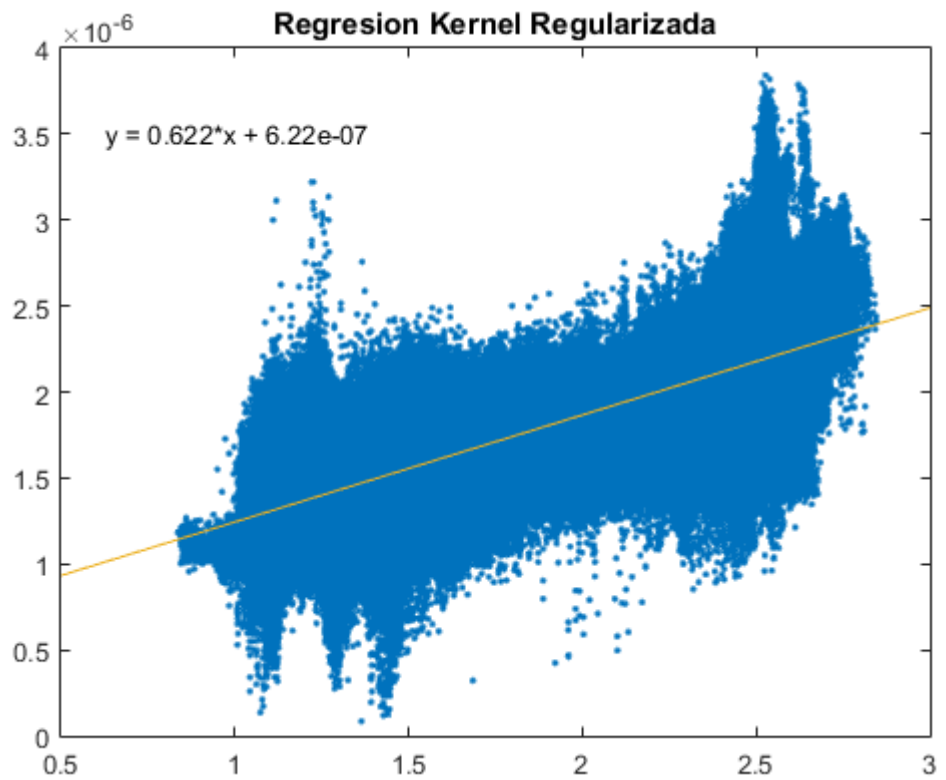


Figura XXII: Diagrama de dispersión entre la Regresión Kernel Regularizada y los datos ECMWF a altura de final de atmósfera, para la humedad. (Fuente: elaboración propia).

Para este caso ambos métodos presentan distintos grupos de anomalías, siendo la regresión kernel la que mejor resultado obtiene.

	Humedad
Regresión Lineal	$0.2740 \cdot 10^{-6}$
Regresión Kernel	$0.2423 \cdot 10^{-6}$
Diferencia	$0.0317 \cdot 10^{-6}$

Tabla IX: Errores medios para lo alto de la atmósfera, para la humedad en ambos modelos de regresión. (Fuente: elaboración propia).

13. MAPAS

Por último, se han obtenido mapas globales para un día, para analizar visualmente el resultado ideal esperado, el obtenido y el error, y poder obtener una mejor idea de la causa de el error en nuestros modelos de predicción:

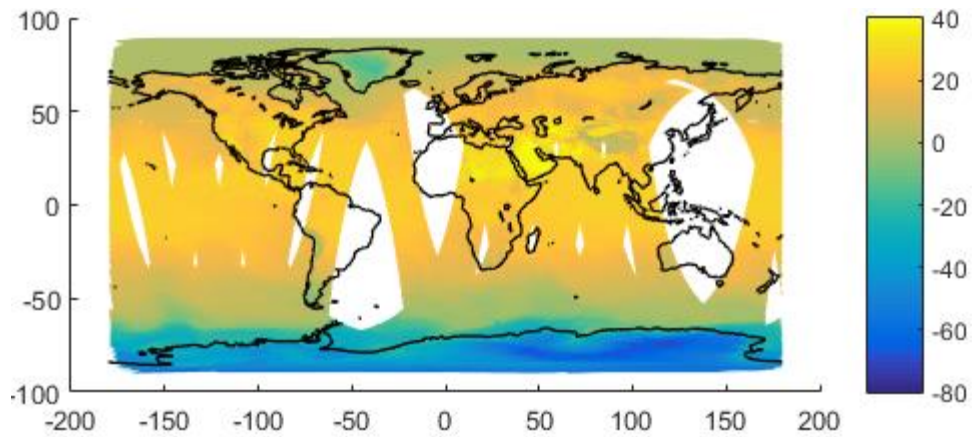


Figura XXIII: Mapa de temperaturas predicho por ECMWFNWP. (Fuente: elaboración propia)

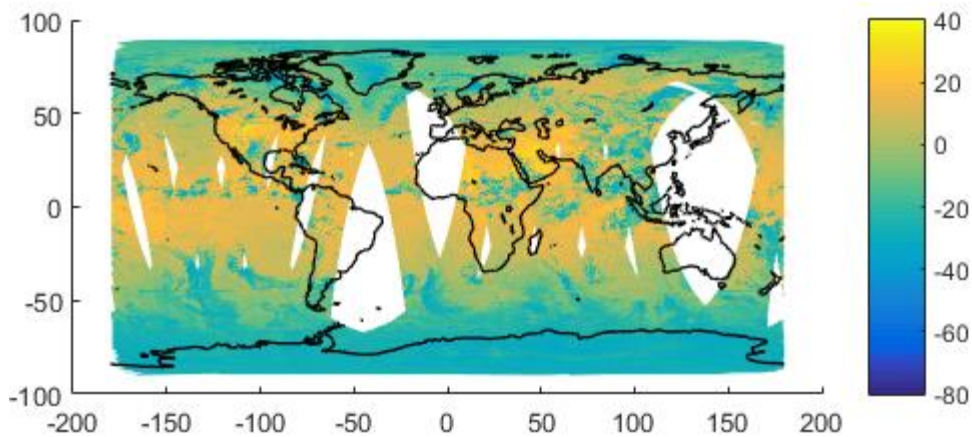


Figura XXIV: Mapa de temperaturas predicho por el modelo de regresión lineal. (Fuente: elaboración propia)

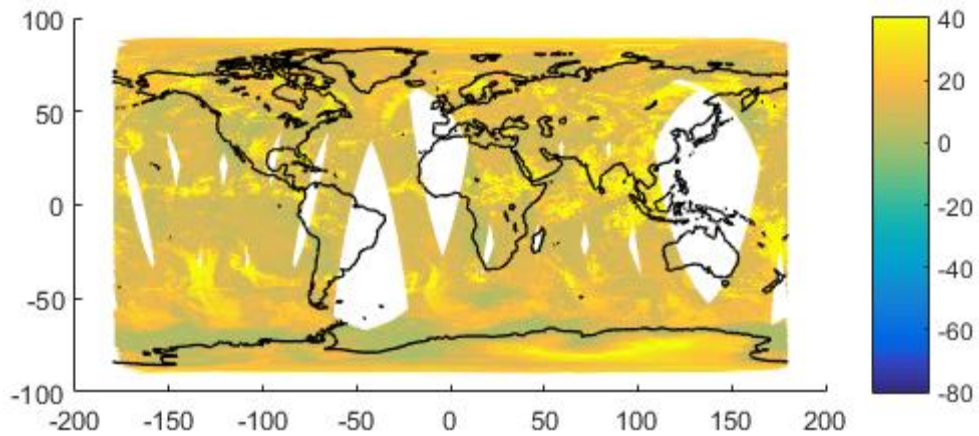


Figura XXV: Mapa de error en temperaturas como resultado de restar el ideal y el predicho. (Fuente: elaboración propia)

Tal y como puede verse, para la temperatura, los resultados guardan grandes similitudes, a excepción de las zonas donde hay nube, en las cuales el modelo falla y resulta en grandes errores en dichas zonas, lo cual aumenta el error global obtenido y contribuye negativamente en nuestro error.

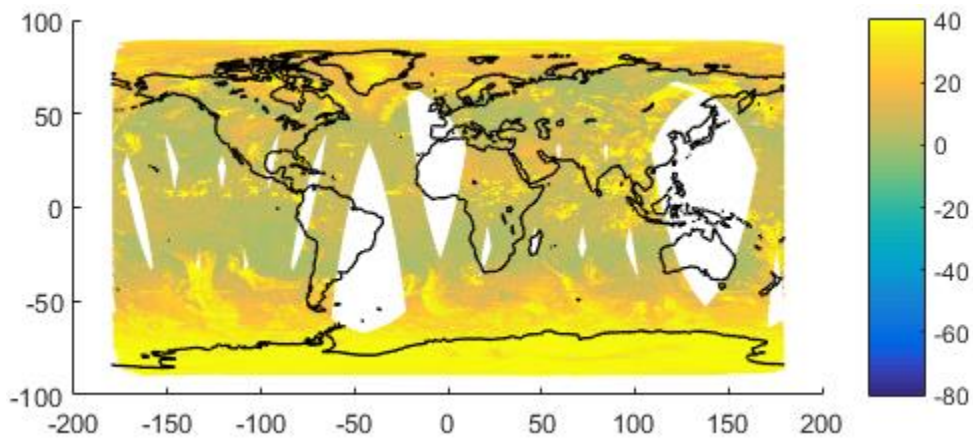


Figura XXVI: Mapa de error humedad. (Fuente: elaboración propia)

En el caso de la humedad, también presenta mayor error en las zonas nubosas, y además en los polos, especialmente en el Sur.

14. CONCLUSIONES

Altura	Modelo de Regresión	Temperatura	Ozono	Humedad
Superficie	Lineal	3.6453	$0.7332 \cdot 10^{-8}$	0.0021

	Kernel	3.4844	$0.6827 \cdot 10^{-8}$	0.0018
Media atmósfera	Lineal	2.0134	$0.1485 \cdot 10^{-5}$	$0.1492 \cdot 10^{-5}$
	Kernel	1.9873	$0.1440 \cdot 10^{-5}$	$0.1437 \cdot 10^{-5}$
Final de la atmósfera	Lineal	3.0512	$0.2736 \cdot 10^{-6}$	$0.2740 \cdot 10^{-6}$
	Kernel	2.9590	$0.2405 \cdot 10^{-6}$	$0.2423 \cdot 10^{-6}$

Tabla X: Errores para cada método de regresión, altura, y variable. (Fuente: elaboración propia)

Cabe destacar que la regresión kernel mejora los resultados obtenidos por la regresión lineal en todos los casos. Aunque teniendo en cuenta que la regresión lineal es menos compleja y requiere menor tiempo computacional y las diferencias entre los resultados de ambos métodos de regresión son en la mayoría de los casos mínimas, se puede considerar un método igual de apto.

Por otra parte, se han obtenido graficado los errores para todas las alturas de la atmósfera y para cada variable, para observar el comportamiento de estos modelos en los distintos niveles de la atmosfera:

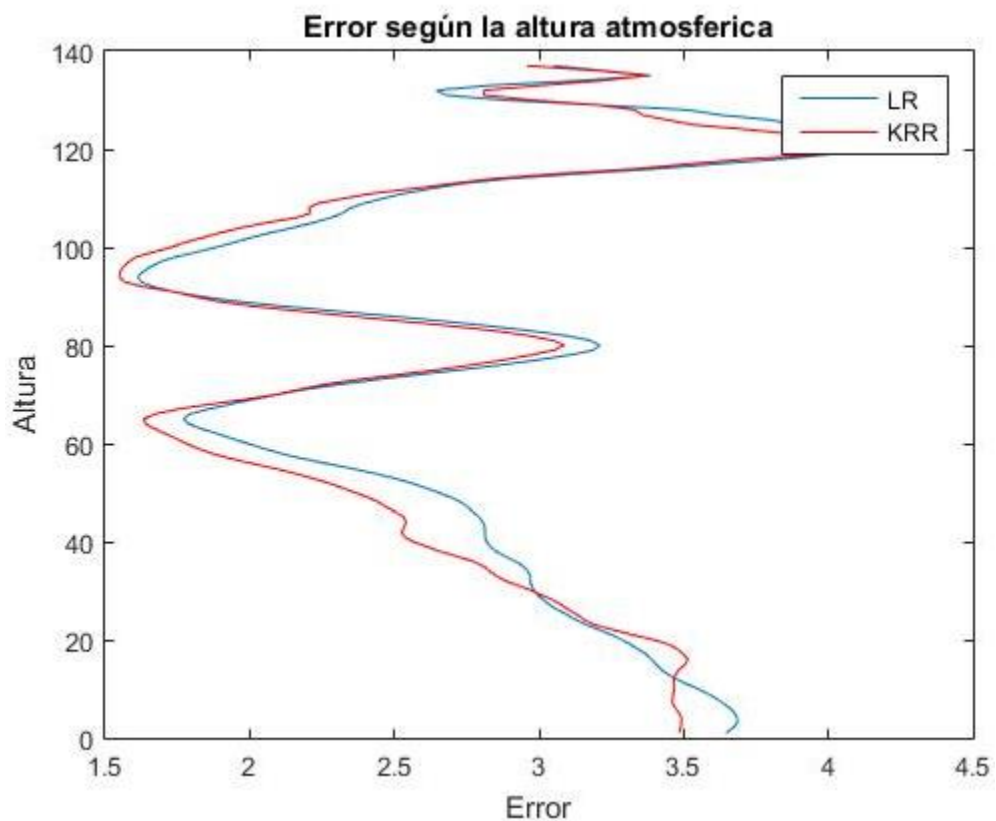


Figura XXVII: Variación del error en temperatura predicha con la altura. (Fuente: elaboración propia)

Al variar la altura atmosférica, en el caso de la predicción de temperaturas, destaca una mejor precisión de los modelos a la altura de media atmosfera y cerca del inicio del último cuarto de atmosfera, no superando el error medio los 4,5º Kelvin en ningún momento.

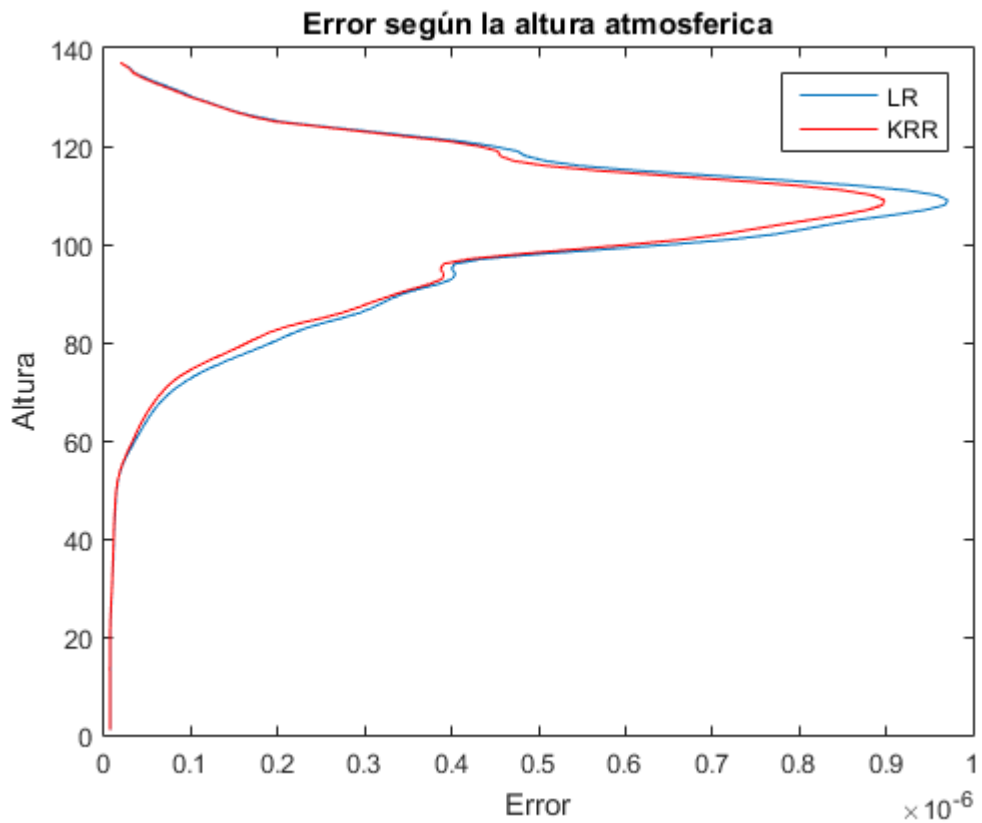


Figura XXVIII: Variación del error en la presencia de ozono predicha con la altura. (Fuente: elaboración propia)

En el caso de la presencia de ozono, observamos que existe una buena predicción para los niveles más bajos y va empeorando con la altura de forma logarítmica, hasta que finalmente el error disminuye justo en el último tramo. Es decir, el cálculo es bueno principalmente para alturas medias o bajas.

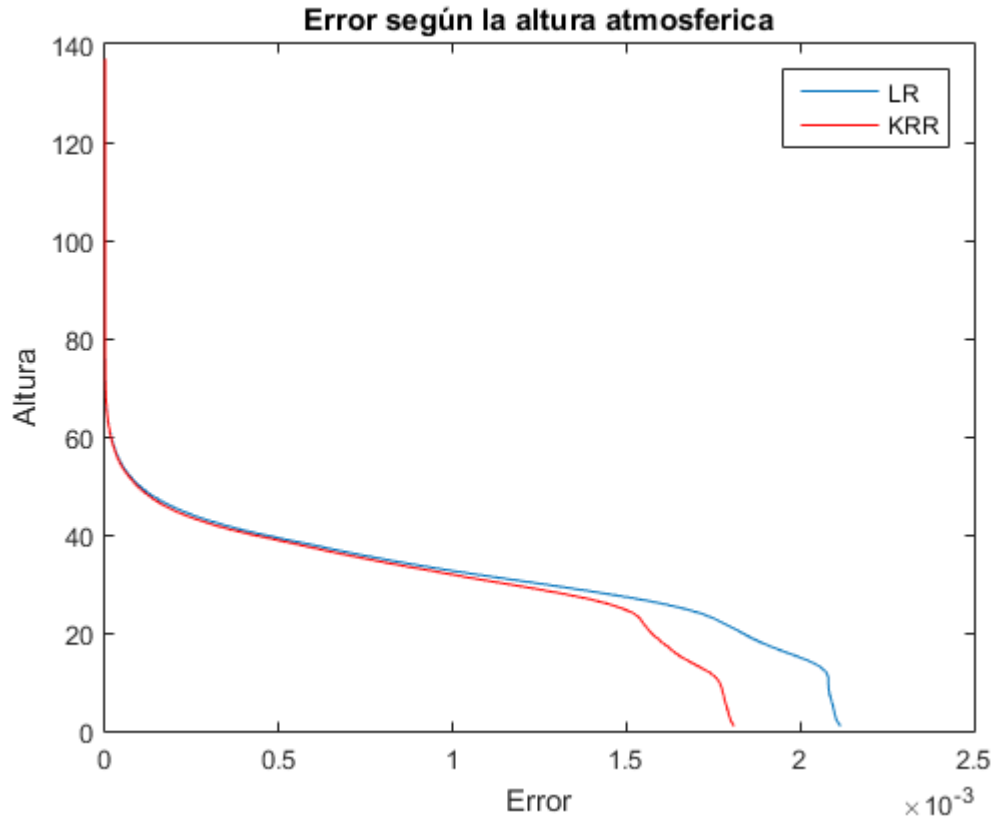


Figura XXIX: Variación del error en humedad predicha con la altura. (Fuente: elaboración propia)

Para el caso de la humedad, existe mayor error en su predicción en alturas bajas, mientras que mejora progresivamente su predicción para alturas altas.

Tras observar estas gráficas concluimos que:

- La precisión varía de forma similar en ambos modelos y para todas las variables, al cambiar la altura, es decir, ambos métodos tienen las mismas tendencias cuando queremos predecir a distintas alturas.
- La regresión kernel tiene mejor precisión que la lineal en la amplia mayoría de los casos.
- Ambos métodos de regresión pueden ser perfectamente válidos para obtener una predicción de los fenómenos climáticos estudiados de forma más rápida que la habitual, pero solo resultan óptimos en determinadas alturas, en la que los datos de la predicción de los modelos y la habitual ofrecen prácticamente el mismo resultado:
 - Para la temperatura, a altura de media atmósfera aproximadamente.
 - Para el ozono, a alturas inferiores a media atmósfera.
 - Para la humedad, a alturas superiores a media atmósfera.

Una posible mejora es utilizar métodos de regresión más complejos, como SVM, Neuronal Networks, Random Forest, etc. Esto mejoraría posiblemente los resultados obtenidos, aunque afectando directamente al coste computacional.

BIBLIOGRAFIA

- ¹Revista de Teledetección (2002). S.B.Ferrero, M.G.Palacio y O.R.Campanella Análisis de componentes principales en teledetección. Obtenido en: https://s3.amazonaws.com/academia.edu.documents/39895531/Anlisis_de_componentes_principales_en_te20151111-820-2sca50.pdf?
- ² ESA. ¿Qué es la teledetección?. Obtenido en: http://www.esa.int/SPECIALS/Eduspace_ES/SEMO1U3FEXF_0.html
- ³Universidad de Murcia. Fundamentos físicos de la teledetección. Obtenido en: <http://www.um.es/geograf/sig/teledet/fundamento.html>
- ⁴ Centro de Astrobiología. Meteorología y Climatología. Obtenido en: <https://cab.inta-csic.es/uploads/culturacientifica/adjuntos/20130121115236.pdf>
- ⁵ Casanova, J.L. (1992). La teledetección en la enseñanza de la meteorología.
- ⁶ Universidad de Murcia. Fundamentos físicos de la teledetección. Obtenido en: <http://www.um.es/geograf/sig/teledet/fundamento.html>
- ⁷ Zancajo Jimeno, J. (2006). Teledetección: nociones y aplicaciones
- ⁸ FisicaLab. Temperatura. Obtenido en: <https://www.fisicalab.com/apartado/temperatura#contenidos>
- ⁹ L.G.Barrón. Universidad de Andalucía. (2009). Cambios climáticos y efectos ambientales. Obtenido en: https://www.researchgate.net/profile/Pablo_Garcia_Murillo/publication/260417035_Elementos_vegetales_del_paisaje_referentes_de_situaciones_climaticas_concretas/links/02e7e5310d233119bf000000.pdf#page=10
- ¹⁰ Revista de Teledetección, (2002). S.B.Ferrero, M.G.Palacio y O.R.Campanella. Análisis de componentes principales en teledetección: Consideraciones estadísticas para optimizar su interpretación. Obtenido en: https://s3.amazonaws.com/academia.edu.documents/39895531/Anlisis_de_componentes_principales_en_te20151111-820-2sca50.pdf
- ¹¹ De la Fuente Crespo, Laura. Componentes Principales. Obtenido en: http://www.fuenterrebollo.com/Master-Econometria/Componentes_Principales.pdf
- ¹² Daniel Peña. Análisis de datos multivariantes (Capítulo 5: Componentes Principales). Obtenido en: https://www.mhe.es/universidad/ciencias_matematicas/pena/home/CAPITULO.PDF
- ¹³ European Organisation for Meteorological Satellites (EUMETSAT). Meteosat Third Generation. Obtenido en: <https://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/MTGDesign/index.html>

PARTE V: ANEXOS

ANEXO I I: CÓDIGO MATLAB

SCRIPT PRINCIPAL: PCA

Este script lee los datos, realiza el Análisis de Componentes Principales y separa los datos en dos grupos, los de entrenamiento y los de testeo:

```

clear all
close all

%% FOLDERS
% donde están las funciones
addpath(genpath('/simpleR-master'))
% donde se guardan los resultados
out_fold = './Out';
% donde están los datos
in_fold2 = 'E:/IASI Data/Data/';
%in_fold2 = 'C:\Users\Máster\Desktop\Data\';

%% DATA IRS
% guarda los nombres de los ficheros
aa = dir(in_fold2);
kk = 1;
for nnn = 1:length(aa)
    bbb = aa(nnn).name;
    if length(bbb)>7
        if strcmp(bbb(end-10:end-4), 'OBS_IRS')
            FILE(kk).name = bbb(1:14);
            kk = kk+1;
        end
    end
end

rand('seed',1)
ii = randperm(length(FILE));

N_train = ii(1:70); % indices para los datos de train
N_test = ii(70:end); % indices para los datos de test

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% PREPARAR LOS DATOS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Temp_m2 = [];
RAD_m3 = [];
iii_n_f = [];
k=0;
len = length(N_train);
for n_f=1:len
    tic
    file_name = FILE(N_train(n_f)).name; % Utilizamos solo los
archivos TRAIN
    %% ASV
    % (de aqui sacamos la mascara de nubes)
    load([in_fold2 file_name '_ASV.mat']);
    %% RADIANCIAS
    load([in_fold2 file_name '_OBS_IRS.mat']);
    RAD_m = RES_OBS_IRS.TOA';

```

```

%%
ii_cld = RES_ASV.BAS.C_A==0;
RAD_m2 = RAD_m(ii_cld,:); %se queda solo con una parte de la
imagen, sin nubes
ii_out = [];
for nb=1:size(RAD_m2,2)
zs = RAD_m2(:,nb) - median(RAD_m2(:,nb));
zs2 = zs./mad(zs(:),1); %% busca los outliers (Valores muy
raros)
ii_out(:,nb) = abs(zs2)>10 | isnan(zs2); %% guarda los indices
de los outliers
end
ii_out2 = sum(ii_out,2)==0; %% guarda las filas que no tienen
outliers

%% TEMPERATURAS
aaa = find(ii_cld);
iii = aaa(ii_out2);

Temp = RES_ASV.PRP.T(:,iii)'; %temperaturas

ii_out = [];
for nb=1:size(Temp,2)
zs = Temp(:,nb) - median(Temp(:,nb));
zs2 = zs./mad(zs(:),1); %% busca los outliers (Valores muy
raros)
ii_out(:,nb) = abs(zs2)>10 | isnan(zs2); %% guarda los indices
de los outliers
end
ii_out3 = sum(ii_out,2)==0; %% guarda las filas que no tienen
outliers

iii_tot = iii(ii_out3);

Temp_m2 = [Temp_m2; RES_ASV.PRP.T(:,iii_tot)']; % ??

RAD_m3 = [RAD_m3; RAD_m(iii_tot,:)];
iii_n_f = [iii_n_f; n_f*ones(length(iii_tot),1)]; %% ???
[toc n_f]
Arch=[n_f len]
end

mm_Temp = mean(Temp_m2); % media global de temperaturas ( mY )
Temp_m3 = Temp_m2 - repmat(mm_Temp,size(Temp_m2,1),1);

%% CALCULA PCA
mm_IRS = mean(RAD_m3); % saca la media de todo el conjunto train (
mRAD )

RAD_m4 = RAD_m3 - repmat(mm_IRS,size(RAD_m3,1),1); % guarda la
normalizada de todo train

% C = cov(RAD_m4); % PCA de todas las de train
%
% [V D] = eig(C);
% dd = diag(D);
% [ss ii] = sort(dd,'descend');
% V = V(:,ii);
% D = diag(ss);

load('./RES/IRS_PCA')
Xrm = RAD_m4*V(:,1:100); %Obtenemos la PCA

%% GUARDA RESULTADOS
% filePCA = ['./Out/IRS_PCA'];

```

```

%Se separa:
fileMed = ['./Out/MediasNum']
fileTempY = ['./Out/Temp_m2']
fileRADx = ['./Out/RAD_m4']

save(fileTempY, 'Temp_m3');
save(fileMed, 'mm_Temp', 'mm_IRS');
save(fileRADx, 'Xrm');

% save(filePCA, 'V', 'D', 'FILE', 'N_train', 'N_test');

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% TEST
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Temp_m2 = [];
RAD_m3 = [];
iii_n_f = [];
k=0;
len = length(N_test);
for n_f=1:len
    tic
TRAIN    file_name = FILE(N_test(n_f)).name; % Utilizamos solo los archivos

        %% ASV
        % (de aqui sacamos la mascara de nubes)
        load([in_fold2 file_name '_ASV.mat']);

        %% RADIANCIAS
        load([in_fold2 file_name '_OBS_IRS.mat']);
        RAD_m = RES_OBS_IRS.TOA';

        %%
        ii_cld = RES_ASV.BAS.C_A==0;
        RAD_m2 = RAD_m(ii_cld,:); %se queda solo con una parte de la
imagen, sin nubes
        z
        ii_out = [];
        for nb=1:size(RAD_m2,2)
            zs = RAD_m2(:,nb) - median(RAD_m2(:,nb));
            zs2 = zs./mad(zs(:),1); %% busca los outliers (Valores muy
raros)
            ii_out(:,nb) = abs(zs2)>10 | isnan(zs2); %% guarda los indices
de los outliers
        end
        ii_out2 = sum(ii_out,2)==0; %% guarda las filas que no tienen
outliers

        %% TEMPERATURAS
        aaa = find(ii_cld);
        iii = aaa(ii_out2);

        Temp = RES_ASV.PRP.T(:,iii); %temperaturas

        ii_out = [];
        for nb=1:size(Temp,2)
            zs = Temp(:,nb) - median(Temp(:,nb)); %% por que la mediana?
            zs2 = zs./mad(zs(:),1); %% busca los outliers (Valores muy
raros)
            ii_out(:,nb) = abs(zs2)>10 | isnan(zs2); %% guarda los indices
de los outliers
        end
        ii_out3 = sum(ii_out,2)==0; %% guarda las filas que no tienen
outliers

        iii_tot = iii(ii_out3);

        Temp_m2 = [Temp_m2; RES_ASV.PRP.T(:,iii_tot)'];

```

```

        RAD_m3 = [RAD_m3; RAD_m(iii_tot,:)];
        iii_n_f = [iii_n_f; n_f*ones(length(iii_tot),1)];
        [toc n_f]
        Arch=[n_f len]
    end

    Temp_m3_test = Temp_m2 - repmat(mm_Temp,size(Temp_m2,1),1);

    %% CALCULA PCA

    RAD_m4 = RAD_m3 - repmat(mm_IRS,size(RAD_m3,1),1); % guarda la
normalizada

    Xrm_test = RAD_m4*V(:,1:100); %Obtenemos la PCA

    %% GUARDA RESULTADOS
    % filePCA = ['./Out/IRS_PCA'];

    % fileOthers = ['./Out/Others']
    % save(fileOthers,'mm_Temp','Temp_m2','mm_IRS','RAD_m4'); %PROBLEMA:
Todo junto ocupa 12gb, solo hay 8 de ram...

    %Se separa:
    fileTempY = ['./Out/Temp_m2_test']
    fileRADx = ['./Out/RAD_m4_test']

    save(fileTempY,'Temp_m3_test');
    save(fileRADx,'Xrm_test');

```

SCRIPT: MODELOS DE REGRESIÓN

Este script carga los datos de la ACP y entrena y prueba los modelos de regresión y los guarda.

```

clear all
close all

%% FOLDERS
in_fold2 = 'C:\Users\Máster\Desktop\Data\';
in_fold2 = 'E:\IASI Data\Data';
%%Cargar funciones:
addpath(genpath('./simpleR-master'))
%% PCA
% ya tenemos los datos para realizar PCA y las radiancias anteriores
guardadas
for j=3:3
if j==1
    arch='Temp';
    load('./Out/Temp/MediasNum');
    load('./Out/Temp/Temp_m2');
    Ym = Temp_m3;
    load('./Out/Temp/Temp_m2_test');
    load('./Out/Temp/RAD_m4_test');
    YYY = Temp_m3_test;
    load('./Out/Temp/RAD_m4')
    MM=mm_Temp;
    MMT=Temp_m3_test;
elseif j==2
    arch='OZ';
    load('./Out/OZ/MediasNumOZ');
    load('./Out/OZ/OZ_m2');
    Ym = OZ_m3;

```

```

        load('./Out/OZ/OZ_m2_test');
        load('./Out/OZ/RAD_m4_test_OZ');
        YYY = OZ_m3_test;
        load('./Out/OZ/RAD_m4_OZ');
        MM=mm_OZ;
        MMTT=OZ_m3_test;
    elseif j==3
        arch='Hum';
        load('./Out/Hum/MediasNumHUM');
        load('./Out/Hum/HUM_m2');
        Ym = HUM_m3;
        load('./Out/Hum/HUM_m2_test');
        load('./Out/Hum/RAD_m4_test_HUM');
        YYY = HUM_m3_test;
        load('./Out/hUM/RAD_m4_HUM');
        MM=mm_HUM;
        MMTT=HUM_m3_test;
    end

    %% Entrenamos los modelos de regresión:
    Ntrain_LR = 10000;
    Ntrain_KRR = 5000; % max 5000
    % Ntrain_RF = 10;

    i2 = randperm(size(Xrm,1));
    ii_LR = i2(1:Ntrain_LR);
    ii_KRR = i2(1:Ntrain_KRR);
    % ii_RF = i2(1:Ntrain_RF);

    model_LR = trainRLR(Xrm(ii_LR,:),Ym(ii_LR,:));
    model_KRR = trainKRR(Xrm(ii_KRR,:),Ym(ii_KRR,:));
    % model_RF = trainRF1(Xrm(ii_RF,:),Ym(ii_RF,:));

    %%GUARDAR MODELOS
    % if j==1
    %     fileModels = ['./Out/SavedModels/TempModels'];
    % elseif j==2
    %     fileModels = ['./Out/SavedModels/OZModels'];
    % elseif j==3
    %     fileModels = ['./Out/SavedModels/HumModels'];
    % end
    % save(fileModels,'model_LR','model_KRR');

    %% Aplicamos los modelos sobre el conjunto de test:
    Ym = YYY + repmat(MM,size(MMTT,1),1);

    NNN = size(Ym,1);
    KKK = round(NNN/2000);
    Indices = crossvalind('Kfold', NNN, KKK);
    y_hat_LR = zeros(size(Ym));
    y_hat_KRR = zeros(size(Ym));
    % y_hat_RF = zeros(size(Ym));
    for ii2 = 1:KKK
        ii_aux = find(Indices==ii2);
        y_hat_LR(ii_aux,:) = testRLR(model_LR,Xrm_test(ii_aux,:));
        y_hat_KRR(ii_aux,:) = testKRR(model_KRR,Xrm_test(ii_aux,:));
        % y_hat_RF(ii_aux,:) = testRF(model_RF,Xrm_test(ii_aux,:));
        ii2
    end

    y_hat_LR = y_hat_LR + repmat(MM,size(Ym,1),1);
    y_hat_KRR = y_hat_KRR + repmat(MM,size(Ym,1),1);
    % y_hat_RF = y_hat_RF + repmat(MM,size(Ym,1),1);

    % plotear scatters predicciones
    figure,plot(Ym(:,1),y_hat_LR(:,1),'.'),title('Regresión
Lineal'),ylabel('Cantidad de Ozono predicho desde IRS'),xlabel('Cantidad de
Ozono predicho por ECMWF')

```

```

figure,plot(Ym(:,1),y_hat_KRR(:,1),'.'),title('Regresion Kernel
Regularizada'),ylabel('Cantidad de Ozono predicho desde
IRS'),xlabel('Cantidad de Ozono predicho por ECMWF')
% figure,plot(Ym(:,137),y_hat_RF(:,137),'.')
% clacular el error medio para una altura
EMedAlturaUnicaLR(j) = mean(abs(Ym(:,68)-y_hat_LR(:,68)))
EMedAlturaUnicaKRR(j) = mean(abs(Ym(:,68)-y_hat_KRR(:,68)))
% mean(abs(Ym(:,1)-y_hat_RF(:,1)))
% clacular el error medio para todas las alturas
E_LR = mean(abs(Ym-y_hat_LR));
E_KRR = mean(abs(Ym-y_hat_KRR));

desvLR_T = std(abs(y_hat_LR));
desvKRR_T= std(abs(y_hat_KRR));
% E_RF = mean(abs(Ym-y_hat_RF));
figure
plot(E_LR,137:-1:1),title('Error según la altura atmosferica')
hold on
plot(E_KRR,137:-1:1,'r'),
xlabel('Error'),ylabel('Altura')
legend('LR','KRR')
% hold on
% plot(E_RF,1:137,'b')
% calcular el error medio total

mean(E_LR(j))
mean(E_KRR(j))
% mean(E_RF)
end

```

SCRIPT: MAPEADO

Este script realiza los mapas globales.

```

clear all
close all

%% FOLDERS
data_fold = 'E:/IASI Data/Data/';
%addpath(genpath('/media/disk/vista/TFMs/2018_Antonio/'))
addpath(genpath('./simpleR-master'))
%% LOAD DATA
dats = dir(data_fold);
data = dats(3:20);
clear dats
j=1;
for i=1:length(data)
    if rem(i,2)==0,
        date(j).name = data(i).name(1:14);
        j=j+1;
    end
end
clear data i j

for i=1:1%length(date)
    load([data_fold date(i).name '_OBS_IRS']) % X
    load([data_fold date(i).name '_ASV']) % Y

    RAD(i).TOA = RES_OBS_IRS.TOA;

```

```

ASV(i).T = RES_ASV.PRP.T;
ASV(i).O = RES_ASV.PRP.O;
ASV(i).W = RES_ASV.PRP.W;
ASV(i).LAT = RES_ASV.PRP.Latitude;
ASV(i).LON = RES_ASV.PRP.Longitude;
i
end
clear RES_OBS_IRS
clear RES_ASV
% fileRADmap = ['./Out/DataMap/RADmap']
% fileASVmap = ['./Out/DataMap/ASVmap']
%
% save(fileRADmap,'RAD','-v7.3');
% save(fileASVmap,'ASV','-v7.3');

load('./Out/DataMap/RADmap');
load('./Out/DataMap/ASVmap');
%CARGAMOS LOS MODELOS:
load('./Out/SavedModels/TempModels1');
clear model_KRR
load('C:/Users/Usuario/Desktop/TFM/TFM/RES/IRS_PCA')%PCA - La cargamos

%% Bucle de analisis
for i=1:length(date)
i
%%TEMPERATURA
YT = ASV(i).O;
mYT = mean(YT,2);
YTm = YT - repmat(mYT,1,size(YT,2)); %Restamos la media para centrar
los datos

RADm = RAD(i).TOA-repmat(mm_IRS',1,size(RAD(i).TOA,2));
Xrm = V(:,1:100)*RADm;

YTts_h_LR = testRLR(model_LR,Xrm);
%%Des-normalizamos
YTts_h_LR = YTts_h_LR' + repmat(mYT,1,size(YT,2));
YTts_h_LR = YTts_h_LR'-273.15;
YTm = YTm + repmat(mYT,1,size(YT,2));
YTm = YTm -273.15;

%%MAPAS
i
% for h_lev = 1:10:137 %Escoger altura
%figure,
h_lev = 137;
% %ERROR
Err_LR = abs(YTm(h_lev,:) - YTts_h_LR(:,h_lev)');
% Err_KRR = abs(YTm(h_lev,:)-YTts_h_KRR(:,h_lev)');
LAT = ASV(i).LAT/2;
LON = ASV(i).LON/2;
coast=load('coast');
% real
%figure
% axesm robinson
% subplot(3,1,1)
figure(1)
LL = size(LAT,2)/120;

geoshow(reshape(LAT,120,LL), reshape(LON,120,LL), reshape(YTm(h_lev,:),120,LL),
'DisplayType', 'texturemap')
colorbar, caxis([-80,40])
hold on

geoshow(coast.lat,coast.long,'color','black','DisplayType','line','linewidth',
1)

```



```
    % predicha LR
    figure(2),
    %subplot(3,1,2)
    LL = size(LAT,2)/120;

geoshow(reshape(LAT,120,LL), reshape(LON,120,LL), reshape(YTts_h_LR(:,h_lev),120,LL), 'DisplayType', 'texturemap')
        colorbar, %caxis([-80,40])
        hold on

geoshow(coast.lat,coast.long,'color','black','DisplayType','line','linewidth',1)

    % error
    figure(3)
    %subplot(3,1,3)
    LL = size(LAT,2)/120;

geoshow(reshape(LAT,120,LL), reshape(LON,120,LL), reshape(Err_LR,120,LL), 'DisplayType', 'texturemap')
        colorbar, caxis([-80,40])
        hold on

geoshow(coast.lat,coast.long,'color','black','DisplayType','line','linewidth',1)

    %end
end
```