

# **TRABAJO FIN DE MÁSTER**

## **MODELIZACIÓN DE TEMPERATURAS MENSUALES ENTRE 2005 Y 2006 EN EL SECTOR CENTRAL DE ARAGÓN COMBINANDO DATOS DE SATÉLITE**

***Autor: Darío Sabater Segurana***

***Director: Miguel Ángel Saz Sánchez***

**Máster Universitario en**

**Tecnologías de la información geográfica para la ordenación del territorio: sistemas de información geográfica y teledetección**

**Noviembre de 2016**



**Universidad  
Zaragoza**

**Departamento de Geografía  
y Ordenación del Territorio**





## Resumen

*La importancia de la temperatura en la superficie terrestre para el desarrollo de multitud de actividades socioeconómicas para el ser humano es notoria. Pero la información climática de la que disponemos posee un carácter puntual y no es continua sobre el territorio. Existen muchos trabajos que se han dedicado a resolver este problema y a generar información climática continua mediante diferentes métodos de interpolación a partir de variables geográficas como la altitud, distancia a los mares o la latitud. Más recientemente, los trabajos están orientados a incluir en los modelos predictivos, variables procedentes de satélite con el fin de mejorar los resultados. En este trabajo se ensaya una metodología para generar modelos de temperaturas mínimas y máximas mensuales entre 2005 y 2006 sobre el sector central de Aragón (España), aportando, además de variables geográficas, información de LST y NDVI procedentes del sensor MODIS. Los modelos han sido creados utilizando el método de interpolación de regresión múltiple por pasos y su posterior validación cruzada. Los resultados presentan valores de  $R^2$  muy diferentes entre los distintos meses y años y entre las mínimas y máximas. El RMSE se sitúa en valores por debajo de 1,5 grados centígrados, si bien no se observan diferencias significativas entre la utilización únicamente de variables geográficas y el uso de datos de satélite.*

**Palabras Clave:** MODIS, LST, interpolación, validación cruzada, RMSE, temperaturas

## Abstract

*The importance of the temperature in the land surface for the development of multitude of socio-economic activities for the human being is notorious. But the climate information we have is timely and not continuous on the territory. There are many works that have been dedicated to solve this problem and to generate continuous climatic information through different methods of interpolation from geographic variables such as altitude, distance to the seas or latitude. More recently, the works are oriented to include in the predictive models, variables from satellite in order to improve the results. In this work, a methodology to generate models of minimum and maximum monthly temperatures between 2005 and 2006 on the central sector of Aragon (Spain) is presented, contributing, in addition to geographic variables, LST and NDVI information from the MODIS sensor. The models have been created using the stepwise multiple regression interpolation method and its subsequent cross validation. The results show very different  $R^2$  values between the different months and years and between the minimum and maximum. The RMSE is in values below 1.5 degrees centigrade, although there are no significant differences between the use of only geographic variables and the use of satellite data.*

**Key Words:** MODIS, LST, interpolation, cross-validation, RMSE, temperatures

---



## Índice

1. Introducción .....	1
2. Objetivos.....	2
3. Área de Estudio .....	3
3.1 Caracterización climática .....	4
3.1.1 Sector central .....	4
3.1.2 Somontanos .....	4
3.2 Caracterización biogeográfica .....	5
4. Fuentes de información y metodología .....	6
4.1. Datos de Satélite .....	6
4.2. Datos meteorológicos y geográficos .....	7
4.3 Métodos de análisis.....	10
4.3.2 Validación del modelo .....	12
5. Resultados y Discusion .....	12
6. Conclusiones.....	19
7. Bibliografía .....	20
8. Anexos.....	22

---



## 1-INTRODUCCIÓN

La importancia de la temperatura en la superficie terrestre para el desarrollo de multitud de actividades socioeconómicas para el ser humano es notoria. Resulta muy útil para una amplia gama de aplicaciones que incluyen el estudio de enfermedades biológicas (Mount and Haile, 1989, Focks et al., 1995, Bertrand and Wilson, 1996, Thomson et al., 1996 and Goetz et al., 2000), usos veterinarios (Flasse et al., 1995), predicciones meteorológicas (Smith et al., 1988), desarrollo de epidemias (Connor et al., 1998) o cambio climático (IPCC, 2001).

La temperatura del aire es un parámetro importante de un lugar, utilizado para determinar la idoneidad de la localización de cultivos agrícolas y forestales (Hudson y Wackernagel, 1994), y se utiliza para parametrizar el hábitat de las especies vegetales (Sánchez-Palomares et al., 2003) y en la determinación de patrones de zonificación vegetal (Richardson et al., 2004). Además, es un factor relacionado con la productividad de la planta, ya que está ligada a la duración del período vegetativo y la evapotranspiración.

Antes del desarrollo y generalización de los Sistemas de Información Geográfica (SIG) ya existía información de carácter cartográfico sobre la distribución espacial de muchas variables ambientales, en especial de variables climáticas, de un país o región. Para hacer esos mapas, la información puntual se espacializaba mediante técnicas de interpolación manuales basadas en el cálculo y representación de isolíneas, obtenidas en función de los valores puntuales medidos y una corrección posterior de estas cartografías mediante el que podríamos denominar como “criterio experto”. Este criterio, basado en el conocimiento geográfico del territorio y de los factores que pueden modular el reparto espacial de los valores climáticos, introducía modificaciones en los mapas realizados a partir del cálculo de las isolíneas por la que trataba de aproximar más la cartografía a la realidad climática del territorio.

Esas modificaciones se realizaban teniendo en cuenta hechos tales como el relieve, la orientación o el conocimiento de la mayor o menor continentalidad de los distintos espacios regionales. En cualquier caso se trataba de una aproximación cualitativa y subjetiva a la representación de un atributo climático sobre un territorio dado. Técnica que sin embargo ha producido ejercicios cartográficos interesantes en cuanto a su calidad, pero que se enfrentaban con los habituales problemas de las cartografías que se han desarrollado fuera de un entorno SIG de trabajo: la posible existencia de errores vinculados a esa apreciación subjetiva que se deriva del criterio experto y la imposibilidad de realizar actualizaciones de esos mapas que no pasen por una nueva elaboración en la que se ha de desarrollar desde el principio todo el proceso metodológico de elaboración del mapa.

Pero en la actualidad, el problema más importante al que se enfrentan esas cartografías es el de la imposibilidad de integrarlas junto a otras variables en un análisis cuantitativo, en el que el clima pueda convertirse en un factor explicativo más de distintos fenómenos naturales o antrópicos. Podían servir para conocer de forma estática la distribución espacial de las variables climáticas sobre un territorio a partir de un proceso inicial de interpolación de registros puntuales y corrección a partir del criterio experto, pero las dificultades de actualización y la imposibilidad de poder operar sobre ellas por medio de un SIG limitaban su uso y funcionalidad en la gestión territorial y la investigación actual.

Con la incorporación de los SIG, la posibilidad de aplicar las potencialidades de los éstos en la elaboración de cartografías de atributos relacionados con las variables ambientales, ha supuesto una auténtica revolución.

La potencia de cálculo de estas herramientas, auxiliada en ocasiones por software estadístico específico, nos permite obtener información georreferenciada sobre distintas variables ambientales en forma de superficies continuas.

Esto nos posibilita ir más allá de lo que sería un mero ejercicio cartográfico destinado a la representación de la distribución espacial de una variable. En el caso de la información climática, por ejemplo, una cartografía de alta resolución sobre las temperaturas o las precipitaciones de una región determinada puede ser un producto ya de por sí interesante. Pero lo son además porque, a partir de esa información, podemos calcular en un SIG productos derivados, como índices de sequía, de aridez,

valores de ETP, cuantificación espacial de tendencias, variabilidad, anomalías, seleccionar áreas de interés para la extracción de datos, etc.

Además, el desarrollo de los SIG permite disponer además de información georreferenciada a cualquier escala de resolución espacial, escala que dependerá tanto de la información en origen como de las necesidades del usuario final de esa cartografía, hace que podamos integrarla junto a otras variables del medio físico o humano para el desarrollo de análisis cuantitativos.

Así, la cartografía pasa de ser un documento meramente informativo sobre uno de los caracteres que contribuyen a definir el medio físico de un territorio y algunas de sus potencialidades, a ser información susceptible de ser integrada en otras herramientas de gestión y evaluación del territorio.

El principal problema de la información climática reside en el carácter puntual de las temperaturas de las que se dispone, a través de los observatorios cuya localización no es homogénea sobre el territorio. Por regla general, existe una mayor densidad de esta información puntual en zonas llanas y dedicadas a la agricultura, en contraste con las áreas de montaña, con escasa o nula información.

Para resolver el problema de la información climática puntual se han utilizado diferentes métodos de interpolación con el fin de generar información climática de forma continua. En consecuencia, se han desarrollado muchos métodos distintos para crear mapas regionales a partir de datos puntuales, basados en la continuidad de la temperatura y su fuerte dependencia de la elevación; En promedio global, la temperatura disminuye alrededor de 0.65 8C / 100 m de altitud (Barry y Chorley, 1987, Gandullo, 1994, Lutgens y Tarbuck, 1995), aunque esta tasa puede variar con la estación y la situación geográfica (Goodale et al. , 1998), y en relación con los efectos diurnos (Richardson et al., 2004).

Existen numerosos trabajos en este ámbito sobre el territorio español (Ninyerola et al. 2000, Vicente et al, 2003 y Saz-Sánchez et al., 2010). También, es necesario citar el “Atlas climático Ibérico. Temperatura del aire y precipitación (1971-2000)”, elaborado por la Agencia Estatal de Meteorología (AEMET, 2011) para la temperatura media, máxima y mínima mensual con resolución de 250 m, en el cual, se utilizó un método de interpolación con regresión multivariada (utilizando como variables la distancia al mar, la latitud y la altitud) y con la corrección de los residuales en unos casos con un método local y en otros mediante técnicas geoestadísticas.

Existen distintos trabajos que han abordado la creación de modelos climáticos incorporando, además de variables geográficas que pueden explicar la distribución de las temperaturas sobre un territorio, información de temperatura a través de satélite. Podemos encontrar trabajos en España (Cristóbal et al., 2006, Cristóbal et al., 2008). Estos trabajos incorporan datos de satélite provenientes de distintas plataformas satelitales como LANDSAT, NOAA AVHRR o MODIS y en los que la descarga, procesado y transformación de los datos es diferente en cada uno de ellos.

En el caso de MODIS, los datos de Temperatura de la Superficie Terrestre (LST) se pueden descargar desde su servidor y su procesamiento conlleva una menor inversión de tiempo. Numerosos trabajos centrados en la estimación de la temperatura del aire incorporan como datos de satélite, únicamente, productos MODIS (Neteler, 2010, Fu et al., 2011, Benali et al., 2012, Hengl et al., 2012). Aunque normalmente la Temperatura del Aire (Tair) se mide desde las estaciones meteorológicas, la disponibilidad de los datos están limitadas por la distribución y la densidad de la red de estaciones de las que disponen (Vogt et al., 1997).

En este trabajo, se ha procedido a utilizar variables geográficas como son la altitud, latitud, longitud, distancia al mar Mediterráneo, distancia al mar Cantábrico, orientación, radiación y rugosidad, para incluirlas en los modelos de regresión, incorporando, por otro lado, + información de satélite como es la Temperatura de Superficie Terrestre (LST) y el NDVI procedentes de productos MODIS, con el fin de observar si existen diferencias en los resultados de los modelos.

## 2. OBJETIVOS

El objetivo principal de este trabajo es analizar el efecto que sobre la modelización espacial de las temperaturas de superficie registradas en observatorios meteorológicos tradicionales tiene el uso de

información obtenida a partir de sensores remotos de forma complementaria a las variables geográficas y estructurales habituales.

Objetivos parciales:

- Comparar los resultados de la modelización de las temperaturas de superficie a partir de datos instrumentales y de variables geográficas y con modelos combinados que incluyen variables obtenidas a través de la teledetección.

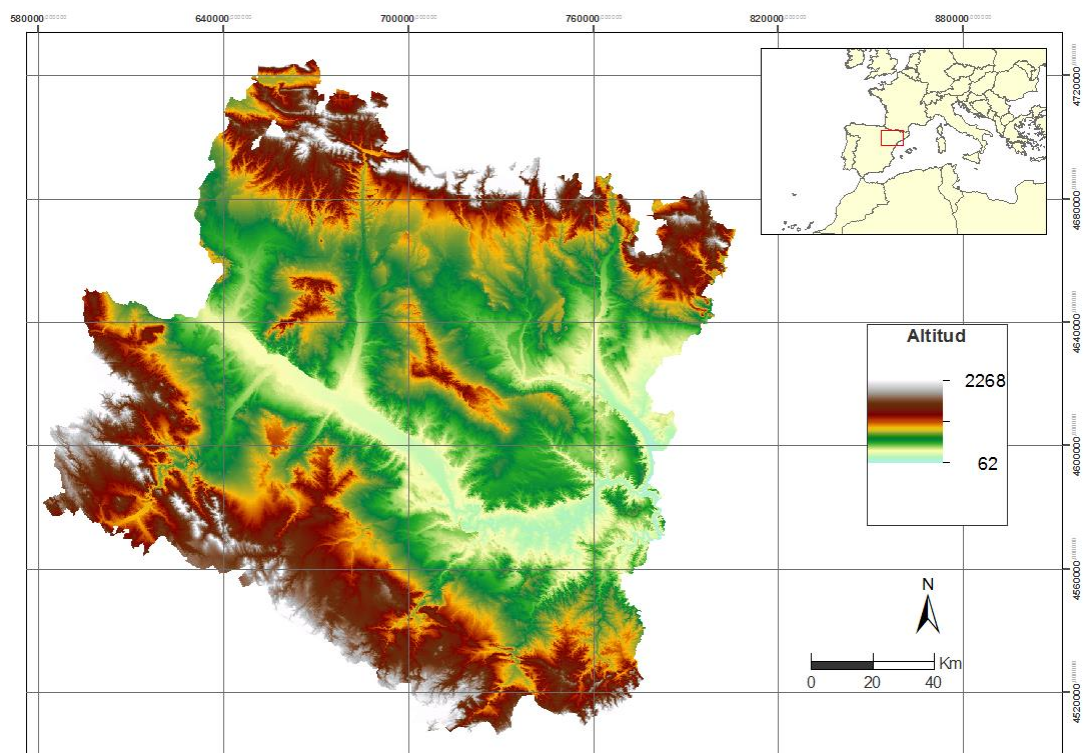
- Analizar el comportamiento de las dos metodologías en las temperaturas máximas y mínimas y observar si existen patrones que expliquen dicho comportamiento.

- Comparar los resultados entre las distintas estaciones del año para identificar posibles errores procedentes de situaciones sinópticas (inversiones térmicas etc.).

### 3. ÁREA DE ESTUDIO

El área de estudio se localiza entre el somontano Oscense y el somontano Ibérico que incluye entre esas dos formaciones, el Sector Central de la Depresión del Ebro, dentro del límite autonómico de Aragón. Tiene una extensión de 29.049 km<sup>2</sup>, pertenecientes a las provincias de Huesca, Zaragoza y Teruel, entre los paralelos 40° 40' 19,953" N y 42° 39' 45,988" N y meridianos 0° 42' 8,703" E y 2° 1' 41,512" W.

Esta área de estudio tiene dos zonas bien diferenciadas como son los somontanos y la depresión del Ebro, pero que a su vez son homogéneas dentro de cada zona desde el punto de vista geográfico. Esto resulta oportuno a la hora de realizar operaciones que incluyen imágenes de satélite, que poseen ciertos problemas a la hora de discriminar cambios bruscos en el paisaje por la resolución espacial de los datos. Su variación altitudinal parte desde los 62 metros en sus zonas más bajas hasta los 2.268 metros en sus cotas más altas.



**Figura1.** Área de Estudio

### 3.1 Caracterización climática

#### 3.1.1 Sector central

El clima del centro de la Depresión del Ebro responde perfectamente al de una cuenca mediterránea con marcado carácter de continentalidad. La presencia de un sistema montañoso continuo que le cierra por el Norte y por el Sur contribuye a exagerar los contrastes térmicos entre el verano y el invierno, y a obstaculizar la entrada de las borrascas portadoras de lluvia, lo que motiva su tendencia a la aridez, y refuerza por su disposición topográfica la acción del viento dominante, el cierzo, muy frío en invierno, fresco en verano y, por su subsidencia, siempre desecante.

La disposición del relieve subraya esa condición, al impedir que hasta esa zona llegue la influencia suavizadora que sobre las temperaturas ejercen las masas oceánicas. Además provoca que en situaciones de estabilidad atmosférica, el aire quede estancado sobre el territorio, favoreciendo su caldeo en verano y su enfriamiento en invierno, incidiendo sobre el contraste térmico que se da entre ambas estaciones.

Esa aridez es, sin duda, el elemento que caracteriza y unifica este espacio central aragonés. Las lluvias son escasas en cualquiera de sus comarcas, en especial en la zona central, pero son sobre todo irregulares. Nos aparecen dos máximos, en primavera y otoño, ganando importancia el segundo conforme avanzamos hacia levante al cobrar mayor importancia la influencia mediterránea. Ambos quedan separados por dos mínimos, acusados, en invierno y verano, estaciones que derivado de la presencia casi constante de situaciones anticiclónicas ven reducidos de forma notable sus volúmenes de precipitación.

El carácter continental del clima se refleja en la fuerte variación de las temperaturas a lo largo del año. En verano, la disposición en cubeta condicionada por el relieve favorece, el progresivo calentamiento de las masas de aire y el aumento de las temperaturas, con medias que en julio y agosto llegan a superar los 24 °C e incluso los 25 °C en los sectores más cálidos del Bajo Aragón y máximas absolutas que superan fácilmente los 35 °C. En invierno aún bajo la misma configuración sinóptica, la situación térmica se invierte, condicionando las altas presiones la presencia de frío intenso, con valores medios en enero inferiores a 5 °C. Son frecuentes las heladas e inversiones térmicas relacionadas con el estancamiento de aire frío invernal, situaciones que provocan además la aparición de nieblas de irradiación que sumergen al valle en un desagradable e incómodo ambiente.

Se superan, no sin cierta dificultad, los 300 mm de media al año, un valor que como ya se ha señalado está sometido a una elevada irregularidad interanual. Su reparto es una clara muestra de la influencia mediterránea, señalando un máximo primaveral y otro otoñal que gana en importancia conforme nos desplazamos hacia levante.

Las temperaturas son extremas tanto en verano como en invierno. Si el rango térmico promedio roza los 20 °C, diferencia que aparece entre la media del mes más cálido y la del más frío, la amplitud térmica absoluta, es decir, la diferencia entre los días más cálidos del verano y los más fríos del invierno, puede superar cualquier año los 40 °C, un valor que denota el elevado grado de continentalidad térmica que caracteriza este clima seco estepario del eje del Ebro (Cuadrat, 1999).

#### 3.1.2 Somontanos

El clima de transición de los somontanos es de tipo Mediterráneo continental del mismo orden que el centro de la Depresión. Sin embargo, los somontanos pirenaico e ibérico están condicionados por la altitud, que les proporciona mayor humedad, les aleja gradualmente del agobiante calor estival, de las inversiones térmicas y de las fuertes ráfagas de cierzo y, por tanto, les hace menos áridos, pero mantienen la irregularidad pluviométrica y elevadas amplitudes térmicas.

El carácter marcadamente seco del clima se suaviza conforme nos desplazamos hacia el norte y el sur del sector más central y deprimido de la depresión del Ebro. La progresiva mayor altitud media que se deriva del ascenso hacia los somontanos pirenaico e ibérico, condiciona un moderado incremento de los volúmenes pluviométricos, que obliga a delimitar un tipo de clima distinto al del eje cen-

tral y que se clasifica, en función de sus características termopluviométricas, como de Mediterráneo continental. Es también un clima de fuertes contrastes térmicos entre el invierno y el verano y unas precipitaciones que, aunque aumentan respecto a las zonas más secas de la depresión del Ebro, siguen ofreciendo unos claros máximos equinocciales y una elevada irregularidad interanual.

Hacia el norte la transición climática es gradual. En ese camino hacia y por el somontano pirenaico, los totales anuales de precipitación pronto sobrepasan el umbral de los 400 mm, rebasando los 500 y aún los 600 mm anuales al pie ya de las Sierras Exteriores. Cabe reseñar, sin embargo, que la pluviometría disminuye desde poniente, zona más favorecida por las perturbaciones atlánticas, hacia levante, donde aquellas llegan con más dificultad, ganando progresivamente en importancia a lo largo de este gradiente oeste-este que nos lleva hasta espacios cada vez más mediterráneos, el máximo pluviométrico otoñal característico de este tipo de clima.

Las temperaturas también se apartan, poco a poco, del carácter extremo que es propio del eje del Ebro, rondando los promedios anuales los 13 °C en el límite con las Sierras del Prepirineo. Pero con todo, el fuerte calor estival sigue siendo un carácter esencial del clima de esta zona, con valores medios de entre 24 y 25 °C en los meses más cálidos y máximas absolutas por encima de los 40 °C. En invierno, es lógico pensar que el progresivo incremento en la altitud media del terreno hace descender los valores térmicos. Sin dejar de ser cierto este hecho, la orientación sur general de toda esta vertiente y el descenso en la frecuencia de fenómenos de inversión térmica en algunas zonas, sobre todo en las comarcas más occidentales, y con ello en el número de días de niebla, matiza en parte ese esperado gradiente térmico negativo.

El desarrollo topográfico hacia el somontano ibérico es menos gradual y así la expresión superficial de este tipo de clima Mediterráneo continental es menor. Aparece compartimentado en dos zonas. La más oriental, que ocupa buena parte de las comarcas del Bajo Aragón Turolense y del Matarraña, ofrece los caracteres más cercanos al del clima mediterráneo puro. La distancia a la costa levantina es escasa y su continentalidad térmica está más relacionada con la barrera topográfica que supone la presencia de los Puertos de Beceite que con una posición interior acusada. Las precipitaciones, en torno a los 400-450 mm, muestran un claro máximo en los meses de otoño, con un verano de acusado estiaje sólo interrumpido por fenómenos de convección local que aportan tormentas importantes. Las temperaturas son las más altas de toda la provincia de Teruel, en torno a los 14 °C, resultando de unos inviernos más benignos, con promedios superiores a los 4 °C, pero sobre todo de unos veranos cálidos, en los que como en cualquier punto de esta franja central de Aragón se superan con facilidad los 35 °C.

Por otro lado, el Campo de Cariñena, la Comarca del Aranda, el Campo de Borja y una parte de la Comunidad de Calatayud, comparten esa clasificación para sus condiciones climáticas de Mediterráneo continentales, si bien su posición más occidental introduce matices significativos en el reparto anual de las precipitaciones. Éstas siguen siendo exiguas, de entre 400 y 450 mm, inferiores incluso a las que recibe buena parte de la vertiente septentrional en función de la presencia de un relieve mucho más compartimentado, mientras que el máximo deja sus preferencias por los meses otoñales para localizarse de forma clara en primavera. Los volúmenes anuales siguen estando sometidos a la misma irregularidad interanual, tan característica de esta franja central de Aragón. Veranos calurosos, con medias en julio y agosto por encima de los 23 °C, e inviernos frescos, con heladas frecuentes y nieblas en el fondo de los valles, completarían la imagen del clima de la zona (Cuadrat, 1999).

### **3.2 Caracterización biogeográfica**

La Depresión del Ebro y los Somontanos pirenaico e ibérico, se localizan desde el punto de vista biogeográfico en la región mediterránea. En ella predominan las comunidades de matorral de medio y bajo porte, en especial romerales, muy presentes en toda la región y presentando variaciones respecto a su composición en función del sustrato y la altitud, junto a él se observan enebrales, matorrales gipsófilos con un importante número de endemismos, matorrales nitrófilos y halófilos, ligados a las extensos sistemas endorreicos y las litologías arcillosas y margo-yesíferas. En cuanto a las formaciones boscosas, aparecen masas de quercíneas (coscoja, carrasca, quejigo y rebollo), por lo general con una es-

pecie dominante, así como también formaciones de pinar xérico (*Pinus halepensis* y *P. nigra*) y formaciones arbóreas de sabinar.

Lo más característico del ambiente mediterráneo aragonés es el carrascal, enriquecido en ocasiones con quejigo, apareciendo entorno a los 500 m de altitud. Por debajo, podemos localizar masas de pino carrasco (*Pinus halepensis*) con coscoja (*Quercus coccifera*), propias de los ambientes más xéricos y que se observan en los relieves destacados de la depresión del Ebro (Zuera, Castejón, Alcubierre, Serreta Negra), evitando la cota en la que se sitúan las frecuentes inversiones térmicas (300 m), del centro de la Depresión que provocan fuertes y prolongadas heladas, dejando tan solo a los sabinars albares (*Juniperus thurifera*) como formación arbórea de las cotas más bajas, ya que parecen ser los únicos adaptados a esta situación (extraído de Aladrén, 2004).

## **4. FUENTES DE INFORMACIÓN Y METODOLOGÍA**

### **4.1. Datos de Satélite**

Para este trabajo se han utilizado distintos productos procedentes de los satélites AQUA-MODIS y TERRA-MODIS, para el período 2005 y 2006. Cada imagen consta de varias capas en la que recogen la temperatura diurna y nocturna y, asociada a las capas anteriores, dos capas que informan de la calidad de cada píxel de temperatura.

En un principio se procedió a trabajar con imágenes de LST procedentes del producto MYD11A1 del satélite AQUA, obtenido a través del portal (<https://ladsweb.nascom.nasa.gov>), con una resolución temporal diaria y 1000 metros de tamaño de píxel (Wan et al., 2015). Para nuestro área de estudio, ha sido necesario utilizar dos imágenes que comprenden una selección espacial compuesto por un Tile de 17 y 18 en la horizontal y de 04 en la vertical. La temperatura de superficie terrestre MODIS (Ts) se deriva de dos canales de banda infrarroja térmica, es decir, 31 (10,78-11,28  $\mu\text{m}$ ) y 32 (11,77-12,27  $\mu\text{m}$ ) usando el algoritmo de ventana dividida (Wan y Dozier, 1996). La validación detallada del producto C6 MODIS LST se da en el documento publicado más recientemente (Wan, 2014).

El producto de la versión 6 de MOD11A1 proporciona diariamente por píxel la temperatura superficial de la tierra (LST) en una rejilla de 1200 x 1200 kilómetros. El valor de temperatura de píxel se deriva del producto MOD11\_L2. Por encima de los 30 grados de latitud, algunos píxeles pueden tener múltiples observaciones donde se cumplen los criterios para el cielo despejado. Cuando esto ocurre, el valor de píxel es el resultado de la media de todas las observaciones que califican. Proporcionadas junto con las bandas de temperatura superficial diurna y nocturna y sus capas de indicador de calidad, están las bandas MODIS 31 y 32 y seis capas de observación.

Posteriormente se ha utilizado imágenes de LST procedentes del producto MOD11B3 del satélite TERRA, obtenido a través del portal (<https://ladsweb.nascom.nasa.gov>), con una resolución temporal mensual y un tamaño de píxel de 5600m (Wan et al., 2015).

El producto de la versión 6 de MOD11B3 proporciona una temperatura promedio mensual por píxel de la superficie terrestre (LST) en un mosaico de 1200 X 1200 (km) con un tamaño de píxel de 5600 metros (m). Cada valor de píxel de temperatura de superficie terrestre (LST) y emisividad en el MOD11B3 es un promedio simple de todos los valores correspondientes de los valores LST del MOD11B1 recogidos durante el período del mes. Cada imagen MOD11B3 consta de 19 capas, incluyendo la capa LST día y noche y sus capas asociadas de indicadores de calidad, seis capas de emisividad y siete capas de observación. Únicamente al conjunto de productos MOD11B se añaden capas LST adicionales de 5600 m de día y de noche, que se agregan de la banda de 1km 31 de la franja MOD11\_L2 correspondiente.

El satélite MODIS/TERRA que obtiene las temperaturas mensuales registra la temperatura durante el día (asociada a las temperaturas máximas) entre las 12:00 y las 13:00 (hora local) y durante la noche (asociada a las temperaturas mínimas) entre las 22:00 y las 24:00.

Por otro lado, se ha utilizado imágenes de NDVI procedentes del producto MOD13C2, obtenido a través del portal (<https://landsweb.nascom.nasa.gov>), con una resolución temporal mensual y un tamaño de píxel de 5600m (Didan, 2015).

El índice de vegetación de diferencias normalizadas (NDVI) (Tucker, 1979) es el índice más común utilizado para parametrizar el estado de la vegetación a partir de satélites (Hope et al., 2003, Stow et al., 2004). Utiliza el principio subyacente de que, para la superficie con vegetación, las longitudes de onda roja (RED) e infrarrojo cercano (NIR) se caracterizan por una absorción alta y baja, respectivamente (Chen et al., 2003) y citado en Zhu (2013).

El NDVI se define como:  $NDVI = (NIR - RED) / (NIR + RED)$

El producto MOD13C2 versión 6 proporciona un valor de índice de vegetación (VI) en una base por píxel. Hay dos capas de vegetación primaria. El primero es el Índice de Vegetación por Diferencia Normalizada (NDVI), el cual se conoce como el índice de continuidad del NDVI derivado de la Administración Nacional Oceánica y Atmosférica (NOAA-AVHRR). La segunda capa de vegetación es el Enhanced Vegetation Index (EVI), que ha mejorado la sensibilidad sobre las regiones de biomasa elevada.

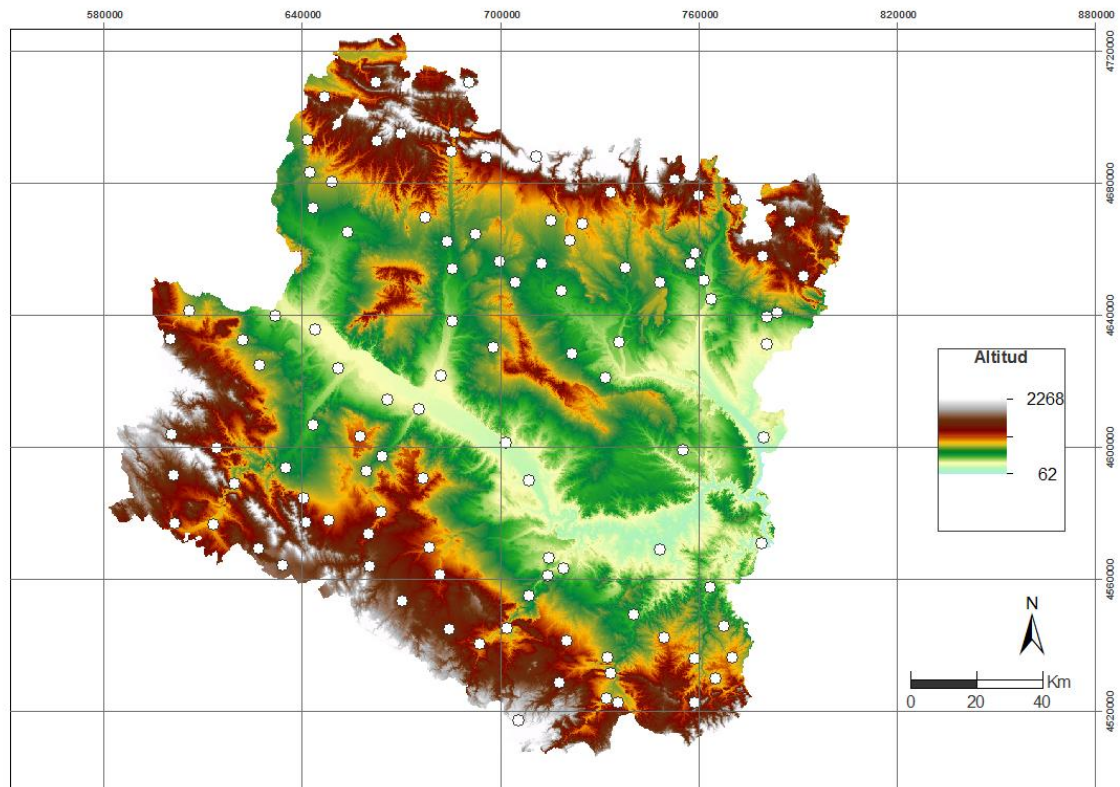
La Red de Modelado Climático (CMG) consta de 3.600 filas y 7.200 columnas de 5,600 metros (m) de píxeles. El producto MOD13A2 (1 kilómetro) se utiliza para producir este producto VI mensual. Nominalmente, cada CMG consta de 286 tiles que se reproyectan en la proyección geográfica, promediado espacialmente y control de calidad (QA). El MOD13C2 tiene campos de datos para NDVI, EVI, VI QA, datos de reflectancia, información angular y estadísticas espaciales tales como media, desviación estándar y número de píxeles de entrada usados en la resolución CMG de 0,05 grados.

#### **4.2. Datos meteorológicos y geográficos**

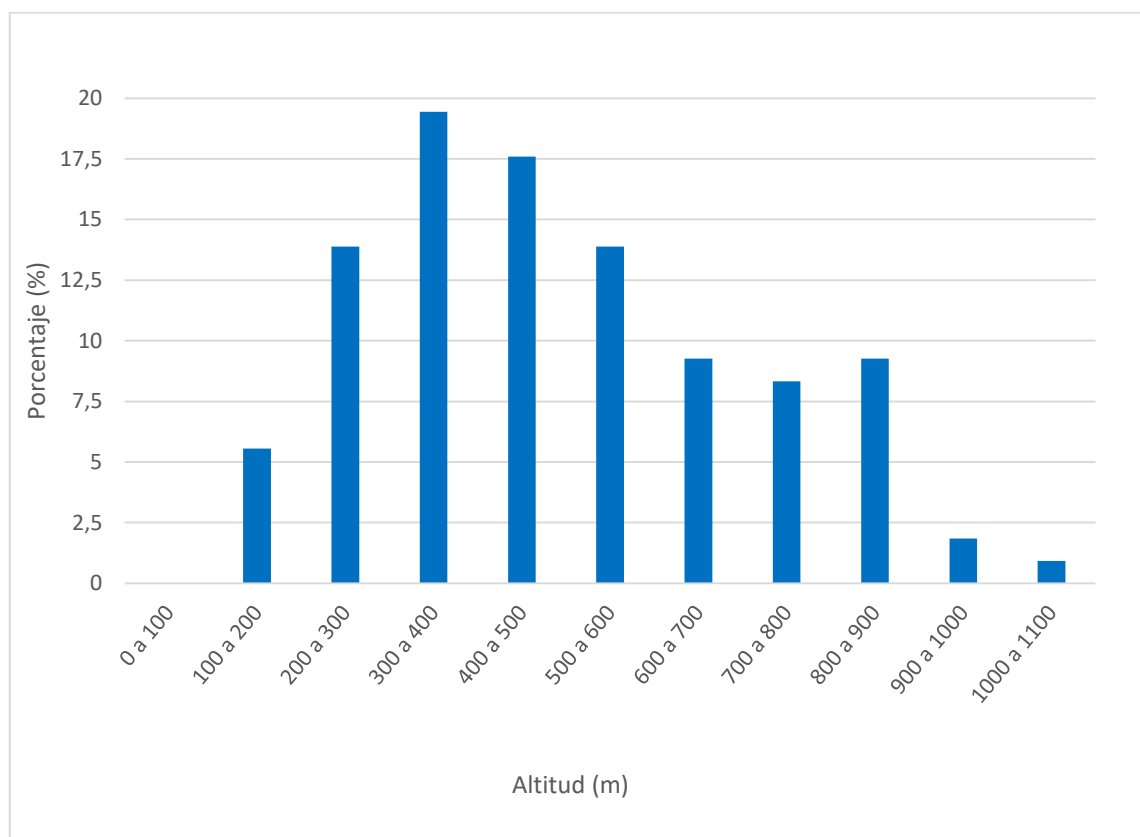
Las variables geográficas que se han incluido son las utilizadas anteriormente en otros estudios (Ninyerola et al., 2000, Cristóbal et al., 2006) como la altitud, la latitud, la continentalidad y la radiación solar. En estos estudios, esta metodología ha sido útil para obtener la temperatura mensual.

Como datos geográficos, se han generado una serie de variables que pueden explicar el comportamiento que tienen las temperaturas espacialmente. Para los modelos de regresión se utilizó un modelo digital de elevaciones (MDE) con una resolución espacial de 100 metros distribuido por el Instituto Geográfico Nacional (IGN) y capas digitales de la misma resolución de la orientación, rugosidad del terreno (según el método descrito por Allmaras et al., 1966, Currence y Lovely, 1977), de distancia al Mar Cantábrico y Mar Mediterráneo, latitud (Y) y longitud (X), generadas en ArcGIS® 10.3 a partir del MDE. La radiación potencial se obtuvo en ArcGIS 10.3

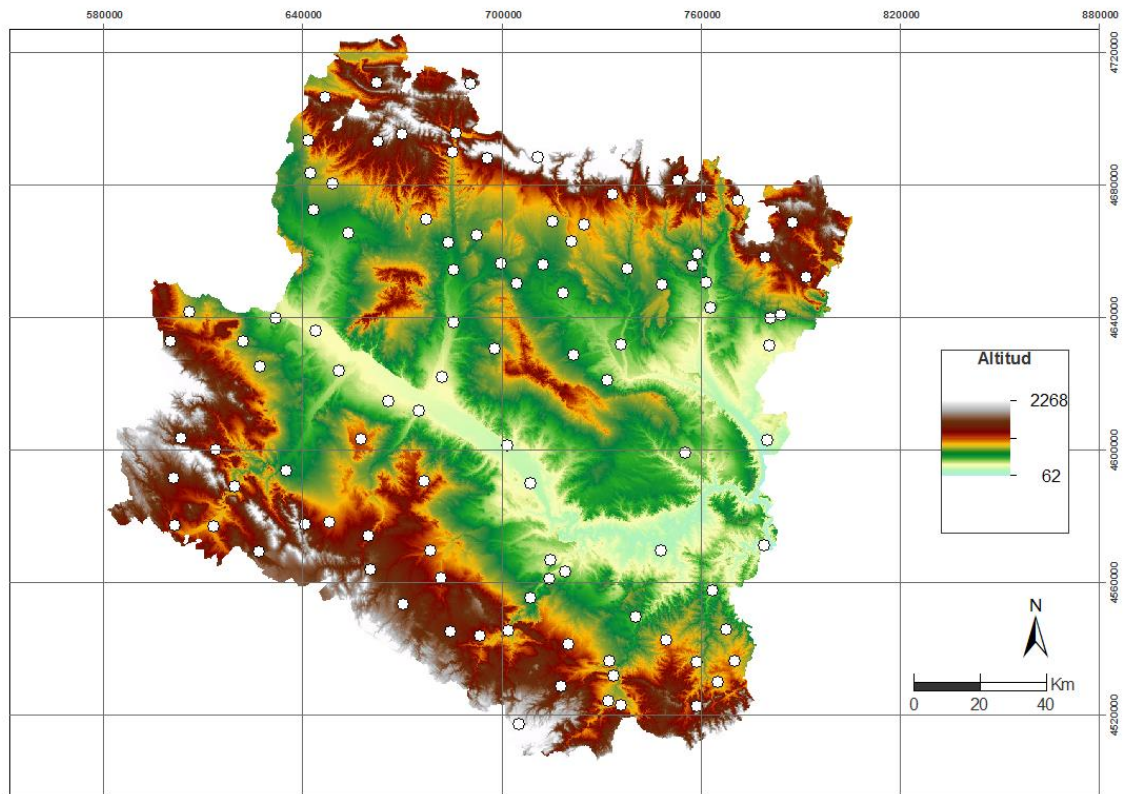
Los datos de temperaturas máximas y mínimas absolutas mensuales entre 2005 y 2006 pertenecen a 391 observatorios instrumentales obtenidas de la Agencia Estatal de Meteorología (AEMET). De todos ellos, se realizó una depuración de los datos, eliminando aquellos observatorios en los que no hubiese registro durante el período descrito anteriormente. La base de datos finalmente se redujo en 108 estaciones en el caso de las temperaturas máximas y de 101 estaciones para las mínimas. El 70,2% de los observatorios se localizan en cotas por debajo de los 600m, el 28,8% entre los 600 y los 1000 metros y solo un 1% por encima de los 1000 metros.



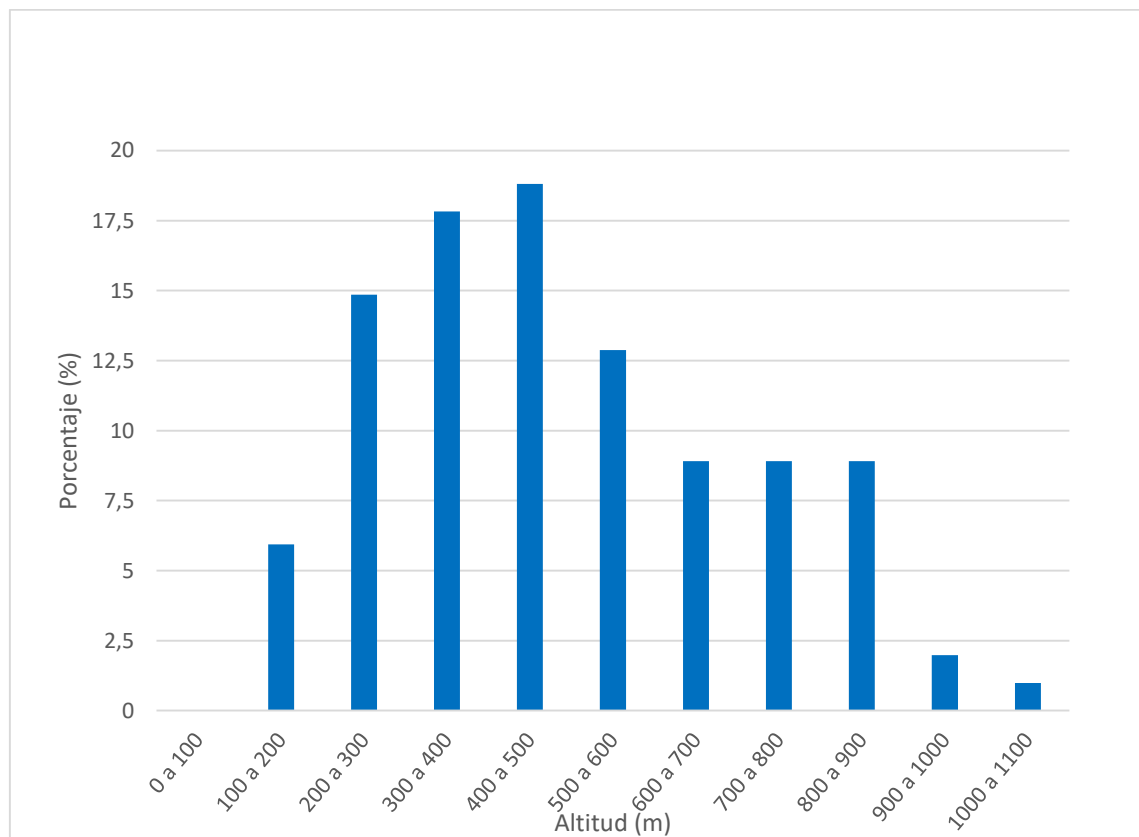
**Figura 2:** Distribución de los observatorios con datos sobre las temperaturas medias de las máximas



**Figura 3:** Distribución altitudinal de los observatorios con datos sobre las temperaturas medias de las máximas.



**Figura 4:** Distribución de los observatorios con datos sobre las temperaturas medias de las mínimas.



**Figura 5.** Distribución altitudinal de los observatorios con datos sobre las temperaturas medias de las mínimas.

### 4.3 Métodos de análisis

En base a otros trabajos anteriores, LST y NDVI han sido elegidos frecuentemente como predictores en la literatura de modelado de la temperatura del aire debido a su relación directa con esta variable (Goward et al., 1994, Prihodko y Goward, 1997, Vogt et al., 1997, Prince et al., 1998, Recondo y Pérez-Morandeira, 2002, Cristóbal et al., 2006, Riddering y Queen, 2006).

Para ello, se concretó que los productos MODIS que otros trabajos han utilizado podían ser una buena herramienta porque no necesitaban un excesivo post-procesamiento. Además, MODIS tiene una resolución temporal diaria, muy acorde con los datos de los observatorios instrumentales.

La fase de descarga de la información ha necesitado de la inversión de un tiempo notable ya que el volumen de datos descargado es elevado y muy escasas las posibilidades de automatización de los procesos. Una vez obtenida la información se han desarrollado todos los cálculos y análisis en R-Studio®.

Para buscar y seleccionar todos los productos de MODIS, se ha utilizado el siguiente enlace de la NASA (<https://ladsweb.nascom.nasa.gov/search/>). Posteriormente, los productos seleccionados se almacenaban en el servidor ftp de la NASA (<https://ladsweb.nascom.nasa.gov/archive/>). Debido al gran volumen y la cantidad de archivos que tenía que descargar, he manejado el programa Filezilla®, un cliente FTP multiplataforma con el fin de optimizar la descarga de los mismos.

Los archivos procedentes del servidor se encontraban en formato hdf, de manera que cada imagen tuvo que ser convertida a formato tif para su análisis, realizar el mosaico (ya que cada fecha se compone de dos imágenes porque el área de estudio se encuentra entre los tiles 17 y 18 h) y reproyectarlas al sistema de coordenadas “ETRS 1989 UTM Zona 30 N”. Por otro lado, estas imágenes convertidas y reproyectadas se recortaron en base a la extensión del área de estudio delimitada. Se ha implementado un código en R-Studio® para optimizar esta tarea (Anexo 1).

Una vez hecho, cada imagen del producto MODIS MYD11A1, que incorpora datos de LST diaria y 1km de tamaño de píxel ha tenido que ser analizada mediante una capa asociada a ella que contenía la información de la calidad de cada píxel, de manera que solo han sido utilizadas aquellas imágenes que tienen un porcentaje de píxeles aptos superior al 50%, al 75% y al 90% de la imagen. Los píxeles marcados con LST Error > 3 K o cuya calidad de descrita en cada píxel no fuera “Good Quality”, fueron eliminados con gran impacto en la disponibilidad de los datos. El código implementado en R-Studio® se puede consultar en el Anexo 2.

De las imágenes que tienen un porcentaje de píxeles aptos superior al 50 % han sido analizadas y se ha comprobado que existían muchas áreas sin datos, de manera que con esta información de satélite no era posible incorporarla al modelo de regresión como una variable explicativa independiente.

El siguiente paso fue intentar “rellenar” esos píxeles sin datos de las imágenes de satélite realizando modelos de interpolación a partir de las variables geográficas y observar si los resultados eran lo suficientemente satisfactorios para poder realizar este “relleno”. Los resultados no alcanzaban unos criterios mínimos exigidos (con un R2 superior a 0,5), con lo cual, la información de satélite procedente del producto MODIS MYD11A1 de LST diaria tuvo que ser desechada.

Existen otros productos MODIS de LST que tienen información temporal cada 8 días y cada 16 días que han sido descargados y analizados pero que tenían, aunque en menor medida, el mismo problema que los datos diarios de LST. Estos productos presentaban amplias zonas sin datos, lo que hacía muy complicado realizar el “relleno” de estos píxeles.

Después se utilizó el producto LST de MODIS MOD11B3 que tiene una resolución temporal mensual y un tamaño de píxel de 5.600 metros. Es cierto que se ha perdido resolución temporal y espacial con respecto a los productos anteriores, pero en este caso la información de los píxeles tenía una calidad apta y no se generaban esos “vacíos” espaciales detectados con los productos citados anteriormente. En Cristóbal et al. (2008) realizan una modelización de las temperaturas mínimas y máximas mensuales utilizando en un apartado de su trabajo imágenes MODIS con buenos resultados.

Además de la LST mensual, se ha descargado los datos de NDVI mensuales para incorporarlos como variable explicativa independiente.

Por otro lado, para los datos de temperaturas de las estaciones ha sido necesaria un trabajo de filtrado de los datos para observar y eliminar aquellas estaciones sin datos o con datos erróneos. Después, se han convertido los datos diarios a datos mensuales para poder compararlos con los datos de satélite.

Con los datos de los observatorios convertidos a datos espaciales y todas las variables geográficas junto con las de satélite, se han integrado en ArcGIS® 10.3 y exportadas en formato dbf para su posterior análisis de regresión por pasos en R-Studio®.

Una vez procesados y preparados los datos, se ha procedido a realizar el análisis utilizando el método de interpolación de regresión lineal. Burrough y Mcdonell (1998) clasifican los métodos de interpolación en locales, geoestadísticos, globales y mixtos, existiendo en la literatura científica numerosos ejemplos de aplicación de estos métodos a variables climáticas.

Los métodos locales se basan en la utilización de los puntos más cercanos al punto a interpolar para estimar el valor de la variable  $V$  en éste. Asumen la existencia de una importante autocorrelación espacial de los datos. O lo que es lo mismo, que detrás de los interpoladores locales se esconde la idea de que los puntos más próximos presentan valores más parecidos entre sí que con los que se encuentran más lejos, por lo que tratarán de estimar el valor de un punto a partir de los puntos más próximos con valor conocido. Es además un interpolador exacto.

Los interpoladores globales utilizan toda la muestra para estimar el valor en cada nuevo punto y asumen la dependencia de la variable a interpolar de otras variables de apoyo. Se trata de un método inexacto, en el sentido de que el valor estimado en un punto en el que sí existe registro no tiene por qué coincidir con el valor allí medido.

Los métodos mixtos combinan interpoladores globales y locales, es decir, plantean la ventaja de considerar las tendencias generales y relaciones físicas que se producen entre la variable climática y los factores geográficos del medio y tener en cuenta la correlación y dependencia espacial de la información climática disponible. Los métodos globales son interpoladores inexactos, de modo que el valor estimado no se corresponde con el valor real de la variable en el punto de observación. Ninyerola et al. (2000) desarrollan un corrector relacionado con la magnitud de las diferencias en cada punto de muestreo:  $\text{Corrector} = \text{observado} - \text{predicho}$ .

Los valores de corrección obtenidos se utilizan para refinar los resultados obtenidos mediante el modelo de regresión.

En los métodos geoestadísticos, las estimaciones realizadas mediante los diferentes métodos de kriging se basan en una suma ponderada de las  $n$  observaciones más cercanas al punto de estimación. El peso de cada uno de los puntos será escogido para que la estimación no se encuentre sesgada y tenga la menor varianza posible en relación a otras posibles combinaciones lineales entre el conjunto de observaciones disponibles. Existen diferentes técnicas geoestadísticas que incluyen una o más variables en la interpolación (simple kriging, cokriging...) con una complejidad creciente cuanto mayor es el número de variables introducidas.

En Vicente Serrano y Saz-Sánchez (2002), concluye que los métodos globales (modelos de regresión) son mejores en la estimación de los valores de temperaturas medias anuales, ya que se trata de una variable con un elevado control ambiental por parte de variables como la elevación o la latitud.

Tratándose de un trabajo de la misma área de estudio, se ha procedido a utilizar el método global de regresión múltiple por pasos e implementado también en otros trabajos (Agnew y Palutikof, 2000, Ninyerola et al, 2000, Vicente et al, 2003, Vicente et al, 2007).

La regresión "stepwise", traducida habitualmente como regresión por pasos, es una versión modificada del proceso de regresión hacia adelante en la que en cada nuevo paso, cuando se incluye una nueva variable, además se reconsidera el mantener las que ya se había añadido previamente, es decir que no sólo puede entrar una nueva variable en cada paso sino que puede salir alguna de las que ya

estaban en la ecuación. El proceso finaliza cuando genera un modelo en el que aunque introduzcamos más variables no mejora la varianza explicada de forma significativa o bien deja fuera variables que muestran colinealidad con otras o no están relacionadas con la variable a cartografiar.

El primer modelo de regresión para la Península Ibérica fue desarrollado por Sánchez-Palomares et al. (1999) y se basó en el análisis de regresión múltiple por etapas, que se llevó a cabo cada mes y para cada una de las 10 cuencas principales en la Península Ibérica.

El coeficiente de determinación ( $R^2$ ) informa sobre la calidad del modelo, su capacidad de predecir y la proporción de variación de los resultados que puede explicarse por el modelo. En un modelo de regresión lineal el coeficiente de determinación es el cuadrado del coeficiente de correlación de Pearson (Steel y Torrie, 1960) y expresa que el comportamiento temporal es similar entre los valores predichos y los observados.

#### 4.3.2 Validación del modelo

El método de validación de los modelos ha consistido en aplicar el sistema de validación cruzada dejando uno fuera o Leave-one-out cross-validation (LOOCV). El número limitado de observatorios disponibles para la modelización podía condicionar mucho la modelización a la hora de extraer un número establecido de datos de muestra de forma aleatoria y dejando la otra parte para el entrenamiento.

La validación cruzada "leave-one-out" o "jackknifing" es un método comúnmente aplicado en geoestadística (Isaaks y Srivastava, 1989, Cressie, 1993, Nalder y Ross, 1998) porque no se requieren datos reservados para la validación de datos. El número de sitios muestreados con datos climáticos no suele ser muy grande y son escasos en toda el área de estudio, por lo que todos los datos muestreados se utilizan para la predicción espacial con el fin de mejorar la precisión de las predicciones.

De manera que se ha considerado el método de validación cruzada (Anexo 3) dejando uno fuera como el que mejor se adaptaba a este trabajo ya que posee varias ventajas. Por ejemplo el error estimado suele ser más estable en este tipo de validación y no se tiende a sobreestimar el error, es decir, es mucho mayor el error en la tabla de test que en toda la tabla de datos.

Este algoritmo implica dividir los datos de tal forma que para cada iteración obtengamos una sola muestra para los datos de prueba y todo el resto forme parte de los datos de entrenamiento.

En la validación cruzada dejando uno fuera, se realizan tantas iteraciones como muestras ( $N$ ) tenga el conjunto de datos. De forma que para cada una de las  $N$  iteraciones se realiza un cálculo de error. El resultado final lo obtenemos realizando la media aritmética de los  $N$  valores de errores obtenidos, según la fórmula:

$$E = \frac{1}{N} \sum_{i=1}^N E_i.$$

## 5. RESULTADOS Y DISCUSION

En color verde se encuentran aquellos meses cuyo  $R^2$  mejora con la incorporación de las variables procedentes de satélite, en naranja los meses en los que su  $R^2$  permanece igual, y en rojo cuando el  $R^2$  empeora al incluir las variables de satélite.

**Tabla 1.** Descripción de los modelos por validación cruzada en temperaturas mínimas.

Año	Mes	Variables Incluidas	R <sup>2</sup>	RMSE	Variables
2005	Enero	No Satélite	0.57	1.13	CANT,MED,ALT,X
		Satélite	0.61	1.12	LST,CANT,ALT,Y,RUG
	Febrero	No Satélite	0.58	1.22	CANT,MED,ALT,Y
		Satélite	0.60	1.17	LST,CANT,ALT,Y
	Marzo	No Satélite	0.42	1.29	CANT,MED,ALT,Y
		Satélite	0.42	1.31	CANT,MED,ALT,Y
	Abril	No Satélite	0.51	1.02	CANT,ALT,Y
		Satélite	0.51	1.02	MED,ALT,X
	Mayo	No Satélite	0.53	1.14	MED,ALT,X
		Satélite	0.53	1.14	MED,ALT,X
	Junio	No Satélite	0.51	1.21	ALT,Y
		Satélite	0.51	1.25	ALT,Y
2006	Julio	No Satélite	0.46	1.42	ALT,Y
		Satélite	0.46	1.42	ALT,Y
	Agosto	No Satélite	0.44	1.40	ALT
		Satélite	0.44	1.42	ALT
	Septiembre	No Satélite	0.47	1.38	ALT,RUG
		Satélite	0.47	1.40	ALT,RUG
	Octubre	No Satélite	0.52	1.08	ALT
		Satélite	0.52	1.14	ALT
	Noviembre	No Satélite	0.47	1.05	CANT,ALT,Y
		Satélite	0.47	1.06	CANT,ALT,Y
	Diciembre	No Satélite	0.49	1.29	CANT,ALT,Y
		Satélite	0.59	1.13	LST,CANT,Y,RUG
2006	Enero	No Satélite	0.59	0.98	CANT,MED,ALT,X
		Satélite	0.59	1.00	CANT,MED,ALT,Y
	Febrero	No Satélite	0.34	1.29	MED,ALT,X
		Satélite	0.37	1.27	LST,RUG
	Marzo	No Satélite	0.58	0.93	CANT,ALT,Y
		Satélite	0.58	0.96	CANT,ALT,Y
	Abril	No Satélite	0.47	1.15	ALT,Y
		Satélite	0.47	1.15	ALT,Y
	Mayo	No Satélite	0.50	1.23	ALT,Y
		Satélite	0.50	1.23	ALT,Y
	Junio	No Satélite	0.52	1.13	ALT
		Satélite	0.52	1.13	ALT
2006	Julio	No Satélite	0.49	1.27	ALT
		Satélite	0.53	1.24	NDVI,ALT
	Agosto	No Satélite	0.67	1.20	CANT,MED,ALT,Y
		Satélite	0.67	1.22	CANT,MED,ALT,Y
	Septiembre	No Satélite	0.58	1.20	ALT,RUG
		Satélite	0.58	1.22	ALT,RUG
	Octubre	No Satélite	0.35	1.33	ALT
		Satélite	0.39	1.29	NDVI,ALT
	Noviembre	No Satélite	0.42	1.21	MED,ALT,Y,RUG
		Satélite	0.42	1.20	MED,ALT,Y,RUG
	Diciembre	No Satélite	0.09	1.34	RUG
		Satélite	0.21	1.23	LST,CANT,Y,RUG

ALT (Altitud), RUG (Rugosidad), MED (Distancia al mar Mediterráneo), CANT (Distancia al mar Cantábrico), Y (Latitud), X (Longitud), LST (Temperatura Superficie Terrestre), NDVI (Índice de vegetación).

**Tabla 2.** Descripción de los modelos por validación cruzada en temperaturas máximas.

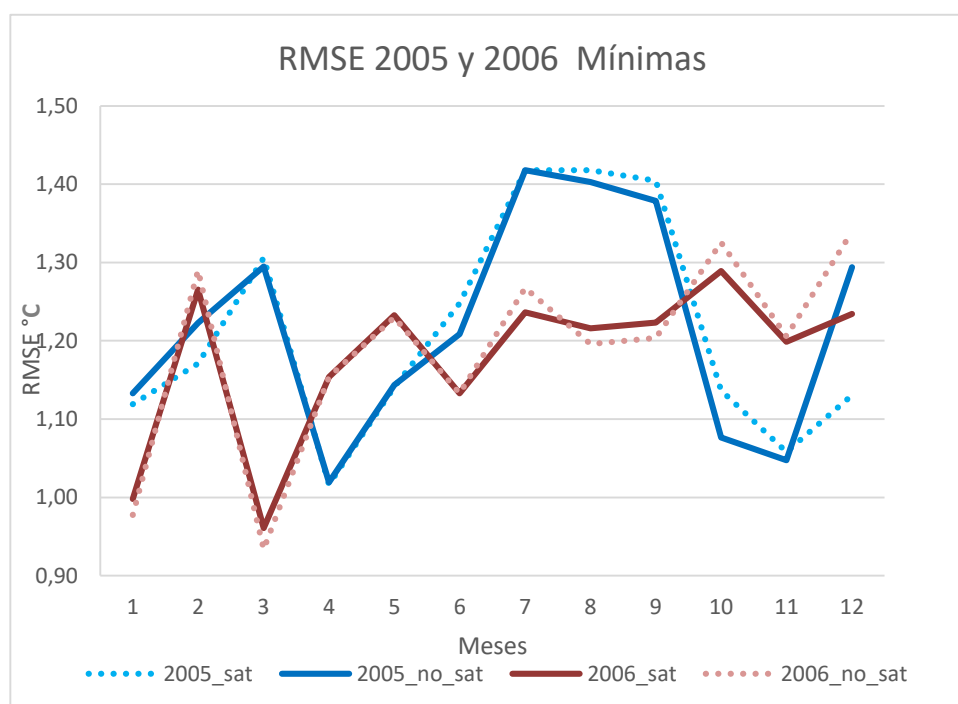
Año	Mes	Variables Incluidas	R <sup>2</sup>	RMSE	Variables
2005	Enero	No Satélite	0.34	1.12	MED,Y,X,ORI,RUG
		Satélite	0.39	1.10	LST,NDVI,MED,Y,X,ORI
	Febrero	No Satélite	0.71	1.17	CANT,ALT,Y,X,ORI,RAD,RUG
		Satélite	0.71	1.19	CANT,ALT,Y,X,ORI,RAD,RUG
	Marzo	No Satélite	0.67	1.11	CANT,ALT,Y,X,ORI,RAD
		Satélite	0.67	1.11	CANT,ALT,Y,X,ORI,RAD
	Abril	No Satélite	0.77	1.06	CANT,ALT,Y,X,ORI,RAD
		Satélite	0.77	1.06	CANT,ALT,Y,X,ORI,RAD
	Mayo	No Satélite	0.69	1.22	CANT,MED,ALT,Y,ORI,RAD
		Satélite	0.69	1.22	CANT,MED,ALT,Y,ORI,RAD
	Junio	No Satélite	0.65	1.12	CANT,MED,ALT,ORI,RAD
		Satélite	0.65	1.14	CANT,MED,ALT,ORI,RAD
2006	Julio	No Satélite	0.53	1.14	CANT,ALT,ORI,RUG
		Satélite	0.54	1.18	LST,ALT,ORI,RUG
	Agosto	No Satélite	0.47	1.22	ALT,ORI,RUG
		Satélite	0.47	1.23	ALT,ORI,RUG
	Septiembre	No Satélite	0.56	1.17	ALT,Y,ORI,RUG
		Satélite	0.56	1.19	ALT,Y,ORI,RUG
	Octubre	No Satélite	0.65	1.00	CANT,ALT,ORI,RAD
		Satélite	0.65	0.99	CANT,ALT,ORI,RAD
	Noviembre	No Satélite	0.63	1.02	CANT,ALT,Y,X,ORI,RAD
		Satélite	0.63	1.03	CANT,ALT,Y,X,ORI,RAD
	Diciembre	No Satélite	0.44	1.14	MED,ALT,X,ORI,RUG
		Satélite	0.47	1.14	LST,MED,ALT,X,ORI,RAD
2006	Enero	No Satélite	0.51	0.99	MED,ALT,X,ORI,RAD,RUG
		Satélite	0.51	1.01	MED,ALT,X,ORI,RAD,RUG
	Febrero	No Satélite	0.54	1.11	MED,ALT,X,ORI,RAD,RUG
		Satélite	0.54	1.10	MED,ALT,X,ORI,RAD,RUG
	Marzo	No Satélite	0.78	0.93	CANT,ALT,Y,X,ORI,RAD
		Satélite	0.78	0.93	CANT,ALT,Y,X,ORI,RAD
	Abril	No Satélite	0.68	1.20	CANT,ALT,Y,X,ORI,RAD
		Satélite	0.68	1.20	CANT,ALT,Y,X,ORI,RAD
	Mayo	No Satélite	0.73	1.08	CANT,ALT,Y,X,ORI,RAD
		Satélite	0.73	1.10	CANT,ALT,Y,X,ORI,RAD
	Junio	No Satélite	0.63	1.12	ALT,Y,ORI,RAD
		Satélite	0.63	1.12	ALT,Y,ORI,RAD
	Julio	No Satélite	0.70	0.99	CANT,ALT,ORI,RAD
		Satélite	0.70	0.98	CANT,ALT,ORI,RAD
	Agosto	No Satélite	0.69	1.11	CANT,MED,ALT,Y,ORI,RUG
		Satélite	0.69	1.11	CANT,MED,ALT,Y,ORI,RUG
	Septiembre	No Satélite	0.64	1.12	CANT,MED,ALT,ORI,RAD
		Satélite	0.64	1.12	CANT,MED,ALT,ORI,RAD
	Octubre	No Satélite	0.71	1.00	ALT,Y,ORI,RAD
		Satélite	0.71	0.99	ALT,Y,ORI,RAD
	Noviembre	No Satélite	0.63	0.94	CANT,ALT,X,ORI,RAD
		Satélite	0.65	0.91	NDVI,ALT,Y,ORI,RAD
	Diciembre	No Satélite	0.35	1.19	CANT,MED,Y,X,RUG
		Satélite	0.32	1.21	CANT,MED,Y,X

ALT (Altitud), RUG (Rugosidad), MED (Distancia al mar Mediterráneo), CANT (Distancia al mar Cantábrico), Y (Latitud), X (Longitud), LST (Temperatura Superficie Terrestre), NDVI (Índice de vegetación), ORI (Orientación), RAD (Radiación).

Para las temperaturas mínimas, la altitud aparece como variable incluida en el proceso de validación en prácticamente todos los casos, en concreto en el 91,6% casos, seguido de la latitud (54,17%), distancia al Cantábrico (37,50%), distancia al Mediterráneo (29,17%), rugosidad (22,92%), longitud (12,50%), LST (10,42%) y NDVI (4,17%). La orientación y radiación no aparecen en ninguno de los casos.

Si observamos las temperaturas máximas, la orientación aparece en el mayor número de los casos con un 95,83%, seguido muy cerca de la altitud (91,67), radiación (72,92%), distancia al Cantábrico (62,50%), latitud (60,42%), longitud (52,08%), distancia al Mediterráneo (37,50%), rugosidad (35,42%), LST (6,25%) y NDVI (4,17%).

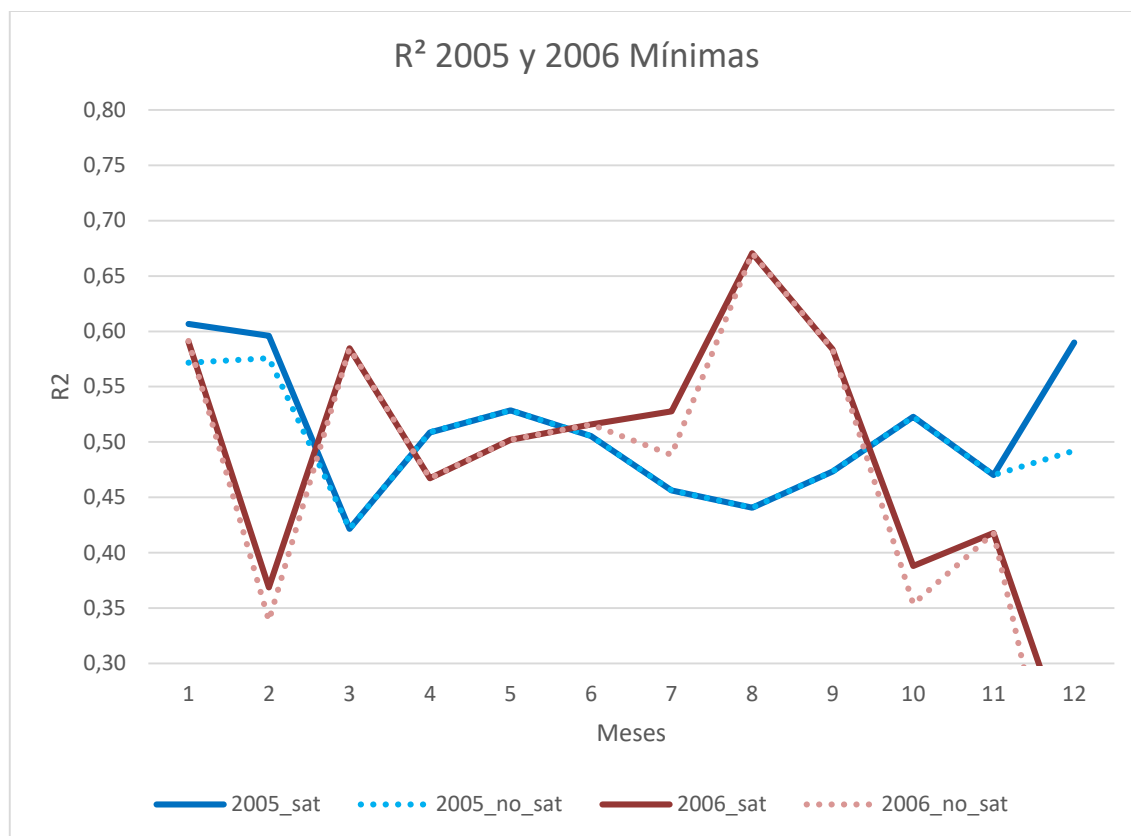
En el caso de los resultados de las temperaturas mínimas en 2005, podemos observar que el RMSE no supera 1,5° grados y en muchos casos ronda valores en torno a un 1°. El  $R^2$  presenta unos valores más bajos en los meses de verano (Julio, agosto y septiembre) aunque existe algún otro mes aislado con valores similares. Se puede observar que los tres meses de verano también tienen asociado un RMSE más alto y que solo intervienen dos variables en los modelos, siendo la altitud la variable incluida en estos tres meses y en general, en todos los modelos de las temperaturas mínimas de 2005. Por otro lado, podemos destacar la presencia de las variables de distancia a los mares (MED y CANT) prácticamente en todas las estaciones menos en verano. La variable NDVI no aparece en ningún mes, mientras que el LST aparece en 3 meses, en color verde, lo que significa que cuando entran las variables de satélite los resultados de los modelos mejoran.



**Figura 6:** Distribución del RMSE entre 2005 y 2006 para las temperaturas mínimas.

*2005\_sat (RMSE de los modelos incluyendo variables geográficas y de satélite en el año 2005), 2005\_no\_sat (RMSE de los modelos incluyendo solo variables geográficas en el año 2005), 2006\_sat (RMSE de los modelos incluyendo variables geográficas y de satélite en el año 2006), 2006\_no\_sat (RMSE de los modelos incluyendo solo variables geográficas en el año 2006).*

En el caso de las temperaturas mínimas de 2006, los resultados son muy distintos a los del año anterior y no se observa esa distinción en la estación de verano. En general, la altitud (ALT) es la variable que permanece inmóvil en prácticamente todos los meses, pero en este año, el NDVI aparece en julio y octubre, mejorando los resultados. El  $R^2$  varía según los meses pero se asemeja en general a los de 2005, exceptuando el mes de diciembre, el cual presenta un  $R^2$  muy bajo, tanto el modelo sin variables de satélite como en el combinado. También en febrero el  $R^2$  es particularmente bajo.

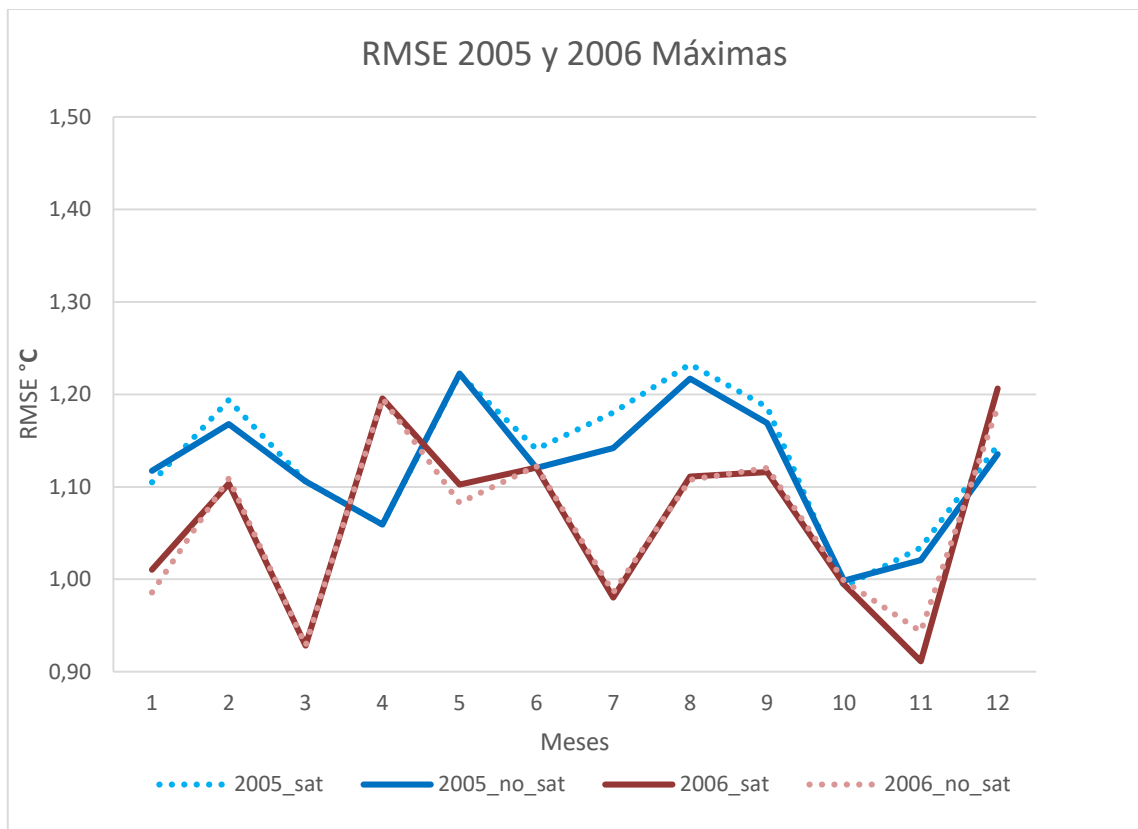


**Figura 7:** Distribución del  $R^2$  entre 2005 y 2006 para las temperaturas mínimas.

2005\_sat ( $R^2$  de los modelos incluyendo variables geográficas y de satélite en el año 2005), 2005\_no\_sat ( $R^2$  de los modelos incluyendo solo variables geográficas en el año 2005), 2006\_sat ( $R^2$  de los modelos incluyendo variables geográficas y de satélite en el año 2006), 2006\_no\_sat ( $R^2$  de los modelos incluyendo solo variables geográficas en el año 2006).

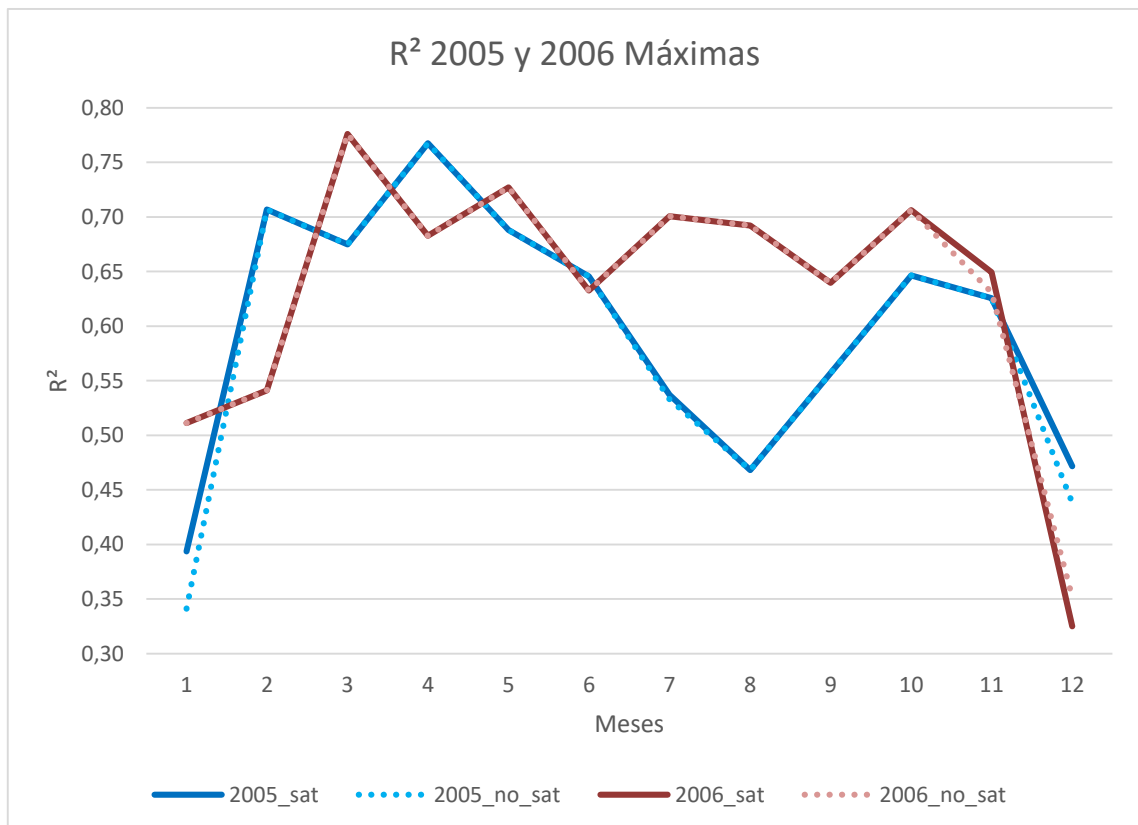
Existen meses donde el  $R^2$  y las variables incluidas son idénticas pero que el RMSE es distintos entre el modelo que incluye a priori solo variables geográficas y el que combina variables geográficas y de satélite. La explicación se debe en la fase de validación cruzada, en la que se estima el valor predicho de un dato mediante la interpolación basada en el conjunto de datos restantes. Es posible que en alguna de las iteraciones, alguna variable de satélite entre a formar parte del modelo por pasos, alterando en parte, el valor del dato predicho y modificando el RMSE global del modelo.

En el caso de las temperaturas máximas de 2005, se aprecia a primera vista la inclusión en general de las variables orientación y radiación (ORI y RAD, respectivamente). Resulta lógico ya que durante el día, estas variables condicionan en gran medida la temperatura. El  $R^2$  y el RMSE presentan unos valores mejores que las mínimas, aunque la inclusión de las variables de satélite aparece de forma puntual, del mismo modo que las mínimas. La variable altitud está incluida en casi todos los modelos.



**Figura 8:** Distribución del RMSE entre 2005 y 2006 para las temperaturas máximas.

Para las temperaturas máximas en el año 2006, todos los meses poseen un  $R^2$  superior a 0,5, exceptuando diciembre de 2006, el cual ya había dado problemas en las temperaturas mínimas. Además, el RMSE se acerca a 1 grado centígrado e incluso se sitúa por debajo de ese umbral en algunos meses. Sin embargo, sólo aparecen variables de satélite en el mes de noviembre, en concreto el NDVI, de modo que en todo 2006, la LST no se considera significativa en los modelos de regresión por pasos.



**Figura 9:** Distribución del  $R^2$  entre 2005 y 2006 para las temperaturas máximas.

El caso del mes de diciembre de 2006 resulta llamativo ya que genera unos índices de correlación muy bajos, tanto para las máximas como para las mínimas. Esto requiere un análisis más profundo tanto de los datos como de la situación climática que se produjo en ese espacio de tiempo.

A la vista de los resultados obtenidos, no es fácil explicar la distribución tanto del  $R^2$  como del RMSE y las variables que intervienen en cada uno de los modelos. No se puede afirmar que aquellos modelos que posean un  $R^2$  más elevado tengan un RMSE más bajo ya que está condicionado por los valores de partida. Por otra parte, la inclusión en los modelos de unas u otras variables tampoco esclarece esta distribución, que adquiere un carácter aleatorio tanto entre los distintos meses de un año concreto como en años diferentes. El único hecho verificado aparece entre los modelos combinados y los modelos en los que solo se utilizan variables geográficas, en los que al incluir variables de satélite mejoran los resultados de  $R^2$  si bien este hecho no ocurre en gran cantidad de meses.

Aunque LST y Tair están fuertemente correlacionados, ambos tienen diferentes significados físicos, magnitudes técnicas de medición, la respuesta a las condiciones atmosféricas y la fase diurna (Jin y Dickinson, 2010).

El gradiente entre LST y Tair es controlado por una superficie compleja de balance de energía dependiente de la información no siempre proporcionada por los satélites (Prince et al, 1998; Vancutsem et al., 2010). El gradiente de temperatura puede variar mucho en un ciclo diurno y los patrones pueden estar influenciados temporalmente por las variaciones en la duración de los días y las noches (Zaksek y Schroedter-Homscheidt, 2009).

Durante el día, la Temperatura de la Superficie Terrestre (LST) es generalmente más alta que la temperatura del aire, y por la noche, ocurre lo contrario, lo que lleva a la sobreestimación y subestimación de aire temperatura, respectivamente (Cresswell et al., 1999).

El hecho de que la información de satélite no aparezca de manera habitual en los modelos de regresión por pasos puede residir en la variación de la temperatura durante un día. Como se ha descrito anteriormente, el satélite MODIS/TERRA registra la temperatura durante el día entre las 12:00 y las 13:00 (hora local) y durante la noche entre las 22:00 y las 24:00. Por regla general, se asume que los datos registrados entre las 12:00 y 13:00 por el satélite están asociados a las temperaturas máximas de ese día y que, aunque no coincidiese exactamente, existe una relación. De la misma forma sucede para los datos registrados por la noche por el satélite y las temperaturas mínimas.

Esto a priori resulta razonable, pero si analizamos en profundidad las posibles situaciones sinópticas de algunas zonas del área de estudio a tratar en este trabajo, es posible que los datos captados por el satélite no estén tan relacionados con las temperaturas recogidas por las estaciones meteorológicas. Una de las situaciones podría ser la aparición de inversiones térmicas en invierno, donde las zonas con menor altitud presentan temperaturas inferiores otras localizadas a mayor altitud. Este razonamiento podría ser perfectamente válido sin embargo, la inclusión de la variable altitud en la mayoría de los modelos tanto de las máximas como de las mínimas, establece que las posibles inversiones térmicas no afectan en tal grado para desestimar la variable altitud en alguno de los modelos.

Otro aspecto a tener en cuenta a la hora de analizar la escasa influencia de las variables de satélite en los modelos es el modo de recogida de datos por parte del satélite. Los datos de temperaturas mensuales satelitales proceden de otros productos compuestos de temperaturas cada 8 o 16 días que a su vez, se extraen de los productos de LST diarios. En este trabajo se ha mencionado anteriormente el problema de utilizar datos diarios de temperatura procedentes de satélites por la ausencia de datos en distintas zonas y distintos días, ya sea por la presencia de nubes o por que la información captada no posee una calidad mínima exigida. En los productos compuestos de 8 o 16 días, si existen días sin datos en los píxeles se realiza la media aplicando unos procedimientos estadísticos e incluyendo otro tipo de información. Este método soluciona en parte el problema, pero es posible que altere los datos finales mensuales y con ello, la variable explicativa de satélite que contiene estos datos, no resulte significativa para ser incluida al aplicar el modelo de regresión por pasos.

Por último, se ha de mencionar el tamaño del pixel que tiene asociado cada dato, que en este caso tiene 5600 m y que condiciona sobre manera la modelización de las temperaturas incluyendo esta va-

riable independiente. Como es lógico, en un radio de más de 5km, las variaciones de temperatura pueden ser sensiblemente distintas y por consiguiente, diferentes a los datos registrados por los observatorios meteorológicos.

## **6. CONCLUSIONES**

Los resultados obtenidos muestran diferencias tanto entre las temperaturas máximas y mínimas al aplicar los modelos incorporando solo variables geográficas y los modelos combinados, incluyendo variables geográficas y de satélite.

Los datos de satélite aparecen en los modelos un número limitado de veces, pero cuando están incluidos como variables explicativas, mejoran los modelos.

Los resultados de  $R^2$  de las temperaturas máximas son sensiblemente mejor en general que para las temperaturas mínimas.

Las posibles situaciones sinópticas (por ejemplo, inversiones térmicas) que se pueden dar en algunas estaciones del año sobre el área de estudio, no afectan de forma sensible al comportamiento de la distribución de las temperaturas. Este hecho no produce alteraciones en la inclusión de las variables geográficas a los modelos.

Las variables de satélite no aparecen de forma continua en los distintos modelos porque se encuentran condicionadas, en parte, a la reducida resolución espacial, que produce datos muy generales.

## 7. BIBLIOGRAFÍA

- AEMET, I. (2011). Atlas climático ibérico/Iberian climate atlas. Agencia Estatal de Meteorología, Ministerio de Medio Ambiente y Rural y Marino, Madrid, Instituto de Meteorologia de Portugal.
- Agnew, M. D., & Palutikof, J. P. (2000). GIS-based construction of baseline climatologies for the Mediterranean using terrain variables. *Climate Research*, 14(2), 115-127.
- Aladrén, L. L. VARIEDAD BIOGEOGRÁFICA DEL TERRITORIO ARAGONÉS.
- Allmaras, R. R., Burwell, R. E., Larson, W. E., & Holt, R. F. (1966). Total porosity and random roughness of the interrow zone as influenced by tillage.
- Barry, R.G., Chorley, R.J., 1987. Atmosphere, Weather and Climate, fifth ed. Methuen and Co., London, 459 pp
- Benali, A., Carvalho, A. C., Nunes, J. P., Carvalhais, N., & Santos, A. (2012). Estimating air surface temperature in Portugal using MODIS LST data. *Remote Sensing of Environment*, 124, 108-121.
- Bertrand, M. R., & Wilson, M. L. (1996). Microclimate-dependent survival of unfed adult *Ixodes scapularis* (Acari: Ixodidae) in nature: life cycle and study design implications. *Journal of Medical Entomology*, 33(4), 619-627.
- Chen, P. Y., Srinivasan, R., Fedosejevs, G., & Kiniry, J. R. (2003). Evaluating different NDVI composite techniques using NOAA-14 AVHRR data. *International Journal of Remote Sensing*, 24(17), 3403-3412.
- Connor, S. J., Thomson, M. C., Flasse, S. P., & Perryman, A. H. (1998). Environmental information systems in malaria risk mapping and epidemic forecasting. *Disasters*, 22(1), 39-56.
- Cressie, N. (1993). Statistics for spatial data: Wiley series in probability and statistics. *Wiley-Interscience*, New York, 15, 105-209.
- Cresswell, M. P., Morse, A. P., Thomson, M. C., & Connor, S. J. (1999). Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model. *International Journal of Remote Sensing*, 20(6), 1125-1132.
- Cristóbal, J., Ninyerola, M., & Pons, X. (2008). Modeling air temperature through a combination of remote sensing and GIS data. *Journal of Geophysical Research: Atmospheres*, 113(D13).
- Cristóbal, J., Ninyerola, M., Pons, X., & Pla, M. (2006). Improving air temperature modelization by means of remote sensing variables. In *2006 IEEE International Symposium on Geoscience and Remote Sensing* (pp. 2251-2254). IEEE.
- Cuadrat, J. M., Saz, M. A., & Vicente-Serrano, S. M. (2007). Atlas climático de Aragón. *Gobierno de Aragón*.
- Cuadrat, J.M. (1999). El clima de Aragón. *CAI-100*. Zaragoza. 109 pág.
- Currence, H. D., & Lovely, W. G. (1970). The analysis of soil surface roughness. *Transactions of the ASAE*, 13(6), 710-0714.
- Didan, K. (2015). *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250 m SIN Grid V006*. tech. rep., NASA EOSDIS Land Processes DAAC.
- Flasse, S., Stephenson, P., Hutchinson, P., & Walker, C. (1995, September). Direct use of Meteosat in veterinary applications. In *Proceedings of the EUMETSAT Meteorological Satellite Data Users' Conference: Polar Orbiting Systems* (pp. 453-459).
- Focks, D. A., Daniels, E., Haile, D. G., & Keesling, J. E. (1995). A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *American Journal of Tropical Medicine and Hygiene*, 53(5), 489-506.
- Fu, G., Shen, Z., Zhang, X., Shi, P., Zhang, Y., & Wu, J. (2011). Estimating air temperature of an alpine meadow on the Northern Tibetan Plateau using MODIS land surface temperature. *Acta Ecologica Sinica*, 31(1), 8-13.
- Gandullo Gutiérrez, J. M. (1994). Climatología y Ciencias de suelo. *Fundación Conde del Valle de Salazar*. Madrid.

- Goetz, S. J., Prince, S. D., & Small, J. (2000). Advances in satellite remote sensing of environmental variables for epidemiological applications. *Advances in Parasitology*, 47, 289-307.
- Goodale, C. L., Aber, J. D., & Ollinger, S. V. (1998). Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model. *Climate Research*, 10(1), 35-49.
- Goward, S. N., Waring, R. H., Dye, D. G., & Yang, J. (1994). Ecological remote sensing at OTTER: satellite macroscale observations. *Ecological applications*, 4(2), 322-343.
- Hengl, T., Heuvelink, G. B., Tadić, M. P., & Pebesma, E. J. (2012). Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. *Theoretical and Applied Climatology*, 107(1-2), 265-277.
- Hope, A. S., Boynton, W. L., Stow, D. A., & Douglas, D. C. (2003). Interannual growth dynamics of vegetation in the Kuparuk River watershed, Alaska based on the Normalized Difference Vegetation Index. *International Journal of Remote Sensing*, 24(17), 3413-3425.
- Hudson, G., & Wackernagel, H. (1994). Mapping temperature using kriging with external drift: theory and an example from Scotland. *International journal of Climatology*, 14(1), 77-91.
- IPCC (2001). *Climate change 2001: The scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 881). Cambridge, United Kingdom: Cambridge University Press Retrieved from the World Wide Web: [http://www.grida.no/publications/other/ipcc\\_tar/?src=/CLIMATE/IPCC\\_TAR/wg1/052.htm](http://www.grida.no/publications/other/ipcc_tar/?src=/CLIMATE/IPCC_TAR/wg1/052.htm)
- Isaaks, E. H. (1989). *Applied geostatistics* (No. 551.72 I86). Oxford University Press.
- Jin, M., & Dickinson, R. E. (2010). Land surface skin temperature climatology: benefitting from the strengths of satellite observations. *Environmental Research Letters*, 5(4), 044004.
- Kloog, I., Chudnovsky, A., Koutrakis, P., & Schwartz, J. (2012). Temporal and spatial assessments of minimum air temperature using satellite surface temperature measurements in Massachusetts, USA. *Science of the total environment*, 432, 85-92.
- Lutgens, F.K., Tarbuck, E.J., 1995. *The Atmosphere: An Introduction to Meteorology*, sixth ed. Prentice-Hall, Englewood Cliffs, NJ, 462 pp.
- Mount, G. A., & Haile, D. G. (1989). Computer simulation of population dynamics of the American dog tick (Acari: Ixodidae). *Journal of medical entomology*, 26(1), 60-76.
- Nalder, I. A., & Wein, R. W. (1998). Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. *Agricultural and forest meteorology*, 92(4), 211-225.
- Neteler, M. (2010). Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data. *Remote sensing*, 2(1), 333-351.
- Ninyerola, M., Pons, X., & Roure, J. M. (2000). A methodological approach of climatological modeling of air temperature and precipitation through GIS techniques. *International Journal of Climatology*, 20(14), 1823-1841.
- Prihodko, L., & Goward, S. N. (1997). Estimation of air temperature from remotely sensed surface observations. *Remote Sensing of Environment*, 60(3), 335-346.
- Prince, S. D., Goetz, S. J., Dubayah, R. O., Czajkowski, K. P., & Thawley, M. (1998). Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using Advanced Very High-Resolution Radiometer satellite observations: comparison with field observations. *Journal of Hydrology*, 212, 230-249.
- Recondo, C., & Pérez-Morandeira, C. (2002). Obtención de la temperatura del aire de Asturias a partir de la temperatura de superficie calculada con imágenes NOAA-AVHRR. *Revista de Teledetección*, 17, 5-12.
- Richardson, A. D., Lee, X., & Friedland, A. J. (2004). Microclimatology of treeline spruce–fir forests in mountains of the northeastern United States. *Agricultural and Forest Meteorology*, 125(1), 53-66.

- Riddering, J. P., & Queen, L. P. (2006). Estimating near-surface air temperature with NOAA AVHRR. *Canadian Journal of Remote Sensing*, 32(1), 33-43.
- Vicente Serrano, S. M., & Saz Sánchez, M. Á. (2002). Cartografía de precipitaciones y temperaturas en el valle medio del Ebro mediante la utilización de diferentes técnicas estadísticas. *Geographica*, (42), 73-92.
- Sánchez, O., Rubio, A., Blanco, A., Elena, R., & Gómez, V. (2003). Autoecología paramétrica de los hayedos de Castilla y León. *Invest Agrar: Sist Recur For*, 12(1), 87-110.
- Saz-Sánchez, M. A., Serrano-Notivoli, R., de Luis-Arrillaga, M., & Longares-Aladrén, L. A. (2010). Comparación de métodos de interpolación y de regresión para la cartografía de temperaturas máximas y mínimas absolutas: el caso de navarra (Norte de España) en 2009. In *VII Congreso AEC: Clima, ciudad y ecosistemas*.
- Smith, W. L., Leslie, L. M., Diak, G. R., Goodman, B. M., Velden, C. S., Callan, G. M., ... & Wade, G. S. (1988). The integration of meteorological satellite imagery and numerical dynamical forecast models. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 324(1579), 317-323.
- Steel, R. G., & Torrie, J. H. (1960). Principles and procedures of statistics. *Principles and procedures of statistics*.
- Stow, D. A., Hope, A., McGuire, D., Verbyla, D., Gamon, J., Huemmrich, F., ... & Hinzman, L. (2004). Remote sensing of vegetation and land-cover change in Arctic Tundra Ecosystems. *Remote sensing of environment*, 89(3), 281-308.
- Thomson, M. C., Connor, S. J., Milligan, P. J., & Flasse, S. P. (1996). The ecology of malaria--as seen from Earth-observation satellites. *Annals of tropical medicine and parasitology*, 90(3), 243-264.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2), 127-150.
- Vancutsem, C., Ceccato, P., Dinku, T., & Connor, S. J. (2010). Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, 114(2), 449-465.
- Vicente Serrano, S. M., Sánchez, S., & Cuadrat, J. M. (2003). Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature.
- Vicente-Serrano, S. M., Lanjeri, S., & López-Moreno, J. I. (2007). Comparison of different procedures to map reference evapotranspiration using geographical information systems and regression-based techniques. *International Journal of Climatology*, 27(8), 1103-1118.
- Vogt, J. V., Viau, A. A., & Paquet, F. (1997). Mapping regional air temperature fields using satellite-derived surface skin temperatures. *International Journal of Climatology*, 17(14), 1559-1579.
- Wan, Z., Hook, S., & Hulley, G. (2015). MOD11 L2 MODIS/Terra Land Surface Temperature/Emissivity 5-Min L2 Swath 1 km V006. NASA EOSDIS Land Processes DAAC.
- Zakšek, K., & Schroedter-Homscheidt, M. (2009). Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4), 414-421.
- Zhu, W., Lü, A., & Jia, S. (2013). Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sensing of Environment*, 130, 62-73.

## 8. ANEXOS

**Anexo1:** Código en R para cambiar el formato, hacer el mosaico, reproyectar y recortar las imágenes.

```
#PRIMERO LISTO LOS ARCHIVOS EN FORMATO
HDF
setwd("C:/Users/Dario/Desktop/TFM/Nueva carpe-
ta/imagenes_modis/MENSUALES/2006")
out.files <- list.files(getwd(), pattern="hdf$",
full.names=FALSE)
print(out.files)
x=0
y=".tif"
imagenes=out.files
imagenes2=imagenes
tt=imagenes
ss=imagenes
rr=imagenes
#ESTE BUCLE ME SELECCIONA EL LST_DAY(1)
o EL LST_NIGHT(5) Y LO CONVIERTE A TIF

for(i in 1:length(imagenes)) {
  x=x+1
  tt[x]=paste(imagenes[x],".tif",sep="")

  gdal_translate(imagenes[x],tt[x],sd_index=1)
}
# AHORA LISTO LOS ARCHIVOS TIF PARA DES-
PUES TRABAJAR CON ELLOS
setwd("C:/Users/Dario/Desktop/TFM/Nueva carpe-
ta/imagenes_modis/MENSUALES/2006")
out.files <- list.files(getwd(), pattern=".tif$",
full.names=FALSE)
print(out.files)
imagenes=out.files
destino1="C:/Users/Dario/Desktop/TFM/Nueva carpe-
ta/imagenes_modis/MENSUALES/2006/LST_DAY/"
setwd("C:/Users/Dario/Desktop/TFM/Nueva carpe-
ta/imagenes_modis/MENSUALES/2006")
x=1
tt=imagenes
ss=imagenes
rr=imagenes
imagenes[1]
numero=length(imagenes)
numero=numero/2

for(i in 1:numero) {      #ESTE BUCLE ME HARÁ
SOLO LOS MOSAICOS DE LOS TIF DEL BUCLE
ANTERIOR

  tt[i]=paste(imagenes[x],"mosaic.tif",sep="")
  dest1=paste(destino1,tt[i],sep="")
  im1=raster(imagenes[x])
  im2=raster(imagenes[x+1])
  m=merge(im1,im2)
  writeRaster(m,filename=dest1)
  x=x+2
}
```

```
# AHORA REPROYECTAMOS LOS MOSAICOS Y
RECORTAMOS LA IMAGEN EN BASE AL AREA
DE ESTUDIO
setwd("C:/Users/Dario/Desktop/TFM/Nueva carpe-
ta/imagenes_modis/MENSUALES/2006/LST_DAY")
destino1="C:/Users/Dario/Desktop/TFM/Nueva carpe-
ta/imagenes_modis/MENSUALES/2006/LST_DAY/"
destino2="C:/Users/Dario/Desktop/TFM/Nueva carpe-
ta/imagenes_modis/MENSUALES/2006/LST_DAY/fin
ales/"
out.files <- list.files(getwd(), pattern=".tif$",
full.names=FALSE)
print(out.files)
imagenes=out.files
xx <- readShape-
Poly(fn="C:/Users/Dario/Desktop/TFM/carto_tfm/area_
estudio.shp", IDvar="D_MUNI_INE")
tt=imagenes
ss=imagenes
rr=imagenes
for(i in 1:length(imagenes)) {

  tt[i]=paste(imagenes[i],"trans.tif",sep="")
  ss[i]=paste(imagenes[i],"proy.tif",sep="")
  dest1=paste(destino1,ss[i],sep="")
  gdal_translate(a_srs="+proj=sinu +lon_0=0 +x_0=0
+y_0=0 +a=6371007.181 +b=6371007.181 +units=m
+no_defs",imagenes[i],tt[i])

  gdalwarp(s_srs="+proj=sinu +lon_0=0 +x_0=0
+y_0=0 +a=6371007.181 +b=6371007.181 +units=m
+no_defs",t_srs="EPSG:25830 +proj=utm +zone=30
+ellps=GRS80 +datum=ETRS89 +units=m +no_defs
",tt[i],ss[i], overwrite = TRUE)
  dest1=paste(destino1,ss[i],sep="")
  im=raster(dest1)
  rr[i]=paste(ss[i],"mask.tif",sep="")
  dest2=paste(destino2,rr[i],sep="")
  pp=crop(im,xx)
  pp <- mask(pp, xx, updateNA=FALSE)
  writeRaster(pp,filename=dest2)
}
```

## Anexo2: Código en R para verificar la calidad de los pixeles de LST diarios.

```

imagenes <-
list.files("C:/Users/Dario/Desktop/TFM/Nueva carpeta/imagenes_modis/2006_buenas", pattern = ".tif" ,
full.names=FALSE)
imagenes2=imagenes
tt=imagenes
ss=imagenes
rr=imagenes
xx <- readShapePoly(fn="C:/Users/Dario/Desktop/TFM/carto_tfm/area_estudio.shp", IDvar="D_MUNI_INE")

destino1= "C:/Users/Dario/Desktop/TFM/Nueva carpeta/imagenes_modis/2006_buenas/"
destino2="C:/Users/Dario/Desktop/TFM/Nueva carpeta/imagenes_modis/2006_buenas/QC_DAY/"

QC_Data <- data.frame(Integer_Value = 0:255,
                      Bit7 = NA,
                      Bit6 = NA,
                      Bit5 = NA,
                      Bit4 = NA,
                      Bit3 = NA,
                      Bit2 = NA,
                      Bit1 = NA,
                      Bit0 = NA,
                      QA_word1 = NA,
                      QA_word2 = NA,
                      QA_word3 = NA,
                      QA_word4 = NA
)
for(i in QC_Data$Integer_Value){
  AsInt <- as.integer(intToBits(i)[1:8])
  QC_Data[i+1,2:9]<- AsInt[8:1]
}
QC_Data$QA_word1[QC_Data$Bit1 == 0 &
QC_Data$Bit0==0] <- "LST GOOD"
QC_Data$QA_word1[QC_Data$Bit1 == 0 &
QC_Data$Bit0==1] <- "LST Produced,Other Quality"
QC_Data$QA_word1[QC_Data$Bit1 == 1 &
QC_Data$Bit0==0] <- "No Pixel,clouds"
QC_Data$QA_word1[QC_Data$Bit1 == 1 &
QC_Data$Bit0==1] <- "No Pixel, Other QA"

QC_Data$QA_word2[QC_Data$Bit3 == 0 &
QC_Data$Bit2==0] <- "Good Data"
QC_Data$QA_word2[QC_Data$Bit3 == 0 &
QC_Data$Bit2==1] <- "Other Quality"
QC_Data$QA_word2[QC_Data$Bit3 == 1 &
QC_Data$Bit2==0] <- "TBD"
QC_Data$QA_word2[QC_Data$Bit3 == 1 &
QC_Data$Bit2==1] <- "TBD"

QC_Data$QA_word3[QC_Data$Bit5 == 0 &
QC_Data$Bit4==0] <- "Emiss Error <= .01"
QC_Data$QA_word3[QC_Data$Bit5 == 0 &
QC_Data$Bit4==1] <- "Emiss Err >.01 <=.02"
QC_Data$QA_word3[QC_Data$Bit5 == 1 &
QC_Data$Bit4==0] <- "Emiss Err >.02 <=.04"
QC_Data$QA_word3[QC_Data$Bit5 == 1 &
QC_Data$Bit4==1] <- "Emiss Err > .04"

QC_Data$QA_word4[QC_Data$Bit7 == 0 &
QC_Data$Bit6==0] <- "LST Err <= 1"
QC_Data$QA_word4[QC_Data$Bit7 == 0 &
QC_Data$Bit6==1] <- "LST Err > 2 LST Err <= 3"
QC_Data$QA_word4[QC_Data$Bit7 == 1 &
QC_Data$Bit6==0] <- "LST Err > 1 LST Err <= 2"
QC_Data$QA_word4[QC_Data$Bit7 == 1 &
QC_Data$Bit6==1] <- "LST Err > 4"

# FINAL3 ES PARA QUITAR TAMBIEN LOS
ERRORES ENTRE 2 Y 3 GFRADOS
# FINAL4 INCLUYE DATOS BUENOS HASTA 2 Y
3 GRADOS INCLUSIVE
FINAL3 <- QC_Data[(QC_Data$Bit1 == 1 &
QC_Data$Bit6 ==1) | (QC_Data$Bit1 == 1) |
(QC_Data$Bit6 == 1 & QC_Data$Bit7 ==1) |
(QC_Data$Bit7 == 0 & QC_Data$Bit6 ==1),]
FINAL4 <- QC_Data[(QC_Data$Bit1 == 1 &
QC_Data$Bit6 ==1) | (QC_Data$Bit1 == 1) |
(QC_Data$Bit6 == 1 & QC_Data$Bit7 ==1),]

vv<-FINAL3["Integer_Value"]
malos=array(dim=2)
for(t in 1:192){
  malos[t]=vv[t,1]
}
# IMAGENES BUENAS AL
50%////////////////////////////////////

# LOS PÍXELES FUERA DEL RECORTE CON VA-
LOR "NA" SON 20.400

malas=array(dim = 2)
buenas=array(dim = 2)
malas1=array(dim = 2)
buenas1=array(dim = 2)
setwd("C:/Users/Dario/Desktop/TFM/Nueva carpeta/imagenes_modis/2005_buenas/QC_DAY")
SDS <- list.files("C:/Users/Dario/Desktop/TFM/Nueva carpeta/imagenes_modis/2005_buenas/QC_DAY",
pattern=".tif$", full.names=FALSE)
SDS

bb=1
mm=1
y=1

for(y in 1:length(SDS)){
  DayQC <- raster(SDS[y])
  m <- DayQC

  for(w in 1:192){
    m[m == malos[w]]<-NA
  }
}

```

```

freq(m)
valores=getValues(m)
contar=0
contar2=0
valores
length(valores)
valores
for(i in 1:length(valores)){
  # print(i)
  if (is.na(valores[i])==TRUE){
    # print(valores[i])
    contar=contar+1
  }

}
contar=contar-20400
total=length(valores)
total1=total-20400
cincuenta=total1*0.5

# contar
print(contar)
if(contar > cincuenta) {

  malas[mm]=SDS[y]
  malas1[mm]=contar

  mm=mm+1
} else {

  print("dentro")## Si la condición no se cumple
(FALSE)
  buenas[bb]=SDS[y]
  buenas1[bb]=contar
  bb=bb+1
}
y=y+1
}
write.table(malas,"img_malas_50_0.txt")
write.table(malas1,"img_malas_50_1.txt")
write.table(buenas,"img_buenas_50_0.txt")
write.table(buenas1,"img_buenas_50_1.txt")
# //////////////////////////////////////
# IMAGENES BUENAS AL
75%////////////////////////////////////

# LOS PÍXELES FUERA DEL RECORTE CON VA-
LOR "NA" SON 20.400
setwd("C:/Users/Dario/Desktop/TFM/Nueva carpe-
ta/imagenes_modis/2005_buenas/QC_DAY")
malas=array(dim = 2)
buenas=array(dim = 2)
malas1=array(dim = 2)
buenas1=array(dim = 2)
SDS=read.table("img_buenas50.txt")
bb=1
mm=1
y=1

```

```

g1=seq(2,62,4)
g2=seq(130,190,4)
g3=seq(66,126,4)
g4=seq(194,254,4)
g5=seq(3,63,4)
g6=seq(131,191,4)
g7=seq(67,127,4)
g8=seq(195,255,4)

valores_malos=c(g1,g2,g3,g4,g5,g6,g7,g8)
valores_malos
for(x in SDS){
  DayQC <- raster(SDS[y])
  m <- DayQC
  for(w in 1:128){
    m[m == valores_malos[w]]<-NA
  }
  valores=getValues(m)
  contar=0
  contar2=0
  valores
  length(valores)
  for(i in 1:length(valores)){
    # print(i)
    if (is.na(valores[i])==TRUE){
      # print(valores[i])
      contar=contar+1
    }
  }
  contar=contar-20400
  total=length(valores)
  total1=total-20400
  cincuenta=total1*0.75

  # contar
  print(contar)

  if(contar > cincuenta) {

    malas[mm]=SDS[y]
    malas1[mm]=contar
    mm=mm+1
  } else {
    print("dentro")## Si la condición no se cumple
    (FALSE)
    buenas[bb]=SDS[y]
    buenas1[bb]=contar
    bb=bb+1
  }
  y=y+1
}
write.table(malas,"img_malas_75.txt",ncol=1)
write.table(malas1,"img_malas_75.txt",ncol=2)
write.table(buenas,"img_buenas_75.txt",ncol=1)
write.table(buenas1,"img_buenas_75.txt",ncol=2)

write.table(malas,"img_malas_75.txt")
write.table(buenas,"img_buenas_75.txt")

```

### Anexo3: Código en R aplicando la validación cruzada dejando uno fuera.

```
prediccion1=data.frame()
# ESTE BUCLE ME HACE LA REGRESION DE
2005

for(j in 0:11) {

  setwd("C:/Users/Dario/Desktop/TFM/carto_tfm/img_m
odelos/MAXIMAS2")

  datos <- read.dbf(file = "MAXI-
MAS_FINAL_BIEN2.dbf")
  head(datos) # primeras filas de la tabla
  str(datos) # estructura de la tabla
  names(datos) # nombres de las columnas

  names(datos) # ver en qué columnas están las varia-
bles predictoras
  fijas=paste(names(datos)[45:52], collapse = " + ")
  var1=paste(names(datos)[j+9], "~", collapse = " + ")
  var2=paste(names(datos)[j+33], "+", collapse = " + ")
  var3=paste(names(datos)[j+65], "+", collapse = " + ")
  todas=paste(var1,var2,var3,fijas)
  todas

  Folds <- nrow(datos)
  Folds1=unique(datos$Codigo)
  Folds1
  for (i in 1:Folds)
  {
    Test <- subset(datos, Codigo == Folds1[i])
    Entrenamiento <- subset(datos, !Codigo ==
Folds1[i])
    Modelo <- step(lm(todas, family = gaussian, data =
Entrenamiento))
    Modelo <- modelTrim(Modelo)

    # Modelo <- rpart(todas, data = Entrenamiento)
    Prediccion <- predict(Modelo, Test)
    prediccion1[i,1]=Prediccion
    prediccion1[i,2]=Test[1,j+9]
  }
  rmse <- sqrt(mean((prediccion1$V1 - predic-
cion1$V2)^2))
  rmse

  setwd("C:/Users/Dario/Desktop/TFM/carto_tfm/img_m
odelos/MAXIMAS2/RMSE")
  destino=paste("Maximas_2005_RMSE_sat.txt", sep =
"")
  write.table(rmse,destino, sep = "\n \n", quote =
FALSE,row.names = FALSE,col.names = FALSE,
append =TRUE)
}
```

```
# ESTE BUCLE ME HACE LA REGRESION DE
2006
# //////////////////////////////////////

for(j in 0:11) {

  setwd("C:/Users/Dario/Desktop/TFM/carto_tfm/img_m
odelos/MAXIMAS2")

  datos <- read.dbf(file = "MAXI-
MAS_FINAL_BIEN2.dbf")
  head(datos) # primeras filas de la tabla
  str(datos) # estructura de la tabla
  names(datos) # nombres de las columnas

  names(datos) # ver en qué columnas están las varia-
bles predictoras
  fijas=paste(names(datos)[45:52], collapse = " + ")
  var1=paste(names(datos)[j+21], "~", collapse = " + ")
  var2=paste(names(datos)[j+53], "+", collapse = " + ")
  var3=paste(names(datos)[j+77], "+", collapse = " + ")
  todas=paste(var1,var2,var3,fijas)
  todas
  # Folds <- nrow(datos)
  # datos$skfold <- seq(1,nrow(datos))

  Folds <- nrow(datos)
  Folds1=unique(datos$Codigo)

  for (i in 1:Folds)
  {
    Test <- subset(datos, Codigo == Folds1[i])
    Entrenamiento <- subset(datos, !Codigo ==
Folds1[i])
    Modelo <- step(lm(todas, family = gaussian, data =
Entrenamiento))
    Modelo <- modelTrim(Modelo)
    # Modelo <- rpart(todas, data = Entrenamiento)
    Prediccion <- predict(Modelo, Test)
    prediccion1[i,1]=Prediccion
    prediccion1[i,2]=Test[1,j+21]
    Test
  }
  rmse <- sqrt(mean((prediccion1$V1 - predic-
cion1$V2)^2))
  rmse

  setwd("C:/Users/Dario/Desktop/TFM/carto_tfm/img_m
odelos/MAXIMAS2/RMSE")
  destino=paste("Maximas_2006_RMSE_sat.txt", sep =
"")
  write.table(rmse,destino, sep = "\n \n", quote =
FALSE,row.names = FALSE,col.names = FALSE,
append =TRUE)
}
```