



Universidad de Valladolid

FACULTAD DE CIENCIAS
DEPARTAMENTO DE FÍSICA APLICADA

TESIS DOCTORAL

Corrección atmosférica de imágenes de satélite por métodos de aprendizaje automático

Presentada por JAVIER RODRÍGUEZ MARTÍN para optar al grado de
Doctor por la Universidad de Valladolid

Dirigida por:
José-Luis Casanova Roque y María Julia Sanz Justo

D. JOSÉ-LUIS CASANOVA ROQUE, Catedrático de la Universidad de Valladolid y
Dña. MARIA JULIA SANZ JUSTO, Profesor Titular de la Universidad de Valladolid:

CERTIFICAN: Que la presente memoria titulada “**Corrección atmosférica de imágenes de satélite por métodos de aprendizaje automático**” ha sido realizada bajo nuestra dirección por **D. JAVIER RODRIGUEZ MARTIN**

Y PARA QUE CONSTE y en cumplimiento de la legislación vigente, lo firmamos en Valladolid a 15 de Septiembre de 2014

Fdo: José-Luis Casanova Roque

Fdo: Julia Sanz Justo

*Dedicado a mi familia,
en especial, a la pequeña Elisa.*

AGRADECIMIENTOS

Seguramente, nunca hubiera llegado a terminar este trabajo, ni quizá ninguno de los proyectos que emprendí, sin el apoyo incondicional de mi familia y amigos, por tanto, el primer y más importante de los agradecimientos es para ellos. En particular, doy las gracias a mis padres por su enorme paciencia y todo el esfuerzo realizado, y a Mariana, por estar siempre conmigo, escuchándome y animándome, en todo momento.

Quiero dar las gracias a los directores de esta tesis, el Prof. José-Luis Casanova y la Dra. Julia Sanz, por aceptarme en el laboratorio dándome la oportunidad de realizar este trabajo, por su apoyo durante estos años y por sus orientaciones. De igual forma, agradezco a mis compañeros de laboratorio su amistosa acogida y los muchos buenos momentos que hemos pasado juntos. De forma especial, agradezco al Dr. Pablo Salvador su permanente disposición para ayudarme a resolver las dificultades que han ido surgiendo.

Me gustaría agradecer al Prof. Holger Kantz, del Instituto Max Planck (MPIPKS), la gran oportunidad que me brindó aceptándome en su grupo los 6 meses de mi 'estancia breve'.

Por último, agradezco al programa FPI del Ministerio de Ciencia e Innovación la financiación de este trabajo.

Contenidos

Lista de Figuras	XIII
Lista de Tablas	XV
Resumen	XVII
Glosario de símbolos	XVIII
1. Introducción	1
2. Modelo meteorológico	7
2.1. Introducción	7
2.2. Componentes principales del modelo WRF	8
2.2.1. El núcleo dinámico ARW	9
2.2.2. WPS: pre-procesamiento	10
2.2.3. Funciones del WPS	10
2.2.3.1. Geogrid	11
2.2.3.2. Ungrib	11
2.2.3.3. Metgrid	11
2.3. Ecuaciones	11
2.3.1. Coordenada vertical y condición de borde inferior WRF-ARW.	12
2.3.2. Ecuaciones de Euler en forma de flujo	13
2.3.3. Humedad	13
2.3.4. Ecuaciones incluyendo el factor de mapa y los términos rotacionales	15
2.4. Discretización	16
2.4.1. Temporal	17
2.4.2. Método de integración en el tiempo: RK3	17
2.4.2.1. Paso de tiempo dividido (“Split time step”).	18
2.4.3. Discretización espacial: Retícula C de Arakawa.	18
2.4.3.1. Filtros incluidos en el paso de tiempo	19

2.4.3.2.	Interacciones físicas	19
3.	Modelo de gases	21
3.1.	Introducción	21
3.2.	Características del modelo WRF/Chem	22
3.2.1.	Transporte	23
3.2.2.	Deposición seca	23
3.2.3.	Química de la fase gaseosa	23
3.2.4.	Emisiones	24
3.2.5.	Parametrización de aerosoles	24
3.2.5.1.	Distribución de tamaños	25
3.2.5.2.	Nucleación, condensación y coagulación	25
3.2.5.3.	Química de aerosoles	25
3.2.5.4.	Frecuencias de fotólisis	26
3.2.6.	GOCART	26
3.2.6.1.	Transporte	26
3.2.6.2.	Emisiones de sulfuros	27
3.2.6.3.	Emisiones de polvo	27
3.2.6.4.	Emisiones de OC y BC	27
3.2.6.5.	Emisiones de sales marinas	28
3.2.6.6.	Química	28
3.2.6.7.	Deposición seca	28
3.2.6.8.	Absorción de humedad	29
3.2.6.9.	Propiedades ópticas	29
3.2.7.	PREP-CHEM-SRC	30
3.2.8.	Bases de datos de emisiones antropogénicas globales	30
3.3.	Configuración utilizada	31
4.	Asimilación de datos	33
4.1.	Introducción	33
4.2.	Métodos de asimilación	38
4.2.0.1.	Interpolación óptima	39
4.2.1.	Asimilación de datos secuencial	40
4.2.1.1.	Filtro de Kalman estándar y extendido (EKF)	40
4.2.1.2.	EnKF	42
4.2.2.	Asimilación de datos variacional	42
4.2.2.1.	3DVAR	42
4.2.2.2.	4DVAR	45
4.2.3.	Ejemplo: asimilación de datos en el modelo de Lorenz	46
4.3.	Asimilación de datos en el modelo WRF	48
4.3.1.	Metodología	49
4.3.2.	Sensores ATOVS utilizados	49
4.3.2.1.	HIRS/3, HIRS/4	49
4.3.2.2.	AMSU-A, AMSU-B	50

5. Aprendizaje automático	53
5.1. Introducción	53
5.2. Aprendizaje supervisado	54
5.2.1. Regresión	54
5.2.1.1. Regresión lineal	54
5.2.1.2. Ridge	54
5.2.1.3. Lasso	55
5.2.1.4. Bayesian Ridge	55
5.2.1.5. SVR (Support Vector Regression)	57
5.2.1.6. Red neuronal	57
5.2.2. Clasificación	62
5.2.2.1. Vecinos más próximos	62
5.2.2.2. Análisis discriminante lineal	62
5.2.2.3. Árboles de decisión	63
5.2.2.4. Random Forests	64
5.2.2.5. SVM (Support vector machine)	65
5.2.2.6. Redes neuronales	67
5.3. Validación	68
5.3.1. Teoría del aprendizaje	68
5.3.1.1. Caso de \mathcal{H} finito	70
5.3.1.2. Caso \mathcal{H} infinito.	72
5.3.2. Métodos indirectos	73
5.3.3. Métodos directos: Cross validation	74
5.3.3.1. K-fold cross validation	74
5.3.3.2. Leave one out cross validation.	74
5.4. Selección del modelo.	74
5.4.1. Selección de parámetros de ajuste	75
5.4.2. Selección de características.	75
5.4.2.1. Búsqueda secuencial	75
5.4.3. Filtrado de características	75
5.4.3.1. Información y probabilidad.	76
5.4.3.2. Entropía	78
5.4.3.3. Información Mutua	79
5.4.3.4. Divergencia de Kullback-Leibler	79
5.5. Compensación bias/varianza y regularización	80
5.6. Curvas de aprendizaje	80
6. Corrección atmosférica	81
6.1. Introducción	81
6.1.1. Datos de entrada en el modelo 6S	82
6.2. Datos de entrenamiento	83
6.2.1. Parámetros tomados del WRF	84
6.2.1.1. Altitud en el modelo WRF	85

6.2.1.2.	Presión en el modelo WRF	85
6.2.1.3.	Temperatura en el modelo WRF	85
6.2.1.4.	Vapor de agua (g/m^3)	86
6.2.1.5.	Densidad de O_3 (g/m^3)	87
6.2.1.6.	Espesor óptico de aerosoles	87
6.2.2.	Parámetros tomados de la imagen	87
6.2.3.	Variables	87
6.3.	Selección del modelo: cross validation	87
6.3.1.	Algoritmos de regresión comparados	88
6.4.	Selección de características	90
6.4.1.	Forward Stepwise Selection	90
6.4.2.	Backward Stepwise Selection	91
6.4.3.	Correlación	91
6.4.4.	Información mutua e información mutua conjunta	92
6.4.5.	Métodos shrinkage	93
6.5.	Parámetros de ajuste	93
6.6.	Selección del modelo	94
6.7.	Curvas de aprendizaje	96
6.8.	Resultados	98
6.8.1.	Error cometido	98
6.8.2.	Tiempo de cálculo	98
6.8.3.	Impacto sobre las radiancias en cada banda	99
6.8.4.	Impacto sobre NDVI	101
6.9.	Algoritmos de clasificación comparados	103
6.9.1.	Variables más influyentes	104
Conclusiones		109
A. Imágenes Landsat		111
B. Modelo de transferencia radiativa 6S		117
C. Ficheros de configuración		121
C.1.	WRF	121
C.2.	WRFDA	122
C.3.	PREP-CHEM-SRC	124
D. Resultados intermedios		125
Bibliografía		127

Índice de figuras

2.1. Componentes del sistema WRF. Imagen tomada de Skamarock et al. (2005)	9
2.2. Flujo del WPS. Gráfico tomado de la guía de usuario	10
2.3. Coordenada η . Gráfico tomado de Skamarock et al. (2005)	12
3.1. Perfiles de τ , g y ω_0	30
4.1. Evolución real y de la predicción de la componente X sin asimilación	47
4.2. Evolución real y de la predicción con asimilación EKF	47
4.3. Evolución real y de la predicción con asimilación EnKF	48
4.4. Perfil de temperaturas medido por el sensor HIRS	50
4.5. Perfil de vapor de agua medido por el sensor AMSU	51
4.6. Sensor cross-track, IFOV, SWATH	51
5.1. Neurona	58
5.2. (a) Red neuronal multicapa, (b) Red recurrente, (c) Red competitiva.	59
5.3. Red neuronal multicapa	60
5.4. Red neuronal multicapa con k salidas	67
6.1. Salida del modelo 6S	83
6.2. Comparación preliminar de los modelos de regresión	89
6.3. Evolución del error de validación con el número de variables	91
6.4. Errores de validación y entrenamiento para cada algoritmo en cada banda	96
6.5. Curvas de aprendizaje en cada banda	98
6.6. Histogramas de correcciones en cada banda	100
6.7. Diferencias de NDVI corregido y sin corregir	101
6.8. NDVI corregido arriba, sin corrección abajo	102
6.9. Comparativa de modelos de regresión	104
6.10. Predictores más importantes en cada banda	105
6.11. Variables más relevantes en LR	107
6.12. Variables más relevantes en clasificación	108
A.1. Bandas de los sensores OLI y TIRS en Landsat 8 y ETM en Landsat 7. Figura tomada de la web landsat.gsfc.nasa.gov	112

A.2. Respuesta espectral del sensor OLI. Figura tomada de la web landsat.gsfc.nasa.gov	113
A.3. Composición RGB de imagen Lansat 8 corregida.	114
A.4. Composición RGB de imagen Lansat 8 sin corrección.	115

Índice de tablas

1.1. Resumen de técnicas de corrección atmosférica	3
6.1. Variables	88
6.2. Variables más relevantes según el algoritmo forward stepwise en LR	90
6.3. Correlación lineal entre predictores y reflectancia	92
6.4. MI between variables and target	93
6.5. JMI Información mutua conjunta con $\lambda = 0,05$	94
6.6. Variables más relevantes en LR y clasificación	106
A.1. Bandas Landsat	112
C.1. Opciones parametrización física	121
C.2. Opciones parametrización física	122
C.3. Opciones de asimilación de datos	123
C.4. Opciones de emisiones	124

RESUMEN

El objetivo de la corrección atmosférica es determinar el valor real de reflectancia de la superficie de la Tierra, a partir de la reflectancia aparente medida por el sensor. Constituye la parte más importante del pre-procesamiento de datos obtenidos mediante teledetección por satélite. Los procedimientos clásicos de corrección utilizan modelos de transferencia radiativa que suponen el estado atmosférico estático, sin atender a sus condiciones cambiantes, tanto espaciales como temporales. De esta manera, el coste computacional que se ahorra es muy significativo pero, dicha simplificación, difiere notablemente de la situación real. Suponiendo conocidos los campos dinámicos que determinan el estado de la atmósfera, la metodología clásica de corrección atmosférica mediante estos modelos, involucraría la ejecución de los mismos en cada punto, resultando inabordable debido al coste computacional que conllevaría. En este trabajo se propone una alternativa que considera la dependencia espacial y temporal de las condiciones atmosféricas sustituyendo, parcialmente, el uso del modelo de transferencia por el algoritmo de aprendizaje automático más adecuado. De esa manera, conseguimos reducir el tiempo de cálculo sustancialmente. En esta tesis evaluamos la habilidad de seis algoritmos de aprendizaje automático diferentes para resolver el problema. Paralelamente, estudiamos los conjuntos de variables más adecuados, para la resolución del problema en cada banda de la imagen. Finalmente, tras haber elegido el método más adecuado para sustituir el modelo de transferencia radiativa, y haciendo uso del subconjunto de variables seleccionado, llevamos a cabo la corrección de la imagen en cada una de las bandas.

GLOSARIO DE SÍMBOLOS

AMSU	Advanced Microwave Sounding Unit
AOT	Aerosol Optical Thickness
ATOVS	Advanced TIROS Operational Vertical Sounder
BC	Black Carbon
BRR	Bayesian Ridge Regression
EKF	Extended Kalman filter
EnKF	Ensemble Kalman filter
FOV	Field Of View
GOCART	Goddard Chemistry Aerosol Radiation and Transport
HIRS	High Resolution Infrared Radiation Sounder
IFOV	Instantaneous Field Of View
JMI	Joint Mutual Information
KF	Kalman Filter
LR	Linear Regression
MI	Mutual Information
ML	Machine Learning
NDVI	Normalized Difference Vegetation Index
OC	Organic Carbon
ROC	Receiver Operating Characteristic
RTM	Radiative Transfer Model
SVM	Support Vector Machine
SVR	Support Vector Regression
TOA	Top Of Atmosphere
WRF	Weather Research and Forecasting

Capítulo 1

Introducción

Se denomina teledetección, en el contexto de este trabajo, al proceso de medida de propiedades de objetos que se encuentran sobre la superficie terrestre mediante sensores aerotransportados, ya sean aviones, satélites u otros sistemas. En definitiva, nos referimos a medidas tomadas desde cierta distancia en lugar de *in situ*, es decir, realizadas sin contacto directo con el objeto en cuestión. Dichas medidas pueden ser puntuales, refiriéndose a un punto en concreto de la superficie o al perfil de cierta magnitud en una localización concreta; o pueden tomarse sobre una rejilla de dos dimensiones, dando lugar a imágenes. Esta definición, tomada de Schowengerdt (2006), no es la única que hemos encontrado revisando la literatura, en Campbell (2002) se recopilaron algunas más y a continuación presentamos otras:

La teledetección se define como la adquisición de información acerca de un objeto sin estar en contacto físico con él. La información se adquiere mediante la detección y medición de los cambios que el objeto produce en el campo circundante, ya sea electromagnético, acústico, o potencial. Esto podría incluir el campo electromagnético emitido o reflejado por el objeto, las ondas acústicas reflejadas o perturbadas por el objeto, o las perturbaciones, en los alrededores, del campo gravitatorio o magnético debido a la presencia del objeto. (Elachi and Van Zyl, 2006)

La teledetección es la ciencia y arte de obtener información sobre un objeto, superficie o fenómeno a través del análisis de los datos adquiridos por un dispositivo que no está en contacto con el objeto, superficie o fenómeno bajo investigación. Definición tomada de Lillesand et al. (2004), Yang (2011) y Colwell (1997).

En el marco de la teledetección desde satélite, la adquisición de datos se realiza de forma periódica, permitiendo la monitorización, a corto y largo plazo, de los cambios medioambientales y del impacto de la actividad humana. Algunas aplicaciones interesantes de esta tecnología son: la evaluación de cambios globales como la deforestación, la desaparición de la capa de ozono o el calentamiento global; la monitorización y gestión de recursos naturales; la vigilancia y predicción de desastres como incendios, erupciones volcánicas, inundaciones, etc.; en agricultura, el seguimiento de cultivos, la estimación de producción o la erosión del suelo; en meteorología, el estudio de la dinámica atmosférica y la predicción; etcétera.

Para llevar a cabo este tipo de tareas, se desarrollan sistemas que ofrecen diferentes posibilidades de muestreo de acuerdo a los parámetros espaciales, temporales o espectrales, según las necesidades de cada una.

El estudio de las imágenes capturadas por los sensores puede centrarse en la fotointerpretación de las mismas o en el análisis de los datos. En la interpretación de la escena, el objetivo se focaliza en el reconocimiento de objetos mediante la identificación de las características que lo distinguen de su entorno. En este ámbito, tradicionalmente, resultaba imprescindible la participación de un experto. Actualmente, el reconocimiento automático de objetos en las imágenes, constituye un campo de constante evolución que ha sufrido un desarrollo muy importante en los últimos años. Desde el punto de vista del análisis de los datos capturados por los sensores, más que relación de los objetos con su entorno, se valora la calidad de la medida realizada. Por esta razón, los procedimientos de calibración radiométrica y de corrección de los efectos atmosféricos cobran una gran importancia, especialmente cuando se estudian series temporales largas. En esos casos, es imprescindible que las medidas sean consistentes independientemente de la variación del estado de la atmósfera, la degradación del sensor o, incluso, si se realizan mediante aparatos distintos.

En este trabajo, se aborda el problema de la corrección del efecto de la atmósfera teniendo en cuenta la variación espacial y temporal de su estado.

Corrección atmosférica

El objetivo de la corrección atmosférica es determinar el valor real de reflectancia de la superficie de la Tierra, a partir de la reflectancia aparente medida por el sensor (Du et al., 2009). Constituye la parte más importante del pre-procesamiento de datos obtenidos mediante teledetección por satélite, (Hadjimitsis et al., 2010).

En la tabla 1.1 tomada de Schowengerdt (2006), presentamos un resumen de las técnicas de corrección utilizadas. Algunas de ellas se basan únicamente en los datos de la imagen sin atender al estado atmosférico. Otras hacen uso de modelos de transferencia radiativa, como en nuestro caso. Los procedimientos clásicos de corrección mediante modelos de transferencia, trabajan con las denominadas atmósferas estándar sin atender a las condiciones

Sensor	Técnica	Referencia
MSS	Regresion banda a banda	(Potter and Mendlowitz, 1975)
MSS	Covarianza espectral de todas las bandas	(Switzer et al., 1981)
airborne MSS	Regresion banda a banda	(Potter, 1984)
AVHRR	Estimación iterativa	(Singh and Cracknell, 1986)
MSS, TM	DOS (Dark object subtraction) más modelo de difusión	(Chavez Jr, 1988)
TM	DOS más modelo de difusión, medidas de penetración de radiación en la atmósfera	(Chavez Jr, 1989)
TM	Parámetro 'tasseled cap haze' en cada pixel	(Lavreau, 1991)
AVHRR	DOS, NDVI, AVHRR band 3	(Holben et al., 1992)
airborne TMS	medidas de radiación mediante sensores en tierra y aerotransportados, código de simulación atómsférica Landsat TM	(Wrigley et al., 1992)
TM	comparación de diez métodos DOS y simulación atmosférica más datos de campo	(Moran et al., 1992)
TM	dark target, código de simulación	(Teillet and Fedosejevs, 1995)
TM	código de simulación atmosférica, igualado de histogramas	(Richter, 1996b); (Richter, 1996a)
TM	DOS con estimación de transmitancia	(Chavez Jr, 1996)
TM	dark target, código de simulación atómsférica	(Ouaidrari and Vermote, 1999)
TM, ETM+	método EL (empirical line), medidas en tierra	(Moran et al., 2001)
TM	comparación de 7 métodos en 12 fechas	(Hadjimitsis et al., 2004)
AVHRR	2 bandas PCT usadas para separar las componentes de aerosoles	(Salama et al., 2004)

TABLA 1.1: Resumen de técnicas de corrección atmosférica

cambiantes de esta. De esta manera, el coste computacional que se ahorra es muy significativo pero, dicha simplificación, difiere notablemente de la situación real. Entendemos que, en caso de tener una resolución espacial baja, el argumento es aún válido; el error cometido al despreciar tales efectos, es sustancialmente menor que el debido a la impureza de los píxeles, la mínima entidad que podemos observar. Sin embargo, en el caso de la alta resolución espacial, la situación es radicalmente diferente. Por ejemplo, la medida del NDVI (índice de vegetación normalizado) de píxeles puros en las proximidades de una zona urbana, en la costa o en una zona rural, alejada de ambas, no debe realizarse sin tener en cuenta el efecto de tales condiciones. Obviamente, medir empíricamente el estado de la atmósfera no

es posible ya que necesitaríamos conocer el valor de las variables que lo definen en cada punto del espacio. La opción factible es trabajar sobre una rejilla, luego, en nuestro problema, necesitamos especificar en cada punto de esa rejilla el valor de los campos que determinan la dinámica así como las concentraciones de los componentes que entran en juego en el modelo de transferencia radiativa. Para ello, haremos uso de los modelos WRF y WRF/Chem. Con el fin de optimizar la aproximación que supone, tomaremos el análisis obtenido mediante la asimilación de datos de radiancia.

Asimilación de datos

El objetivo de un esquema de asimilación de datos es usar las observaciones junto con los resultados del modelo del sistema dinámico con el fin de obtener estimaciones más precisas del estado actual y futuro del sistema. Los distintos métodos conocidos operan de forma secuencial o variacional, como veremos en el capítulo 4. Trataremos de minimizar la varianza de las diferencias entre las medidas y los resultados del modelo. Para ello, se introduce el operador de medida que nos lleva del espacio del modelo (rejilla) a los puntos dónde se realizan las observaciones. De forma similar, algunos esquemas solo aceptarán medidas en los instantes de tiempo de la rejilla temporal o en puntos intermedios.

Suponiendo conocidos los campos que determinan el estado de la atmósfera, la metodología clásica de corrección atmosférica mediante el modelo 6S (Vermote et al., 1997a), involucraría la ejecución del mismo en cada punto, resultando inabordable debido al coste computacional que eso conllevaría. El tiempo que necesita este RTM seleccionado para calcular la reflectancia de un punto es de aproximadamente segundos en promedio. Es el único que considera la influencia del entorno en los cálculos, razón por la que se utilizó dicho modelo y no otro. En este trabajo se propone una alternativa que considera la dependencia espacial de las condiciones atmosféricas sustituyendo, parcialmente, el uso del modelo de transferencia por el algoritmo de aprendizaje automático más adecuado. De esta manera, conseguimos reducir el tiempo de cálculo sustancialmente pasando a ser del orden $20,6\mu s$ por píxel pero sin aplicar cálculo matricial.

Bajo el nombre de aprendizaje automático (machine learning) o aprendizaje estadístico se engloba un conjunto de técnicas estadísticas destinadas a generalizar las respuestas de un sistema a partir de una información suministrada en forma de ejemplos. Actualmente, las aplicaciones de esta tecnología son muy numerosas y van en aumento. Como ejemplo, citamos una lista amplia con algunas de las más conocidas: visión por computador, reconocimiento de objetos, procesamiento del lenguaje natural, reconocimiento de patrones sintácticos, motores de búsqueda, diagnóstico médico, bioinformática, interfaces cerebro-máquina, quimioinformática, detección de fraude en operaciones bancarias, análisis bursátil, clasificación de secuencias de ADN, minería de datos, reconocimiento del habla y escritura, ingeniería

de software, locomoción de robots, publicidad, finanzas, vigilancia del estado estructural de edificios, recuperación de la información, sistemas de recomendación, etcétera.

Para poder llevar a cabo la corrección atmosférica mediante estas herramientas, necesitamos el conjunto de entrenamiento que nos permita aprender los parámetros del nuevo modelo y sustituir, así, el uso del RTM en cada punto por el método de aprendizaje automático. Lo construimos de la siguiente manera: sobre la imagen, seleccionamos 3000 puntos distribuidos de manera uniforme. Realizamos la corrección atmosférica vía 6S conservando tanto los datos de entrada como los resultados. Sobre ese conjunto, entrenaremos nuestros métodos de regresión. Dividimos dicho conjunto en tres partes: el conjunto de entrenamiento propiamente dicho, conteniendo el 60 % de los datos, servirá para determinar los parámetros del modelo; el conjunto de validación, con el 20 % de los datos restantes, nos servirá para estimar el error de generalización; el 20 % que queda, constituirá el conjunto de test, sin involucrarse en el proceso de selección del modelo.

Una vez obtenido el conjunto de entrenamiento, evaluamos la habilidad de seis algoritmos de aprendizaje automático diferentes para resolver el problema de regresión y otros seis para clasificar las atmósferas más influyentes. Para la regresión utilizamos: regresión lineal, regresión Lasso, Ridge, SVR, regresión Ridge Bayesiana y una red neuronal. En la clasificación, hicimos uso del análisis de discriminante lineal, vecinos más próximos, árboles de decisión, random forest, SVM y otra red neuronal. La selección del modelo se llevó a cabo mediante la estimación del error cometido y el área bajo la curva ROC sobre el conjunto de validación. Previamente, optimizamos cada uno de los algoritmos ajustando parámetros los móviles siguiendo el mismo procedimiento: construimos una rejilla para cada uno de ellos y calculamos el error de validación.

Paralelamente, en el paso anterior, estudiamos los conjuntos de variables más adecuados, tanto para realizar la regresión, como la clasificación. Evaluamos la capacidad predictora de los conjuntos de variables utilizando cada uno de los métodos y tomamos el que producía menor error sobre los datos del conjunto de validación.

Finalmente, tras haber elegido el método más adecuado para sustituir el modelo de transferencia radiativa, y haciendo uso del subconjunto de variables seleccionado, llevamos a cabo la corrección de la imagen en cada una de las bandas.

Los errores cometidos en la corrección rondaron el 0.5 % en todas las bandas. Fuimos capaces de reducir el tiempo de cálculo hasta hacerlo operativo: pasamos de una media de 3 seg por operación en cada píxel (dependiendo de la banda) a una media de $0.039\mu s$, ya que tarda $117\mu s$ en corregir los 3000 píxeles del conjunto de entrenamiento, en promedio. Notemos que la vectorización de las operaciones reduce sustancialmente el tiempo de cálculo, pasando de $20,6\mu s$ por píxel aislado a $0,039\mu s$.

Evaluamos el impacto de la corrección atmosférica en la estimación del NDVI, concluyendo que, sin ella, se subestima entre 1 y 6 % de su valor, dependiendo del estado de la vegetación.

En el capítulo 2, presentamos el modelo numérico de predicción meteorológica WRF así como un breve repaso de las ecuaciones que integra, el método numérico utilizado y el proceso de discretización del sistema. En el capítulo 3, se introduce el módulo químico WRF-Chem y las bases de datos de emisiones que se utilizaron para definir las concentraciones de los componentes atmosféricos. Al final del mismo, se incluye una pequeña descripción de la herramienta que utilizamos para adaptar dichas emisiones a la rejilla del modelo. En el capítulo 4, exponemos algunos métodos de asimilación de datos así como los conceptos fundamentales involucrados. También incluimos un par de ejemplos de aplicación de alguno de los métodos en el modelo de Lorenz. Sucintamente, presentamos los sensores de radiancia (ATOVS) que utilizamos en la asimilación. En el capítulo 5, presentamos los algoritmos de aprendizaje automático que fueron comparados en la tesis; la teoría de validación; algunos métodos de selección de variables y selección del modelo; y una herramienta de diagnóstico de la descompensación entre el bias/varianza. En el capítulo 6, describimos la metodología llevada a cabo durante el desarrollo de la tesis. En la primera parte, expusimos cómo se calculan las variables de entrada del modelo de transferencia a partir de las salidas de los modelos meteorológicos y de composición química. Posteriormente, desarrollamos el procedimiento de selección del modelo y de las variables más adecuadas. Finalmente, presentamos los resultados. El capítulo 7 recoge las conclusiones del trabajo. Hemos incluido cuatro Apéndices que muestran las imágenes Landsat utilizadas, la descripción del modelo 6S, los conjuntos de parámetros de configuración y una serie de tablas a modo de ilustración del procedimiento y los resultados.

Capítulo 2

Modelo meteorológico

2.1. Introducción

El modelo WRF (Weather Research and Forecasting) (Skamarock et al., 2005), es un modelo numérico de predicción meteorológica y simulación de la dinámica atmosférica de escala regional, diseñado tanto para fines científicos como para aplicaciones operacionales.

Comenzó a desarrollarse a finales de la década de los años 90 por diferentes agencias, principalmente estadounidenses, que compartían el propósito de construir la siguiente generación del modelo mesoescalar MM5. Algunas de las más importantes son:

- NCAR National Center for Atmospheric Research
- MMM Mesoscale and Microscale Meteorology
- NOAA National Oceanic and Atmospheric Administration
- NCEP National Centers for Environmental Prediction
- ESRL Earth System Research Laboratory
- AFWA Air Force Weather Agency
- NRL Naval Research Laboratory
- CAPS Center for Analysis and Prediction of Storms
- FAA Federal Aviation Administration

- Diferentes universidades de Estados Unidos

No obstante, una gran cantidad de usuarios e investigadores de todo el mundo colaboran en la implementación de nuevas parametrizaciones de fenómenos físicos y otras mejoras. El modelo WRF es de código abierto, portable, es decir puede ser instalado y ejecutado en una gran variedad de plataformas y puede ejecutarse en paralelo. Entre las aplicaciones de este modelo se encuentran:

- Investigación en el desarrollo de nuevas parametrizaciones (gracias a que el código permite incorporación de nuevos esquemas de forma sencilla).
- Investigación y desarrollo en la asimilación de datos.
- Investigación y desarrollo en el pronóstico meteorológico.
- Implementación operativa para el pronóstico en diferentes escalas (la nueva versión incluye la posibilidad de realizar modelado global de la atmósfera hasta pronósticos con convección explícita)
- Educación

Está compuesto por dos núcleos dinámicos que se encargan de resolver las ecuaciones, el ARW (Advanced Research WRF), gestionado por NCAR, y el NMM (Nonhydrostatic Mesoscale Model) dirigido por NCEP.

2.2. Componentes principales del modelo WRF

- WPS: Sistema de pre-procesamiento
- WRF (Núcleos dinámicos ARW y NMM)
- Programas de preparación de los datos reales e idealizados (real.exe/ideal.exe)
- Programa de integración numérica (wrf.exe)
- WRFDA módulo de asimilación de datos (opcional)
- WRF-Chem módulo químico (opcional)
- WRF-Fire - modelización de propagación de incendios

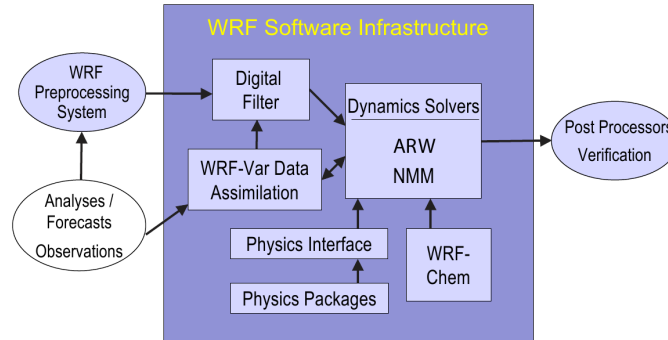


FIGURA 2.1: Componentes del sistema WRF.
Imagen tomada de Skamarock et al. (2005)

2.2.1. El núcleo dinámico ARW

A continuación se describen las principales características del ARW

- Coordenada vertical adaptada al terreno η , representando la presión hidrostática del aire seco. Presenta un valor constante en el techo de la atmósfera.
- Ecuaciones compresibles, no hidrostáticas e hidrostáticas, conservativas para las variables escalares
- Variables: velocidades $u, v, w = \dot{\eta}$, potencial de temperatura θ , geopotencial ϕ, p y $\alpha = 1/\rho$
- Rejilla horizontal: Arakawa tipo C
- Integración en tiempo: Runge-Kutta de segundo y tercer orden, con paso temporal menor para los modos acústicos y gravitacionales. Puede modificarse el paso temporal.
- Resolución de los términos advectivos
- Filtros
- Condiciones de borde
- Representación del terreno
- Proyecciones
- Anidamientos

Todos estos aspectos forman parte del núcleo dinámico del modelo (en contraposición al conjunto de parametrizaciones a las que generalmente se denomina como la física del modelo).

2.2.2. WPS: pre-procesamiento

El cometido del sistema de pre-procesamiento es preparar la entrada del modelo WRF para realizar las simulaciones.

2.2.3. Funciones del WPS

- Define el dominio principal, dónde se llevará a cabo la simulación, y los dominios anidados.
- Interpola los parámetros invariantes en el tiempo que definen el terreno como el uso de suelo, altura del terreno, etc. sobre las rejillas de simulación. (Campos estáticos)
- Decodifica y da formato adecuado a los ficheros en formato GRIB y GRIB2 que contienen los datos meteorológicos de entrada (u, v, T, q, presión superficial, datos de suelo, datos de nieve, temperatura de la superficie del mar, etc.)
- Interpola horizontalmente los campos meteorológicos variables en el tiempo a la rejilla del modelo WRF. Normalmente, estos campos proceden de modelos globales.

En definitiva, es un conjunto de programas cuyo objetivo es generar las condiciones iniciales para el modelo WRF.

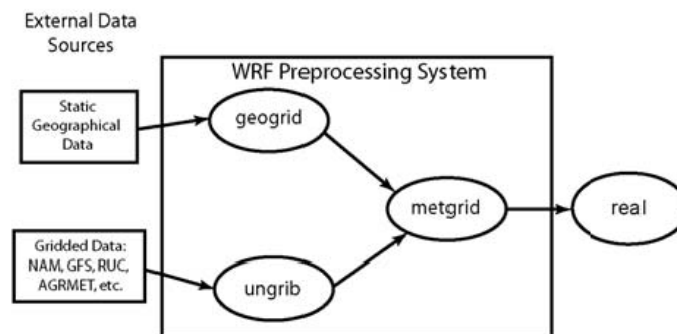


FIGURA 2.2: Flujo del WPS. Gráfico tomado de la guía de usuario

2.2.3.1. Geogrid

Este programa calcula la posición de los puntos de retícula de acuerdo a las especificaciones de resolución, número de puntos, posición del centro, etc. provistas por el usuario. Una vez posicionados los puntos de la retícula, interpola los datos de topografía, uso de suelo, tipo de suelo, albedo, cobertura vegetal, máscara tierra-agua y el resto de datos que en el futuro puedan ser utilizados por el modelo.

2.2.3.2. Ungrib

Permite decodificar datos que vienen en formato grib y escribirlos en el formato que necesita el programa metgrid.exe para que puedan ser interpolados a la retícula del modelo WRF. Extrae los campos meteorológicos y, si es necesario deriva otros relacionados.

2.2.3.3. Metgrid

Este programa interpola horizontalmente los datos de atmósfera extraídos por el programa ungrid.exe a la retícula definida por el programa geogrid.exe. Por ejemplo puede tomar un análisis y un pronóstico global e interarlo a la retícula del modelo.

2.3. Ecuaciones

Las ecuaciones que gobiernan la dinámica atmosférica forman un conjunto completo de siete ecuaciones con siete incógnitas (Bjerknes, 1904).

- La segunda ley de Newton o conservación del momento;
- La ecuación de conservación de la masa;
- La ecuación de estado para los gases ideales;
- La primera ley de la termodinámica o conservación de la energía
- La conservación de la masa de agua.

En el modelo WRF, las ecuaciones se escriben en forma de flujo usando una coordenada adaptada al terreno. En los siguientes apartados, presentamos estas ecuaciones así como el método de integración utilizado.

2.3.1. Coordenada vertical y condición de borde inferior WRF-ARW.

Las ecuaciones que resuelve el núcleo dinámico ARW se escriben utilizando una coordenada de presión hidrostática vertical adaptada al terreno y denotada por η

$$\eta = (p_h - p_{ht})/\mu \quad (2.1)$$

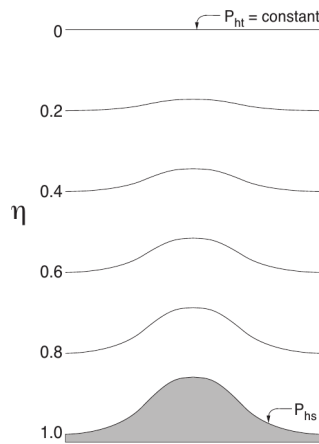


FIGURA 2.3: Coordenada η .
Gráfico tomado de Skamarock et al. (2005)

donde $\mu = (p_{hs} - p_{ht})$, p_h es la componente hidrostática de la presión; p_{hs} y p_{ht} son los valores en la superficie y en la frontera superior. La definición de la coordenada fue propuesta por Laprise (1992) y también es denominada como coordenada de masa vertical.

Como $\mu(x, y)$ representa la masa por unidad de área dentro de la columna en (x, y) , las variables del modelo se acoplan a la masa (para poder escribir las ecuaciones en forma de flujo). Las letras mayúsculas designan a las variables “acopladas” con la masa.

$$V = \mu \cdot v = (U, V, W), \quad \Omega = \mu \dot{\eta}, \quad \Theta = \mu \theta \quad (2.2)$$

donde $\mathbf{v} = (u, v, w)$ es la velocidad covariante, $w = \dot{\eta}$ es la velocidad contravariante vertical y θ , la temperatura potencial.

2.3.2. Ecuaciones de Euler en forma de flujo

Haciendo uso de las expresiones anteriores, el sistema de ecuaciones sin humedad, escrito en forma de flujo es

$$F_U = \partial_t U + (\nabla \cdot Vu) - \partial_x(p\partial_\eta\phi) + \partial_\eta(p\partial_x\phi) \quad (2.3)$$

$$F_V = \partial_t V + (\nabla \cdot Vv) - \partial_y(p\partial_\eta\phi) + \partial_\eta(p\partial_y\phi) \quad (2.4)$$

$$F_W = \partial_t W + (\nabla \cdot Vw) - g(\partial_\eta p - \mu) \quad (2.5)$$

$$F_\Theta = \partial_t \Theta + (\nabla \cdot V\Theta) \quad (2.6)$$

$$0 = \partial_t \mu + (\nabla \cdot V) \quad (2.7)$$

$$0 = \partial_t \phi + \mu^{-1}[(V \cdot \nabla \phi) - gW] \quad (2.8)$$

donde p es la presión, $\phi = gz$ el geopotencial y $\alpha = 1/\rho$, el inverso de la densidad.

La relación de balance hidrostático, viene dada por la expresión

$$\partial_\eta \phi = -\alpha\mu, \quad (2.9)$$

y la ecuación de estado

$$p = p_0(R_d\theta/p_o\alpha)^\gamma. \quad (2.10)$$

$\gamma = c_p/c_v = 1.4$, es el ratio de capacidades caloríficas para el aire seco, R_d es la constante de los gases para el aire seco, p_0 la presión de referencia (típicamente 10^5 Pascales). F_U, F_V, F_W y F_Θ son los términos de fuerza que provienen del modelo físico, la mezcla turbulenta, las proyecciones esféricas y la rotación terrestre.

En las ecuaciones anteriores, se adoptó la siguiente notación

$$\nabla \cdot Va = \partial_x(Ua) + \partial_y(Va) + \partial_\eta(\Omega a)$$

$$V \cdot \nabla a = U\partial_x a + V\partial_y a + \Omega\partial_\eta a$$

siendo a una variable genérica.

2.3.3. Humedad

Cuando se incluye la humedad, la coordenada vertical se define con respecto a una atmósfera seca e hidrostática. Es decir, la presión que se utiliza para calcular el nivel sigma, es la presión

hidrostática calculada a partir del perfil de temperatura sin tener en cuenta la humedad.

$$\eta = (p_{dh} - p_{dht}/\mu_d) \quad (2.11)$$

donde μ_d representa la masa de aire seco en la columna y p_{dh} y p_{dht} representa la presión hidrostática en el techo de la atmósfera seca. Las variables acopladas se definen como

$$V = \mu_d \mathbf{v}, \quad \Omega = \mu_d \dot{\eta}, \quad \Theta = \mu_d \theta. \quad (2.12)$$

Con estas definiciones, las ecuaciones de Euler con humedad se escriben como

$$F_U = \partial_t U + (\nabla \cdot V u) - \mu_d \alpha \partial_x p + (\alpha/\alpha_d) \partial_\eta p \partial_x \phi \quad (2.13)$$

$$F_V = \partial_t V + (\nabla \cdot V v) - \mu_d \alpha \partial_y p + (\alpha/\alpha_d) \partial_\eta p \partial_y \phi \quad (2.14)$$

$$F_W = \partial_t W + (\nabla \cdot V w) - g[(\alpha/\alpha_d) \partial_\eta p - \mu_d] \quad (2.15)$$

$$F_\Theta = \partial_t \Theta + (\nabla \cdot V \Theta) \quad (2.16)$$

$$0 = \partial_t \mu_d + (\nabla \cdot V) \quad (2.17)$$

$$0 = \partial_t \phi + \mu_d^{-1} [(V \cdot \nabla \phi) - gW] \quad (2.18)$$

$$F_{Q_m} = \partial_t Q_m + (\nabla \cdot V Q_m) \quad (2.19)$$

con la relación de balance hidrostático para el inverso de la densidad de aire seco

$$\partial_\eta \phi = -\alpha_d \mu_d \quad (2.20)$$

y la ecuación de estado con la presión total (aire seco más vapor de agua)

$$p = p_0 (R_d \theta / p_0 \alpha_d)^\gamma. \quad (2.21)$$

α_d es el inverso de la densidad del aire seco ($1/\rho_r$) y α es el inverso de la densidad que incluye la presencia de agua en forma vapor y condensada

$$\alpha = \alpha_d (1 + q_v + q_c + q_r + q_i + \dots)^{-1} \quad (2.22)$$

donde q_* son ratios de mezcla (masa por masa de aire seco) de vapor de agua q_v , nube q_c , lluvia q_r , hielo q_i , etc. Además, $\theta_m = \theta(1 + (R_v/R_d)q_v) \equiv \theta(1+1,61q_v)$ y $Q_m = \mu_d m$; $q_m = q_v, q_c, q_r, q_i, \dots$

Notar que las variables se acoplan con la componente no hidrostática y seca de la presión.

2.3.4. Ecuaciones incluyendo el factor de mapa y los términos rotacionales

El núcleo dinámico ARW admite cuatro proyecciones: Lambertiana conforme, estereográfica polar, Mercator y latitud-longitud. Estas se describen en Haltiner and Williams (1980). La transformación es isótropa en las tres primeras, es decir, $(\Delta x, \Delta y)|_{Tierra}$ es constante en todos los puntos de la rejilla. El ARW precisa que Δx y Δy sean constantes, así que introduce los llamados factores de mapa m_x y m_y , para hacerlo aplicable al caso no isótropo,

$$(m_x, m_y) = \frac{(\Delta x, \Delta y)}{\text{distancia sobre la Tierra}} \quad (2.23)$$

De esa forma, las variables acopladas se redefinen como

$$U = \mu_d u / m_y, \quad V = \mu_d v / m_x, \quad W = \mu_d w / m_y, \quad \Omega = \mu_d \dot{\eta} / m_y. \quad (2.24)$$

Y, usando estas variables, las ecuaciones de la dinámica se escriben como

$$F_U = \partial_t U + m_x [\partial_x(Uu) + \partial_y(Vu)] + \partial_\eta(\Omega u) + \mu_d \alpha \partial_x p + (\alpha / \alpha_d) \partial_\eta p \partial_x \phi \quad (2.25)$$

$$F_V = \partial_t V + m_y [\partial_x(Uu) + \partial_y(Vu)] + (m_y / m_x) \partial_\eta(\Omega v) + \mu_d \alpha \partial_y p + (\alpha / \alpha_d) \partial_\eta p \partial_y \phi \quad (2.26)$$

$$F_W = \partial_t W + (m_x m_y / m_y) [\partial_x(Uu) + \partial_y(Vu)] + \partial_\eta(\Omega w) - m_y^{-1} g [(\alpha / \alpha_d) \partial_\eta p - \mu_d] \quad (2.27)$$

$$F_\Theta = \partial_t \Theta + m_x m_y [\partial_x(U\theta) + \partial_y(V\theta)] + m_y \partial_\eta(\Omega \theta) \quad (2.28)$$

$$0 = \partial_t \mu_d + m_x m_y [U_x + V_y] + m_y \partial_\eta(\Omega) \quad (2.29)$$

$$0 = \partial_t \phi + \mu_d^{-1} [(V \cdot \nabla \phi) - gW] \quad (2.30)$$

$$F_{Q_m} = \partial_t Q_m + m_x m_y [\partial_x(Uq_m) + \partial_y(Vq_m)] + m_y \partial_\eta(\Omega) q_m \quad (2.31)$$

con la relación de balance hidrostático

$$\partial_\eta \phi = -\alpha_d \mu_d, \quad (2.32)$$

y la ecuación de estado

$$p = p_0 (R_d \theta_m / p_0 \alpha_d)^\gamma. \quad (2.33)$$

En las ecuaciones 2.25, 2.26 y 2.27 los términos de curvatura y Coriolis están incluidos en los términos de fuerza. Para los términos isótropos, donde $m = m_x = m_y$, se escriben de la

siguiente manera:

$$F_{U_{cor}} = + \left(f + u \frac{\partial m}{\partial y} - v \frac{\partial m}{\partial x} \right) V - eW \cos \alpha_r - \frac{uW}{r_e} \quad (2.34)$$

$$F_{V_{cor}} = + \left(f + u \frac{\partial m}{\partial y} - v \frac{\partial m}{\partial x} \right) U - eW \sin \alpha_r - \frac{vW}{r_e} \quad (2.35)$$

$$F_{W_{cor}} = +e(U \cos \alpha_r - V \sin \alpha_r) + \frac{uV + vU}{r_e} \quad (2.36)$$

En el caso de proyección no isotropa, los términos de curvatura y de Coriolis, se escriben como

$$F_{U_{cor}} = + \frac{m_x}{m_y} \left[fV + \frac{uV}{r_e} \tan \varphi \right] V - eW \cos \alpha_r - \frac{uW}{r_e} \quad (2.37)$$

$$F_{V_{cor}} = + \frac{m_x}{m_y} \left[-fU + \frac{uU}{r_e} \tan \varphi + eW \sin \alpha_r - \frac{uW}{r_e} \right] \quad (2.38)$$

$$F_{W_{cor}} = +e \left(U \cos \alpha_r - \frac{m_x}{m_y} V \sin \alpha_r \right) + \frac{uV + (m_x m_y) vU}{r_e} \quad (2.39)$$

donde α_r es el ángulo de rotación local entre el eje y , y los meridianos, φ es la latitud, $f = 2\Omega_e \sin \varphi$, $e = 2\Omega_e \cos \varphi$, Ω_e es la velocidad angular de rotación de la Tierra, r_e el radio medio de la Tierra.

2.4. Discretización

Las ecuaciones diferenciales no lineales que describen la dinámica atmosférica no tienen solución analítica, aun estando el problema bien planteado, en términos de la existencia y unicidad de la solución. Por ello, tenemos que recurrir a métodos numéricos para encontrar una aproximación a la solución verdadera del sistema de ecuaciones.

La solución numérica se basa en cálculos finitos, de modo que es necesario representar las variables meteorológicas por medio de un número finito de valores en el espacio y el tiempo. Es decir, los campos quedaran definidos por sus valores en cada punto de la rejilla definida en el espacio-tiempo.

Podemos discretizar las variables dependientes de diferentes maneras:

Diferencias finitas

$$f(x) \sim f(x_j) \quad j = 1, \dots, N \quad (2.40)$$

Métodos espectrales

$$f(x) \sim \sum_{j=1}^M f_j \omega_j \quad j = 1, \dots, N \quad (2.41)$$

2.4.1. Temporal

El ARW usa un esquema de integración de tiempo dividido. En términos generales, esto significa que los modos de baja frecuencia y los de alta frecuencia se integran con diferentes pasos de tiempo para mantener la estabilidad numérica. La integración de ecuaciones en forma de flujo con tiempo dividido se analiza en Klemp et al. (2007) y su adaptación al esquema Runge-Kutta de orden 3 (RK3), en Wicker and Skamarock (2002).

2.4.2. Método de integración en el tiempo: RK3

El esquema RK3 integra un conjunto de ecuaciones diferenciales ordinarias. Definiendo las variables de pronóstico como

$$\Phi = (U, V, W, \Theta, \phi', \mu', Q_m) \quad (2.42)$$

y las ecuaciones del modelo como

$$\Phi_t = R(\Phi) \quad (2.43)$$

el RK3 lleva a cabo tres pasos para mover la solución de $\Phi(t)$ a $\Phi(t + \Delta t)$:

$$\Phi^* = \Phi^t + \frac{\Delta t}{3} R(\Phi^t) \quad (2.44)$$

$$\Phi^{**} = \Phi^t + \frac{\Delta t}{2} R(\Phi^*) \quad (2.45)$$

$$\Phi^{t+\Delta t} = \Phi^t + \Delta t R(\Phi^{**}) \quad (2.46)$$

RK3 permite pasos de tiempo el doble de largos que Leapfrog, es estable para esquemas centrados, atrasados y adelantados, pero requiere tres evaluaciones del término advectivo, mientras que el de Leapfrog requiere solo una evaluación.

Las frecuencias altas de los modos acústicos no tienen significado meteorológico pero pueden limitar significativamente el paso Δt . Para evitar este inconveniente, se usa el “time-split” que se describe a continuación.

2.4.2.1. Paso de tiempo dividido (“Split time step”).

En las ecuaciones, los términos responsables de los modos rápidos, se integran con un paso de tiempo y los de los modos lentos con otro.

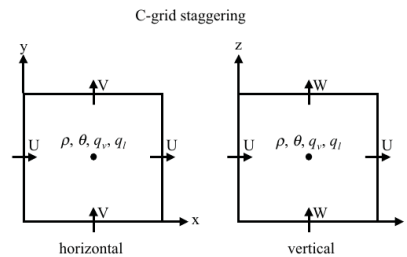
$$U_t = L_s(U) + L_f(U) \quad (2.47)$$

Por cada uno de los pasos de RK3 se hacen varios pasos de tiempo más cortos. En el ejemplo que muestra la figura, para el primer paso se hace 1 paso de tiempo corto, para el segundo 2 y para el tercero 4. Además, en los pasos de tiempo cortos se utilizan métodos semi-implícitos e implícitos para incrementar la estabilidad de las soluciones.

En cada paso de tiempo corto, se actualiza el valor de los términos “rápidos” de la ecuación. Los términos “lentos” solo se actualizan siguiendo los pasos del método de Runge Kutta tal cual se mostró antes.

2.4.3. Discretización espacial: Retícula C de Arakawa.

La discretización espacial en el ARW se lleva a cabo mediante una rejilla de tipo C, es decir, las variables son evaluadas en el centro de la misma, como se indica en la figura. Los índices dan cuenta de la localización de las variables $(x, y, \eta) = (i\Delta x, j\Delta y, k\Delta \eta)$. Los puntos dónde se sitúa θ se denominan puntos de masa y, de igual manera, llamamos puntos u, v o w los puntos dónde encontramos dichas variables. La columna de masa, μ se define en los puntos de masa (i, j) , el geopotencial ϕ en los puntos w , y las variables de humedad q_m , también en los puntos de masa. Las variables de diagnóstico, presión p y la densidad α , se computan en los puntos de masa. Las longitudes Δx y Δy son constantes en la formulación del modelo, sin embargo, la longitud $\Delta \eta$ de la rejilla vertical no lo es y puede especificarse en la entrada.



2.4.3.1. Filtros incluidos en el paso de tiempo

El modelo WRF incluye una serie de filtros para diferentes propósitos que pueden ser activados o no en las simulaciones. Filtro de la divergencia tridimensional: para filtrar las ondas acústicas. Filtro a la integral de la divergencia horizontal en la vertical (o a la divergencia total en la columna): para filtrar el modo externo en las soluciones. Tener en cuenta que al estar el tope ubicado en un nivel de P constante el tope no es rígido. Filtro para reducir las ondas sonoras que se desplazan verticalmente. Filtro para controlar la velocidad vertical en el modo no hidrostático: controla la amplitud de W detectando valores anómalamente grandes y reduciéndolos, evitando que el modelo explote. Capa de absorción de ondas de gravedad: este filtro consiste en una capa adyacente al tope del modelo, en donde se incrementa el valor de los coeficientes de intercambio turbulento de manera de reducir la amplitud de las ondas de gravedad. Capa de absorción de Rayleigh: persigue el mismo fin que el anterior, pero lo hace relajando el perfil de diferentes variables a un perfil de referencia cerca del tope del modelo.

2.4.3.2. Interacciones físicas

A pesar de que las parametrizaciones físicas son categorizadas como modulares, debemos saber que existen interacciones entre ellas a través de las variables de estado del modelo (temperatura potencial, humedad, viento, etc.) y sus tendencias y a través de los flujos de superficie. Las parametrizaciones de la radiación, del cúmulus y de la capa límite dan como resultado cambios en las tendencias de las variables de estado del modelo. Además, estos esquemas físicos no tienen que ser llamados con la misma frecuencia o paso de tiempo del modelo. Esto típicamente es usado con la parametrización de la radiación, ya que es muy costoso llamarla en cada paso de tiempo y con la parametrización del cúmulus, ya que no es necesario llamarla cada paso de tiempo. Sin embargo, los esquemas de suelo y de capa límite son llamados en general, cada paso de tiempo ya que esto lleva a mejores resultados. La radiación es llamada primero, debido a que el modelo de suelo necesita los flujos radiativos como entrada. A su vez, el modelo de suelo también necesita la precipitación de la parametrización de la microfísica y convección, pero usa las salidas del paso de tiempo anterior. El esquema de capa límite es llamado siempre después del de suelo ya que requiere los flujos de calor y humedad que son salidas del modelo de suelo.

Capítulo 3

Modelo de gases

3.1. Introducción

En este capítulo se presentan los modelos e inventarios utilizados en la adquisición de los datos de la composición atmosférica así como una breve introducción teórica a las propiedades ópticas que de ella se derivan. A su vez, se estudiará la influencia de estas en los procesos de transferencia radiativa.

En el marco general de la tesis, este capítulo aborda la consideración de condiciones atmosféricas con variabilidad espacial dentro de nuestro dominio de interés. Nos motiva el tamaño de las imágenes capturadas por la mayoría de los satélites de alta resolución, frecuentemente con swath superiores a 600 km. Pese a que las imágenes que abarcan una gran extensión no son una novedad en teledetección, la consideración de dicha variabilidad no suponía una mejora notable en la calidad de las medidas debido a la escasa resolución espacial de las mismas. Esta característica está cambiando drásticamente en los últimos años con el desarrollo de nuevos sensores y dispositivos. En la actualidad se dispone de multitud de sensores con resoluciones espaciales del orden del metro.

En el primer apartado presentamos el módulo químico del modelo WRF que puede considerarse como el esqueleto del sistema de adquisición de datos en este ámbito. En el apartado 2, se describen los inventarios utilizados y las parametrizaciones utilizadas en el modelo meteorológico. También en dicho apartado, se presenta el software utilizado para llevar dichos datos sobre la rejilla de nuestro dominio. A continuación se describen las propiedades ópticas derivadas de la composición y, finalmente, cómo se introducen en el modelo de transferencia radiativa.

3.2. Características del modelo WRF/Chem

La predicción y simulación de la composición química de la atmósfera involucra, tanto factores meteorológicos, tales como la velocidad y dirección del viento, la turbulencia, radiación, presencia de nubes y precipitación, como químicos, tales como la deposición y los procesos de transformación de componentes.

En la atmósfera real, los procesos químicos y meteorológicos están acoplados, pero habitualmente los modelos han tratado su fenomenología de forma independiente. La ventaja de estos modelos 'offline' frente a los que sí consideran la interacción de ambos mecanismos es meramente computacional.

El módulo químico WRF/Chem, (Grell et al., 2005), se trata de un modelo online, es decir, interacciona con el núcleo del WRF que resuelve la dinámica, y está totalmente embebido en él. Es consistente con los procesos de transporte que realiza el modelo meteorológico, utiliza las mismas coordenadas verticales y horizontales así como idénticas parametrizaciones físicas para los fenómenos de escala subrejilla. No realiza interpolaciones ni espaciales ni temporales. Está adaptado para el estudio de los feedbacks entre la composición química y la meteorología.

El WRF/Chem es, actualmente, parte del Weather Research and Forecasting. Su desarrollo fue llevado a cabo por las siguientes instituciones:

- Department of Commerce/National Oceanic and Atmospheric Administration
- The Cooperative Institute for Research in Environmental Sciences
- The University Corporation for Atmospheric Research (UCAR)
- The National Center for Atmospheric Research (NCAR)
- The Max Plank Institute
- The University of Chile
- Centro de Previsão de Tempo e Estudos Climáticos

A continuación se describen algunos de los procesos presentes en el modelo y expuestos en (Grell et al., 2005).

3.2.1. Transporte

El transporte de las especies químicas se realiza online. El núcleo dinámico utilizado es el Advanced Research WRF (ARW), que utiliza la descripción euleriana del movimiento del fluido. En la aproximación de variables conservadas, el modelo ARW integra las ecuaciones de conservación de masa y conservación escalar de la forma

$$\mu_t + \nabla \cdot (V\mu) = 0 \quad (3.1)$$

$$(\mu\phi)_t + \nabla \cdot (V\mu\phi) = 0 \quad (3.2)$$

Donde μ es la masa de la columna de aire seco, V la velocidad (u, v, ∇) , y ϕ es el escalar del ratio de la mezcla.

El mecanismo químico utilizado pronostica 39 especies. El modulo de aerosoles añade otras 34 incluyendo el número total de partículas dentro de cada modo así como las especies orgánicas e inorgánicas para los modos de Aitken y de acumulación. También se incluyen las especies antropogénicas, marinas y aerosoles derivados del suelo, en el modo ordinario.

3.2.2. Deposición seca

El flujo de los gases traza y de partículas desde la atmósfera hasta la superficie se calcula multiplicando las concentraciones en la capa inferior por las velocidades de deposición, con dependencia espacial y temporal. Esta es proporcional a tres resistencias características, la resistencia aerodinámica, subcapa y superficial. La resistencia superficial se deriva del tipo de superficie y cobertura vegetal, también se considera su dependencia con el coeficiente de difusión, reactividad y solubilidad en agua del gas traza en cuestión. Cuando se emplea la parametrización de aerosoles, la velocidad de deposición \hat{v}_{dk} del k-ésimo momento de un aerosol polidispersivo es, (Slinn and Slinn, 1980)

$$\hat{v}_{dk} = (r_a + \hat{r}_{dk} + r_a \hat{r}_{dk} \hat{v}_{Gk})^{-1} + \hat{v}_{Gk} \quad (3.3)$$

r_a es la resistencia superficial, \hat{v}_{Gk} la velocidad de asentamiento, \hat{r}_{dk} la difusividad Browniana.

3.2.3. Química de la fase gaseosa

Este mecanismo atmosférico fue desarrollado originalmente por Stockwell et al. (1990) para el modelo RADM2 (Chang et al., 1989) (Regional Acid Deposition Model, version 2). RADM2

se utiliza para predecir las concentraciones de los oxidantes y contaminantes químicos del aire. Incluye 14 especies inorgánicas estables, 4 reactivos intermedios y 3 especies estables abundantes: oxígeno, nitrógeno y agua. Los componentes orgánicos vienen representados por 26 especies estables y 16 radicales peróxidos. En RADM2, los procesos orgánicos se tratan a través de agregados moleculares siguiendo el enfoque de (Middleton et al., 1990).

3.2.4. Emisiones

Las emisiones producidas por la actividad humana se tratan en el modelo WRF/Chem de forma similar a como se hace en McKeen et al. (2002). Las emisiones de tipo biológico se tratan en base a la descripción de Guenther et al. (1993), Guenther et al. (1994) y Simpson et al. (1995). Se consideran emisiones de isoprenos, monoterpenos, otros compuestos volátiles de origen orgánico (VOC) y nitrógeno procedente del suelo.

Las emisiones de isoprenos procedentes de bosques dependen de la temperatura y de la radiación fotosintéticamente activa y se parametriza mediante la fórmula desarrollada por Guenther et al. (1993). Los isoprenos procedentes de cultivos y áreas de hierba se estiman considerando solo su dependencia con la temperatura. Las emisiones de monoterpenos, OVOC (Oxygenated Volatile Organic Compounds) y nitrógeno se tratan como función, solamente, de la temperatura.

Debemos destacar que las categorías de uso de suelo se toman del sistema de clasificación USGS de 24 clases y no se hace referencia a las especies que componen cada tipo de cubierta.

3.2.5. Parametrización de aerosoles

El modelo WRF/Chem dispone de cinco posibles esquemas de aerosoles, The Modal Aerosol Dynamics Model for Europe (MADE/SORGAM) (Ackermann et al., 1998), (Schell et al., 2001), The Modal Aerosol Dynamics Model for Europe with the Volitivity Basis Set (MADE/VBS), The Modal Aerosol Module (MAM), The Model for Simulating Aerosol Interactions and Chemistry (MOSAIC) sectional model aerosol parameterization y Global Ozone Chemistry Aerosol Radiation and Transport (GOCART), en este caso, hicimos uso de este último que presentamos más abajo.

3.2.5.1. Distribución de tamaños

La distribución por tamaño de los aerosoles se presenta en dos intervalos solapados llamados modos. Asumiendo que dicha distribución es log-normal dentro de cada modo

$$n(\ln d_p) = \frac{N}{\sqrt{2\pi \ln \sigma_g}} \exp \left[-\frac{1}{2} \frac{(\ln d_p - \ln d_{p_g})^2}{\ln^2 \sigma_g} \right] \quad (3.4)$$

donde N es el número de concentración (m^{-3}), d_p el diámetro de la partícula, d_{p_g} el diámetro medio, σ_g la desviación estándar de la distribución. El momento k -ésimo de la distribución se define como

$$M_k = \int_{-\infty}^{\infty} d_p^k n(\ln d_p) d(\ln d_p) \quad (3.5)$$

con la solución

$$M_k = N d_{p_g}^k \exp \left[\frac{k^2}{2} \ln^2 \sigma_g \right] \quad (3.6)$$

M_0 es el número total de partículas de aerosoles suspendidas en una unidad de volumen de aire, dentro del modo correspondiente. M_2 es proporcional al área total formado por las superficies las partículas suspendidas por unidad de volumen, y M_3 es proporcional al volumen total de las partículas suspendidas por unidad de volumen.

3.2.5.2. Nucleación, condensación y coagulación

El proceso más importante en la creación de aerosoles secundarios es el de nucleación homogénea en el sistema ácido sulfúrico - agua. En el modelo WRF/Chem este proceso se evalúa mediante el método dado por (Kulmala et al., 1998). El crecimiento de los aerosoles por condensación ocurre en dos procesos: producción de materia condensable en reacciones químicas y condensación y evaporación de especies volátiles sobre aerosoles (Binkowski and Shankar, 1995). Durante los procesos de coagulación, las distribuciones mantienen su forma log-normal. Solo se consideran los efectos debidos al movimiento Browniano. La formulación matemática puede encontrarse en (Binkowski and Shankar, 1995).

3.2.5.3. Química de aerosoles

El sistema que simula los procesos de química inorgánica está basado en MARS (Saxena et al., 1986) y las modificaciones realizadas por (Binkowski and Shankar, 1995). Calcula la composición química de los aerosoles sulfato-nitrato-amonio-agua en el equilibrio termodinámico. Se consideran dos regímenes en función del ratio molar de sulfatos y amónicos.

La química de los aerosoles de procedencia orgánica está basada en SORGAM (Schell et al., 2001).

3.2.5.4. Frecuencias de fotólisis

Se calculan las frecuencias de fotólisis para 21 reacciones fotoquímicas en cada punto de rejilla de acuerdo con (Madronich, 1987). La frecuencia de fotólisis del gas i , J_i , viene dada por la integral en longitud de onda del producto del flujo actínico $I_A(\lambda)$, la sección eficaz de absorción $\sigma(\lambda)$ y el rendimiento cuántico de la disociación $\varphi(\lambda)$

$$J_i = \int I_A(\tau, \lambda) \cdot \sigma_i(\lambda) \cdot \varphi_i(\lambda) d\lambda \quad (3.7)$$

Para el cálculo del flujo actínico se utiliza un modelo de transferencia radiativa que considera las absorciones debidas a O_2 y O_3 , el scattering Rayleigh y la absorción y scattering provocadas por los aerosoles de acuerdo a (Chang et al., 1989). Las secciones eficaces de absorción y el rendimiento cuántico vienen dados por (Stockwell et al., 1990). La integral se resuelve en el rango de 186 a 730 nm.

3.2.6. GOCART

Seleccionamos el modelo GOCART, del Instituto Tecnológico Goddard de Georgia, principalmente, porque simula el parámetro de espesor óptico de aerosoles para 5 tipos distintos: sulfuros, polvo, carbono orgánico (OC), carbono negro (BC) y sales marinas (Chin et al., 2002). Tiene una resolución de 2° de latitud por $2,5^\circ$ de longitud y 20-30 niveles verticales sigma, dependiendo de la versión de GEOS DAS que utilice (Chin et al., 2000). Cada uno de estos 5 grandes tipos está dividido en otros atendiendo a los tamaños, haciendo un total de 14. Hace uso de los campos meteorológicos proporcionados por el sistema GEOS DAS, Goddard Earth Observing System Data Assimilation System, (Schubert et al., 1993). Incluye los siguientes módulos de simulación: emisión, químico, advección, mezcla turbulenta en capa límite atmosférica, convección con humedad, deposición seca y con humedad.

3.2.6.1. Transporte

La advección se computa por el método semi-Lagrangiano en forma de flujo (Lin and Rood, 1996). La convección con humedad se parametriza usando el archivo de campos de flujo de masas de nubes de GEOS DAS. La mezcla turbulenta en la capa límite se calcula usando el esquema descrito en Helfand and Labraga (1988).

3.2.6.2. Emisiones de sulfuros

Incluye emisiones de sulfuro de dimetilo (DMS) procedente del océano (se considera dependiente de la temperatura y la velocidad del viento (Liss and Merlivat, 1986)), SO_2 y sulfuros de la actividad humana, biomasa quemada, emisiones de aviones y erupciones volcánicas. Las emisiones procedentes de combustibles fósiles o bio-combustibles se toman de la base de datos EDGAR (Emission Data Base for Global Atmospheric Research) (Olivier et al., 1996). En Europa se considera una variación anual debida a la demanda de calor en las casas (Sandnes and Styve, 1993).

Las emisiones volcánicas tienen en cuenta tanto los volcanes continuamente activos como los que erupcionan esporádicamente. Los continuamente activos se toman de la base de datos de (Andres and Kasgnoc, 1998). Incluye datos de 49 volcanes que han permanecido activos durante 25 años. Los datos de las erupciones esporádicas se toman del Global Volcanism Program (GVP), (Simkin and Siebert, 1994).

3.2.6.3. Emisiones de polvo

Las partículas de polvo de radio entre 0.1 y 10 μm se consideran en el modelo en 8 grupos de tamaños (0.1-0.18, 0.18-0.3, 0.3-0.6, 0.6-1, 1-1.8, 1.8-3, 3-6, and 6-10 μm). El flujo de emisión para cada grupo p se expresa como

$$F_p = S \cdot s_p \cdot u^2(u - u_t) \quad \text{if } u > u_t \quad (3.8)$$

Donde S es la probabilidad de emisión, s_p es la fracción del grupo p en el suelo, u es la velocidad del viento en la superficie, u_t es la velocidad del viento a la que comienza la erosión, viene determinada por el tamaño de las partículas y la humedad del suelo.

3.2.6.4. Emisiones de OC y BC

La biomasa quemada de OC y BC se estima a partir de la base de datos (Duncan et al., 2003), desarrollada a partir de observaciones de satélite, fuegos, índice de aerosoles y biomasa media quemada anualmente. Los datos de emisiones debidas a la actividad humana se toman (Cooke and Wilson, 1996). Además de las emisiones directas, la estimación de la producción de OC por fuentes terrestres se realiza en base a los componentes orgánicos volátiles (Guenther et al., 1995).

3.2.6.5. Emisiones de sales marinas

Las emisiones de sales marinas procedentes del océano dependen en gran medida de la velocidad del viento en la superficie y se calcula con la expresión:

$$dF/dr = 1,37 \cdot u^{3,41} r^{-3} \cdot (1 + 0,057 \cdot r^{1,05}) \cdot 10^{1,19 \cdot \exp(-B^2)} \quad (3.9)$$

donde F es el flujo de emisión, r es el radio de la partícula, u es la velocidad a 10 m de altura, y $B = (0,380 - \log r)/0,65$. Se consideran cuatro rangos de tamaños con radios de 0.1-0.5, 0.5-1.5, 1.5-5, y 5-10 μm .

3.2.6.6. Química

Las reacciones químicas incluidas en el modelos son: Oxidación *DMS* por *OH* durante el día para formar SO_2 y ácido metanosulfónico (*MSA*); por la noche, oxidación por nitratos (NO_3) para formar SO_2 ; oxidación de SO_2 por *OH* por H_2O_2 para formar sulfatos. Los ratios de reacciones se toman de DeMore et al. (1987).

Debido a que el contenido de agua no está disponible en GEOS, se parametriza la oxidación de SO_2 por *OH* por H_2O_2 dentro de las nubes como función de la fracción de nube siguiendo Chin and Jacob (1996). La fracción de nube f_c para cada caja de rejilla se estima mediante una función empírica de la humedad relativa, de acuerdo con Sundqvist et al. (1989):

$$f_c = 1 - \sqrt{1 - \frac{r - r_0}{1 - r_0}} \quad (3.10)$$

donde r es la humedad relativa, r_0 es el umbral de condensación como función de la presión para la humedad relativa (Xu and Krueger, 1991).

3.2.6.7. Deposición seca

Las velocidades de deposición seca para el SO_2 , sulfatos y *MSA* se calculan usando el esquema (Wesely and Hicks, 1977). En ese esquema, las velocidades se calculan a partir de la resistencia aerodinámica, la resistencia superficial y subcapa. La resistencia aerodinámica se toma del archivo de DEOS DAS. La resistencia superficial y subcapa para el SO_2 se calculan usando la formulación de Walcek et al. (1986) y Wesely (1989). Se asume que la velocidad de deposición seca del *MSA* es la misma que la del sulfato.

3.2.6.8. Absorción de humedad

Se considera tanto la absorción que ocurre dentro de las nubes como la que ocurre por debajo de ellas cuando se produce precipitación (Chin et al., 2000).

3.2.6.9. Propiedades ópticas

Conociendo la composición y distribución de los aerosoles en la atmósfera, se calcula el parámetro AOT, espesor óptico de aerosoles, a partir del índice de refracción, la distribución por tamaño, radio efectivo y masa por unidad de área, (Chin et al., 2002). Se asume que la distribución de los tamaños es lognormal para los 5 grandes tipos estudiados. Un parámetro clave en las propiedades radiativas es la sección eficaz pesada con el radio efectivo (Hansen and Travis 1974)

$$r_e = \frac{\int r \pi r^2 f(r) dr}{\int \pi r^2 f(r) dr} \quad (3.11)$$

donde $f(r)$ es la fracción de partículas con radio entre r y $r + dr$. La relación entre el espesor óptico de aerosoles y la masa por unidad de área se expresa por (Lacis and Mishchenko 1994),

$$\tau = \frac{3QM}{4\rho r_e} \quad (3.12)$$

donde ρ es la densidad de partículas y Q es el coeficiente de extinción calculado con la teoría del scattering de Mie.

La salida del modelo WRF-Chem, nos proporciona, además de los perfiles de AOT en cada punto y en las longitudes de onda, 300, 400, 600 y 1000nm, los correspondientes al factor de asimetría g y al 'single scattering albedo' ω_0 . Este tiene en cuenta los coeficientes de absorción y scattering $\omega_0 = \frac{k_s}{k_a + k_s}$, k_a es el coeficiente de absorción, directamente relacionado la cantidad de carbono negro; k_s es el coeficiente de scattering, depende del tamaño del aerosol. A partir de estos y de los perfiles de temperatura y presión, definiremos nuestras condiciones atmosféricas con dependencia espacial que serán utilizadas para calcular las radiancias TOA. En el siguiente gráfico vemos los perfiles de estos parámetros en las 4 bandas anteriormente mencionadas en colores violeta, azul, amarillo y rojo. Figura 1. Perfiles de espesor óptico, factor de asimetría y single scattering albedo en 300, 400, 600 y 1000nm (azul, verde, rojo y negro)

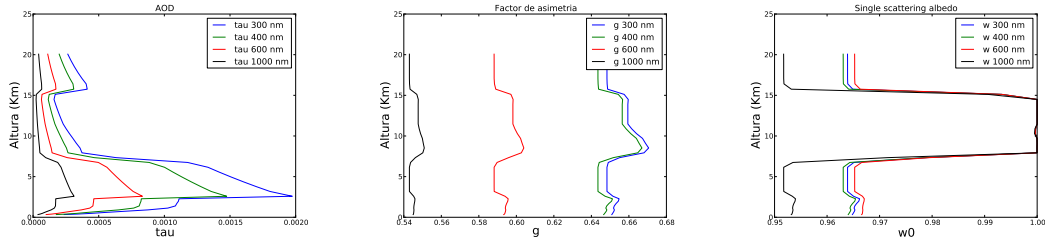


FIGURA 3.1: Perfiles de τ , g y ω_0

3.2.7. PREP-CHEM-SRC

Para realizar las simulaciones del modelo meteorológico con la opción de componentes químicos, necesitamos incluir datos adicionales que describan las diferentes especies presentes en la atmósfera. La preparación de estos datos se realiza fuera del modelo WRF. Para ello, hicimos uso de la herramienta PREP-CHEM-SRC (Freitas et al., 2011), que adapta los campos de emisiones de gases traza y aerosoles a la rejilla del modelo meteorológico. En este trabajo se utilizaron los datos de emisiones proporcionados por los proyectos RETRO y EDGAR, que se presentan más abajo, así como los de aerosoles, procedentes del modelo GOCART, ya descrito.

Esta herramienta posee varios mecanismos químicos disponibles: RACM, RADM2, CB07 y RELACS. Aquí, se utilizó la opción RADM2 (Stockwell et al., 1990) de la agencia NOAA (<http://ruc.noaa.gov/wrf/WG11/radm2.htm>).

3.2.8. Bases de datos de emisiones antropogénicas globales

Los datos de emisiones antropogénicas que se utilizaron en el modelo WRF/Chem fueron tomados de las bases de datos RETRO (Schultz et al., 2007) y EDGAR (Van Aardenne et al., 2005) principalmente. A continuación se describen de forma breve.

Inventario de emisiones RETRO.

Dentro del proyecto RETRO (<http://retro.enes.org/index.shtml>), se generan datos de emisiones globales debidas a la actividad humana e incendios en zonas de vegetación. Cubre el período de 1960 a 2000 con una resolución temporal mensual. Las emisiones antropogénicas en el inventario RETRO se derivan de la base de datos de TNO y el inventario VERITAS de las emisiones del tráfico internacional de aviones. Las emisiones de los incendios de vegetación se construyen a partir de una gran variedad de fuentes con el objetivo de proporcionar una

estimación razonable de las emisiones. Se considera su variabilidad estacional e interanual en las principales regiones del mundo.

EDGAR (Emission Database for Global Atmospheric Research)

El sistema de información EDGAR es un proyecto conjunto de numerosas organizaciones de Países Bajos, Italia y Alemania. Almacena los inventarios de las emisiones globales de gases de efecto invernadero procedentes de fuentes antropogénicas, incluyendo halocarbonos y aerosoles, en función de cada país y región. El sistema EDGAR genera datos sobre las emisiones mundiales, regionales y nacionales en varios formatos. La actual base de datos ha sido desarrollada con el apoyo financiero del Ministerio de Medio Ambiente de los Países Bajos (VROM) y el Programa Nacional de Investigación holandés sobre la Contaminación Atmosférica Global y Cambio Climático (PNR), en estrecha cooperación con el Inventario de Emisiones de la Actividad Global (GEIA). EDGAR consta de: fuentes relacionadas con el combustible fósil y, biocombustibles, en función de cada país; la producción industrial y el consumo, también en función de cada país; fuentes relacionadas con el uso de la tierra, incluyendo el tratamiento de residuos, en función de cada país; fuentes naturales sobre una rejilla.

3.3. Configuración utilizada

En el Apéndice C presentamos un esquema de las configuraciones que se han utilizado en las simulaciones realizadas. En resumen, los pasos que hemos llevado a cabo han sido

1. Se realiza la primera simulación WRF en el dominio de interés
2. Se generan las emisiones para dicho dominio así como los mecanismos químicos en un fichero intermedio que convertiremos en la entrada del modelo meteorológico.
3. Se realiza la segunda simulación WRF con los datos químicos anteriores.

Capítulo 4

Asimilación de datos

4.1. Introducción

"The aim of a data assimilation scheme is to use measured observations in combination with a dynamical system model in order to derive accurate estimates of the current and future states of the system, together with estimates of the uncertainty in the estimated states."
(Nichols, 2010)

El propósito de un esquema de asimilación de datos es usar las observaciones junto con los resultados del modelo del sistema dinámico para obtener estimaciones más precisas del estado actual y futuro del sistema. Además, se obtiene información de las incertidumbres de los estados estimados. Básicamente, existen dos enfoques posibles. El primero usa un observador dinámico, dando lugar a un esquema secuencial; el segundo, usa un observador directo y produce un esquema de asimilación variacional (Evensen, 2007).

- Asimilación de datos secuencial
 - Dinámica lineal: Filtro de Kalman
 - Dinámica no lineal: Filtro de Kalman extendido
 - Ensemble Kalman filter
- Asimilación de datos variacional
 - Dinámica lineal
 - Dinámica no lineal

- Métodos de minimización

En el contexto de la asimilación de datos se usan los términos *análisis* y *background* en el siguiente sentido:

Análisis: se dice que es el proceso de estimación del estado real del sistema en un momento dado.

Background: es la estimación del estado a priori; habitualmente es un pronóstico dado por el modelo en cuestión.

A modo de introducción a los métodos de asimilación, presentamos el siguiente ejemplo que ilustra la incorporación de una observación a nuestro pronóstico (background), con el fin de obtener el análisis óptimo. En primera instancia, damos por hecho que la medida se toma en un punto del espacio y en un momento dado. Por ejemplo, podría corresponder a un pronóstico de temperatura en un punto concreto del espacio en un momento determinado.

Sea T_b nuestro pronóstico o background, T_o el valor de la medida (observación), T el valor o estado real y T_a el análisis buscado, es decir, la mejor estimación posible de T , que será aquella que consigue el mínimo error.

Podríamos comenzar con una combinación lineal de ambos valores

$$T_a = \alpha T_o + \beta T_b + \gamma \quad (4.1)$$

Los errores asociados con cada uno serían

$$\epsilon_o = T_o - T \quad (4.2)$$

$$\epsilon_b = T_b - T \quad (4.3)$$

Considerando que ambos tienen bias nulo $\bar{\epsilon}_o = \bar{\epsilon}_b = 0$ y que el error en el análisis será $\epsilon_a = T_a - T$, tenemos

$$T_a = \epsilon_a + T = \alpha(\epsilon_o + T) + \beta(\epsilon_b + T) + \gamma \quad (4.4)$$

$$\epsilon_a = (\alpha + \beta - 1)T + (\alpha\epsilon_o + \beta\epsilon_b) + \gamma \quad (4.5)$$

tomando el valor medio y teniendo en cuenta que pretendemos que el error asociado al análisis cumpla $\bar{\epsilon}_a = 0$

$$\bar{\epsilon}_a = (\alpha + \beta - 1)T + \gamma \quad (4.6)$$

y debido a que la igualdad debe cumplirse para cualquier valor de la temperatura real T , tendríamos que

$$\gamma = 0 \quad (4.7)$$

$$\alpha + \beta - 1 = 0 \quad (4.8)$$

De esta forma llegaríamos a que la estimación lineal con bias nulo queda así

$$T_a = \alpha T_o + (1 - \alpha) T_b \quad (4.9)$$

y daría lugar a una relación similar para los errores

$$\epsilon_a = \alpha \epsilon_o + (1 - \alpha) \epsilon_b \quad (4.10)$$

Ahora, calculamos la varianza de dicha estimación

$$\overline{\epsilon_a^2} = \alpha^2 \overline{\epsilon_o^2} + 2\alpha(1 - \alpha) \overline{\epsilon_o \epsilon_b} + (1 - \alpha)^2 \overline{\epsilon_b^2} \quad (4.11)$$

El término $\overline{\epsilon_o \epsilon_b}$ representa la covarianza entre el error de la medida y el error asociado con el pronóstico. Ya que no hay razón para que ambos estén relacionados, suponemos $\overline{\epsilon_o \epsilon_b} = 0$. Buscamos el valor de α que hace mínima esta cantidad

$$\frac{d(\overline{\epsilon_a^2})}{d\alpha} = 2\alpha \overline{\epsilon_o^2} - 2(1 - \alpha) \overline{\epsilon_b^2} \quad (4.12)$$

$$\Rightarrow \alpha = \frac{\overline{\epsilon_b^2}}{\overline{\epsilon_o^2} + \overline{\epsilon_b^2}}, \quad \overline{\epsilon_a^2} = \left(\frac{1}{\overline{\epsilon_b^2}} + \frac{1}{\overline{\epsilon_o^2}} \right)^{-1} \quad (4.13)$$

el α optimiza la combinación lineal de partida consiguiendo ese valor mínimo de la varianza del error $\overline{\epsilon_a^2}$

Extensión a varias dimensiones

En la siguiente exposición, repetiremos el razonamiento anterior para aplicarlo al caso de tener campos escalares en lugar de un valor en un punto solamente. El estado de un campo escalar, en un momento determinado, puede venir dado por los valores de dicho campo sobre una rejilla, o un conjunto de coeficientes armónicos esféricos, etc. Dicho estado vendrá representado por un vector de \mathbb{R}^n , construido con los valores del campo en cada punto. La dimensión n de dicho espacio podría ser muy alta, del orden de 10^7 , considerando que vendrá dada por la multiplicación del número de puntos de la rejilla en el caso de campo escalar.

Supongamos que x_b denota los valores pronosticados de un campo escalar sobre los puntos de una rejilla; podría tratarse de temperatura, presión, etc. x_o es el conjunto de observaciones, x el estado real del campo y x_a el análisis.

La principal diferencia que encontramos con el caso puntual es que, ahora, no necesariamente tendremos una observación en cada punto de la rejilla sobre la que definimos el estado pronosticado. En el caso que nos ocupa resulta obvio, ya que deseamos incorporar las medidas realizadas por el sensor del satélite en el modelo meteorológico. Notemos que ni tendremos medidas en cada punto, ni coincidirán los puntos de ambas rejillas en las que hay dato disponible.

Para resolver este problema, se acude al operador de medida o de observación \mathcal{H} , que va del espacio de estados al espacio de observaciones. En ocasiones, esta operación consistirá en la interpolación del campo a los puntos en que existe observación disponible, en otras será una mera selección de un subconjunto, etc. Por el momento, pensaremos que esta aplicación es perfecta y asocia a un estado real x , perfecto, su correspondiente medida $\mathcal{H}(x) = x_o^*$, sin introducir error alguno.

$$\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}^m \quad (4.14)$$

Como hicimos anteriormente, probamos una combinación lineal de pronóstico y medida para construir el análisis

$$x_a = Fx_b + G\mathcal{H}(x_b) + Kx_o + \bar{\gamma} \quad (4.15)$$

donde F, G, K son matrices y $\bar{\gamma}$ un vector.

Si \mathcal{H} es lineal, consideramos la estimación lineal con bias nulo, como hicimos en el caso puntual. En el caso más general de \mathcal{H} no lineal, exigimos que las entradas sin error ($x_b = x$, $x_o = x_o^*$) produzcan un análisis sin error $x_a = x$

$$x = Fx + G\mathcal{H}(x) + K\mathcal{H}(x) + \bar{\gamma} \quad (4.16)$$

Ya que la ecuación debe cumplirse para cualquier x , $\bar{\gamma}$ tiene que ser 0 y, a nivel de operadores, resulta la siguiente equivalencia

$$F + G\mathcal{H}(\cdot) \equiv I - K\mathcal{H}(\cdot) \quad (4.17)$$

donde I es la matriz identidad. Por tanto, el análisis quedaría así

$$x_a = Fx_b + G\mathcal{H}(x_b) + Kx_o + \bar{\gamma} = Ix_b - K\mathcal{H}(x_b) + Kx_o + \bar{\gamma} \quad (4.18)$$

$$\Rightarrow x_a = x_b + K(x_o - \mathcal{H}(x_b)) \quad (4.19)$$

Recordemos que en el caso puntual teníamos

$$T_a = \alpha T_o + (1 - \alpha)T_b \quad (4.20)$$

$$T_a = T_b + \alpha(T_o - T_b) \quad (4.21)$$

K se denomina matriz de ganancia y juega un papel similar al α del caso puntual.

Definimos los errores asociados

$$\epsilon_a = x_a - x \quad (4.22)$$

$$\epsilon_b = x_b - x \quad (4.23)$$

$$\epsilon_o = x_o - x_o^* \quad (4.24)$$

Suponemos que los errores son pequeños, luego

$$\mathcal{H}(x_b) = \mathcal{H}(x) + H\epsilon_b + O(\epsilon_b^2) \quad (4.25)$$

H es el Jacobiano de \mathcal{H} . De tal manera que

$$\epsilon_a + x = \epsilon_b + x + K(\epsilon_o + x_o^* - (\mathcal{H}(x) + H\epsilon_b)) \quad (4.26)$$

$$\Rightarrow \epsilon_a = \epsilon_b + K(\epsilon_o - H\epsilon_b) \quad (4.27)$$

Como anteriormente, volvemos a suponer que $\bar{\epsilon}_o = \bar{\epsilon}_b = 0$ y eso implica $\bar{\epsilon}_a = 0$.

En el caso multidimensional, tratamos con las matrices de covarianza en lugar de varianzas.

Para el error asociado al análisis, teníamos

$$\epsilon_a = \epsilon_b + K(\epsilon_o - H\epsilon_b) \quad (4.28)$$

$$\epsilon_a = (I - KH)\epsilon_b + K\epsilon_o \quad (4.29)$$

luego la covarianza será

$$\overline{\epsilon_a \epsilon_a^T} = \overline{[(I - KH)\epsilon_b + K\epsilon_o][(I - KH)\epsilon_b + K\epsilon_o]^T} \quad (4.30)$$

$$\begin{aligned} &= (I - KH)\overline{\epsilon_b \epsilon_b^T}(I - KH)^T + (I - KH)\overline{\epsilon_b \epsilon_o^T}K^T \\ &\quad + K\overline{\epsilon_o \epsilon_b^T}(I - KH)^T + K\overline{\epsilon_o \epsilon_o^T}K^T \end{aligned} \quad (4.31)$$

y suponiendo que los errores no están correlacionados, $\overline{\epsilon_b \epsilon_o^T} = \overline{\epsilon_o \epsilon_b^T} = 0$, resulta

$$\overline{\epsilon_a \epsilon_a^T} = (I - KH)\overline{\epsilon_b \epsilon_b^T}(I - KH)^T + K\overline{\epsilon_o \epsilon_o^T}K^T \quad (4.32)$$

que vuelve a tener una forma similar a la del caso puntual

$$\overline{\epsilon_a^2} = \alpha^2 \overline{\epsilon_o^2} + 2\alpha(1-\alpha)\overline{\epsilon_o\epsilon_b} + (1-\alpha)^2 \overline{\epsilon_b^2} \quad (4.33)$$

El siguiente paso en el caso puntual fue encontrar α que minimizaba la varianza del error. En este caso, hemos utilizado la matriz de covarianzas y esta contiene las varianzas sobre la diagonal. Paralelamente al caso anterior, buscamos el K que minimiza la suma de las varianzas, es decir, la traza de la matriz de covarianzas.

$$\frac{\partial(\text{tr}(\overline{\epsilon_a\epsilon_a^T}))}{\partial K} = 0 \quad (4.34)$$

Utilizando las siguientes identidades,

$$\frac{\partial(\text{tr}(KAK^T))}{\partial K} = K(A + A^T), \quad \frac{\partial(\text{tr}(KA))}{\partial K} = A^T \quad (4.35)$$

$$\frac{\partial(\text{tr}(AK^T))}{\partial K} = A \quad (4.36)$$

obtenemos

$$\frac{\partial(\text{tr}(\overline{\epsilon_a\epsilon_a^T}))}{\partial K} = 2K[H\overline{\epsilon_b\epsilon_b^T}H^T + \overline{\epsilon_o\epsilon_o^T}] - 2\overline{\epsilon_b\epsilon_b^T}H^T = 0 \quad (4.37)$$

$$\Rightarrow K = \overline{\epsilon_b\epsilon_b^T}H^T[H\overline{\epsilon_b\epsilon_b^T}H^T + \overline{\epsilon_o\epsilon_o^T}]^{-1} \quad (4.38)$$

K es la matriz de ganancia de Kalman y la covarianza del análisis sería

$$\overline{\epsilon_a\epsilon_a^T} = \left[(\overline{\epsilon_b\epsilon_b^T})^{-1} + H^T(\overline{\epsilon_o\epsilon_o^T})^{-1}H \right]^{-1} \quad (4.39)$$

4.2. Métodos de asimilación

En lo sucesivo, adoptaremos la notación estándar para las matrices que aparecen en la teoría de Kalman

$$P^a \equiv \overline{\epsilon_a\epsilon_a^T} \quad (4.40)$$

$$P^b \equiv \overline{\epsilon_b\epsilon_b^T} \quad (4.41)$$

$$R \equiv \overline{\epsilon_o\epsilon_o^T} \quad (4.42)$$

La matriz de ganancia de Kalman se escribe ahora como

$$K = P^b H^T [HP^b H^T + R]^{-1} \quad (4.43)$$

y multiplicando a ambos lados por $[(P^b)^{-1} + H^T R^{-1} H]$, obtenemos

$$\begin{aligned} [(P^b)^{-1} + H^T R^{-1} H]K &= [H^T + H^T R^{-1} HP^b H^T][HP^b H^T + R]^{-1} \\ &= H^T R^{-1} [R + HP^b H^T][HP^b H^T + R]^{-1} \\ &= H^T R^{-1} \end{aligned} \quad (4.44)$$

$$K = [(P^b)^{-1} + H^T R^{-1} H]^{-1} H^T R^{-1} \quad (4.45)$$

4.2.0.1. Interpolación óptima

Es un método de asimilación de datos estadístico basado en las ecuaciones del análisis que derivamos anteriormente para el caso multi-dimensional. Consiste en dividir el análisis en un conjunto de cajas que pueden ser analizadas de forma independiente.

$$x_a^{(i)} = x_b^{(i)} + K^{(i)}(x_o^{(i)} - \mathcal{H}^{(i)}(x_b)) \quad (4.46)$$

siendo

$$x_a = \begin{pmatrix} x_a^{(1)} \\ x_a^{(2)} \\ \vdots \\ x_a^{(M)} \end{pmatrix} \quad x_b = \begin{pmatrix} x_b^{(1)} \\ x_b^{(2)} \\ \vdots \\ x_b^{(M)} \end{pmatrix} \quad K = \begin{pmatrix} K^{(1)} \\ K^{(2)} \\ \vdots \\ K^{(M)} \end{pmatrix}$$

Aunque, en principio, todas las observaciones disponibles tendrían que tenerse en cuenta en el análisis de cada caja, se consideran solamente las del entorno para reducir el coste computacional. Así, la dimensión de $[(HP^b H^T)^{(i)} + R^{(i)}]$ es la correspondiente al número de observaciones que se tienen en cuenta y el problema queda simplificado. En cualquier caso, seguimos necesitando explícitamente $P^b H^T$ y $(HP^b H^T)$, que limita considerablemente la complejidad del operador de observación si queremos mantener un bajo coste computacional.

4.2.1. Asimilación de datos secuencial

En el caso de problemas dependientes del tiempo, los métodos de asimilación de datos secuenciales utilizan el esquema presentado en la sección anterior para actualizar, en cada paso de tiempo, el estado del sistema.

4.2.1.1. Filtro de Kalman estándar y extendido (EKF)

Vamos a exponer directamente la teoría del filtro de Kalman extendido que generaliza la versión estándar al caso no lineal. Además de las referencias citadas anteriormente, se tuvo en cuenta la exposición de Welch and Bishop (1995). Una vez encontradas las ecuaciones del filtro EKF, señalaremos los operadores que, en el caso estándar, son simplemente lineales. En muchas ocasiones, en lugar de asimilar la medida en un momento determinado, necesitamos hacerlo en diferentes instantes, de forma secuencial.

El filtro de Kalman aplica el análisis visto anteriormente en cada uno de estos momentos t_k

$$x_k^a = x_k^b + K_k(x_k^o - \mathcal{H}_k(x_k^b)) \quad (4.47)$$

La matriz de ganancia de Kalman

$$K_k = P_k^b H_k^T [H_k P_k^b H_k^T + R_k]^{-1} \equiv [(P_k^b)^{-1} + H_k^T R_k^{-1} H_k]^{-1} H_k^T R_k^{-1} \quad (4.48)$$

La matriz covarianza del error de análisis

$$P_k^a = (I - K_k H_k) P_k^b (I - K_k H_k)^T + K_k R_k K_k^T \quad (4.49)$$

Queremos generar P_k^b de forma óptima.

La mejor estimación del estado a tiempo t_k viene dada por la propagación del análisis en t_{k-1}

$$x_k^b = \mathcal{M}_{t_{k-1} \rightarrow t_k}(x_{k-1}^a) \quad (4.50)$$

Según esto, ¿qué matriz de covarianza del error correspondería al background obtenido así?. Veamos qué ecuaciones de evolución aparecen. Partimos del error asociado al background, que sería

$$\epsilon_k^b = \mathcal{M}_{t_{k-1} \rightarrow t_k}(x_{k-1}^a) - x_k \quad (4.51)$$

Ahora escribimos $x_{k-1}^a = x_{k-1} + \epsilon_{k-1}^a$ y suponemos que ϵ_{k-1}^a es suficientemente pequeño para que la siguiente aproximación lineal sea válida

$$\mathcal{M}_{t_{k-1} \rightarrow t_k}(x_{k-1}^a) \approx \mathcal{M}_{t_{k-1} \rightarrow t_k}(x_{k-1}) + M_{t_{k-1} \rightarrow t_k} \epsilon_{k-1}^a \quad (4.52)$$

Entonces

$$\epsilon_k^b = \mathcal{M}_{t_{k-1} \rightarrow t_k}(x_{k-1}) + M_{t_{k-1} \rightarrow t_k} \epsilon_{k-1}^a - x_k \quad (4.53)$$

$$= M_{t_{k-1} \rightarrow t_k} \epsilon_{k-1}^a + \eta_k \quad (4.54)$$

donde $\eta_k = \mathcal{M}_{t_{k-1} \rightarrow t_k}(x_{k-1}) - x_k$ es el error del modelo

Suponemos $\overline{\epsilon_{k-1}^a} = \overline{\eta_k} = 0 \Rightarrow \overline{\epsilon_k^b} = 0$

La covarianza del error del background será

$$\overline{\epsilon_k^b (\epsilon_k^b)^T} = \overline{(M_{t_{k-1} \rightarrow t_k} \epsilon_{k-1}^a + \eta_k)(M_{t_{k-1} \rightarrow t_k} \epsilon_{k-1}^a + \eta_k)^T} \quad (4.55)$$

Asumiendo que los errores del análisis y el modelo no están correlacionados, tendríamos

$$\overline{\epsilon_k^b (\epsilon_k^b)^T} = M_{t_{k-1} \rightarrow t_k} \overline{\epsilon_{k-1}^a (\epsilon_{k-1}^a)^T} M_{t_{k-1} \rightarrow t_k}^T + \overline{\eta_k \eta_k^T} \quad (4.56)$$

es decir,

$$P_k^b = M_{t_{k-1} \rightarrow t_k} P_{k-1}^a M_{t_{k-1} \rightarrow t_k}^T + Q_k \quad (4.57)$$

donde $Q_k = \overline{\eta_k \eta_k^T}$ es la matriz de covarianza del error del modelo

Resumiendo, las ecuaciones del filtro de Kalman extendido son

$$x_k^b = \mathcal{M}_{t_{k-1} \rightarrow t_k}(x_{k-1}^a) \quad (4.58)$$

$$P_k^b = M_{t_{k-1} \rightarrow t_k} P_{k-1}^a M_{t_{k-1} \rightarrow t_k}^T + Q_k \quad (4.59)$$

$$K_k = P_k^b H_k^T [H_k P_k^b H_k^T + R_k]^{-1} \quad (4.60)$$

$$x_k^a = x_k^b + K_k (x_k^o - \mathcal{H}_k(x_k^b)) \quad (4.61)$$

$$P_k^a = (I - K_k H_k) P_k^b (I - K_k H_k)^T + K_k R_k K_k^T \quad (4.62)$$

En el caso estándar, \mathcal{H}_k y $\mathcal{M}_{t_{k-1} \rightarrow t_k}$ son lineales.

4.2.1.2. EnKF

Cuando la dimensión del espacio de fases es tan grande, el cálculo de las covarianzas del error resulta muy costoso computacionalmente. En este sentido, el EnKF (Ensemble Kalman Filter) ofrece una alternativa que aligera notablemente el proceso de cálculo (Kalnay, 2003).

En lugar de actualizar la covarianza del error en cada paso, que equivaldría a realizar un número de integraciones del modelo del orden de la dimensión, vamos a añadir perturbaciones aleatorias a las observaciones asimiladas y estimaremos la covarianza del error de predicción a partir de ese conjunto. Es decir, propagamos el análisis generado con cada observación perturbada

$$x_{k,m}^b = \mathcal{M}_{t_{k-1} \rightarrow t_k}(x_{k-1,m}^a) \quad (4.63)$$

y calculamos la covarianza del error haciendo uso de las distintas evoluciones de los miembros del conjunto, mediante la fórmula

$$P_k^b = \frac{1}{M-1} \sum_{m=1}^{M-1} (x_{k,m}^b - \overline{x_{k,m}^b})(x_{k,m}^b - \overline{x_{k,m}^b})^T \quad (4.64)$$

En lugar de utilizar las ecuaciones de evolución del apartado anterior, que involucraban el operador tangente y adjunto (M , M^T , H y H^T), se utilizan las aproximaciones siguientes

$$P_k^b H_k^T \approx \frac{1}{M-1} \sum_{m=1}^{M-1} (x_{k,m}^b - \overline{x_{k,m}^b})(\mathcal{H}(x_{k,m}^b) - \overline{\mathcal{H}(x_{k,m}^b)})^T \quad (4.65)$$

$$H_k P_k^b H_k^T \approx \frac{1}{M-1} \sum_{m=1}^{M-1} (\mathcal{H}(x_{k,m}^b) - \overline{\mathcal{H}(x_{k,m}^b)})(\mathcal{H}(x_{k,m}^b) - \overline{\mathcal{H}(x_{k,m}^b)})^T \quad (4.66)$$

4.2.2. Asimilación de datos variacional

4.2.2.1. 3DVAR

Exponemos aquí la formulación Bayesiana del método variacional en tres dimensiones, basada en (Lewis et al., 2006).

Los componentes básicos de esta formulación son

- i. **El estado real** x

Sea $x \in \mathbb{R}^n$ la variable que denota el estado real del campo (temperatura, presión, etc.) sobre nuestra rejilla. \mathbb{R}^n es el espacio del modelo o el espacio de la rejilla. La dimensión n de dicho espacio puede ser de orden 10^7 . Nuestro objetivo es estimar este estado.

ii. **El background** x_b

Es la información que poseemos a priori del estado real del sistema, podría decirse que es el pronóstico previo. Denotamos este campo por $x_b \in \mathbb{R}^n$

$$\epsilon_b = x_b - x \quad (4.67)$$

es el error del background. Se asume que tiene media nula $\overline{\epsilon_b} = 0$ y, su covarianza espacial viene dada por

$$P^b = \overline{\epsilon_b \epsilon_b^T} \quad (4.68)$$

donde $P^b \in \mathbb{R}^{n \times n}$ es simétrica y definida positiva.

Vamos a asumir que el estado real tiene una distribución normal $x \sim \mathcal{N}(x_b, P^b)$

$$p(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |P^b|^{\frac{1}{2}}} \exp[-J_b(x)] \quad (4.69)$$

donde J_b es

$$J_b(x) = \frac{1}{2} (x - x_b)^T (P^b)^{-1} (x - x_b) \quad (4.70)$$

la función de coste asociada al background.

iii. **La observación** x_o

Sea $x_o \in \mathbb{R}^m$ la observación del estado desconocido y \mathbb{R}^m el espacio de observación. Sea

$$x_o = \mathcal{H}(x) + v \quad (4.71)$$

donde $\mathcal{H} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathcal{H}(x) = (\mathcal{H}_1(x), \mathcal{H}_2(x), \dots, \mathcal{H}_m(x))^T$ y $v \in \mathbb{R}^m$ es el vector de ruido asociado a la medida. Se supone

$$E(v) = 0 \quad y \quad R = \text{Cov}(v) = \overline{vv^T} \quad (4.72)$$

$R \in \mathbb{R}^{m \times m}$ es una matriz conocida y definida positiva. Se entiende que no hay correlación entre el ruido v y x ni entre v y x_b

$$E[v \epsilon_b^T] = E[\epsilon_b v^T] = E[v \epsilon^T] = E[\epsilon v^T] = 0 \quad (4.73)$$

Se asume que $v \sim \mathcal{N}(0, R)$. por tanto, $x_o \sim \mathcal{N}(\mathcal{H}(x), R)$

$$p(x_o|x) = \frac{1}{(2\pi)^{\frac{m}{2}} |R|^{\frac{1}{2}}} \exp[-J_o(x)] \quad (4.74)$$

y

$$J_o(x) = \frac{1}{2}(x_o - \mathcal{H}(x))^T R^{-1}(x_o - \mathcal{H}(x)) \quad (4.75)$$

la función de coste asociada a la observación.

iv. La regla de Bayes

Pretendemos encontrar la distribución $p(x|x_o)$, aplicando el teorema de Bayes, tenemos

$$p(x|x_o) = \frac{p(x, x_o)}{p(x_o)} = \frac{p(x_o|x)p(x)}{p(x_o)} \quad (4.76)$$

$p(x_o)$ viene dado por

$$p(x_o) = \int_{\mathbb{R}^n} p(x_o|x)p(x)dx \quad (4.77)$$

que puede considerarse como constante de normalización y denotamos por $C = [p(x_o)]^{-1}$, resultando

$$p(x_o|x) = \frac{C}{(2\pi)^{\frac{m}{2} + \frac{n}{2}} |R|^{\frac{1}{2}} |P^b|^{\frac{1}{2}}} \exp[-(J_o(x) + J_b(x))] \quad (4.78)$$

Por lo tanto, la máxima probabilidad -a posteriori- de que, hallándose el sistema en el estado x , observemos x_o , la obtenemos dando con el estado x que minimiza la función coste $J(x)$

$$J(x) = J_o(x) + J_b(x) = \frac{1}{2}(x_o - \mathcal{H}(x))^T R^{-1}(x_o - \mathcal{H}(x)) + \frac{1}{2}(x - x_b)^T (P^b)^{-1}(x - x_b) \quad (4.79)$$

Este problema es de gran escala debido a que la dimensión, generalmente, será muy alta. Abordar el problema sin calcular la derivada suele ser muy lento, ya que evaluar la función da poca información acerca de la forma de la misma: $J(x + \delta v) - J(x)$ solo da información de una componente del gradiente y tiene $\approx 10^8$. Necesitamos conocer el gradiente para saber en que dirección debemos ir para encontrar el mínimo. De ahí que los algoritmos que se usan en la practica calculan el gradiente. Sabemos que en el mínimo de esta función coste, su gradiente será nulo

$$\nabla J(x) = (P^b)^{-1}(x - x_b) + H^T R^{-1}(\mathcal{H}(x) - x_o) = 0 \quad (4.80)$$

Como comentario, decir que si \mathcal{H} es lineal

$$\mathcal{H}(x) = \mathcal{H}(x_b) + H(x - x_b) \quad (4.81)$$

podemos escribir el gradiente como

$$(P^b)^{-1}(x - x_b) + H^T R^{-1}(\mathcal{H}(x_b) + H(x - x_b) - x_o) = 0 \quad (4.82)$$

y manipulando un poco esta expresión, recuperamos la ecuación que encontramos exigiendo un varianza mínima.

Vamos a hacer un repaso breve de los métodos más utilizados para minimizar la función coste. En primer lugar, citaremos el más sencillo de todos: el gradiente descendente. Su algoritmo consistiría en los siguientes pasos:

- Partimos de un punto inicial x_0 , elegido aleatoriamente
- Sustituimos dicho punto de partida por $x_{k+1} := x_k - \alpha \nabla J(x_n)$
- Repetimos el proceso hasta que J sea suficientemente pequeño.

Este algoritmo no es eficiente cuando las curvas de nivel de la función no tienen simetría esférica, ya que no usa la información que da la curvatura. El algoritmo más sencillo que usa dicha información es el de Newton.

4.2.2.2. 4DVAR

En el apartado anterior, suponíamos que las observaciones, el análisis y el background eran válidos en el mismo instante de tiempo. De esa manera, era suficiente considerar la propagación espacial (o interpolación) de los datos de la rejilla a los puntos donde se realizaba la medida ya que ocurrían simultáneamente. En el caso 4DVAR, esta exigencia se relaja y consideramos la posibilidad de que las observaciones se realicen en diferentes momentos. Necesitamos pues, propagar los campos temporalmente para realizar las comparaciones oportunas. Este hecho conlleva un aumento de complejidad en el operador \mathcal{H} aunque, formalmente, el problema tiene la misma forma que el 3DVAR. Para resaltar la capacidad del operador de medida de interpolar en el tiempo, lo denotamos como \mathcal{G}

$$J(x) = \frac{1}{2}(x_o - \mathcal{G}(x))^T R^{-1}(x_o - \mathcal{G}(x)) + \frac{1}{2}(x - x_b)^T (P^b)^{-1}(x - x_b) \quad (4.83)$$

Tiene sentido agrupar medidas en sub-vectores x_{o_k} , correspondientes al mismo instante de tiempo t_k . A su vez, también parece razonable pensar que los errores de las observaciones no estarán correlacionados en el tiempo. De esa manera, la matriz R es diagonal por bloques y podemos reescribir la ecuación de la siguiente manera

$$J(x) = \frac{1}{2}(x_{o_k} - \mathcal{G}_k(x))^T R_k^{-1}(x_{o_k} - \mathcal{G}_k(x)) + \frac{1}{2}(x - x_b)^T (P^b)^{-1}(x - x_b) \quad (4.84)$$

pudiendo expresar cada operador generalizado como

$$\mathcal{G}_k = \mathcal{H}_k \mathcal{M}_{t_o \rightarrow t_k} \quad (4.85)$$

donde $\mathcal{M}_{t_o \rightarrow t_k}$ representa la integración del modelo de predicción entre t y t_o y \mathcal{H}_k representa la interpolación y transformación del espacio del modelo al espacio de medida. La integración puede factorizarse en una sucesión de integraciones más cortas

$$\mathcal{M}_{t_o \rightarrow t_k} = \mathcal{M}_{t_{k-1} \rightarrow t_k} \mathcal{M}_{t_{k-2} \rightarrow t_{k-1}} \cdots \mathcal{M}_{t_1 \rightarrow t_2} \mathcal{M}_{t_o \rightarrow t_1} \quad (4.86)$$

Los estados del modelo x_k definidos en el instante t_k , son

$$x_k = \mathcal{M}_{t_o \rightarrow t_k}(x_0) \quad (4.87)$$

$$= \mathcal{M}_{t_{k-1} \rightarrow t_k}(x_{k-1}) \quad (4.88)$$

así, podemos escribir la función de coste como

$$\begin{aligned} J(x_0, x_1, \dots, x_k) &= \frac{1}{2}(x_b - x_0)^T (P^b)^{-1} (x_b - x_0) \\ &+ \frac{1}{2} \sum_{k=0}^K (x_{o_k} - \mathcal{H}_k(x_k))^T R_k^{-1} (x_{o_k} - \mathcal{H}_k(x_k)) \end{aligned} \quad (4.89)$$

Notemos que, introduciendo los vectores x_k , convertimos el problema de minimización sin restricciones

$$x_a = \arg \min_x (J(x_0)) \quad (4.90)$$

en uno con restricciones

$$x_a = \arg \min_{x_0} (J(x_0, x_1, \dots, x_k)) \quad (4.91)$$

$$\text{donde } x_k = \mathcal{M}_{t_{k-1} \rightarrow t_k}(x_{k-1}) \quad (4.92)$$

Este problema se conoce como 4DVAR con restricción fuerte y en él se asume la validez del modelo perfecto $x_o^* = \mathcal{H}(x)$. Así definido, 4DVAR determina el análisis en cada punto de rejilla y para cada valor del tiempo dentro de la ventana del análisis. Constituye, por tanto, una integración del modelo de predicción (una trayectoria).

4.2.3. Ejemplo: asimilación de datos en el modelo de Lorenz

En este apartado, aplicamos en el modelo de Lorenz 63 algunos de los métodos de asimilación vistos anteriormente. El procedimiento general será integrar el modelo para una condición

inicial dada y considerar esa evolución como la real, es decir, como la sucesión de los estados reales. A partir de dicha trayectoria, simulamos las medidas realizadas en determinados instantes de tiempo, simplemente, sumando una cantidad aleatoria al estado real. El propio método de Euler, que nos sirve para integrar el modelo, nos proporciona una forma de simular nuestro pronóstico, este 'background' estará determinado por el análisis en $(t - 1)$.

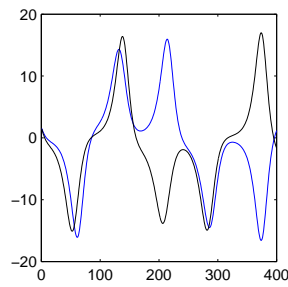


FIGURA 4.1: Evolución real y de la predicción de la componente X sin asimilación

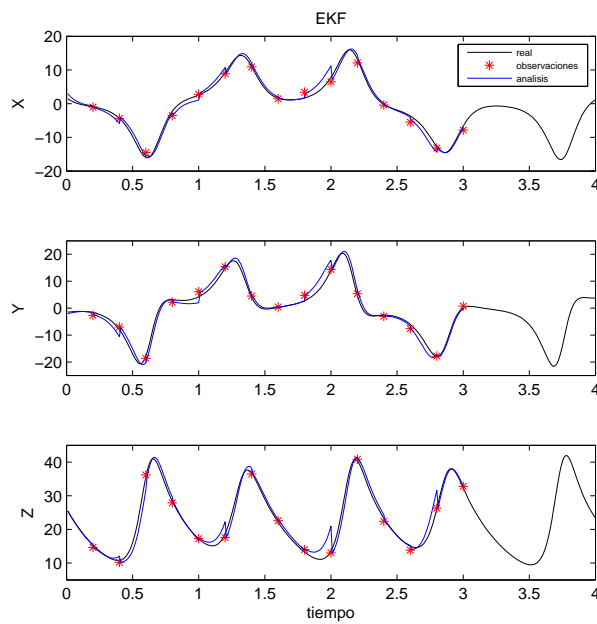


FIGURA 4.2: Evolución real y de la predicción con asimilación EKF

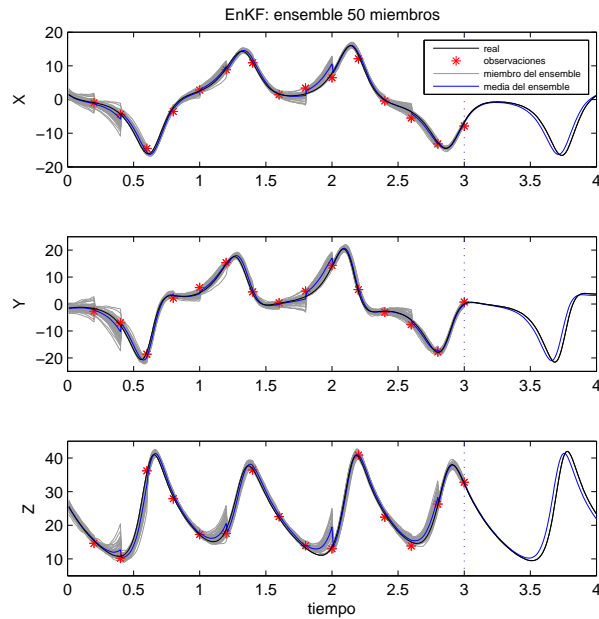


FIGURA 4.3: Evolución real y de la predicción con asimilación EnKF

Si no utilizásemos ningún método de asimilación, estaríamos integrando, sin más, dos trayectorias con diferentes condiciones iniciales. Vemos en el gráfico de la izquierda que, en el modelo de Lorenz, debido a su no linealidad, las trayectorias se separan al cabo de, relativamente, pocos pasos de integración. Sin embargo, veremos en los siguientes ejemplos como, gracias a los esquemas de asimilación, ambas se mantienen cerca una de la otra.

Tal es el caso del filtro de Kalman extendido que se muestra a continuación. Vemos que la incorporación de las medidas mantiene la trayectoria del análisis muy cerca de la evolución real del sistema.

4.3. Asimilación de datos en el modelo WRF

Como hemos visto anteriormente, el objetivo principal de los sistemas de asimilación de datos es producir una estimación óptima del estado real de la atmósfera en el momento del análisis. El problema variacional se puede resumir como la búsqueda de ese objetivo,

encontrar el análisis x mediante la minimización por iteración de la función coste

$$J(x) = J_b(x) + J_o(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y - y_o)^T (E + F)^{-1}(y - y_o) \quad (4.93)$$

Dados los datos *a priori* que describen el estado de la atmósfera: el llamado 'first guess' o 'background' x_b y las observaciones y_o , nuestro objetivo representa el estado de máxima probabilidad *a posteriori*.

En este trabajo, hicimos uso del módulo WRFDA del modelo meteorológico para asimilar los datos de radiancia mediante el método 3DVAR. La implementación del algoritmo en el modelo WRF se describe en el artículo (Barker et al., 2004). Aquí, simplemente, presentamos la metodología que seguimos para realizar las simulaciones.

4.3.1. Metodología

- La función coste $J(x)$ la calcula el módulo WRFDA
- El análisis es la salida del WRFDA
- El background x_b lo introducimos nosotros, proviene de la integración del WRF partiendo de x_a
- B la matriz de covarianza del error de x_b , lo introducimos nosotros tras obtenerlo del programa 'gen_be'
- Las observaciones x_o de los campos fundamentales las obtenemos a partir de los datos de radiancia. Es necesario el uso de un modelo de transferencia radiativa para derivarlas.
- El operador H está incluido en el WRFDA
- R la matriz de covarianza del error la introducimos nosotros

4.3.2. Sensores ATOVS utilizados

4.3.2.1. HIRS/3, HIRS/4

El sensor HIRS (High Resolution Infrared Radiation Sounder) dispone de un escáner de espejo elíptico de 56 pasos de 1.8 grados; escanea líneas perpendiculares a su trayectoria (cross-track scanner); y toma datos en 20 bandas espectrales que permiten el cálculo del perfil vertical de temperaturas, desde la superficie de la Tierra hasta unos 40 km de altura.

Proporciona datos multiespectrales en el canal visible de $0.69 \mu\text{m}$, en siete canales de onda corta entre 3.7 y $4.6 \mu\text{m}$, y en doce canales de onda larga (de 6.5 a $15 \mu\text{m}$). Lleva montado un telescopio con un dispositivo que rota permitiendo la interposición de alguno de los veinte filtros individuales de que dispone. Escanea la escena en dirección perpendicular a la trayectoria con un campo de visión instantáneo (IFOV) para cada canal de, aproximadamente, 1.4 grados en el visible e infrarrojos de onda corta y 1.3 grados en la banda de IR de onda larga. Se encuentra a una altura de 833 kilómetros, por lo que abarca una extensión de unos 20.3 y 18.9 kilómetros por píxel, respectivamente, en el nadir.

Posee tres detectores, un fotodiodo de silicio para el visible, y uno de antimonio de indio y otro de mercurio, cadmio y telurio (HgCdTe) para el infrarrojo de onda corta y larga respectivamente. La calibración de los sensores IR la realiza apuntando a dos blancos radiométricos: uno caliente, típicamente llamado cuerpo negro y que va montado en el satélite, y el frío, apuntando al espacio.

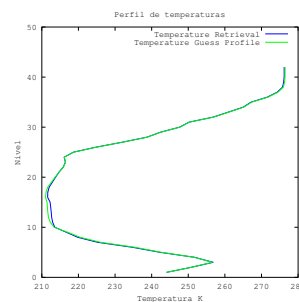


FIGURA 4.4: Perfil de temperaturas medido por el sensor HIRS

4.3.2.2. AMSU-A, AMSU-B

AMSU-A

El sistema AMSU-A (Advanced Microwave Sounding Unit) está compuesto por dos módulos separados AMSU-A1 y AMSU-A2. El AMSU-A1 es un radiómetro multicanal (15 bandas) que se usa para tomar medidas de los perfiles de temperatura y proporcionar datos del contenido de agua en la atmósfera en todas sus formas a excepción de pequeñas partículas de hielo, que son transparentes a las frecuencias de microondas. AMSU-A es cross-track, es decir, escanea la escena en dirección perpendicular a la trayectoria tomando datos del perfil de temperaturas desde la superficie de la tierra hasta una altura de 45 Km. Su campo de visión instantáneo es de unos 48.05 km en el nadir.

AMSU-B

El sistema AMSU-B es un radiómetro con 5 canales (del 16 al 20) en la zona de las microondas. Su cometido es medir la radiación procedente de varias capas distintas de la atmósfera con el fin estimar los perfiles de humedad. Los canales 18, 19 y 20 cubren la zona del espectro de máxima absorción debida al vapor de agua (183 GHz); el 16 y 17, a 89 GHz y 150 GHz son capaces de penetrar hasta la superficie terrestre.

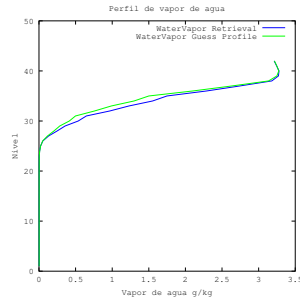


FIGURA 4.5: Perfil de vapor de agua medido por el sensor AMSU

Al igual que los anteriores, es un sensor de tipo cross-track. Es capaz de tomar datos de 90 celdas, cubriendo en cada escaneo 50 grados a cada lado de la trayectoria. La resolución, en diámetro es de 16.3 km en el nadir. Se puede ampliar la información en *The NOAA KLM User's Guide* (revisión de febrero de 2009). Al igual que los anteriores, es un sensor de tipo cross-track. Es capaz de tomar datos de 90 celdas, cubriendo en cada escaneo 50 grados a cada lado de la trayectoria. La resolución, en diámetro es de 16.3 km en el nadir.

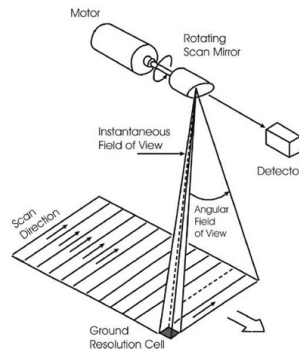


FIGURA 4.6: Sensor cross-track, IFOV, SWATH

Capítulo 5

Aprendizaje automático

5.1. Introducción

El primer objetivo que nos planteamos fue buscar la mejor manera de reproducir las reflectancias que el modelo de transferencia radiativa nos proporcionaba. Tal como lo hemos formulado, se trata de un problema de regresión; partimos de un conjunto de datos atmosféricos y de reflectancias aparentes en un punto y su entorno, y queremos reproducir el valor de la reflectancia real del blanco que hallaríamos con el modelo 6S.

Nuestro segundo objetivo fue clasificar los tipos de atmósferas según su influencia sobre la transferencia de la radiación en las longitudes de onda de interés en teledetección. Para ello, planteamos grados de influencia atmosférica, según la reflectancia corregida difriese de la aparente menos o más de k veces la varianza en las correcciones de la banda. De igual manera, buscaremos el algoritmo idóneo para realizar la tarea y, paralelamente, las variables predictoras de mayor peso en el proceso.

Abordamos este estudio en el marco del aprendizaje automático (Machine Learning) o aprendizaje estadístico. Ambos términos dan nombre a dos ramas de las ciencias de computación y de estadística que, aunque no equivalentes, abordan los mismos tipos de problemas: regresión, clasificación y clustering, principalmente, y comparten muchas de las técnicas de resolución. Podemos decir que ambas consisten en un amplio conjunto de herramientas destinadas a extraer información de los datos (James et al., 2013).

La teoría que se presenta en este capítulo está basada en las exposiciones de (Hastie et al., 2009), (Pedregosa et al., 2011), (Bishop et al., 2006), (Gutiérrez, 2004) y (Duda et al., 1999), además de las citadas en el texto.

5.2. Aprendizaje supervisado

Cuando, en el conjunto de datos de partida, conocemos los resultados o salidas del sistema, decimos que el aprendizaje está supervisado. Los dos grandes tipos de problemas de esta clase son la regresión y la clasificación.

5.2.1. Regresión

En este apartado vamos a presentar, de forma breve, los algoritmos de regresión que se compararon con el fin de seleccionar el método óptimo.

5.2.1.1. Regresión lineal

La regresión lineal ajusta los parámetros del modelo lineal con coeficientes $\theta = (\theta_1, \dots, \theta_p)$ de forma que la suma de los cuadrados de las diferencias entre las observaciones X y las salidas y de dicho modelo sea mínima cuando se evalúan sobre el conjunto de datos de entrenamiento. Matemáticamente, la regresión lineal resuelve el problema de la forma

$$\min_{\theta} \|X\theta - y\|^2 \quad (5.1)$$

Donde $\|\cdot\|$ es la norma euclídea.

5.2.1.2. Ridge

El método de regresión Ridge añade al problema anterior una penalización por el tamaño de los coeficientes. Como antes, pretendemos minimizar la suma de los errores cometidos cuando estimamos las salidas del modelo en cada elemento del conjunto de entrenamiento pero, además, deseamos que dichos coeficientes no se hagan demasiado grandes.

$$\min_{\theta} \|X\theta - y\|^2 + \alpha\|\theta\|^2 \quad (5.2)$$

Vemos que el parámetro α regula la complejidad del modelo. Si α es grande, la penalización también y resultará un modelo más simple, es decir, con alguno de sus coeficientes cercanos a cero.

Este problema puede reformularse como

$$\min_{\theta} \|X\theta - y\|^2 \quad \text{sujeto a} \quad \|\theta\|^2 \leq s \quad (5.3)$$

es decir, la minimización del RSS (Residual sum of squares) sobre el conjunto $\|\theta\|^2 \leq s$, que geoméricamente representan hiper-círculos de radio s .

5.2.1.3. Lasso

La regresión Lasso sigue la misma idea que la regresión Ridge. La diferencia está en la función de penalización o, visto de otra forma, el conjunto sobre el que buscamos la condición de mínimo. La formulación del problema es la siguiente

$$\min_{\theta} \frac{1}{2m} \|X\theta - y\|_2^2 + \alpha \|\theta\|_1 \quad (5.4)$$

Vemos que reemplaza el conjunto de puntos definidos por la circunferencia a los definidos por los valores absolutos, es decir la norma $\ell_1 \equiv \|\cdot\|_1$. De esta manera, consigue que las posibilidades de encontrar la condición de mínimo siendo algunos coeficientes nulos aumenten. En definitiva, el problema replanteado como la minimización sobre el subconjunto podría verse como la búsqueda del primer punto del subconjunto que contacta con las elipses que definen las curvas de nivel de la función coste. Como la definición del subconjunto mediante la norma ℓ_1 sitúa los vértices del mismo sobre los ejes, es más probable que dicha condición se satisfaga en los vértices, anulándose, así, alguno de los coeficientes.

5.2.1.4. Bayesian Ridge

La regresión Ridge Bayesiana se usa para incorporar los parámetros de regularización en el momento de la estimación. Es decir, en lugar de fijarlos a priori, los estimaremos en el ajuste a los datos.

Suponemos que el error cometido con nuestro modelo lineal $y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$, tiene una distribución normal $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$

$$p(\epsilon^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\epsilon^{(i)})^2}{2\sigma^2}\right) \quad (5.5)$$

Por tanto, la respuesta $y^{(i)}$ tendrá una distribución del mismo tipo

$$p(y^{(i)}|x^{(i)}, \theta, \beta) = \mathcal{N}(y^{(i)}|\theta^T x^{(i)}, \beta^{-1}) \quad (5.6)$$

donde β es un parámetro de precisión y corresponde a la inversa de la varianza. Usamos los datos de entrenamiento para hallar los parámetros desconocidos, θ y β . Ya que las muestras tomadas son independientes, tenemos

$$p(y^{(i)}|x^{(i)}, \theta, \beta) = \prod_{i=1}^m \mathcal{N}(y^{(i)}|\theta^T x^{(i)}, \beta^{-1}). \quad (5.7)$$

Buscamos los parámetros del modelo que hacen máxima la probabilidad de reproducir las salidas del conjunto de entrenamiento. Sustituyendo la distribución normal y tomando logaritmos a cada lado, ya que, maximizar la probabilidad o su logaritmo, es equivalente, tendríamos

$$\ln p(y^{(i)}|x^{(i)}, \theta, \beta) = -\frac{\beta}{2} \sum_{i=1}^m \{\theta^T x^{(i)} - y^{(i)}\}^2 + \frac{m}{2} \ln \beta - \frac{m}{2} \ln(2\pi). \quad (5.8)$$

Los dos últimos términos de la expresión anterior no dependen de θ , por tanto, maximizar la parte de la derecha se reduce a encontrar el máximo del logaritmo o el mínimo de este multiplicado por (-1). De ahí, se deduce la equivalencia con el método de los mínimos cuadrados.

Una vez hallados los θ_M que maximizan la probabilidad, podemos buscar el β con el mismo criterio. Derivando en la expresión anterior, tendríamos

$$\frac{1}{\beta_M} = \frac{1}{m} \sum_{i=1}^m \{\theta^T x^{(i)} - y^{(i)}\}^2. \quad (5.9)$$

Habiendo encontrado los parámetros óptimos, podemos predecir las salidas para nuevos datos x .

Si ahora introducimos una distribución a priori para los coeficientes θ , por ejemplo, la normal

$$p(\theta|\alpha) = \mathcal{N}(\theta|0, \alpha^{-1}\mathbf{I}) = \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2} \exp\left\{-\frac{\alpha}{2}\theta^T\theta\right\} \quad (5.10)$$

M , es el orden del polinomio, 1 en este caso. Podríamos aplicar el teorema de Bayes, y tendríamos

$$p(\theta|x, y, \alpha, \beta) \propto p(y|x, \theta, \beta)p(\theta|\alpha) \quad (5.11)$$

Buscando el máximo de esta probabilidad a posteriori, podemos determinar los θ dado el conjunto de entrenamiento. Esta técnica se denomina MAP (Maximum A Posteriori). Tomando

menos el logaritmo de la expresión anterior y combinando con 5.8 y 5.10, encontraríamos que, dicha técnica, es equivalente a minimizar

$$\frac{\beta}{2} \sum_{i=1}^m \{\theta^T x^{(i)} - y^{(i)}\}^2 + \frac{\alpha}{2} \theta^T \theta \quad (5.12)$$

Hemos comprobado que maximizar la probabilidad a posteriori dada la distribución gaussiana de los parámetros θ con precisión β^{-1} siendo $\lambda = \alpha/\beta$ es equivalente a la regresión de Ridge.

5.2.1.5. SVR (Support Vector Regression)

La teoría de los vectores soporte se expone en la sección de clasificación. Aquí, presentamos la formulación del problema de regresión en el espacio de características ampliado \mathcal{H} , siendo la aplicación que va del espacio original a este:

$$\begin{aligned} \phi: \mathcal{X} &\rightarrow \mathcal{H} \\ x &\rightarrow \mathbf{x} := \phi(x). \end{aligned}$$

Dado el conjunto de entrenamiento $\{x_i, y_i\}_{i=1}^m$, buscamos una función lineal $y_i = \theta^T \phi(x_i) + b$. La formulación estándar de SVR usa la función de coste de Vapnik que desestima los errores hasta ε . Necesitamos encontrar el mínimo

$$\min_{\theta} \frac{1}{2} \|\theta\|^2 + C \sum_i (\xi_i + \xi_i^*) \quad (5.13)$$

sujeto a

$$y_i - \theta^T \phi(x_i) - b \leq \varepsilon + \xi_i \quad \forall i = 1, \dots, m \quad (5.14)$$

$$\theta^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \quad \forall i = 1, \dots, m \quad (5.15)$$

$$\xi_i, \xi_i^* \geq 0 \quad \forall i = 1, \dots, m \quad (5.16)$$

donde $i(\xi_i, \xi_i^*)$ representan los márgenes aceptados para tratar con los datos que poseen error mayor que ε .

5.2.1.6. Red neuronal

Las redes neuronales son modelos computacionales que pretenden imitar el funcionamiento del cerebro de los seres vivos. Como tales, están compuestos por unidades simples de proceso, las neuronas que, esquemáticamente, se representan por

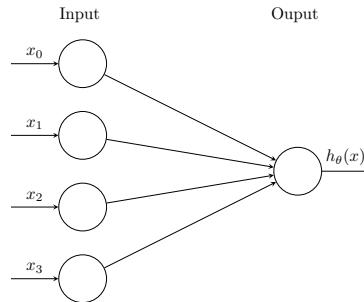


FIGURA 5.1: Neurona

donde las entradas x y los parámetros θ , tienen la forma

$$x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \theta = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}$$

y donde $h_\theta(x)$ es la función de activación que, por ejemplo, podría ser la función sigmoide:

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

Además de la función sigmoide o logística, otras funciones de activación muy utilizadas son:

- *Funciones sigmoidales:*
 1. La función logística, con rango de 0 a 1, $h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$.
 2. La tangente hiperbólica, con rango de -1 a 1, $h_\theta(x) = \tanh(\theta^T x)$.
- *Funciones kernel*, se localizan alrededor de un punto, como la gaussiana.
- *Funciones lineales*, con rango infinito.

Esta función de activación, simplemente, produce una salida para las entradas x . Las neuronas suelen agruparse en capas; la capa de entrada recibe los datos, y la de salida devuelve los resultados. Entre ellas, se encuentran las capas ocultas. El número de neuronas en las diferentes capas, la cantidad de capas ocultas, las funciones de activación y la propia topología

de las redes son variables, es decir, se definen dependiendo del problema que se esté tratando. En general, se decide qué configuración tomar mediante la validación cruzada. En la siguiente figura se muestran algunos ejemplos de topologías diferentes.

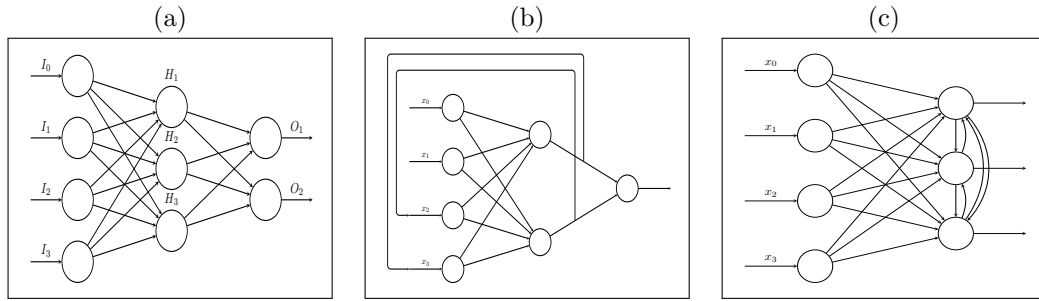


FIGURA 5.2: (a) Red neuronal multicapa, (b) Red recurrente, (c) Red competitiva.

Supongamos una red con 4 neuronas en la capa de entrada, una capa oculta con 3 neuronas (más una entrada para el bias) y una salida, se esquematiza en la figura 5.3. Consideremos como función de activación la sigmoide. Las funciones de activación en la capa oculta serían las siguientes:

$$\begin{aligned}
 a_1^{(2)} &= g(\theta_{10}^{(1)} x_0 + \theta_{11}^{(1)} x_1 + \theta_{12}^{(1)} x_2 + \theta_{13}^{(1)} x_3) \\
 a_2^{(2)} &= g(\theta_{20}^{(1)} x_0 + \theta_{21}^{(1)} x_1 + \theta_{22}^{(1)} x_2 + \theta_{23}^{(1)} x_3) \\
 a_3^{(2)} &= g(\theta_{30}^{(1)} x_0 + \theta_{31}^{(1)} x_1 + \theta_{32}^{(1)} x_2 + \theta_{33}^{(1)} x_3)
 \end{aligned}
 \tag{5.17}$$

La salida producida sería

$$h_{\theta}(x) = g(\theta_{10}^{(2)} a_0^{(2)} + \theta_{11}^{(2)} a_1^{(2)} + \theta_{12}^{(2)} a_2^{(2)} + \theta_{13}^{(2)} a_3^{(2)})
 \tag{5.18}$$

Conocido el funcionamiento básico de la red neuronal, falta conocer el modo de estimar los parámetros de la misma. Ese proceso se conoce como el entrenamiento de la red y consiste, esencialmente, en la minimización de la función coste $J(\theta)$

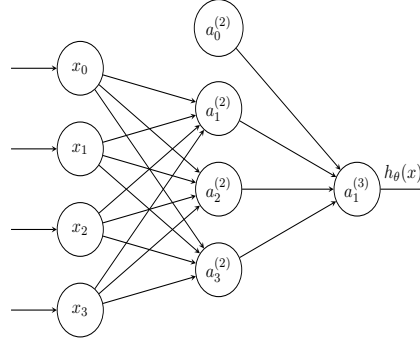


FIGURA 5.3: Red neuronal multicapa

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log((h_\theta(x^{(i)}))_k) + (1 - y_k^{(i)}) \log(1 - (h_\theta(x^{(i)}))_k) \right] \quad (5.19)$$

$$+ \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{n_l} \sum_{j=i}^{n_{l+1}} (\theta_{ij}^{(l)})^2.$$

En la expresión anterior, m es el número de elementos en el conjunto de activación, K el número de neuronas en la capa de salida (igual al número de clases), L el número de capas ocultas y n_l el número de neuronas en la capa l . El primer sumando toma cuenta del error cometido, si la salida y en la neurona k es 1 o 0, penalizamos $\log(h_\theta(x))$ o $(1 - \log(h_\theta(x)))$ respectivamente. El segundo sumando es el término de regularización, penaliza el crecimiento de los parámetros de red. Un λ elevado proporcionaría una red menos compleja o con poca posibilidad de sufrir sobreentrenamiento. Notemos que la regularización no se aplica al término libre (o bias).

Para minimizar la función coste mediante gradiente descendente (u otro método más eficiente), necesitamos conocer el gradiente

$$\frac{\partial J}{\partial \theta_{ij}^{(l)}} \quad \text{siendo} \quad \theta_{ij}^{(l)} \in \mathbb{R} \quad (5.20)$$

En este caso, utilizamos el algoritmo denominado *Backpropagation*, que presentamos a continuación, para computar el gradiente.

Dado un elemento del conjunto de entrenamiento $(x^{(i)}, y^{(i)})$, propagamos hacia adelante los valores de entrada. Podemos pensar que $x^{(i)} = a^{(1)}$ es la activación en la entrada. La

secuencia de cálculos que realiza la red es la siguiente:

$$\begin{aligned}
 a^{(1)} &= x^{(i)} & (5.21) \\
 z^{(2)} &= \theta^{(1)} a^{(1)} \\
 a^{(2)} &= g(z^{(2)}) \quad (\text{más } a_0^{(2)}) \\
 z^{(3)} &= \theta^{(2)} a^{(2)} \\
 a^{(3)} &= g(z^{(3)}) \quad (\text{más } a_0^{(3)}) \\
 z^{(4)} &= \theta^{(3)} a^{(3)} \\
 a^{(4)} &= h_{\theta}(x) = g(z^{(4)})
 \end{aligned}$$

Las activaciones en la primera capa o capa de entrada $a^{(1)}$ son las propias entradas a la red, $x^{(i)}$; en la capa oculta tendremos como señales de activación las resultantes de la función sigmoide $a^{(2)} = \frac{1}{1+e^{-\theta^{(1)} a^{(1)}}}$; en la última capa, tendríamos la activación $a^{(3)} = \frac{1}{1+e^{-\theta^{(2)} a^{(2)}}}$; y en la salida, $\frac{1}{1+e^{-\theta^{(3)} a^{(3)}}}$, según vimos anteriormente.

A continuación, vamos a calcular el error $\delta_j^{(l)}$ en cada nodo j de la capa l partiendo del error en la capa de salida y propagándolo hacia atrás. En la última capa, utilizando las salidas y_k del conjunto de entrenamiento, podemos calcular el error cometido:

$$\delta_k^{(4)} = a_k^{(4)} - y_k \quad (5.22)$$

en la última capa oculta, partiendo del error en la salida, obtendríamos

$$\begin{aligned}
 \delta_i^{(3)} &= [(\theta^{(3)})^T \delta^{(4)}]_i \cdot g'(z^{(3)})_i \quad \text{componente } i\text{-ésima} & (5.23) \\
 \delta_i^{(2)} &= [(\theta^{(2)})^T \delta^{(3)}]_i \cdot g'(z^{(2)})_i \quad \text{componente } i\text{-ésima}
 \end{aligned}$$

Repetiríamos ese procedimiento para todos los elementos del conjunto de entrenamiento $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$. El procedimiento completo se resume en el siguiente esquema

- Definimos los parámetros $(\Delta_{ij}^{(l)})_{k=0} = 0$
- Para $k = 1$ hasta m
 - Tomamos $a^{(1)} = x^{(k)}$
 - Propagamos $a^{(1)}$ hacia adelante para calcular $a^{(l)}$ en las capas $l = 2, 3, \dots, L$
 - Utilizando la salida correspondiente $y^{(k)}$, calculamos el error $\delta^{(L)} = a^{(L)} - y^{(k)}$

- Propagamos hacia atrás el error para computar $\delta^{(L-1)}, \delta^{(L-2)}, \dots, \delta^{(2)}$
- $(\Delta_{ij}^{(l)})_k = (\Delta_{ij}^{(l)})_{k-1} + a_j^{(l)} \delta_i^{(l+1)}$
- $D_{ij}^{(l)} = \frac{1}{m} \Delta_{ij}^{(l)} + \lambda \theta_{ij}^{(l)}$ si $j \neq 0$
- $D_{ij}^{(l)} = \Delta_{ij}^{(l)}$ si $j = 0$
- $\frac{\partial J}{\partial \theta_{ij}^{(l)}} = D_{ij}^{(l)}$

De esa manera obtendríamos el gradiente que se usa en los algoritmos de minimización tales como el gradiente descendente.

Regresión

Hasta ahora, hemos hablado de la función sigmoide como función de activación pero vimos que también existen otras opciones. Haciendo uso de la función lineal en la última capa, podemos abordar el problema de la regresión desde la misma perspectiva de minimización de la función de coste.

5.2.2. Clasificación

A continuación, presentamos los algoritmos que utilizamos para clasificar las atmósferas según la influencia que presentaban sobre la radiación, dentro de la imagen.

5.2.2.1. Vecinos más próximos

En esta clasificación, simplemente asignamos la clase mayoritaria entre los vecinos más próximos. Ajustamos el radio que define la vecindad mediante validación cruzada.

5.2.2.2. Análisis discriminante lineal

En primer lugar, se estudian las distribuciones $p(X|Y = k)$ para cada clase k . Posteriormente, haciendo uso del teorema de Bayes, buscamos las salidas para los valores de entrada de los predictores X : $P(k|X)$. Asignaremos la clase correspondiente a la máxima probabilidad, dividiendo el espacio de características mediante fronteras lineales. En este caso, se supone que las distribuciones $p(X|Y = k)$ son Gaussianas.

5.2.2.3. Árboles de decisión

Este tipo de algoritmos se aplican tanto a problemas de regresión como de clasificación. Segmentan el espacio de características en diferentes regiones simples y asignan, bien el valor medio para la regresión, o bien la clase mayoritaria en la clasificación. De esta forma, se asigna el mismo valor a todos los datos correspondientes a una misma región.

La gran ventaja de los árboles de decisión es que son fácilmente interpretables. La desventaja, es que los resultados no suelen ser tan satisfactorios como los obtenidos con otros algoritmos. Constituyen la base de algoritmos como 'tree-bagging', 'Random Forest' o 'Boosting' que, aun siendo más difíciles de interpretar, suelen dar mejores resultados.

El árbol, normalmente, se construye de arriba hacia abajo. Las regiones en que estratificamos el espacio de características, se denominan 'nodos terminales' u 'hojas' y aparecen en la parte más baja. Los puntos en los que vamos dividiendo el espacio, son los 'nodos internos' y los segmentos que conectan esos nodos, las 'ramas'.

Comenzamos la construcción del árbol dividiendo el espacio en dos bloques. Para ello, consideramos todos los predictores X_j , uno a uno, y todos los umbrales s

$$R_1(j, s) = \{X|X_j < s\} \quad \text{y} \quad R_2(j, s) = \{X|X_j \geq s\} \quad (5.24)$$

de forma que el error cometido sea el mínimo posible

$$\sum_{i:x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \quad (5.25)$$

donde \hat{y}_{R_1} es la respuesta media para todos datos de entrenamiento que caen dentro de la región R_1 y \hat{y}_{R_2} en R_2 .

El proceso de segmentación continúa pero, en lugar de continuar estratificando todo el espacio, solo dividimos una de las dos regiones encontradas en el paso anterior.

Pruning

El proceso anterior suele tener el problema de sobre-entrenamiento (overfitting). Para evitarlo, recurrimos a la poda del árbol. Para ello, construimos un árbol grande T_0 y, posteriormente, vamos obteniendo sub-árboles hasta encontrar aquel que menor RSS produzca. Evaluar todos los posibles es muy costoso computacionalmente, así que, procedemos de la siguiente manera: consideramos una sucesión de valores del parámetro de ajuste α y, para

cada uno de ellos, buscamos aquel sub-árbol $T \subset T_0$ que hace mínima la siguiente suma

$$\sum_{m=1}^{|T|} \sum_{x_i \in R_m} (y_i - \hat{y}_{R_m})^2 + \alpha|T| \quad (5.26)$$

$|T|$ indica el número de hojas o nodos terminales, R_m es la región del espacio de características correspondiente a la hoja m y \hat{y}_{R_m} la respuesta predicha en ese bloque.

Procediendo así, encontraríamos una sucesión de árboles correspondientes a cada α de complejidad decreciente ($\alpha = 0$ corresponde al árbol más complejo). Finalmente, comparamos cada uno de ellos mediante validación cruzada y seleccionamos aquel que menor error de validación genere.

Clasificación

En el caso de la clasificación, se pronostica la clase mayoritaria del bloque correspondiente. El modo de estimación del error también es distinto; la alternativa al RSS es el *ratio de error de clasificación*:

$$E = 1 - \max_k (\hat{p}_{mk}) \quad (5.27)$$

Dado que este no es suficientemente sensible al crecimiento del árbol, se utilizan, generalmente, otros medidores como el índice de Gini

$$G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk}) \quad (5.28)$$

o la entropía cruzada (cross-entropy)

$$D = \sum_{k=1}^K \hat{p}_{mk} \log \hat{p}_{mk} \quad (5.29)$$

ambos índices proporcionan una medida de la pureza de los bloques. Un valor bajo indicaría que están formados, mayoritariamente por elementos de la misma clase.

5.2.2.4. Random Forests

Para entender este algoritmo, necesitamos comentar previamente otro basado en la técnica *'bootstrap'*, el denominado *'tree-bagging'*.

El procedimiento '*Bagging*' o '*Bootstrap aggregation*', se aplica para reducir la varianza de los métodos de aprendizaje estadístico. Recordemos que dado un conjunto de n observaciones independientes Z_1, Z_2, \dots, Z_n , cada una con varianza σ^2 , la varianza de la media será σ^2/n . Por tanto, una manera de reducir la varianza y aumentar la precisión de un algoritmo de aprendizaje, es tomar varios conjuntos de entrenamiento, construir modelos de predicción diferentes usando cada conjunto y, finalmente, promediar las predicciones resultantes. En otras palabras, construimos los modelos $f^1(x), f^2(x), \dots, f^B(x)$ usando B conjuntos separados y promediamos

$$\hat{f}_{\text{avg}}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^b(x) \quad (5.30)$$

En la práctica, generalmente no podemos recurrir a varios conjuntos de entrenamiento. En su lugar, recurrimos a la técnica de '*bootstrap*' tomando repetidas muestras de un único conjunto de entrenamiento. Este es el procedimiento que sigue el algoritmo *tree-bagging*.

Random Forests

Este algoritmo añade una mejora al anterior basada en una modificación que rompe la correlación entre los árboles hallados anteriormente. Cada vez que se lleva a cabo una división del espacio de características, en lugar de considerar todos los predictores p , se toma, solamente, un subconjunto aleatorio de ellos con m elementos. Será uno nuevo cada vez que se realiza una división y, típicamente, el número será $m \approx \sqrt{p}$.

Variables relevantes

Podemos dar una medida de la importancia de cada predictor sumando la cantidad de error que ha sido capaz de reducir en el proceso completo. Es decir, acumulamos la reducción de índice de Gini (o RSS), cada vez que se haya llevado a cabo una división del espacio de características según dicho parámetro.

5.2.2.5. SVM (Support vector machine)

El problema que pretende resolver este algoritmo es el de separar dos clases de objetos en el espacio de características.

Dadas dos clases de objetos $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m) \in \mathcal{X} \times \{\pm 1\}$, necesitamos una manera de medir la similitud de los objetos en el espacio de características. Esa medida será del tipo

$$\begin{aligned} k: \mathcal{X} \times \mathcal{X} &\rightarrow \mathbb{R} \\ (x, x') &\rightarrow k(x, x') \end{aligned}$$

Este tipo de aplicaciones se denominan kernel. El ejemplo más sencillo es el producto escalar:

$$k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$$

$$(x, x') \rightarrow k(x, x') := x \cdot x'$$

Para poder aplicarlo, necesitamos embeber nuestros objetos en un espacio dotado de esa estructura u otra similar

$$\phi: \mathcal{X} \rightarrow \mathcal{H}$$

$$x \rightarrow \mathbf{x} := \phi(x)$$

además de trasladar el objeto hasta un espacio dotado de dicha estructura, nos permitiría extender el espacio de características mediante una aplicación ϕ no lineal, si fuese conveniente.

Una manera sencilla de realizar la clasificación sería asignar a cada elemento nuevo la clase cuyo punto medio se encuentre más cerca.

$$c_+ = \frac{1}{m_+} \sum_{i|y^{(i)}=+1} \mathbf{x}^{(i)} \quad (5.31)$$

$$c_- = \frac{1}{m_-} \sum_{i|y^{(i)}=-1} \mathbf{x}^{(i)} \quad (5.32)$$

$$y = \text{sgn}(\mathbf{x} \cdot c_+ - \mathbf{x} \cdot c_- + b) \quad (5.33)$$

Siendo el conjunto linealmente separable, esto define el hiperplano que, equivalentemente, resolvería el problema. La manera óptima de hacerlo sería localizando aquel que estuviera a la máxima distancia posible de ambos conjuntos, definiendo esta como la mínima existente entre las distancias de todos los puntos del conjunto al plano, es la característica distintiva de los vectores soporte (de ahí que también se denominen clasificadores de margen máximo)

$$\max_{\theta \in \mathcal{H}, b \in \mathbb{R}} \min \left\{ \|\mathbf{x} - x^{(i)}\| \mid \mathbf{x} \in \mathcal{H}, \theta^T \mathbf{x} + b = 0, i = 1, \dots, m \right\} \quad (5.34)$$

A cada lado de la frontera, por los puntos más cercanos a ella \mathbf{x}_+ y \mathbf{x}_- pasan dos hiperplanos que distan entre ellos $\frac{2}{\|\theta\|}$, se tiene

$$\theta \cdot \mathbf{x}_+ + b = +1 \tag{5.35}$$

$$\theta \cdot \mathbf{x}_- + b = -1 \tag{5.36}$$

$$\Rightarrow \theta \cdot (\mathbf{x}_+ - \mathbf{x}_-) = 2 \tag{5.37}$$

$$\Rightarrow \frac{\theta}{\|\theta\|} \cdot (\mathbf{x}_+ - \mathbf{x}_-) = \frac{2}{\|\theta\|} \tag{5.38}$$

Luego, para hacer esa distancia máxima, deberíamos minimizar $\|\theta\|$. La norma involucra la raíz y complica el problema, por ello, suele reformularse de la siguiente manera

$$\min_{\theta \in \mathcal{H}, b \in \mathbb{R}} \tau(\theta) = \frac{1}{2} \|\theta\|^2 \tag{5.39}$$

$$\text{sujeto a } y^{(i)}(\theta^T \mathbf{x}^{(i)} + b) \geq 1, \quad \forall i = 1, \dots, m \tag{5.40}$$

5.2.2.6. Redes neuronales

Las clasificación de los elementos de un conjunto en k clases, puede llevarse a cabo mediante una red neuronal como la que se esquematiza en la figura 5.4. Teniendo sigmoides en las funciones de activación en la salida, asignaríamos la clase correspondiente al máximo valor de $(h_\theta(x))_k$

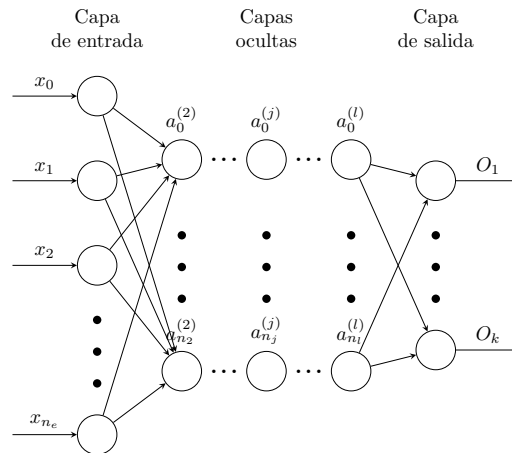


FIGURA 5.4: Red neuronal multicapa con k salidas

5.3. Validación

La generalización de un modelo de aprendizaje automático hace referencia a la capacidad de este para predecir resultados cuando extendemos su aplicación a nuevos datos. Por tanto, da una medida de la calidad del modelo. En este apartado, se verán los métodos más utilizados para evaluar dicha capacidad.

5.3.1. Teoría del aprendizaje

Tratamos de minimizar el error que obtenemos al extender la aplicación del modelo al conjunto total de datos.

El error de generalización de una hipótesis es aquel que cometemos sobre ejemplos que no se encuentran en el conjunto de entrenamiento. Informalmente, definimos el 'bias' de un modelo como aquel error que cometeríamos aun disponiendo de infinitos datos para realizar el ajuste. La varianza del modelo es el error de generalización debido al ajuste del modelo a datos espurios (Andrew Ng). Representa la sensibilidad del modelo con respecto a cambios en el conjunto de entrenamiento inicial. Procuraremos encontrar un punto de complejidad intermedia, de forma que no sea, ni demasiado simple, ya que tendrá bias elevado, ni demasiado complejo, ya que se adaptará bien a los datos de entrenamiento pero mal a los nuevos.

- i. Podemos formalizar la compensación bias/varianza
- ii. Podemos relacionar el error en el conjunto de entrenamiento con el error de generalización.
- iii. Existen condiciones bajo las cuales podemos asegurar que los algoritmos de aprendizaje funcionarán adecuadamente.

Lema 5.1. *The union bound.*

$$P(A_1 \cup \dots \cup A_k) \leq P(A_1) + \dots + P(A_k) \quad (5.41)$$

Podríamos decir: $P(A \cup B)$ 'probabilidad de que ocurra A o B'

Lema 5.2. *Desigualdad de Hoeffding (Chernoff Bound)* Sean Z_1, \dots, Z_m variables aleatorias independientes siguiendo una distribución de Bernoulli ϕ , $P(Z_i = 1) = \phi$ y $P(Z_i = 0) =$

$1 - \phi$. Sea $\hat{\phi}$ la media de ellas $\hat{\phi} = \frac{1}{m} \sum_{i=1}^m Z_i$ y $\gamma > 0$, entonces

$$P(|\phi - \hat{\phi}| \leq 2 \exp(-2\gamma^2 m)) \quad (5.42)$$

Da una cota superior a la probabilidad de que la suma de variables aleatorias se desvíe una cierta cantidad de su valor esperado. Significa que si tomamos $\hat{\phi}$ como estimación de ϕ , la probabilidad de que estemos lejos del valor verdadero es pequeña, siempre y cuando el valor de m sea grande. La probabilidad de que $|\phi - \hat{\phi}|$ sea mayor que α es menor que un valor pequeño $\exp(-2\gamma^2 m)$, siendo m grande.

Usando estos lemas, demostraremos algunos resultados importantes de la teoría del aprendizaje automático. Para simplificar la exposición, nos limitaremos a la clasificación binaria pero es generalizable a la clasificación con mayor número de categorías e incluso a la regresión.

Dado un conjunto de entrenamiento, $S = \{(x^{(i)}, y^{(i)}); i = 1, \dots, m\}$ siendo $(x^{(i)}, y^{(i)})$ muestras de alguna distribución de probabilidad D . Siendo la hipótesis, definimos el 'error de entrenamiento' (o riesgo empírico o error empírico) como la fracción de ejemplos del conjunto de entrenamiento mal clasificados

$$\hat{\varepsilon}(h) = \frac{1}{m} \sum_{i=1}^m 1\{h(x^{(i)}) \neq y^{(i)}\} \quad (5.43)$$

Para hacer explícita la dependencia del conjunto de entrenamiento, escribiremos $\hat{\varepsilon}_s(h)$. Definimos el error de generalización como $\varepsilon(h) = P_{(x,y) \sim D}(h(x) \neq y)$, es decir, la probabilidad de que tomando una muestra (x, y) de la distribución D , nuestra hipótesis la clasifique erróneamente.

PAC (probabilidad aproximadamente correcta): conjunto de entrenamiento formado por muestras de la distribución con la que evaluamos nuestra hipótesis.

Consideramos un clasificador lineal $h_\theta(x) = 1\{\theta^T x \geq 0\}$. Una manera razonable de encontrar θ es tratando de minimizar el error de entrenamiento:

$$\hat{\theta} = \arg \min_{\theta} \hat{\varepsilon}(h_\theta) \quad (5.44)$$

Se determina proceso (ERM) empirical risk minimization. Definimos la clase de hipótesis, $\mathcal{H} = \{h_\theta : h_\theta(x) = 1\{\theta^T x \geq 0\}, \theta \in \mathbb{R}^{n+1}\}$ el conjunto de clasificadores sobre X con frontera de decisión lineal.

Podríamos pensar el proceso (ERM) como minimización sobre la clase de funciones \mathcal{H} en el que el algoritmo de aprendizaje toma las hipótesis.

5.3.1.1. Caso de \mathcal{H} finito

Consideremos una clase de hipótesis finita $\mathcal{H} = \{h_1, \dots, h_k\}$. Queremos dar alguna garantía sobre el error de generalización una vez calculado el de entrenamiento. La estrategia sigue dos fases: primero, demostraremos que error de entrenamiento $\hat{\varepsilon}(h)$ es una estimación buena de $\varepsilon(h)$ para todo $h \in \mathcal{H}$. Seguidamente, demostraremos que implica una cota superior al error de generalización de \hat{h} .

Tomemos cualquier $h_i \in \mathcal{H}$. Consideremos una variable aleatoria de Bernoulli Z , definida como $Z = 1\{h_i(x) = y\}$. Es decir, tomamos una muestra (x, y) de la distribución y vemos si h_i la clasifica correctamente. $\varepsilon(h_i)$, es el valor esperado de Z (y Z_i), además, el error de entrenamiento es $\hat{\varepsilon}(h_i) = \frac{1}{m} \sum_{i=1}^m Z_i$. Por tanto, $\hat{\varepsilon}(h_i)$ es la media de m variables aleatorias Z_j muestreadas de una distribución de Bernoulli con media $\varepsilon(h_i)$. Podemos aplicar la desigualdad de Hoeffding,

$$P(|\varepsilon(h_i) - \hat{\varepsilon}(h_i)| > \gamma) \leq 2 \exp(-2\gamma^2 m) \quad (5.45)$$

Esto indica que, para una hipótesis particular h_i , el error de entrenamiento estará próximo al error de generalización con alta probabilidad, si m es grande. Queremos probar esto para todo $h \in \mathcal{H}$. Llamaremos A_i al evento desfavorable $|\varepsilon(h_i) - \hat{\varepsilon}(h_i)| > \gamma$. Hemos visto que $P(A_i) \leq 2 \exp(-2\gamma^2 m)$, usando la cota de la unión,

$$\begin{aligned} P(\exists h \in \mathcal{H}. |\varepsilon(h_i) - \hat{\varepsilon}(h_i)| > \gamma) &= P(A_1 \cup \dots \cup A_k) \\ &\leq \sum_{i=1}^k P(A_i) \\ &\leq \sum_{i=1}^k 2 \exp(-2\gamma^2 m) \\ &= 2k \exp(-2\gamma^2 m) \end{aligned} \quad (5.46)$$

Lo opuesto es: probabilidad de que no haya hipótesis tal que $|\varepsilon(h_i) - \hat{\varepsilon}(h_i)| > \gamma$.

$$\begin{aligned} P(\nexists h \in \mathcal{H}. |\varepsilon(h_i) - \hat{\varepsilon}(h_i)| > \gamma) &= P(\forall h \in \mathcal{H}. |\varepsilon(h_i) - \hat{\varepsilon}(h_i)| \leq \gamma) \\ &\geq 1 - 2k \exp(-2\gamma^2 m) \end{aligned} \quad (5.47)$$

A esto se llama convergencia uniforme, ya que se cumple simultáneamente para todas las hipótesis.

Hasta ahora, hemos hecho lo siguiente: para valores particulares de m y γ , hemos encontrado un valor máximo de la probabilidad de que la diferencia de errores sea mayor que γ , $|\varepsilon(h) -$

$|\hat{\varepsilon}(h)| > \gamma$. Tenemos 3 variables de interés, γ , m y la probabilidad de error. Podemos jugar con ellas acotando los valores de unas u otras según convenga.

Por ejemplo, podríamos preguntarnos: ¿cuántos elementos debe contener el conjunto de entrenamiento para garantizar que el error de generalización y el de entrenamiento se encuentran a distancia γ con probabilidad $1 - \delta$? Si tenemos m ejemplos, sabremos con probabilidad $1 - \delta$ que $|\varepsilon(h) - \hat{\varepsilon}(h)| \leq \gamma$ para cualquier hipótesis $h \in \mathcal{H}$. 'k' se denomina complejidad del algoritmo.

El número de ejemplos que necesitamos es solo logarítmico en k , el número de hipótesis. Por tanto, aunque aumentemos la complejidad k , no necesitaremos aumentar demasiado el número de elementos m . De igual manera, fijando los valores de m y δ , podríamos decir con probabilidad $1 - \delta$, que para cualquier hipótesis el error de generalización cumple:

$$|\hat{\varepsilon}(h) - \varepsilon(h)| \leq \sqrt{\frac{1}{2m} \log \frac{2k}{\delta}} \quad (5.48)$$

Ahora, asumamos que se cumple la convergencia uniforme $|\varepsilon(h) - \hat{\varepsilon}(h)| \leq \gamma, \forall h \in \mathcal{H}$ entonces. ¿Qué podemos decir acerca de la generalización de nuestro modelo escogido como $\hat{h} = \arg \min_{h \in \mathcal{H}} \hat{\varepsilon}(h)$?

Definimos $h^* = \arg \min_{h \in \mathcal{H}} \varepsilon(h)$ como la mejor hipótesis posible en \mathcal{H} . Ya que h^* es lo mejor a lo que podemos optar, vale la pena comparar el rendimiento. h^* minimiza el error de generalización

$$\varepsilon(\hat{h}) \leq \hat{\varepsilon}(\hat{h}) + \gamma \quad (5.49)$$

$$\leq \hat{\varepsilon}(h^*) + \gamma \quad (5.50)$$

$$\leq \varepsilon(h^*) + 2\gamma \quad (5.51)$$

Si la convergencia uniforme se cumple, el error de generalización de \hat{h} es, a lo sumo, 2γ peor que la mejor hipótesis de \mathcal{H} .

Teorema 5.1. *Sea la clase de hipótesis \mathcal{H} con k elementos. Fijamos m , δ . Entonces, con probabilidad al menos $1 - \delta$ tenemos*

$$\varepsilon(\hat{h}) \leq \left(\arg \min_{h \in \mathcal{H}} \varepsilon(h) \right) + 2 \sqrt{\frac{1}{2m} \log \frac{2k}{\delta}} \quad (5.52)$$

De esta manera, cuantificamos la compensación bias/varianza. Si aumentamos el tamaño del conjunto de hipótesis, conseguiremos reducir el bias. Sin embargo, al aumentar k , el término $\sqrt{\frac{1}{2m} \log \frac{2k}{\delta}}$ aumenta, se corresponde con el aumento de la varianza.

Manteniendo γ , δ fijos y resolviendo para m , obtenemos la siguiente cota para la complejidad del modelo.

Corolario 5.1. *Sea $\mathcal{H} = k$ y δ , γ fijos. Entonces, para que se cumpla $\arg \min_{h \in \mathcal{H}} \varepsilon(h) + 2\gamma$ con probabilidad, al menos, $1 - \delta$, es suficiente que*

$$m \geq \frac{1}{2\gamma^2} \log \frac{2k}{\delta} \quad (5.53)$$

$$\geq O\left(\frac{1}{2\gamma^2} \log \frac{k}{\delta}\right) \quad (5.54)$$

5.3.1.2. Caso \mathcal{H} infinito.

Muchas clases de hipótesis incluyen alguna parametrización con números reales, conteniendo de esta manera infinito número de funciones. ¿Podemos probar resultados similares a los anteriores?. Para comenzar, utilizaremos un argumento no del todo correcto.

Supongamos \mathcal{H} parametrizado por d números reales. Pensando que un número real se representa por 64 bits en un ordenador, tenemos $k = 2^{64d}$ hipótesis. Hemos visto que el error de generalización de nuestra mejor opción sobre el conjunto de datos de entrenamiento \hat{h} es, a lo sumo, el error de h^* (la mejor opción) más 2γ : $\varepsilon(\hat{h}) \leq \varepsilon(h^*) + 2\gamma$ con probabilidad $1 - \delta$. De ahí, podemos estimar que el número de elementos en el conjunto de entrenamiento guarda una relación, como mucho, lineal con el número de parámetros.

$$m \geq O\left(\frac{1}{\gamma^2} \log \frac{2^{64d}}{\delta}\right) = O\left(\frac{2^{64d}}{\delta}\right) = O\left(\frac{d}{\gamma^2} \log \frac{1}{\delta}\right) = O_{\gamma, \delta}(d) \quad (5.55)$$

Es importante destacar que estos resultados son válidos para algoritmos que usan ERM (empirical risk minimization) y no aplican para otros (discriminative learning algorithms) que lo son. Dar garantías teóricas en el ámbito de los algoritmos no ERM es objeto de investigación. Otro aspecto del enfoque que hemos descrito hasta ahora que resulta poco satisfactorio es su (dependencia) apoyo en la parametrización de \mathcal{H} . Intuitivamente, no parece importar.

Derivemos otro argumento más satisfactorio. Definimos unos conceptos necesarios. Dado el conjunto $S = \{x^{(1)}, \dots, x^{(d)}\}$ de puntos $x^{(i)} \in \mathcal{X}$, decimos que ' \mathcal{H} rompe S ' (' \mathcal{H} shatters S ') si \mathcal{H} puede realizar cualquier etiquetado de S . Es decir, para todo conjunto de etiquetas, existe algún $x^{(i)} \in \mathcal{X} | h(x^{(i)}) = y^{(i)}, \forall i = 1, \dots, d$

Dimensión de Vapnik- Chervonenkis $VC(H)$: el tamaño del conjunto más grande roto por \mathcal{H} . El siguiente es un teorema importante de la teoría del aprendizaje debido a Vapnik.

Teorema 5.2. Dado \mathcal{H} , sea $d = VC(H)$, entonces, con probabilidad al menos $1 - \delta$ tenemos que para todo $h \in \mathcal{H}$

$$|\varepsilon(h) - \hat{\varepsilon}(h)| \leq O\left(\sqrt{\frac{d}{m} \log \frac{m}{d} + \frac{1}{m} \log \frac{1}{\delta}}\right) \quad (5.56)$$

Por tanto, con probabilidad al menos $1 - \delta$ tenemos también,

$$\varepsilon(\hat{h}) \leq \varepsilon(h^*) + O\left(\sqrt{\frac{d}{m} \log \frac{m}{d} + \frac{1}{m} \log \frac{1}{\delta}}\right) \quad (5.57)$$

En otras palabras, si una clase de hipótesis tiene dimensión $VC(H)$ finita la convergencia uniforme se tiene a medida que m aumenta. Al igual que anteriormente, nos permite acotar $\varepsilon(h)$ en términos de $\varepsilon(h^*)$.

Corolario 5.2. Es suficiente que $m = O_{\gamma, \delta}(d)$ para que $|\varepsilon(h) - \varepsilon(h^*)| \leq \gamma$ con probabilidad, al menos, $1 - \delta$. Se cumple $\forall h$ y, por tanto, $\varepsilon(\hat{h}) \leq \varepsilon(h^*) + 2\gamma$.

Concluimos, el número de muestras en el conjunto de entrenamiento necesarias para aprender correctamente utilizando la clase \mathcal{H} , es lineal en la dimensión de Vapnik-Chervonenkis. Resulta que, para la mayoría de clases, la dimensión $VC(H)$ es lineal, aproximadamente en el número de parámetros. Por tanto, habitualmente, el número de muestras necesarias es aproximadamente lineal en el número de parámetros.

5.3.2. Métodos indirectos

Podemos estimar el error de generalización del modelo mediante los C_p , AIC , BIC y Adjusted R^2 (James et al., 2013).

$$C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2) \quad (5.58)$$

$$AIC = \frac{1}{n\hat{\sigma}^2}(RSS + 2d\hat{\sigma}^2) \quad (5.59)$$

$$BIC = \frac{1}{n}(RSS + \log(n)d\hat{\sigma}^2) \quad (5.60)$$

$$\text{Adjusted } R^2 = 1 - \frac{RSS/(n-d-1)}{TSS/(n-1)} \quad (5.61)$$

5.3.3. Métodos directos: Cross validation

- i. Dividimos aleatoriamente el conjunto de muestras S , en S_{train} y S_{cv} . Normalmente, un 70 % por ciento de los datos estarán en S_{train} .
- ii. Entrenamos cada modelo M_i en S_{train} para conseguir las hipótesis h_i
- iii. Seleccionaremos la hipótesis h_i con el error más pequeño (error empírico de h en el conjunto de muestras S_{cv})

En algunas ocasiones, si el modelo no es muy sensible a cambios en las condiciones iniciales de los datos, es buena idea entrenar de nuevo en el conjunto S completo.

La desventaja de la validación que hemos descrito es que consume el 30 % de los datos. Existen alternativas más económicas como son: *k-fold cross validation* o *leave one out cross validation*.

5.3.3.1. K-fold cross validation

- i. Dividimos S en k subconjuntos disjuntos de manera aleatoria con m/k elementos.
- ii. Evaluamos cada modelo M_i como sigue:
for $j = 1, \dots, k$ {
Entrenamos cada modelo M_i en $S_1 \cup \dots \cup S_{j-1} \cup S_{j+1} \dots \cup S_k$, todas menos S_j , para conseguir las hipótesis h_{ij}
Evaluamos h_{ij} en S_j para obtener $\hat{\epsilon}_{S_j}(h_{ij})$
El error de generalización estimado será la media de ellos }
iii. Tomamos el M_i con error mínimo y volvemos a entrenarlo en S completo.

5.3.3.2. Leave one out cross validation.

Solo deja un elemento fuera.

5.4. Selección del modelo.

¿Cómo podemos seleccionar automáticamente un modelo con un equilibrio bias/varianza adecuado?

5.4.1. Selección de parámetros de ajuste

Para ajustar los parámetros de cada modelo, elegimos una rejilla para el parámetro en cuestión y estimamos el error de validación en cada punto de ella.

5.4.2. Selección de características.

Dadas n características, tenemos 2^n subconjuntos posibles. Podemos plantear el problema como el de selección de un modelo en un conjunto con 2^n M_i . Ya que, para n grande, el coste computacional es alto, suele usarse algún procedimiento heurístico de búsqueda del subconjunto.

5.4.2.1. Búsqueda secuencial

Forward stepwise:

1. $F = \phi$
2. Repetimos {For $i = 1, \dots, n$ si $i \notin F_i = F \cup \{i\}$ y validamos}
3. Seleccionamos el mejor subconjunto.

Backward stepwise

Funciona de la manera inversa, partimos de todas las características y vamos eliminando.

5.4.3. Filtrado de características

Vamos a computar algún índice que nos de idea de cuánta información contienen las características x_i acerca de las etiquetas y . Elegiremos las características con mayor índice $S(i)$.

Una posible elección del índice $S(i)$ sería la correlación. En la práctica, es más frecuente usar la información mutua. A continuación describiremos dicho concepto así como otros relacionados.

5.4.3.1. Información y probabilidad.

La información mutua deriva del concepto de entropía. Esta se define como la información promedio de un proceso, ¡referencial!, por tanto, antes de nada, vamos a definir el concepto de información en este contexto y exponer la relación entre información y probabilidad.

Para ello, partimos de un sistema simplificado, supongamos que

- el sistema tiene $M = 2^b$ posibles resultados, potencia de dos
- todos tienen la misma probabilidad

Los posibles resultados $S = \{O_1, O_2, \dots, O_M\}$, podrán codificarse mediante la variación con repetición de 2 símbolos $\{0, 1\}$, tomados de b en b . Es decir, cada O_x puede codificarse por la cadena mínima de bits $B_1 B_2 \dots B_b$; solo necesitamos b bits para codificar cada posible resultado. De esa manera, podemos decir que la información que nos proporciona cada resultado es de b bits; la longitud de la cadena que lo representa

$$I(O_x) \equiv I(x) = b$$

Si la probabilidad es igual para todos, cada resultado tiene la probabilidad

$$P = \frac{1}{M} = \frac{1}{2^b} \Rightarrow P = \frac{1}{2^I} \Rightarrow I = -\log_2 P$$

Esa expresión nos da la relación entre la probabilidad y la información recabada en cada resultado.

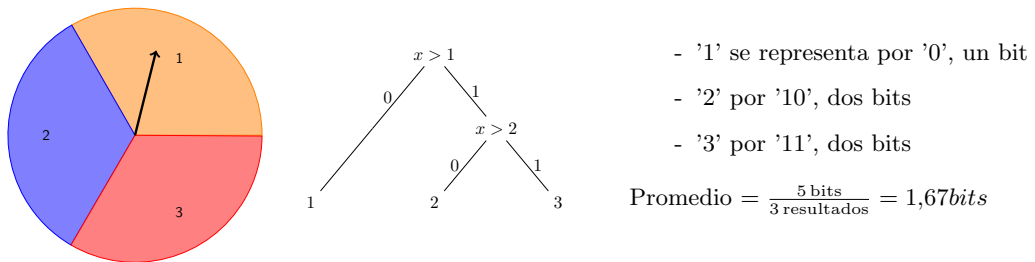
Si, relajando esas condiciones, permitimos $\Rightarrow M = a^k$ posibles resultados igualmente probables, donde $a, k \in \mathbb{Z}$. Entonces, necesitaremos k -'its', es decir, k símbolos de los a que contiene el 'alfabeto' para codificar cada posible resultado. Por ejemplo $M = 3^2$ posibles resultados pueden codificarse mediante cadenas de 2-'trits', 2 símbolos de los 3 posibles, permitiendo repetición.

$$I = k \text{ en 'aits'} \Rightarrow P = \frac{1}{a^k} = \frac{1}{a^I} \Rightarrow -I \log_a(a) = \log_a(P) \quad (5.62)$$

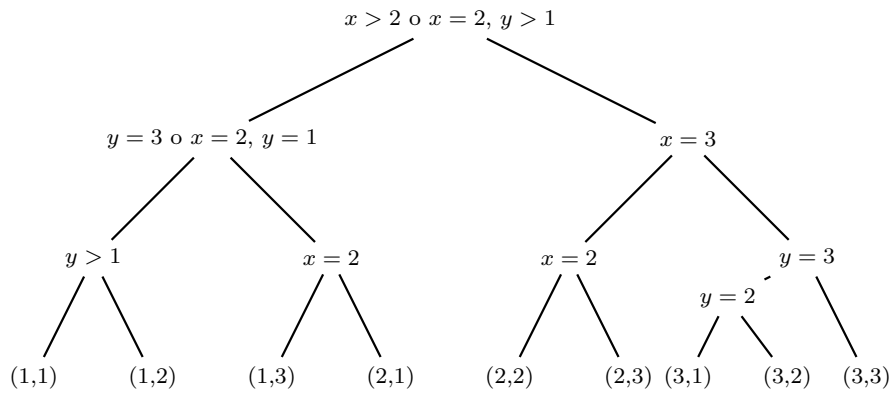
$$\Rightarrow I = -\log_a P \quad (5.63)$$

Si tratamos de codificar esto mismo con el alfabeto binario, la cadena de bits necesaria para representar los posibles resultados no tendrá siempre la misma longitud. Veámoslo en el siguiente ejemplo de una ruleta con tres zonas.

Si hacemos girar la flecha una vez y codificamos el posible resultado mediante un árbol de decisión como el de la figura, veremos que



La longitud promedio de bits necesaria para codificar el resultado es $\frac{5}{3}$. Sin embargo, si realizamos el experimento dos veces, las parejas de valores posibles: (1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3) pueden codificarse mediante el siguiente árbol:



de donde se deduce que la longitud mínima media de las cadenas de bits es $1/2 * (7 * 3 + 2 * 4) / 9 = 1,611 \text{ bits}$.

En definitiva, vemos que esa longitud o información no es constante y por ello se habla de longitud promedio. Si la hacemos girar 'n' veces, el promedio de bits que necesitamos para codificar el resultado, sería $H = \log_2(3)$

Podríamos traducirlo a 'bits', mediante la relación

$$\log_a P = \frac{\log_2 P}{\log_2(a)} \Rightarrow 1 \text{ ait} = \log_2(a) \text{ bits} \quad (5.64)$$

Cuando los resultados no son igualmente probables, tenemos que O_x ocurre alrededor de $P(O_x) \cdot n$ veces. La información asociada es $I(O_x) \equiv I(x)$, $I(x) = -\log_2 P(x)$. El promedio será

$$\begin{aligned} H(x) &= \lim_{n \Rightarrow \infty} \frac{P(O_1) \cdot n \cdot I(O_1) + \dots + P(O_n) \cdot n \cdot I(O_n)}{n} \\ &= \sum_{x=1}^M P(x) \cdot I(x) = - \sum_{x=1}^M P(x) \cdot \log P(x) \end{aligned}$$

que constituye la definición de entropía. En los siguientes apartados, tuvimos en cuenta, además de las referencias citadas, las exposiciones que aparecen en Bishop et al. (1995) y Kantz and Schreiber (2004).

5.4.3.2. Entropía

$H(x)$ es la información promedio recabada al tomar n muestras de la distribución $P(X)$ cuando $n \Rightarrow \infty$

Cuantifica la información: la cantidad de información de un mensaje es el menor número de bits con el que podemos codificarlo. Información (self-information): es la cadena de bits más corta que puede representar cualquier posible resultado.

$$I(x) \equiv \log_2 \frac{1}{P(x)} = -\log_2 P(x) \quad (5.65)$$

$$H(x) \equiv I(x) = \log_2 M(x) = -\log_2 \frac{1}{M} = -\log_2 P(x) \quad (5.66)$$

Cuando todos los resultados son equivalentes la entropía es la self-information. Si tuviera M posibles resultados, tras repetir el experimento n veces, el resultado o_x ocurría $n \cdot p(x)$ veces. El promedio de información recabada:

$$H(x) = \lim_{n \Rightarrow \infty} \frac{P(O_1) \cdot n \cdot I(O_1) + \dots + P(O_m) \cdot n \cdot I(O_m)}{n} = \sum_{x=1}^M P(x) \cdot I(x) \quad (5.67)$$

5.4.3.3. Información Mutua

La información mutua deriva del concepto de entropía y proporciona una medida no lineal de la correlación o dependencia entre dos variables. Mide la cantidad de información que una variable contiene de otra (Tourassi et al., 2001)

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (5.68)$$

$$I(X; Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y) \quad (5.69)$$

5.4.3.4. Divergencia de Kullback-Leibler

También llamada entropía relativa o entropía diferencial. Es una medida estadística que cuantifica en bits la distancia que existe entre dos posibles distribuciones de probabilidad, p y q , de una misma variable aleatoria X ,

distribución de probabilidad $p = \{p_i\}$ y otro modelo $q = \{q_i\}$.

$$D_{KL}(p||q) = \sum_i p_i \log_2 \frac{p_i}{q_i} \quad (5.70)$$

Distribución binomial

Es la distribución de probabilidad de n ensayos de Bernoulli independientes entre sí con una probabilidad fija p

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; \quad \binom{n}{x} \frac{n!}{x!(n-x)!} \quad (5.71)$$

Distribución multinomial

Es una generalización de la distribución binomial. Da la probabilidad de que un suceso $x = (x_1, \dots, x_k)$ ocurra, conocido el modelo P , con probabilidades $P = (p_1, \dots, p_k)$, $\sum p_i = 1$. Equivalentemente, podríamos decir: 'probabilidad del estado (histograma) \mathcal{X} dado el modelo P '

$$f(x_1, \dots, x_k; n, p_1, \dots, p_k) = \Pr(\mathcal{X}_1 = x_1 \cap \dots \cap \mathcal{X}_k = x_k) \quad (5.72)$$

$$= \frac{n!}{\prod x_i!} \prod p_i^{x_i} = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} \quad (5.73)$$

5.5. Compensación bias/varianza y regularización

Hemos visto cómo elegir los parámetros por el método de máxima probabilidad.

$$\theta_{ML} = \arg \max_{\theta} \prod_{i=1}^m P(y^{(i)} | x^{(i)}; \theta) \quad (5.74)$$

El hecho de suponer que θ es un valor constante es propio de la estadística clásica (frecuentista). La idea es estimar su valor por métodos estadísticos.

La aproximación Bayesiana es diferente. Pensamos en θ como una variable aleatoria cuyo valor desconocemos. En este enfoque deberíamos especificar una distribución a priori $p(\theta)$ que expresara nuestras creencias 'a priori' acerca de los parámetros. Dado el conjunto de entrenamiento $S = \{(x^{(i)}, y^{(i)})\}_{i=1}^m$, cuando nos piden hacer una predicción sobre un nuevo valor de x , podemos calcular la distribución a posteriori

$$P(\theta|S) = \frac{P(\theta|S)P(\theta)}{P(S)} = \frac{\prod_{i=1}^m P(y^{(i)}, x^{(i)}, \theta) \cdot P(\theta)}{\int_{\theta} \prod_{i=1}^m P(y^{(i)}, x^{(i)}, \theta) \cdot P(\theta)} \quad (5.75)$$

$P(y^{(i)}, x^{(i)}, \theta)$ proviene del modelo que estemos usando en nuestro problema de aprendizaje.

5.6. Curvas de aprendizaje

Las curvas de aprendizaje representan la evolución de los errores de validación y de entrenamiento con el número de datos que consideramos en el conjunto de entrenamiento. La evolución de ambas, nos permite ver la relación entre el bias y la varianza en el algoritmo en cuestión. Si este posee un bias alto, ambas evoluciones tenderán a un valor alto del error. En tal caso, un número de datos de entrenamiento muy grande no mejoraría los resultados sustancialmente. Si, por el contrario, el modelo posee alta varianza, las curvas quedarán distanciadas y podríamos esperar que los resultados mejorasen al ampliar el conjunto de datos de entrenamiento.

Capítulo 6

Corrección atmosférica

6.1. Introducción

Habitualmente, la corrección atmosférica en las imágenes de satélite no considera la variación espacial ni temporal de las concentraciones de los componentes químicos ni de las variables termodinámicas implicadas. Dependiendo del tamaño de la imagen, realizar dicha tarea mediante los procedimientos clásicos, que involucrarían la ejecución de modelos de transferencia radiativa en cada punto (en adelante, también usaremos blanco o target), resultaría inabordable debido al coste computacional que conllevan. En este trabajo se propone una alternativa que considera la dependencia espacial y temporal de las condiciones atmosféricas y resuelve la corrección en tiempo asequible.

En este capítulo se describe la metodología que hemos seguido para llevar a cabo dicho objetivo. En primer lugar, se realiza la corrección en un conjunto de píxeles, elegidos aleatoriamente en las imágenes de la serie, por el método tradicional. Así, conseguiremos el conjunto de entrenamiento necesario para fijar los parámetros del modelo de regresión seleccionado y finalizar la tarea en el resto de la serie. En el primer apartado, se detalla el conjunto de datos de entrenamiento, es decir, se verá cómo se obtuvieron las variables de entrada y la variable objetivo. Posteriormente, se compara la habilidad de los algoritmos de aprendizaje automático utilizados. En la sección de compensación bias/varianza, se analiza el error de generalización cometido con los distintos modelos al aplicarlos a datos no incluidos en el conjunto de entrenamiento. En el siguiente apartado, estudiamos el número de datos de entrenamiento suficiente para llevar a cabo nuestro cometido. Seguidamente,

haciendo uso de los índices apropiados evaluaremos las características más relevantes. Por último, se presentan los resultados y algunas conclusiones.

6.1.1. Datos de entrada en el modelo 6S

Elegimos el modelo de transferencia radiativa 6S, principalmente, porque tiene en cuenta el entorno del punto a corregir, pero existen otras alternativas. En la actualidad, existen modelos más complejos como Sciatran (Rozañov et al., 2005), libRadtran (Mayer and Kylling, 2005) y LIDORT/VLIDORT (Spurr, 2008), que, si bien implican un mayor tiempo de computación, proporcionan resultados más exactos.

El modelo 6S (Vermote et al., 1997a) predice la reflectancia aparente, es decir, la observada en el TOA (top of atmosphere), usando la información de la reflectancia superficial y las condiciones atmosféricas (Vermote et al., 1997b). La reflectancia TOA se puede estimar usando la fórmula:

$$\rho = (\pi \cdot L_{\text{sat}} \cdot d^2) / (E_0 \cdot \cos SZ) \quad (6.1)$$

donde L es la radiancia espectral que llega al sensor; d , la distancia Tierra-Sol en unidades astronómicas; E_0 , la irradiancia solar media exoatmosférica; y SZ el ángulo cenital solar. El modelo de transferencia radiativa 6S permite invertir el proceso proporcionando los parámetros necesarios, podemos calcular la reflectancia superficial, sin efectos atmosféricos, mediante la fórmula proporcionada en la salida del modelo, figura 6.1.

$$\rho_{\text{cor}} = (A \cdot \rho - B) / [1 + (\gamma(A \cdot \rho - B))] \quad (6.2)$$

donde $A = 1/\alpha\beta$, $B = \rho/\beta$, α es la transmitancia global de los gases; β , la transmitancia total debida al scattering; y γ , el albedo esférico. Las constantes α, β, γ son generadas tras la ejecución del modelo 6S.

```

*****
*                               atmospheric correction result                               *
*                               -----                                                  *
*   input apparent reflectance      :    0.002                                         *
*   measured radiance [w/m2/sr/mic] :    0.131                                         *
*   atmospherically corrected reflectance                                         *
*   Lambertian case :    0.07305                                                    *
*   BRDF case :    0.07305                                                         *
*   coefficients xa xb xc           :    1.06648  0.06716  0.00364                    *
*   y=xa*(measured radiance)-xb; acr=y/(1.+xc*y)                                   *
*****

```

FIGURA 6.1: Salida del modelo 6S

Tiempos de cálculo

Realizamos una corrección preliminar parcial para estimar el tiempo que tardaríamos en corregir una imagen completa. Consideramos 8 bandas de una imagen Landsat con 7861×7991 píxeles. El tiempo que tardamos en corregir 3000 píxeles fue próximo a 9500 segundos, es decir, una media de unos 3 segundos por pixel. Necesitaríamos unos $7861 \times 7991 \times 8 \times 3$ segundos, aproximadamente 47 años, para realizar la corrección completa con nuestro sistema.

El modelo 6S, está pensado para realizar la corrección a la imagen completa sin tener en cuenta, ni las condiciones atmosféricas, ni la posición de cada pixel. El coste computacional de este procedimiento motivó el tratamiento del problema en el marco del aprendizaje automático

6.2. Datos de entrenamiento

En primer lugar, determinamos las variables de entrada del modelo 6S necesarias para realizar la corrección. Algunas de ellas se obtendrán del modelo WRF y otras, directamente de la imagen. El procedimiento llevado a cabo consiste en dividir la imagen en dos conjuntos de píxeles. En el primero, corregiremos las reflectancias de los puntos por el método clásico, es decir, haciendo uso del modelo de transferencia radiativa (en nuestro caso el modelo 6S). En el resto de la imagen, lo haremos usando el algoritmo ML elegido. Es decir, al mismo tiempo que se realiza y se anota la corrección, se escriben los parámetros atmosféricos de

entrada al 6S en el fichero que constituirá el conjunto de entrenamiento. Una vez obtenido y tras elegir el modelo de regresión idóneo, lo utilizaremos para corregir el resto de los datos de reflectancia.

Sobre la ventana de trabajo, se eligen puntos uniformemente distribuidos de manera aleatoria y se recopilan los datos atmosféricos necesarios para ejecutar el modelo de transferencia radiativa. Se discutirá sobre el número adecuado posteriormente. Los datos que contiene son los siguientes:

- Condiciones geométricas: día, mes, hora, longitud y latitud.
- Perfiles de altitud (Km), presión (mb) temperatura (K) densidad de H_2O (g/m^3) O_3 densidad (g/m^3), AOT (espesor óptico de aerosoles) en la banda 550 nm
- Tipo de aerosol según la lista:
 - 1 continental model
 - 2 maritime model
 - 3 urban model
 - 4 shettle model for background desert aerosol
 - 5 biomass burning
 - 7 stratospheric model
- Nivel de altura del blanco y del sensor.
- Función de respuesta del sensor
- Reflectancia aparente del blanco y su entorno
- Radiancia aparente medida por el satélite

La elección inicial de dichos parámetros podría verse modificada si, en lugar de realizar la corrección mediante el modelo libRadtran, se realizara mediante alguno de los citados anteriormente.

6.2.1. Parámetros tomados del WRF

A continuación detallamos cómo fueron calculadas cada una las variables necesarias para ejecutar el modelo de transferencia radiativa. Algunas de ellas se generan a partir de los datos de salida del modelo meteorológico WRF y otras de los datos de la imagen que queremos corregir.

6.2.1.1. Altitud en el modelo WRF

La altitud en el modelo WRF no aparece de forma directa, calculamos la altura geopotencial, en Km, mediante la fórmula

$$h = (ph + phb)/9800 \quad (6.3)$$

donde ph se denomina perturbación geopotencial y phb, estado base geopotencial, ambas en unidades de $m^2 \cdot s^{-2}$.

6.2.1.2. Presión en el modelo WRF

La presión, en mbar, viene dada por la expresión

$$P = (p + pb) * 0,01 \quad (6.4)$$

donde p y pb, en unidades de Pa, son la perturbación de presión y el estado base de presión, respectivamente.

6.2.1.3. Temperatura en el modelo WRF

La temperatura que proporciona el modelo meteorológico es la perturbación de temperatura potencial $\hat{T} = \theta - T_0$, siendo la temperatura T_0 igual a 300 K. La temperatura potencial total θ en K está relacionada con la temperatura que nos interesa mediante la expresión tomada de American Meteorological Society, cited 2014: Climatology. Glossary of Meteorology. [Available online at http://glossary.ametsoc.org/wiki/Potential_temperature.]

$$\theta = T \cdot (p_0/p)^\kappa \quad (6.5)$$

Por tanto, la temperatura será

$$T = \theta \cdot (p/p_0)^\kappa \quad (6.6)$$

donde p_0 tiene el valor de 1000 mbar y p es la presión en mbar.

La constante de Poisson κ , es el ratio de la constante de los gases R para el calor específico a presión constante c_p . Para el aire seco $\kappa = 0,2854$. Para el aire húmedo

$$\kappa = \frac{R_d}{c_{pd}} \frac{1 + r_v/\epsilon}{1 + r_v c_{pv}/c_{pd}} \approx 0,2854 \cdot (1 - 0,24 \cdot r_v)$$

donde R_d y c_{pd} son la constante de los gases y el calor específico del aire seco; ϵ es el ratio de la constante de gases de vapor de agua y aire seco; r_v es el ratio de vapor de agua en la mezcla con aire seco (American Meteorological Society, cited 2014: Climatology. Glossary of Meteorology. [Available online at http://glossary.ametsoc.org/wiki/Poisson_constant.]).

6.2.1.4. Vapor de agua (g/m^3)

En el modelo WRF no tenemos esta variable como tal. Calcularemos el perfil de la densidad de agua a partir de la variable de mezcla de vapor de agua por masa de aire seco, que coincide con la r_v que acabamos de mencionar y viene dada en unidades de $\text{kg} \cdot \text{kg}^{-1}$.

Calcularemos el perfil de volumen de aire seco a través de la ecuación de estado de los gases ideales, ya que podemos considerar el aire seco como tal. Pasaremos de unidades $\text{kg} \cdot \text{kg}^{-1}$ a g/m^3 una vez conozcamos cuanto ocupa 1kg de aire seco en función de la altura. Para ello,

$$V = nRT/P \quad (6.7)$$

donde P es presión del modelo (p+pb) en Pa; T la temperatura del modelo anteriormente calculada en K; R es la constante universal de los gases ideales; y n, el número de moles en un Kg de aire seco.

El número de moles por gramo vendrá dado por

$$n = \frac{1g}{\text{masa molecular}} \quad (6.8)$$

siendo la masa molecular del aire, en promedio, de 28,97 g/mol y R, la constante de los gases, $8.314472 \text{ m}^3 \cdot \text{Pa}/(\text{K} \cdot \text{mol})$, tenemos que el volumen ocupado por 1g de aire seco en m^3 será

$$\frac{1g}{28,97g \cdot \text{mol}^{-1}} \times 8,314472 \frac{\text{m}^3 \cdot \text{Pa}}{\text{K} \cdot \text{mol}} \cdot \frac{T}{P} = \frac{8,314472 T}{28,97 P} \approx 0,287003 \frac{T}{P} \text{ (en } \text{m}^3 \text{)} \quad (6.9)$$

Finalmente, la densidad del vapor de agua en g/m^3 será

$$\text{volumen del aire seco} = \frac{n \cdot R \cdot T}{P} \quad (6.10)$$

y por tanto

$$\text{densidad del vapor de } H_2O = \frac{r_v}{\text{volumen del aire seco}} \quad (6.11)$$

En lo sucesivo, denotaremos la densidad del vapor de agua por H_2O .

6.2.1.5. Densidad de O_3 (g/m^3)

El factor de conversión depende de la temperatura según la ecuación general

$$\frac{mg}{m^3} = \frac{(ppmv)(12,187) \times (\text{masa molecular})}{\text{Temperatura en K}} \quad (6.12)$$

6.2.1.6. Espesor óptico de aerosoles

En nuestro caso, el modelo WRF-Chem solo proporciona de datos de espesor óptico de aerosoles en las bandas 300, 400, 600 y 1000nm, por lo que nos quedaremos con el valor más cercano a la banda en cuestión.

6.2.2. Parámetros tomados de la imagen

El resto de parámetros de entrada son tomados de la imagen Landsat que vamos a corregir. En este caso fue tomada por el sensor TM (Thematic Mapper) a bordo del satélite Landsat 8, el día 2013-08-09 a las 11:02:58. Además de los datos geométricos y de posición, los más relevantes son los de la función filtro de la banda correspondiente, las reflectancias del blanco y la media del entorno así como la radiancia del pixel objetivo.

6.2.3. Variables

Según se ha visto, las magnitudes necesarias para realizar la corrección atmosférica pueden hacer referencia a las condiciones geométricas, atmosféricas o a los valores de radiancia medidos por el sensor. Consideramos como variables independientes aquellas que corresponden a la misma magnitud pero se encuentran en diferente nivel de altura. De esta manera, el conjunto consta de 218 variables predictoras que presentamos en la tabla 6.1. Destacamos, en este punto, que prescindimos de los niveles intermedios 6, 9, 12, 15, 18 y 21 del modelo WRF, por exigencia del número de entradas del modelo 6S, pero no modificamos los índices. De tal manera que, aunque solo tenemos 34 niveles por variable, numerados de 0 a 33, existen variables denotadas con número de nivel superior a 33.

6.3. Selección del modelo: cross validation

Dividimos el conjunto de entrenamiento en tres partes: 60 % Training set, 20 % conjunto de validación, 20 % conjunto test. El procedimiento general consiste en ajustar los parámetros

Geometría	Atmósfera	Radiancia
month	h(0-33)	wavelength inf
day	P(0-33)	wavelength sup
hour	T(0-33)	ρ_t
longitude	$H_2O(0-33)$	ρ_e
latitude	$O_3(0-33)$	Radiance
i	tau(0-33)	...
j
target level
sensor level

TABLA 6.1: Variables

de cada modelo sobre el conjunto de entrenamiento y, posteriormente, estimar el error de generalización en el conjunto de validación correspondiente. Una vez finalizado el proceso de selección, podremos estimar el error real sobre los datos del conjunto de test. De tal manera, podemos asegurar que ninguno de los modelos comparados ha visto previamente ese conjunto de datos.

6.3.1. Algoritmos de regresión comparados

Comparamos 6 algoritmos de 'machine learning' (ML): regresión lineal (LR), regresión Ridge Bayesiana (BRR), regresión Lasso, regresión Ridge, SVR y regresión mediante red neuronal. En la figura 6.2, presentamos los resultados preliminares de cada algoritmos sin optimizar los parámetros ni el conjunto de variables predictoras. Buscamos el algoritmo que mejor reproduzca las reflectancias corregidas tras haber ajustado los parámetros correspondientes. A su vez, elegiremos el subconjunto de variables predictoras que proporcionen el mejor resultado. Con la reducción del número de variables, pretendemos simplificar el modelo y facilitar su interpretación en el marco de la transferencia radiativa.

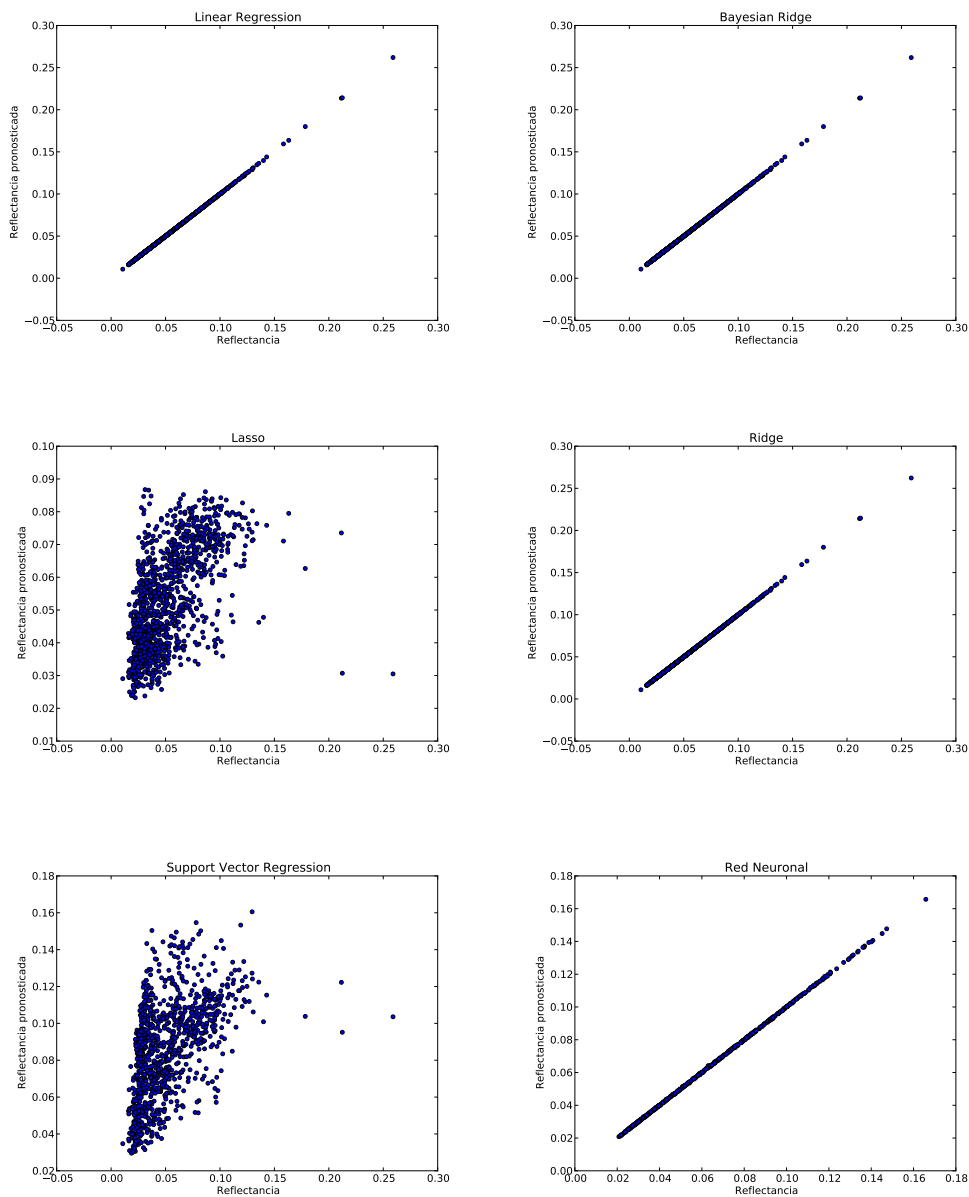


FIGURA 6.2: Comparación preliminar de los modelos de regresión

6.4. Selección de características

Para elegir el subconjunto de variables predictoras seguimos distintos procedimientos. Algunos de ellos, clasificados como filtros de características, atendían a la correlación lineal y no lineal con la variable objetivo y, en ocasiones, con los predictores mismos. Otros métodos que estudiamos, consistieron en la búsqueda directa del subconjunto óptimo. Debido al coste computacional que suponía elegir el mejor subconjunto entre todos los posibles 2^{218} , dichos subconjuntos fueron ampliándose o reduciéndose en cada paso. Presentamos sus resultados a continuación.

6.4.1. Forward Stepwise Selection

Este algoritmo funciona iterativamente, partimos del conjunto vacío y ampliamos el subconjunto, en cada paso, incluyendo la variable predictora que minimiza el error de validación. Realizamos dicha tarea para cada uno de los seis métodos de regresión comparados y para cada banda, en el apéndice mostramos los resultados que obtuvimos. Vemos en la gráfica 6.3 que el error cae rápidamente. Por esa razón, dedujimos que era suficiente considerar alrededor de 10 variables predictoras. La elección del número de variables es un compromiso entre los recursos disponibles, el error admisible y la complejidad del modelo. En nuestro caso, dado que el error cometido era admisible, en relación a la resolución radiométrica del satélite, preferimos optar por un número relativamente pequeño que facilitara la interpretación física de los resultados a la luz de la transferencia radiativa en la atmósfera.

B1	B2	B3	B4	B5	B6	B7	B8
radiance	radiance	ρ_t	radiance	ρ_t	radiance	radiance	radiance
$H_2O(33)$	$H_2O(32)$	$H_2O(32)$	$H_2O(32)$	T(33)	T(33)	j	$H_2O(32)$
$H_2O(20)$	$H_2O(31)$	i	T(20)	tau(29)	$O_3(29)$	T(35)	lon_medida
$H_2O(13)$	$H_2O(8)$	$O_3(8)$	$O_3(8)$	i	$H_2O(10)$	$H_2O(5)$	T(27)
$H_2O(10)$	tau(19)	$H_2O(8)$	$H_2O(28)$	j	T(5)	$H_2O(31)$	tau(25)
tau(0)	$H_2O(34)$	tau(26)	T(26)	tau(34)	$O_3(23)$	T(38)	tau(10)
$O_3(24)$	$O_3(22)$	$O_3(13)$	T(38)	T(38)	lon_medida	T(5)	$O_3(14)$
$O_3(13)$	T(36)	radiance	tau(38)	$H_2O(27)$	$H_2O(37)$	tau(27)	ρ_e
i	i	T(38)	$O_3(20)$	radiance	$O_3(10)$	$O_3(10)$	$O_3(16)$
tau(17)	$O_3(13)$	T(7)	h(0)	$O_3(5)$	$H_2O(13)$	$H_2O(39)$	$H_2O(33)$
$H_2O(39)$	T(28)	tau(34)	$O_3(10)$	$O_3(2)$	tau(8)	tau(36)	$O_3(26)$
...

TABLA 6.2: Variables más relevantes según el algoritmo forward stepwise en LR

Si bien, es cierto, que hubiera sido posible trabajar con el conjunto completo, ya que computacionalmente no suponía un gran coste añadido y el error era menor, preferimos quedarnos con el subconjunto y estudiar qué variables iban apareciendo como más importantes, a la hora de pronosticar la reflectancias que el modelo 6S hubiera proporcionado. En la tabla 6.2 mostramos los resultados obtenidos en el caso de la regresión lineal.

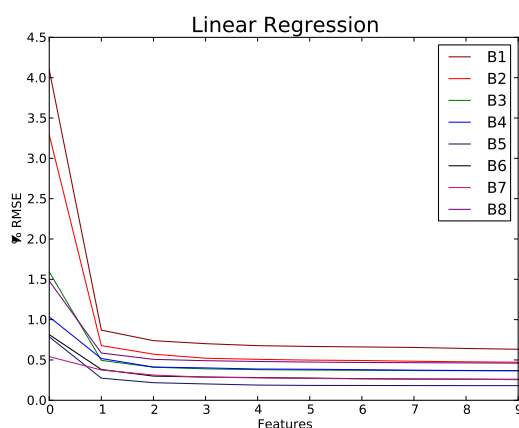


FIGURA 6.3: Evolución del error de validación con el número de variables

6.4.2. Backward Stepwise Selection

En este caso el procedimiento es opuesto al anterior. Partimos del conjunto completo y fuimos descartando, iterativamente, las variable que producía la máxima reducción del error sobre los datos del conjunto de validación. Realizamos este proceso con cada algoritmo de regresión y para cada una de las bandas. En la comparación definitiva, se consideró cada opción como un modelo distinto.

6.4.3. Correlación

Calculamos la correlación lineal de cada variable con la reflectancia dada por el 6S y ordenamos el conjunto en función del valor obtenido, de forma decreciente. Comparamos el subconjunto formado por las 10 variables con mayor correlación con el resto de subconjuntos de predictores obtenidos. 6.3

B1	B2	B3	B4	B5	B6	B7	B8
ρ_t	radiance	ρ_t	ρ_t	ρ_t	ρ_t	ρ_t	radiance
radiance	ρ_t	radiance	radiance	radiance	radiance	radiance	ρ_t
ρ_e	ρ_e	ρ_e	ρ_e	ρ_e	ρ_e	ρ_e	ρ_e
j	T(37)	T(37)	T(37)	j	j	T(37)	T(37)
T(37)	T(38)	T(38)	T(38)	T(37)	T(37)	j	T(38)
T(38)	j	j	T(36)	T(30)	T(30)	T(38)	j
$H_2O(7)$	$H_2O(7)$	$H_2O(7)$	T(31)	$H_2O(10)$	$H_2O(17)$	T(30)	T(36)
tau(10)	tau(7)	T(31)	j	$H_2O(16)$	$H_2O(10)$	T(36)	T(31)
$H_2O(16)$	tau(5)	T(36)	$H_2O(7)$	$H_2O(19)$	$H_2O(16)$	$H_2O(7)$	$H_2O(7)$
tau(7)	tau(10)	tau(7)	T(32)	T(38)	T(38)	$H_2O(10)$	T(30)
...

TABLA 6.3: Correlación lineal entre predictores y reflectancia

6.4.4. Información mutua e información mutua conjunta

Para estudiar la correlación no lineal entre los predictores y la variable objetivo, recurrimos a la información mutua. Con la información mutua conjunta, conseguimos el conjunto de variables que mayor información nos proporcionaba acerca de la variable objetivo compartiendo con el resto de predictores la mínima posible (Tourassi et al., 2001). El procedimiento para conseguir este propósito fue iterativo como vimos en el tema anterior. Partimos de la variable con máxima información de reflectancia y ampliamos dicho conjunto con aquella que maximizaba la suma de la MI menos la penalización correspondiente a la información compartida con el resto del subconjunto. En función de la importancia dada a esa penalización, construimos 10 subconjuntos. Cuando no consideramos penalización alguna, obtenemos la información mutua.

B1	B2	B3	B4	B5	B6	B7	B8
i	i	radiance	radiance	i	i	i	i
j	radiance	i	i	radiance	j	j	j
radiance	j	j	j	j	radiance	radiance	radiance
P(2)	P(1)	P(1)	P(1)	P(1)	P(0)	P(1)	P(0)
P(1)	P(0)	P(0)	P(0)	P(3)	P(1)	P(3)	P(1)
P(0)	P(2)	P(3)	P(2)	P(2)	P(2)	P(2)	P(2)
P(3)	P(3)	P(2)	P(3)	P(0)	P(4)	P(0)	P(3)
P(5)	P(4)	P(4)	P(5)	P(7)	P(3)	P(4)	P(5)
P(4)	P(5)	P(5)	P(4)	P(5)	P(5)	P(5)	P(4)
P(8)	P(7)	P(7)	P(8)	P(4)	P(7)	P(7)	P(7)
...

TABLA 6.4: MI between variables and target

6.4.5. Métodos shrinkage

En los métodos Lasso y Ridge, el parámetro de regularización nos permite seleccionar las variables más importantes. A medida que su valor crece la penalización por el valor de los coeficientes aumenta, de esta forma algunos quedan reducidos o, en el caso de Lasso, incluso anulados.

En nuestro caso los resultados no fueron tan satisfactorios como los obtenidos con las variables seleccionadas mediante otros procedimientos, por ello no aparecen en los gráficos de resultados finales 6.4 que aparecen en este capítulo sino que se han incluido en el anexo.

6.5. Parámetros de ajuste

El procedimiento general que seguimos para ajustar los parámetros de cada modelo, fue elegir una rejilla para el mismo y estimar el error de validación en cada punto de ella. Así, por ejemplo, en el caso del parámetro de regularización λ que aparece en Ridge y Lasso, tomamos 200 puntos entre 10^{-10} y 1, espaciados según la escala logarítmica. En la regresión Bayesiana, las distribuciones a priori para β y α se suponen de tipo gamma. Utilizamos una rejilla logarítmica con 5 puntos entre 10^{-10} y 1 para los cuatro parámetros que las definen: λ_1 , λ_2 , α_1 y α_2 . En SVR, además de los parámetros ϵ , 10 puntos entre 0 y 1; γ , 10 puntos entre 0 y 1; y C que tomó los valores 1,2 o 3; ajustamos el tipo de función: lineal, polinomial, rbf (radial basis function) o sigmoideal y el grado del polinomio (1,2 o 3). En el caso de la red neuronal, modificamos las funciones de activación (logística, tanh o lineal) y el número de capas ocultas (entre 1 y 10).

B1	B2	B3	B4	B5	B6	B7	B8
i	i	radiance	radiance	i	i	i	i
radiance	radiance	i	i	radiance	j	j	radiance
j	j	j	j	j	radiance	radiance	j
lat_medida	T(37)	$H_2O(11)$	P(1)	P(1)	P(0)	P(1)	T(39)
T(30)	$H_2O(16)$	T(37)	lat_medida	T(32)	P(1)	$H_2O(16)$	T(37)
T(37)	lat_medida	$H_2O(16)$	T(37)	h(10)	ρ_t	lat_medida	T(30)
T(31)	T(30)	T(31)	T(38)	h(11)	ρ_e	T(38)	ρ_t
T(38)	T(38)	lat_medida	T(31)	h(22)	T(38)	ρ_t	T(38)
h(29)	T(31)	T(38)	$H_2O(11)$	h(23)	h(13)	h(25)	T(31)
T(35)	$H_2O(11)$	h(25)	h(13)	h(24)	h(17)	T(31)	T(36)
...

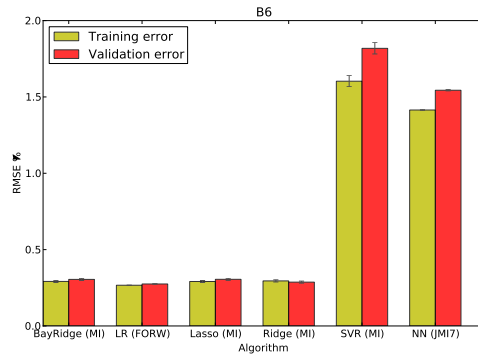
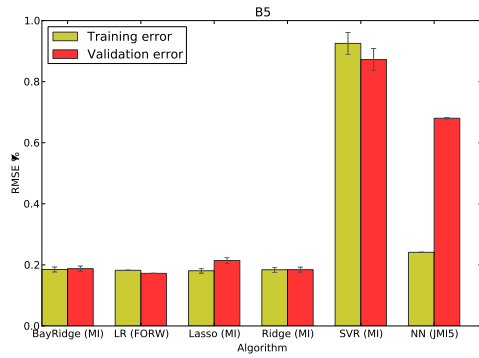
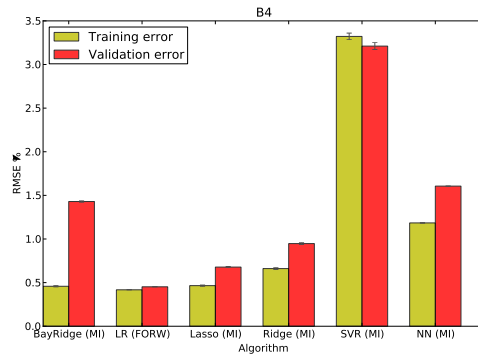
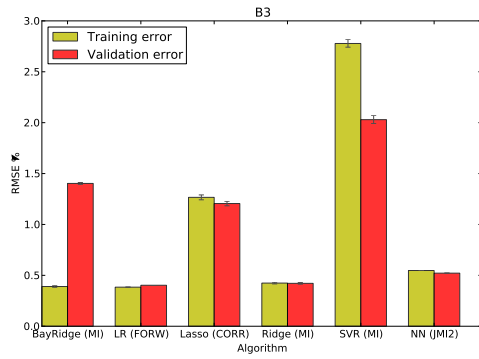
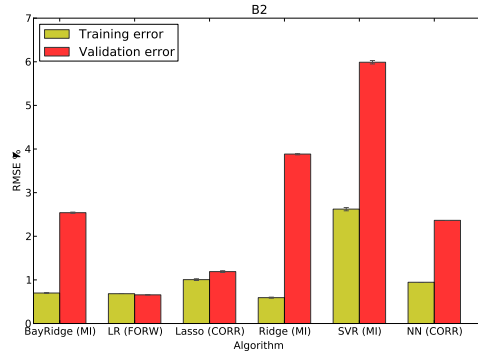
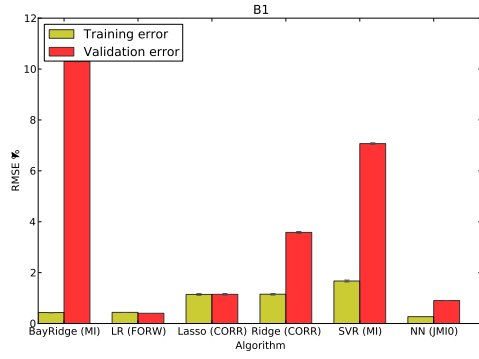
TABLA 6.5: JMI Información mutua conjunta con $\lambda = 0,05$

6.6. Selección del modelo

Para encontrar el modelo más adecuado, además de seguir el criterio de error mínimo sobre el conjunto de validación, dimos importancia a la simplicidad del mismo. Para ello, fue imprescindible la reducción del número de predictores. Observando la evolución del error de validación con el número de variables consideradas, pudimos deducir que un número cercano a diez era suficiente. A la luz de los resultados de cada uno de los algoritmos, las variables más importantes a la hora de predecir las reflectancias fueron las que presentamos en la tabla 6.2.

El estudio final consistió en considerar los diferentes grupos de variables seleccionados y los diferentes algoritmos, cada uno con sus diferentes parámetros de ajuste optimizados. En los gráficos siguientes 6.4, mostramos para cada banda los errores de entrenamiento y validación de cada algoritmo optimizado utilizando el mejor conjunto de datos correspondiente. El algoritmo que mejor funcionó, además de ser el más sencillo, fue el de regresión lineal con el conjunto de datos construido mediante forward stepwise.

Los algoritmos que, en principio, darían cuenta de las relaciones no lineales entre predictores y variable objetivo, no funcionaron tan bien como los lineales en este caso. Deducimos, por tanto, que el modelo 6S operó de forma lineal en los rangos de valores que aparecieron en nuestro problema concreto.



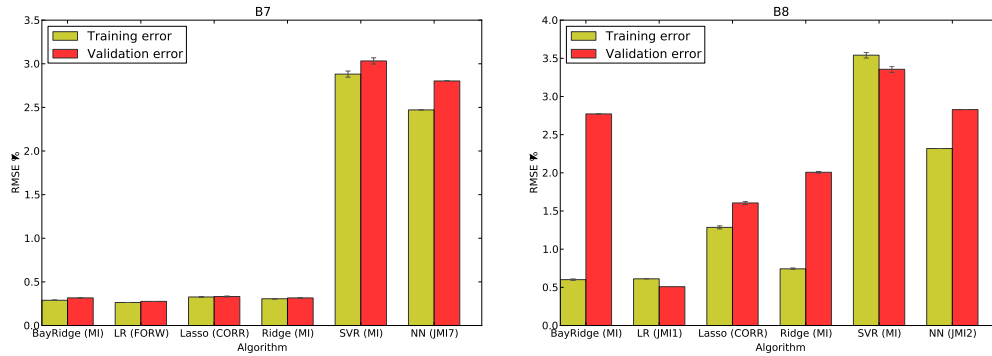
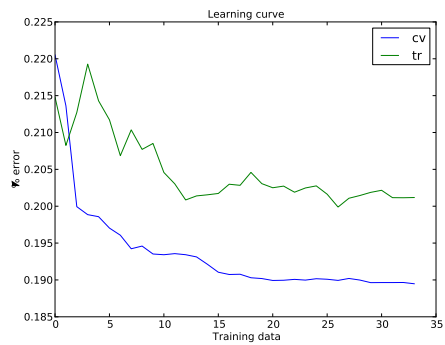
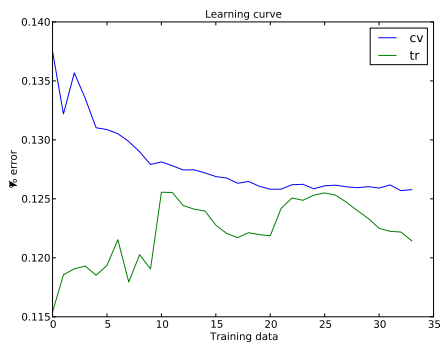
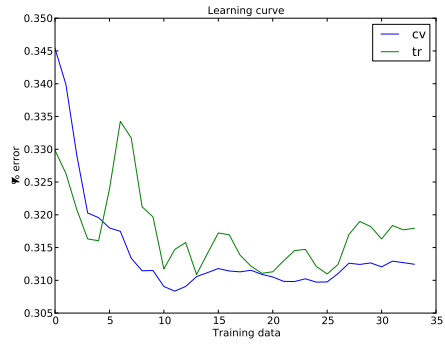
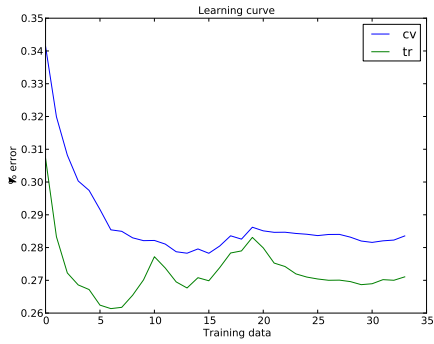
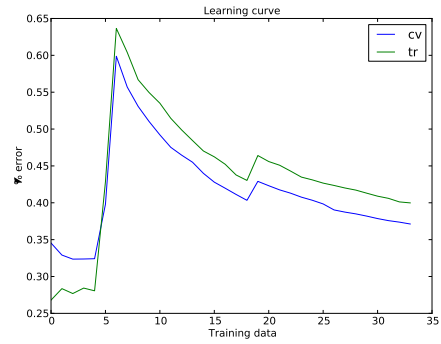
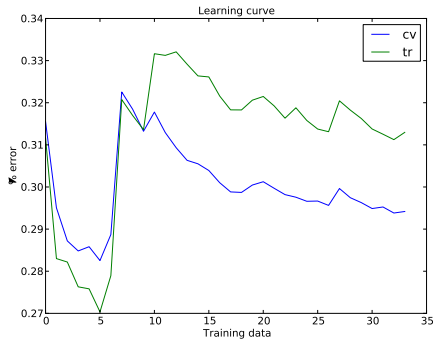


FIGURA 6.4: Errores de validación y entrenamiento para cada algoritmo en cada banda

Salvo el algoritmo Bayesian Ridge, que mostró sobre-entrenamiento en varias de las bandas, el resto, con alguna excepción, resultó compensado en cuanto a los errores cometidos en el conjunto de entrenamiento y validación.

6.7. Curvas de aprendizaje

Las curvas de aprendizaje representan la evolución de los errores de entrenamiento y validación con el número de datos utilizados en el primero. El estudio de estas curvas proporciona una herramienta de diagnóstico de la descompensación entre el bias y la varianza. Por ejemplo, si ambos errores fuesen altos y de valor similar, tendríamos evidencia de bias elevado. En este caso, no serviría de nada aumentar el conjunto de datos de entrenamiento. Sin embargo, si el error de entrenamiento fuese notablemente menor que el de validación y la evolución de esa diferencia disminuyese, estaríamos en el caso contrario; podríamos decir que el sistema tiene varianza elevada y, quizá, el aumento del número de datos de entrenamiento resultaría de utilidad. En nuestro caso, representamos estas curvas en la figura 6.5, vemos que no hay gran descompensación bias/varianza y que los errores se estabilizan, dentro de unos límites admisibles, tomando alrededor de 1000 datos en el conjunto de entrenamiento.



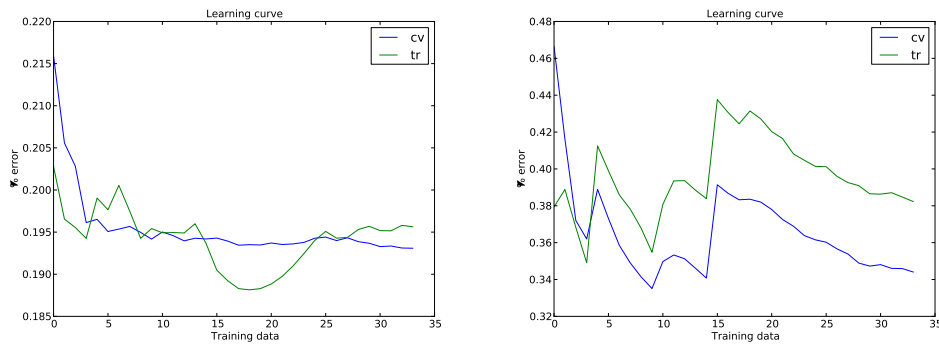


FIGURA 6.5: Curvas de aprendizaje en cada banda

6.8. Resultados

Realizamos la corrección atmosférica en las ocho primeras bandas de la imagen Landsat. En el Apéndice A, presentamos una composición RGB de la imagen corregida.

6.8.1. Error cometido

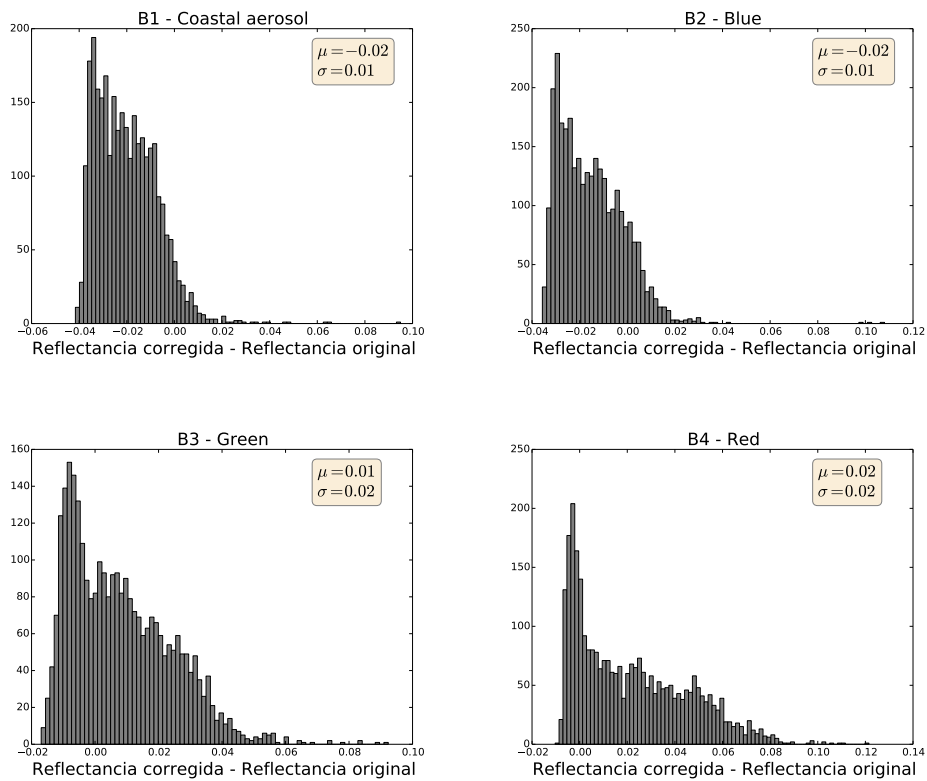
Los errores cometidos en la corrección rondaron el 0.5% en todas las bandas como podemos ver en la figura 6.4. Para evitar que la estimación del error de generalización estuviese influenciada por los datos, calculamos su error de generalización en el conjunto de test, intacto hasta el momento, consiguiendo un error del mismo orden.

6.8.2. Tiempo de cálculo

Fuimos capaces de reducir el tiempo de cálculo hasta hacerlo operativo: pasamos de una media de 3 seg por operación en cada píxel (dependiendo de la banda) a una media de $0.039\mu s$, ya que tarda $117\mu s$ en corregir los 3000 píxeles del conjunto de entrenamiento, en promedio. Notemos que la vectorización de las operaciones reduce sustancialmente el tiempo de cálculo, pasando de $20,6\mu s$ por píxel aislado a $0,039\mu s$.

6.8.3. Impacto sobre las radiancias en cada banda

En los histogramas de las correcciones que presentamos en la figura 6.6, apreciamos cómo afecta la corrección atmosférica a cada una de las bandas. La extensión del histograma da cuenta de la sensibilidad de la corrección en la banda al cambio en las condiciones espaciales. Así, en las bandas cuya varianza es de 0.01, las condiciones atmosféricas y de entorno, no provocan correcciones muy diferentes. Destacamos que en las bandas 1 y 2, gran parte de las correcciones son negativas, debido al efecto de la dispersión Rayleigh.



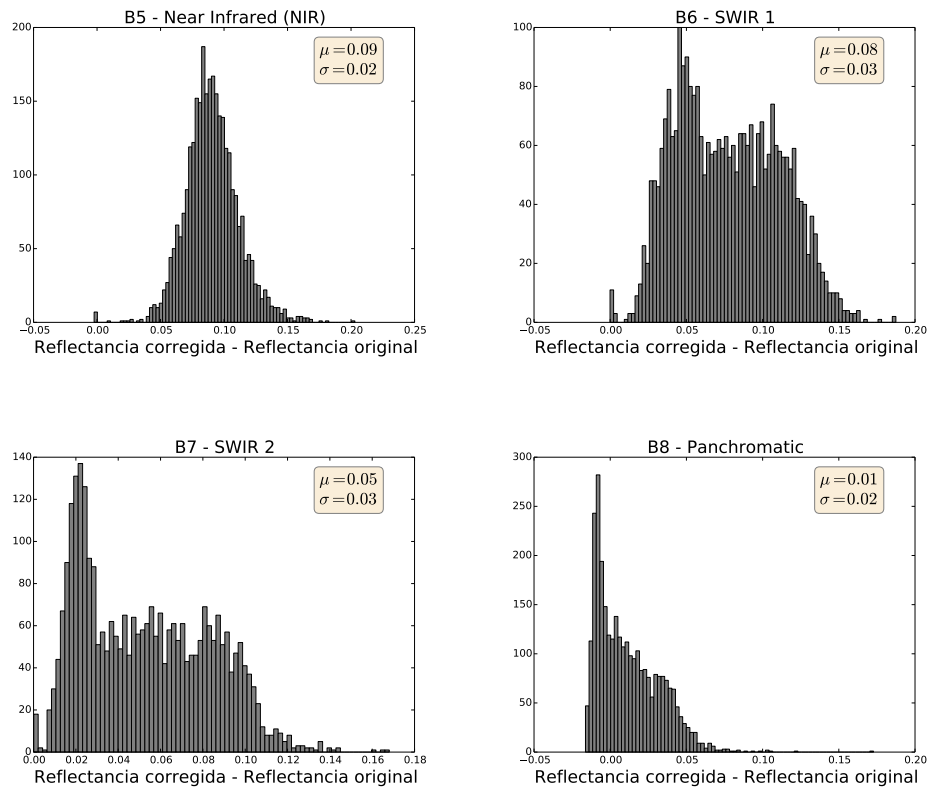


FIGURA 6.6: Histogramas de correcciones en cada banda

6.8.4. Impacto sobre NDVI

Comparamos el NDVI calculado con las bandas de reflectancia de Landsat y las corregidas, en una pequeña ventana de la imagen que abarca una zona regadío. Vemos en que el NDVI de las bandas corregidas es ligeramente mayor que el estimado con las originales.

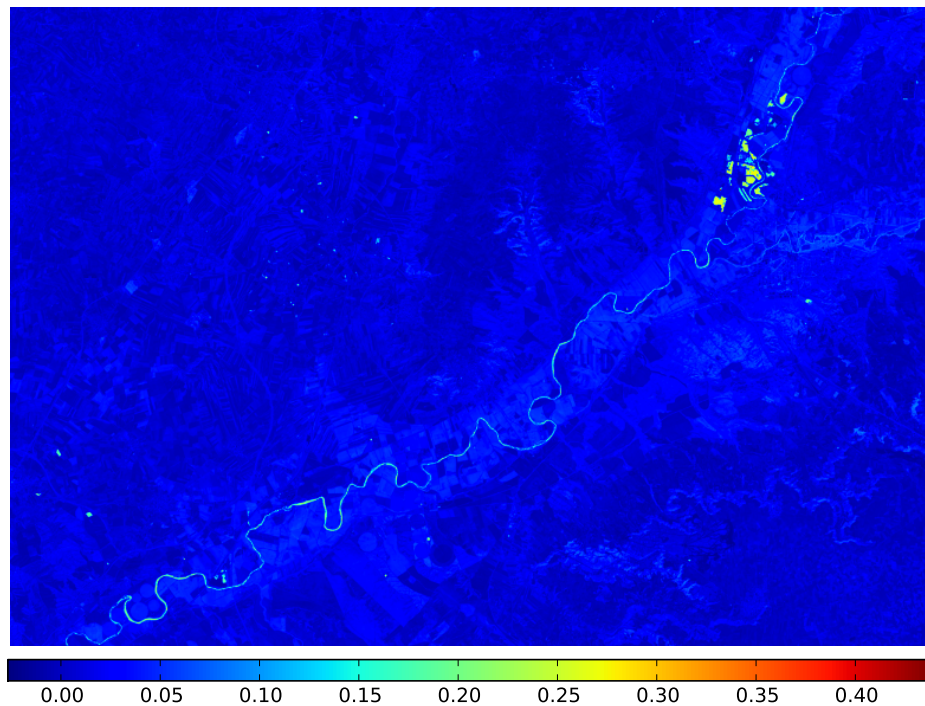


FIGURA 6.7: Diferencias de NDVI corregido y sin corregir

La diferencia es más significativa en las zonas de NDVI muy bajo, como podemos ver en la figura ???. En las zonas de NDVI muy alto, como los pivots de regadío, la diferencia está en torno al 1%. En otras zonas de regadío con NDVI menor, ronda el 3-5%.

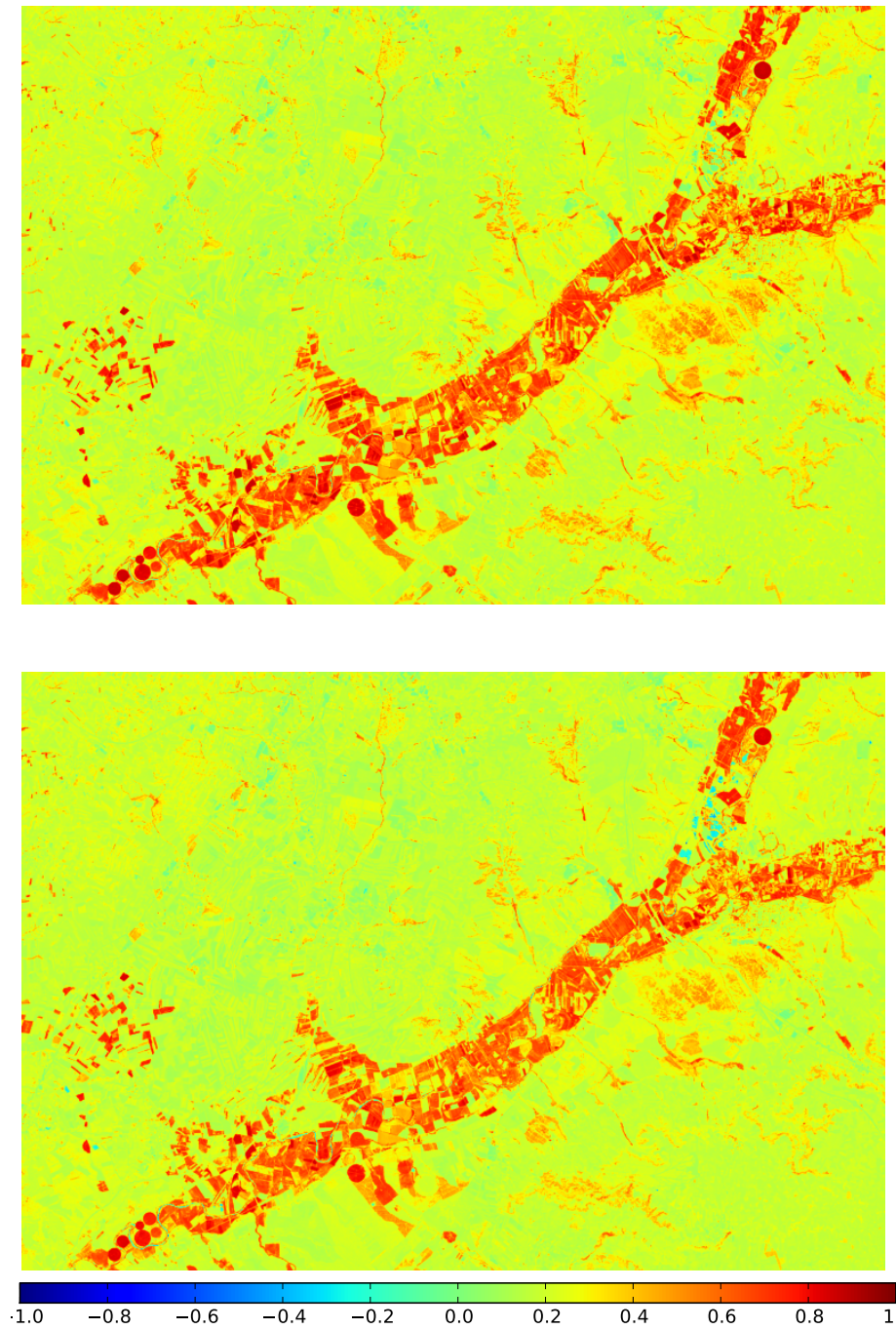
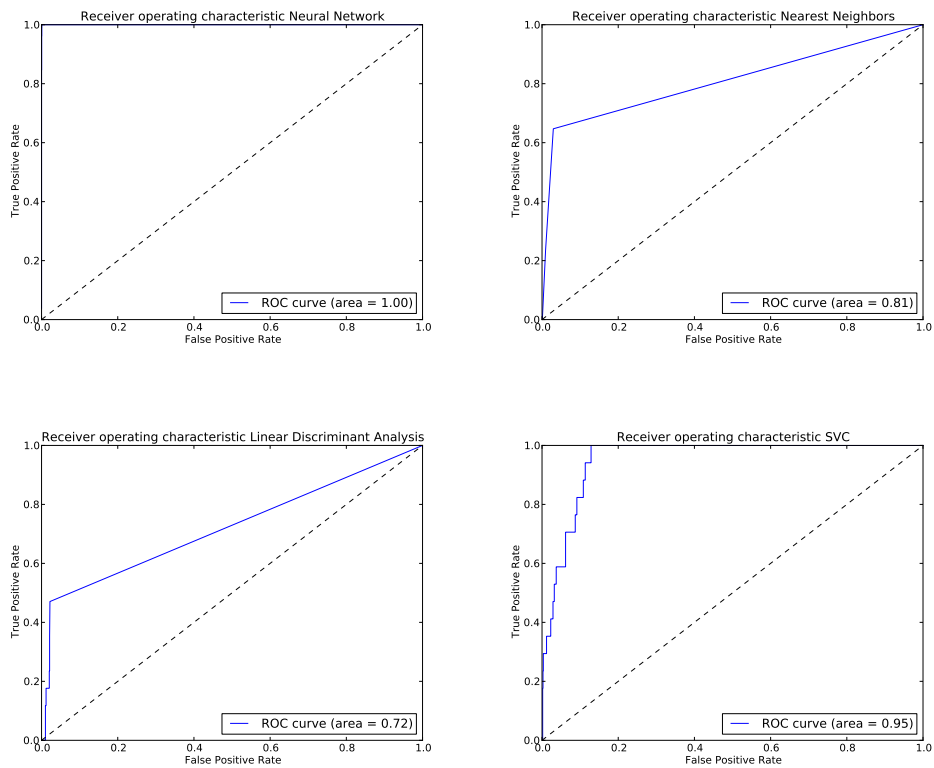


FIGURA 6.8: NDVI corregido arriba, sin corrección abajo

6.9. Algoritmos de clasificación comparados

El estudio para caracterizar las atmósferas más influyentes se llevó a cabo reformulando el problema de la regresión para convertirlo en uno de clasificación. Haciendo uso de un umbral, dividimos la imagen en dos grupos; uno contiene los píxeles en los que la corrección supera dicho umbral y el otro, los que no. De esta manera, mediante el mismo conjunto de entrenamiento, podemos realizar el estudio en el marco de la minería de datos.

Comparamos, como en el caso de la regresión, varios algoritmos de aprendizaje automático destinados a resolver problemas de clasificación. Fueron también seis: Vecinos más próximos, Análisis discriminante lineal, SVM, Random forest, Decision Trees y una red neuronal. En la figura 6.9 presentamos las curvas ROC (Receiver operating characteristic) que nos sirven para estimar la capacidad de cada algoritmo.



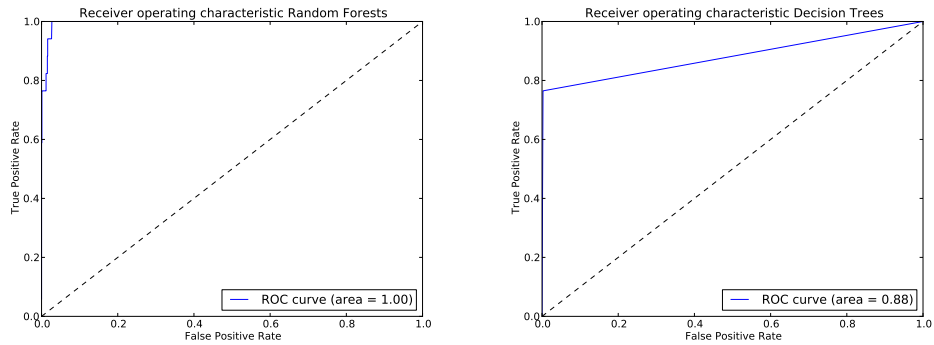
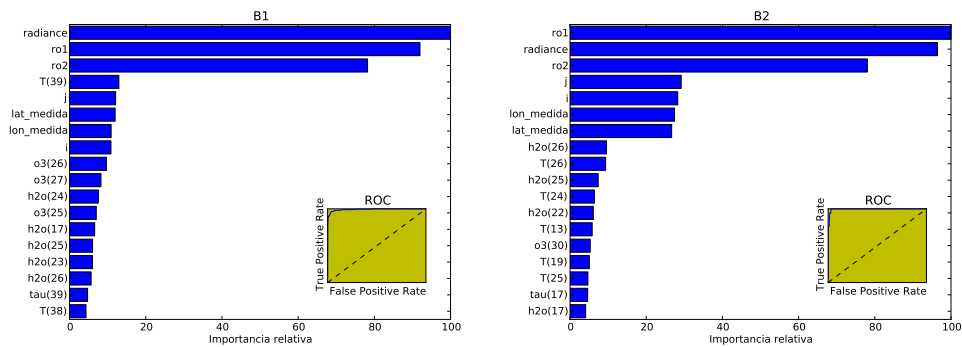


FIGURA 6.9: Comparativa de modelos de regresión

Elegimos el método random forest por ser el que mayor área bajo la curva ROC producía junto con la red neuronal y, además, permitirnos el estudio de las variables más influyentes de una manera fácilmente interpretable.

6.9.1. Variables más influyentes

Vimos en el capítulo 5 que, podemos dar una medida de la importancia de cada predictor sumando la cantidad de error que ha sido capaz de reducir en el proceso completo. Es decir, acumulamos la reducción del error, cada vez que se haya llevado a cabo una segmentación del espacio de características según esa variable. En la figura 6.10 presentamos la relación de variables predictoras más importantes y su importancia relativa, estimada como acabamos de comentar.



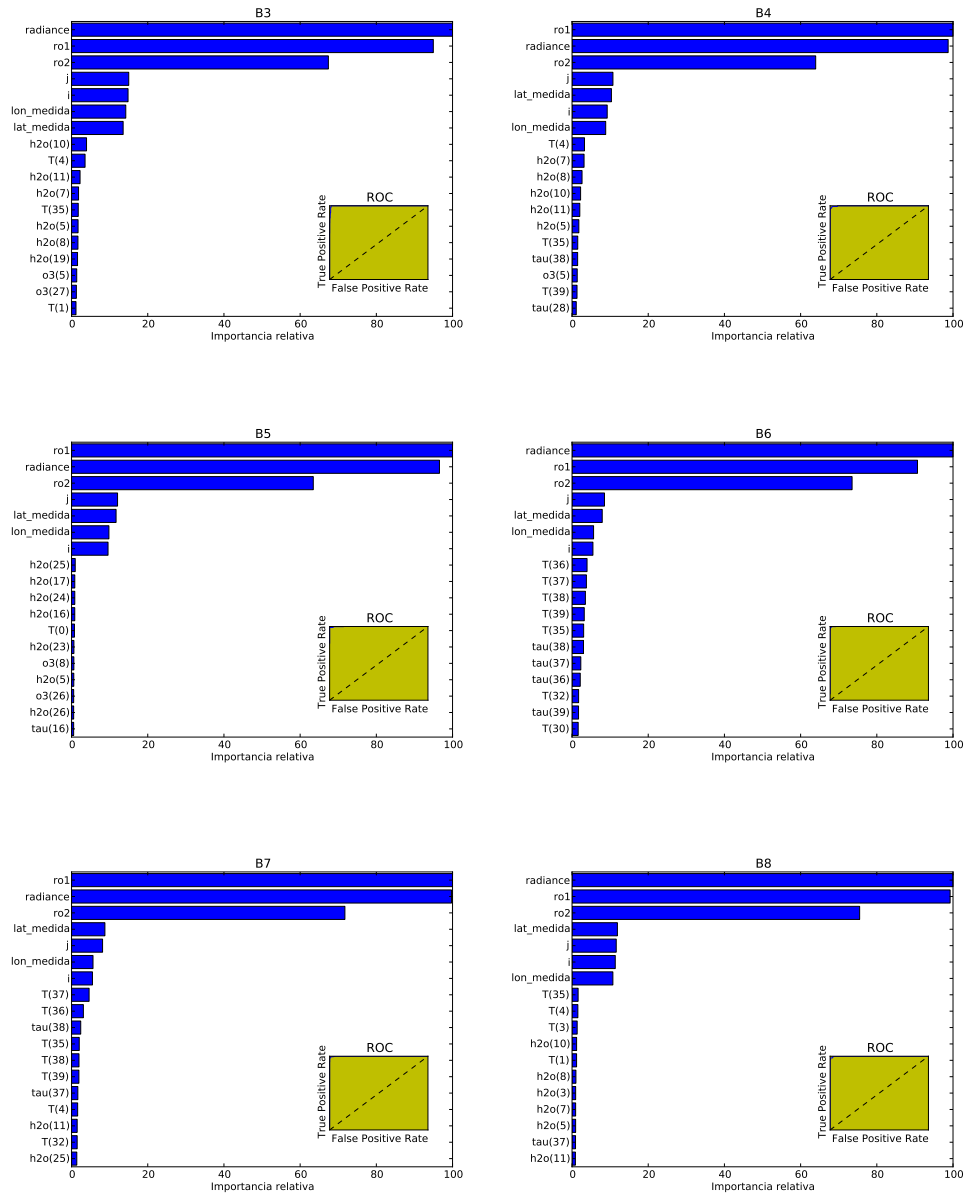


FIGURA 6.10: Predictores más importantes en cada banda

En la tabla 6.6 presentamos una lista de las variables con mayor capacidad de predicción tanto en regresión como en clasificación ordenadas por frecuencia.

Regresión	Clasificación
radiance	radiance
i	ρ_e
T(38)	ρ_t
$H_2O(32)$	lon_medida
$O_3(13)$	lat_medida
$O_3(10)$	$H_2O(11)$
$H_2O(31)$	$H_2O(25)$
T(5)	$H_2O(7)$
ρ_t	T(35)
$H_2O(8)$	T(4)
$H_2O(10)$	T(37)
...	...

TABLA 6.6: Variables más relevantes en LR y clasificación

En los gráficos 6.11 y 6.12, mostramos las correlaciones entre las variables más relevantes en regresión y clasificación

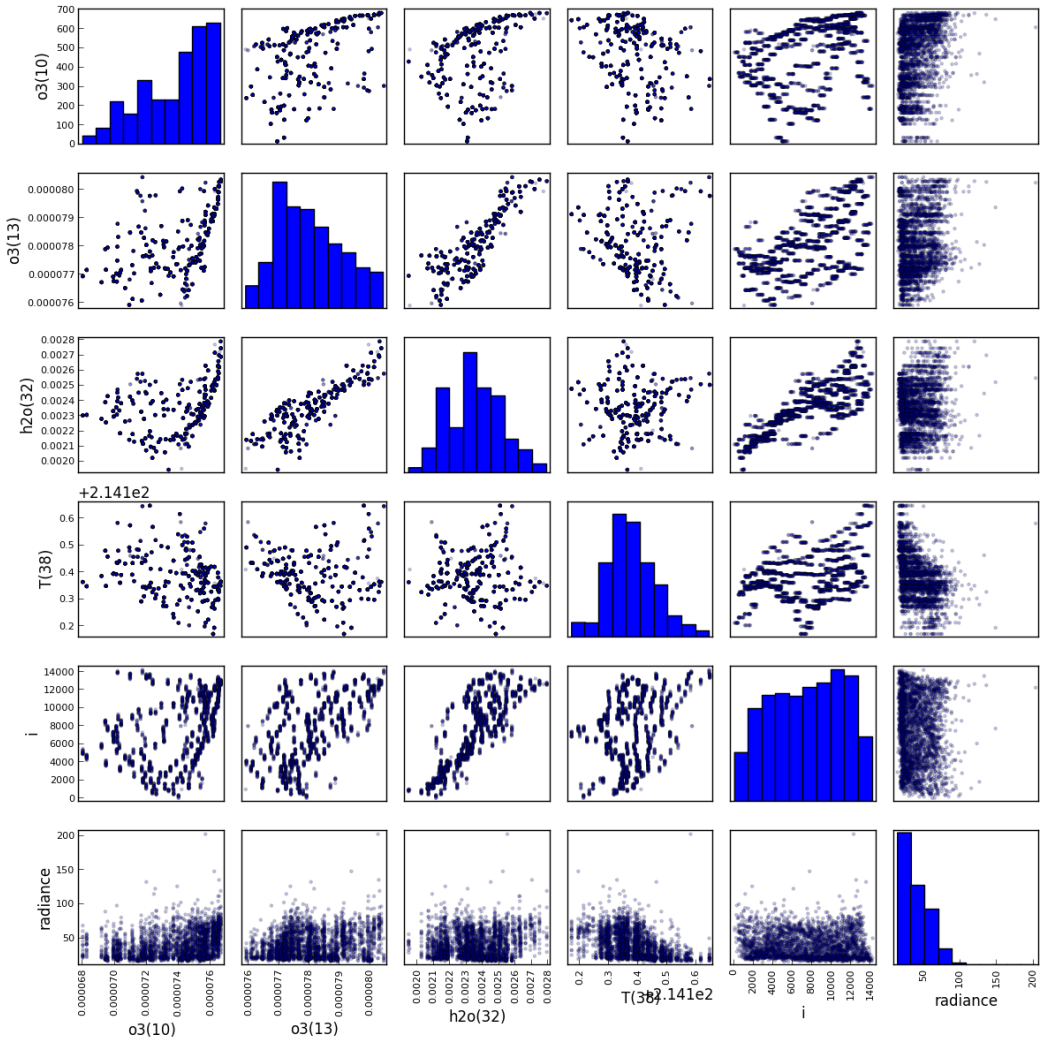


FIGURA 6.11: Variables más relevantes en LR

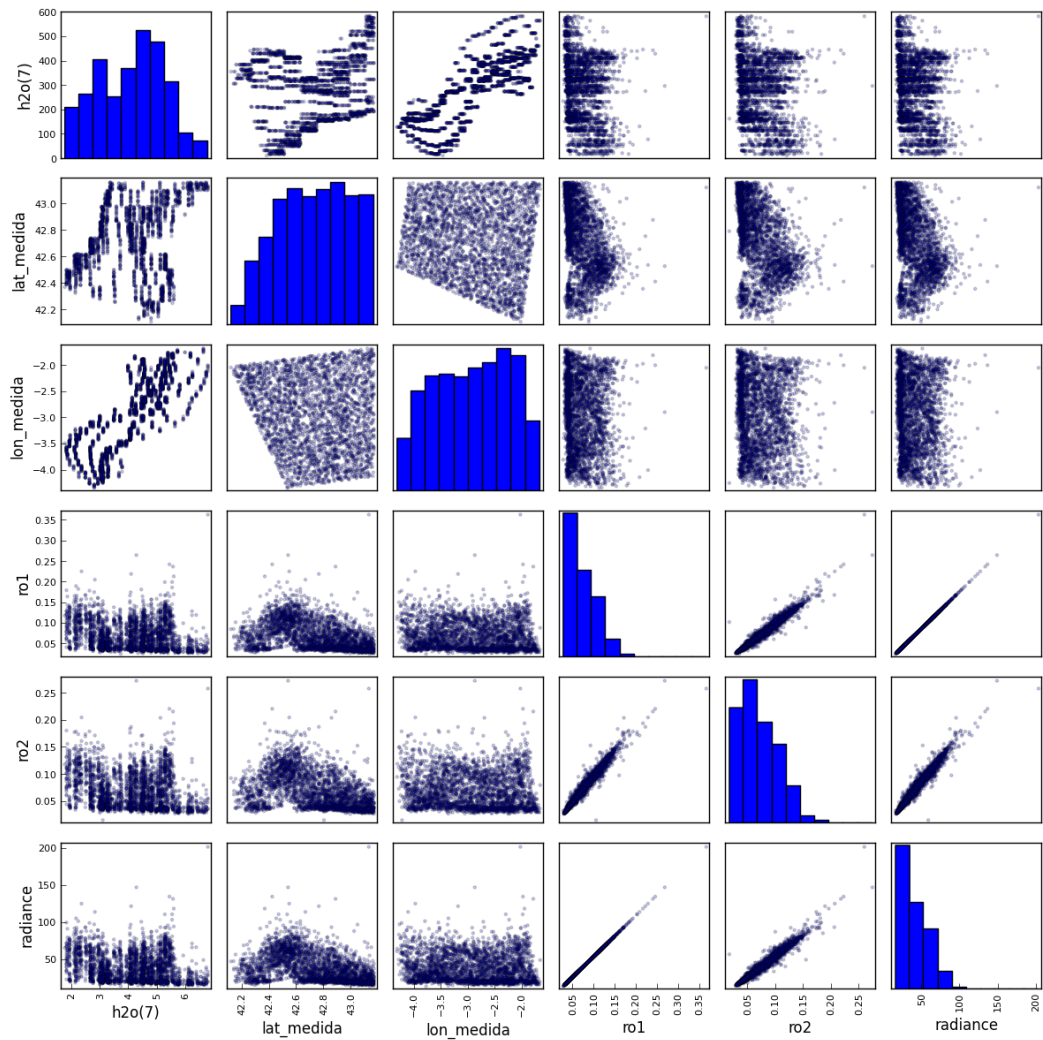


FIGURA 6.12: Variables más relevantes en clasificación

Conclusiones

- Aplicando la metodología expuesta a lo largo de esta tesis, hemos realizado la corrección atmosférica de una imagen Landsat completa, de 7931 líneas por 7801 columnas, considerando las condiciones atmosféricas de cada punto, con un error del 0.5% respecto a la corrección del RTM. De esta forma, hemos podido disminuir el tiempo de cálculo empleado por el modelo de transferencia radiativa de una media de 3 seg por operación en cada píxel, a una media de $0.039\mu s$, abriendo así la posibilidad de corregir atmosféricamente tanto imágenes VHR como largas series temporales de imágenes. Destacamos que, aunque se ha utilizado una imagen Landsat, el procedimiento es igualmente aplicable a cualquier imagen.
- Se ha conseguido que el procedimiento desarrollado en este trabajo sea ampliamente generalizable a otros casos, con independencia de la fuente de datos seleccionada. Por lo tanto, el método es igualmente aplicable para la corrección atmosférica de cualquier imagen y admite la sustitución de los datos de entrada que utilizamos en este trabajo, procedentes del modelo WRF y su módulo químico, por otros procedentes de diferentes modelos atmosféricos o de componentes químicos, incluso sondeos verticales, medidas de satélite, datos de estaciones meteorológicas, red AERONET (AErosol RObotic NETwork), etcétera.
- La técnica desarrollada puede ser aplicada al caso concreto del análisis de series temporales, donde la corrección atmosférica cobra gran importancia. Dado el volumen de datos involucrados en dicho análisis, el coste computacional resulta muy elevado. El procedimiento sería el mismo que el desarrollado en esta tesis y supondría un ahorro aproximado de 10^{-8} veces el tiempo de cálculo necesario para llevar a cabo la corrección pixel a pixel, vía modelo de transferencia radiativa.
- Hemos determinado los predictores y el modelo de regresión más adecuados para reproducir las reflectancias que estimaría el modelo 6S. En este caso, el conjunto de predictores hallado mediante el procedimiento de búsqueda 'forward stepwise', junto

con el método de regresión lineal, resultaron ser los más adecuados, pero el procedimiento es igualmente válido en caso de que el sistema necesite mayor complejidad para reproducir los resultados del RTM. Como se comentó, elegimos el modelo 6S porque tiene en cuenta el entorno del punto a corregir, pero existen otros modelos de transferencia radiativa más complejos como Sciatran, libRadtran o LIDORT/VLI-DORT, que pudieran requerir alguno de los métodos de aprendizaje considerados, por esta razón, se ha mantenido este grado de generalidad.

- También localizamos las variables atmosféricas que más afectan a la radiación, según el modelo 6S, reformulando el problema para convertirlo en uno de clasificación; aquellas atmósferas que implicaban una corrección mayor de 2.5 veces la desviación típica del total de correcciones se denominaron como 'muy influyentes' y, a través del método 'random forest', se detectaron. Paralelamente, encontramos las variables predictoras más importantes que caracterizan tales atmósferas.
- Hemos analizado las capacidades del modelo de transferencia 6S, encontrando que un modelo lineal es capaz de reproducir sus resultados con un error del 0.5% en el caso concreto tratado en esta tesis. También comprobamos que los métodos no lineales han sido menos eficaces en este caso.
- Esperamos que el uso de los algoritmos de 'machine learning', aquí desarrollados, nos sea de utilidad en muchos ámbitos de la teledetección tales como el registro de imágenes, la fusión, corrección geométrica, reconocimiento de patrones, etcétera. Asimismo, esperamos que nos permita abordar otros problemas como por ejemplo la extracción de información de imágenes multispectrales. En el caso concreto de la corrección atmosférica, puede ampliarse el rango de variación de los perfiles de forma controlada para ampliar el análisis de las capacidades de los RTM.

Apéndice A

Imágenes Landsat

El programa Landsat se puso en marcha el 23 de julio 1972 constituyendo, así, el más largo en la historia de la adquisición de imágenes terrestres desde satélite. Ha consistido en una serie de misiones de observación de la tierra gestionadas conjuntamente por la NASA y el Servicio Geológico de Estados Unidos (USGS). La última de ellas, originalmente llamada Misión de Continuidad de Datos de Landsat (LDCM), se inició el 11 de febrero de 2013. Tras los primeros 108 días en órbita, en los que LDCM fue sometido a pruebas por parte de la NASA, el 30 de mayo 2013 el USGS comenzó a estionar sus operaciones y pasó a llamarse oficialmente Landsat 8.

El satélite Landsat 8 incorpora dos instrumentos de barrido: Operational Land Imager (OLI), y un sensor térmico infrarrojo llamado Thermal Infrared Sensor (TIRS).

Las bandas espectrales del sensor OLI, aunque similares a las del sensor Landsat 7 ETM +, han sido mejoradas con respecto a las misiones Landsat anteriores, debido a la incorporación de dos nuevas bandas espectrales: un canal en el azul, banda 1, diseñado específicamente para el estudio de los recursos hídricos e investigación en zonas costeras; y un nuevo canal infrarrojo, banda 9, para la detección de nubes cirrus. Adicionalmente, una nueva banda de control de calidad se incluye con cada producto de datos generado. Esto proporciona información más detallada sobre la presencia de características tales como las nubes, agua y nieve. En la tabla A.1, tomada de <http://landsat.usgs.gov>, vemos dónde se sitúan las bandas de los sensores Landsat 7 y 8 con respecto a la absorción atmosférica.

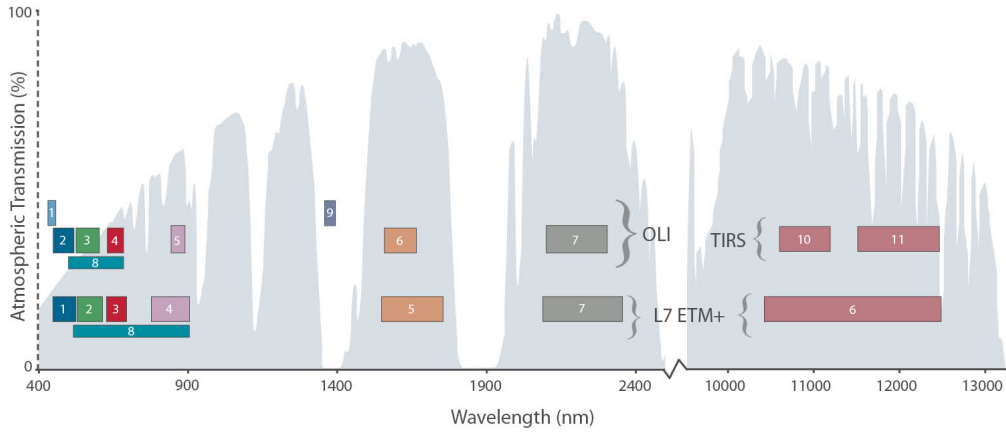


FIGURA A.1: Bandas de los sensores OLI y TIRS en Landsat 8 y ETM en Landsat 7.
Figura tomada de la web landsat.gsfc.nasa.gov

En la tabla A.1, presentamos las bandas del sensor OLI y en la figura A.2 su respuesta espectral. En el presente trabajo, aplicamos la corrección atmosférica en las bandas 1-8.

Banda	μm	Resolución
Band 1	Coastal aerosol 0.43 - 0.45	30
Band 2	Blue 0.45 - 0.51	30
Band 3	Green 0.53 - 0.59	30
Band 4	Red 0.64 - 0.67	30
Band 5	Near Infrared (NIR) 0.85 - 0.88	30
Band 6	SWIR 1 1.57 - 1.65	30
Band 7	SWIR 2 2.11 - 2.29	30
Band 8	Panchromatic 0.50 - 0.68	15
Band 9	Cirrus 1.36 - 1.38	30
Band 10	Thermal Infrared (TIRS) 1 10.60 - 11.19	100
Band 11	Thermal Infrared (TIRS) 2 11.50 - 12.51	100

TABLA A.1: Bandas Landsat

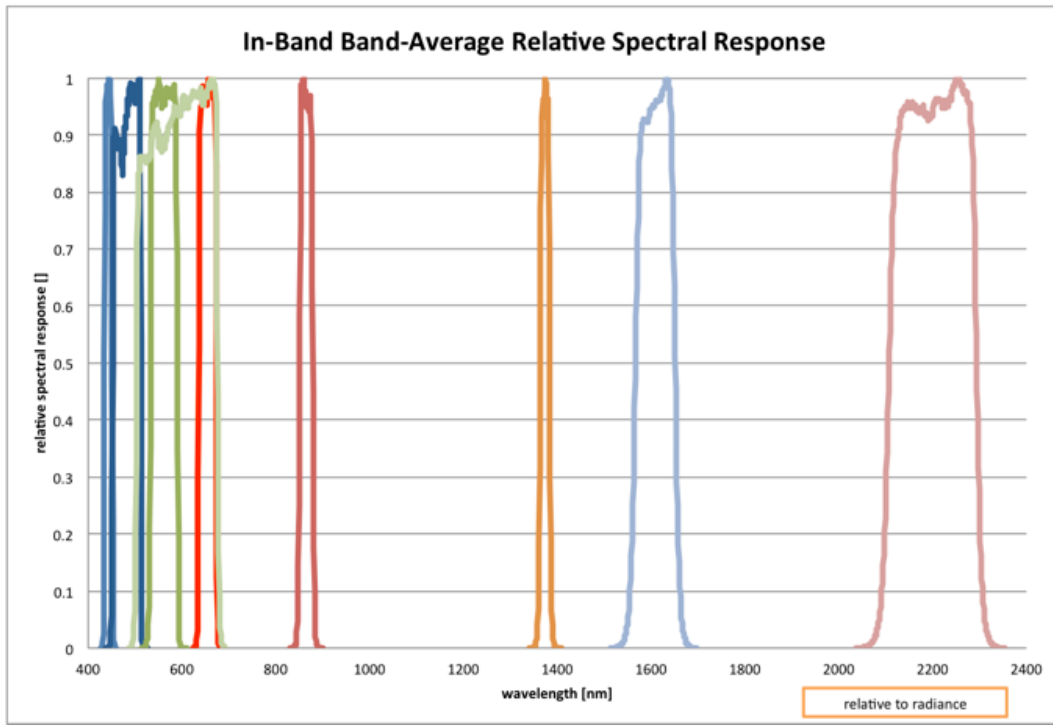


FIGURA A.2: Respuesta espectral del sensor OLI.
 Figura tomada de la web landsat.gsfc.nasa.gov

Las características de los datos de Productos LDCM a Nivel 1 son

- Formato de datos GeoTIFF
- Remuestreo por convolución cúbica (CC)
- Norte arriba (MAP) de orientación
- Proyección cartográfica: Universal Transversal Mercator (UTM) (estereográfica polar de la Antártida)
- Datum al Sistema Geodésico Mundial (WGS) 84
- 12 metros de error circular, 90% de confianza exactitud global para OLI
- 41 metros de error circular, 90% de confianza exactitud global para TIRS
- Los valores de píxel en 16 bits

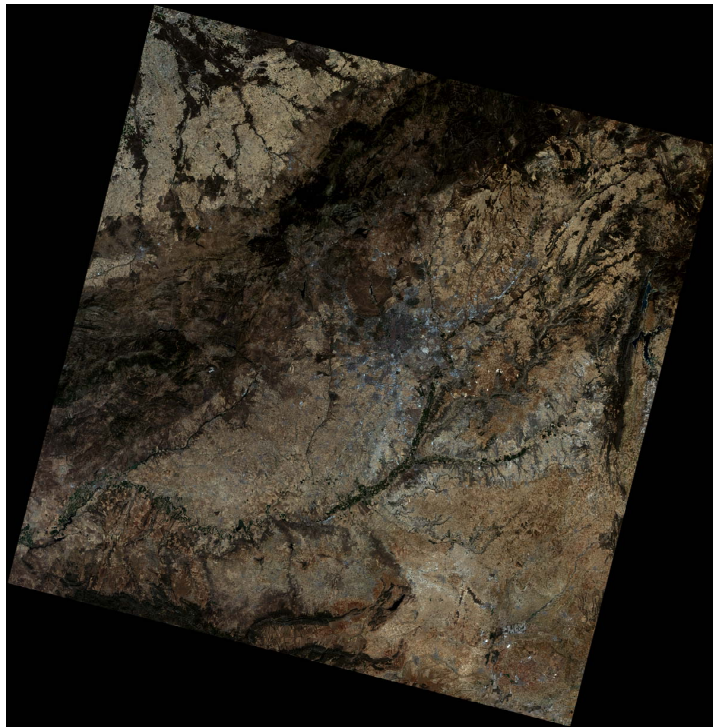


FIGURA A.3: Composición RGB de imagen Lansat 8 corregida.

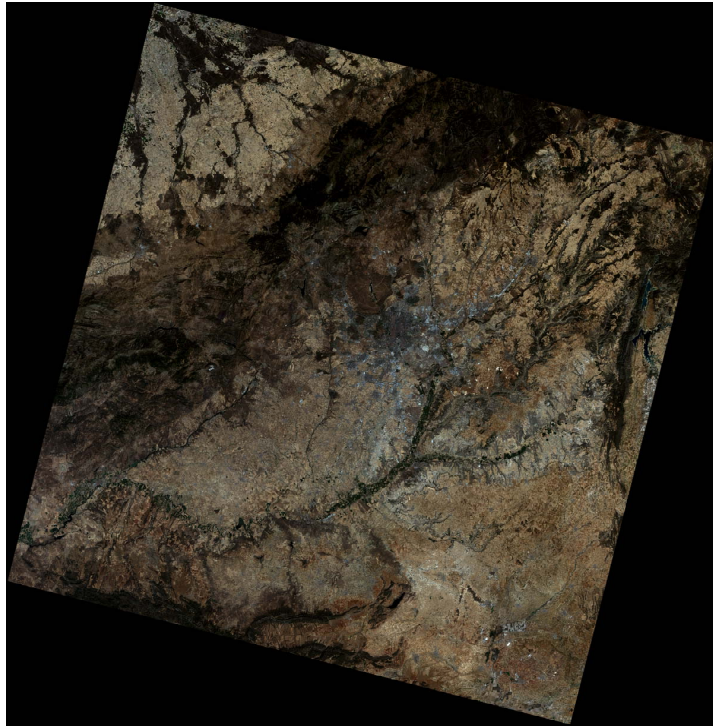


FIGURA A.4: Composición RGB de imagen Landsat 8 sin corrección.

Apéndice B

Modelo de transferencia radiativa 6S

Los procesos atmosféricos que modifican la radiancia solar reflejada por un blanco cuando se observa desde el satélite son dos: la absorción por gases y el scattering por aerosoles. En el modelo 6S, se supone que ambos efectos pueden desacoplarse y considera que los gases se encuentran en una capa por encima de los agentes difusores. De esa manera, la ecuación de transferencia para la radiación solar reflejada por un blanco lambertiano y homogéneo de reflectancia ρ_s , al nivel del sensor, es (Vermote et al., 1997a)

$$\rho_{\text{ap}}(\theta_s, \theta_v, \phi) = \rho_{\text{atm}}(\theta_s, \theta_v, \phi) + T_{\text{atm}}(\theta_s) \cdot T_{\text{atm}}(\theta_v) \cdot \frac{\rho_t}{1 - \rho_t \cdot S} \quad (\text{B.1})$$

donde ρ_{ap} , es la reflectancia aparente al nivel del satélite; ρ_{atm} , la reflectancia en el techo de la atmósfera; $T_{\text{atm}}(\theta_s)$, $T_{\text{atm}}(\theta_v)$ las transmitancias atmosféricas en el ángulo solar y de observación; ρ_t , la reflectancia del suelo; y S , el albedo esférico de la atmósfera. La transmitancia atmosférica tiene dos contribuciones,

$$T_{\text{atm}}(\theta_v) = e^{\tau/\mu_v} + t_d(\theta_v), \quad (\text{B.2})$$

correspondientes a la radiación directa y difusa. La difusa está relacionada con la influencia del entorno ya que, debido al scattering producido por la atmósfera, llegará radiancia al sensor procedente de puntos de los alrededores al blanco.

Si, ahora, asumimos que la superficie reflectora no es uniforme, tendremos

$$\rho_{\text{ap}}(\theta_s, \theta_v, \phi) = \rho_{\text{atm}}(\theta_s, \theta_v, \phi) + \frac{T_{\text{atm}}(\theta_s)}{1 - \rho_e \cdot S} \cdot [\rho_c(M)e^{\tau/\mu_v} + \rho_e(M)t_d(\theta_v)] \quad (\text{B.3})$$

donde, para un pequeño blanco M , $\rho_c(M)$ representa su reflectancia y $\rho_e(M)$ la de su entorno.

Reflector Lambertiano perfecto: $\rho(\theta_s, \theta_v, \phi) = 1$ Reflector Lambertiano: $\rho(\theta_s, \theta_v, \phi) = \rho$

En el camino descendente, la radiación puede reflejarse en la superficie de la atmósfera, ρ_{atm} , o incidir en ella. Siendo la transmitancia de la atmósfera T_{atm} ,

$$T_{\text{atm}} = \frac{E_t}{E_i}, \quad (\text{B.4})$$

la radiancia que alcanza el suelo será E_t . En la superficie de la Tierra, la reflectancia del blanco (target) es ρ_t y se refleja la cantidad E_r .

$$E_t \cdot \rho_t = E_r. \quad (\text{B.5})$$

En el camino ascendente, E_r incide de nuevo en la atmósfera y dependiendo de la transmitancia de ésta en la dirección θ_v de observación $T_{\text{atm}}(\theta_v)$, la radiancia E_0 que alcanzará el sensor será

$$E_0 = T_{\text{atm}}(\theta_v) \cdot E_r \quad (\text{B.6})$$

La reflectancia aparente será la suma de la ocurrida en el techo de la atmósfera y la ocurrida en la superficie terrestre

$$\rho = \rho_{\text{atm}} + \frac{E_0}{E_i} \quad (\text{B.7})$$

Haciendo uso de las expresiones anteriores, podemos reescribir la reflectancia terrestre de la siguiente forma

$$\frac{E_0}{E_i} = \frac{T(\theta_v) \cdot E_r}{E_i} = \frac{T(\theta_v) \cdot E_t \rho_t}{E_i} = T(\theta_s) \cdot T(\theta_v) \cdot \rho_t. \quad (\text{B.8})$$

Si, en lugar de considerar solo una interacción con la atmósfera, aceptamos la posibilidad de que la radiación del entorno llegue al sensor tras sufrir una reflexión en la atmósfera, el proceso se resumiría con los siguientes pasos

(i) $E_i \cdot T(\theta_s)$

(ii) $E_i \cdot T(\theta_s) \cdot \rho_t$

(iii) $E_i \cdot T(\theta_s) \cdot \rho_t \cdot S$

$$(iv) E_i \cdot T(\theta_s) \cdot \rho_t \cdot S \cdot \rho_t$$

$$(v) E_i \cdot T(\theta_s) \cdot \rho_t \cdot S \cdot \rho_t \cdot T(\theta_v)$$

Si considerásemos más interacciones, la expresión general sería

$$\rho(\theta_s, \theta_v, \phi) = \rho_{\text{atm}} + T(\theta_s)T(\theta_v)\rho_t[1 + \rho_t \cdot S + (\rho_t \cdot S)^2 + (\rho_t \cdot S)^3 + \dots]. \quad (\text{B.9})$$

Siendo la suma de la serie

$$1 + r + r^2 + r^3 + \dots + r^{n-1} = \frac{1 - r^n}{1 - r}, \quad (\text{B.10})$$

la expresión entre corchetes quedaría de la siguiente manera

$$\rho_t \cdot S < 1 \Rightarrow \lim_{n \rightarrow \infty} (\rho_t \cdot S)^n = 0 \quad (\text{B.11})$$

$$\Rightarrow [\dots] = \frac{1}{1 - \rho_t \cdot S}. \quad (\text{B.12})$$

De esta forma, recuperamos la expresión que mostramos al principio del apéndice.

Apéndice C

Ficheros de configuración

C.1. WRF

Tomamos un paso temporal de 40 segundos. En la tabla C.1 presentamos la parametrización física elegida para las simulación del modelo WRF.

Archivos de entrada fueron tomados de NCEP FNL (Final) Operational Global Analysis (<http://rda.ucar.edu/datasets/ds083.2/>)

Parámetro	Opción
mp_physics	4, WRF Single-Moment 5-class scheme
ra_lw_physics	1, RRTM scheme: Rapid Radiative Transfer Model
ra_sw_physics	2, Goddard shortwave
sf_sfclay_physics	1, no surface-layer
sf_surface_physics	2, unified Noah land-surface model
bl_pbl_physics	1, YSU scheme, use sf_sfclay_physics=1
cu_physics	5, New Grell scheme (G3)

TABLA C.1: Opciones parametrización física

C.2. WRFDA

Archivos de observaciones tomados de

<http://rda.ucar.edu/data/ds337.0/tarfiles/prepbufr>

Archivos de datos ATOVS tomados de

<http://rda.ucar.edu/data/ds735.0/1bamua/>
<http://rda.ucar.edu/data/ds735.0/1bamub/>
<http://rda.ucar.edu/data/ds735.0/1bhrs3/>
<http://rda.ucar.edu/data/ds735.0/1bhrs4/>

En la tabla C.2 presentamos la parametrización física elegida para la ejecución del modelo WRFDA.

Parámetro	Opción
mp_physics	4, WRF Single-Moment 5-class scheme
ra_lw_physics	1, RRTM scheme: Rapid Radiative Transfer Model
ra_sw_physics	2, Goddard shortwave
sf_sfclay_physics	1, no surface-layer
sf_surface_physics	2, unified Noah land-surface model
bl_pbl_physics	1, YSU scheme, use sf_sfclay_physics=1
cu_physics	5, New Grell scheme (G3)
num_land_cat	24
num_soil_cat	16

TABLA C.2: Opciones parametrización física

En la tabla C.3, presentamos la lista de opciones activadas para los datos asimilados.

Parámetro	Opción
USE_SYNOPOBS	T
USE_SHIPSOBS	T
USE_METAROBS	T
USE_SOUNDROBS	T
USE_MTGIRSOBS	F
USE_TAMDAROBS	F
USE_PILOTROBS	F
USE_AIREPOBS	T
USE_GEOAMVOBS	T
USE_POLARAMVOBS	T
USE_BOGUSOBS	F
USE_BUOYROBS	T
USE_PROFILEROBS	T
USE_SATEMROBS	F
USE_GPSZTROBS	F
USE_GPSPWROBS	F
USE_GPSREFROBS	F
USE_QSCATROBS	T
USE_AIRSRETOBS	F
use_ssmiretrievalobs	F
use_amsuaobs	T
use_amsubobs	T
use_hirs3obs	T
use_hirs4obs	T
use_mhsobs	F
use_airsobs	F
use_eos_amsuaobs	F
USE_OBS_ERRFAC	F

TABLA C.3: Opciones de asimilación de datos

C.3. PREP-CHEM-SRC

A continuación presentamos las opciones de bases de datos elegidas para su uso en el modelo WRF-Chem.

Base de datos	Opción
use_retro	1 ! 1 yes, 0 not
use_edgar	2 ! 0 - not, 1 - Version 3, 2 - Version 4 for some species
use_gocart	1
use_fwbaawb	0
use_bioge	0 ! 1 - geia, 2 - megan
use_gfedv2	0
use_bbem	1
use_bbem_plumerise	1

TABLA C.4: Opciones de emisiones

Apéndice D

Resultados intermedios

Ver CD anexo.

Bibliografía

- Ackermann, I. J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F. S., and Shankar, U. (1998). Modal aerosol dynamics model for europe: development and first applications. *Atmospheric Environment*, 32(17):2981 – 2999.
- Andres, R. J. and Kasgnoc, A. D. (1998). A time-averaged inventory of subaerial volcanic sulfur emissions. *Journal of Geophysical Research: Atmospheres*, 103(D19):25251–25261.
- Barker, D. M., Huang, W., Guo, Y. R., Bourgeois, A. J., and Xiao, Q. N. (2004). A three-dimensional variational data assimilation system for mm5: Implementation and initial results. *Monthly Weather Review*, 132(4):897–914.
- Binkowski, F. S. and Shankar, U. (1995). The regional particulate matter model: 1. model description and preliminary results. *Journal of Geophysical Research: Atmospheres*, 100(D12):26191–26209.
- Bishop, C. M. et al. (1995). Neural networks for pattern recognition.
- Bishop, C. M. et al. (2006). *Pattern recognition and machine learning*, volume 1. springer New York.
- Bjerknes, V. (1904). Weather forecasting as a problem in mechanics and physics. *Meteor. Zeitschr*, 21(1):7.
- Campbell, J. B. (2002). *Introduction to remote sensing*. CRC Press.
- Chang, J., Middleton, P. B., Stockwell, W. R., Walcek, C. J., Pleim, J. E., Lansford, H. H., Madronich, S., Binkowski, F. S., Seaman, N. L., and Stauffer, D. R. (1989). *The regional acid deposition model and engineering model*, volume I- Emissions, Atmospheric Processes and Deposition. National Acid Precipitation Assessment Program.
- Chavez Jr, P. S. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote sensing of environment*, 24(3):459–479.

- Chavez Jr, P. S. (1989). Radiometric calibration of landsat thematic mapper multispectral images. *Photogrammetric Engineering and Remote Sensing*, 55(9):1285–1294.
- Chavez Jr, P. S. (1996). Image-based atmospheric corrections-revisited and improved. *Photogrammetric engineering and remote sensing*, 62(9):1025–1035.
- Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A., and Nakajima, T. (2002). Tropospheric aerosol optical thickness from the gocart model and comparisons with satellite and sun photometer measurements. *Journal of the Atmospheric Sciences*, 59(3):461–483.
- Chin, M. and Jacob, D. J. (1996). Anthropogenic and natural contributions to tropospheric sulfate: A global model analysis. *Journal of Geophysical Research: Atmospheres*, 101(D13):18691–18699.
- Chin, M., Rood, R. B., Lin, S. J., Müller, J. F., and Thompson, A. M. (2000). Atmospheric sulfur cycle simulated in the global model gocart: Model description and global properties. *Journal of Geophysical Research: Atmospheres*, 105(D20):24671–24687.
- Colwell, R. N. (1997). History and place of photographic interpretation. *Manual of photographic interpretation*, 2:33–48.
- Cooke, W. F. and Wilson, J. J. N. (1996). A global black carbon aerosol model. *Journal of Geophysical Research: Atmospheres*, 101(D14):19395–19409.
- DeMore, W. B., Molina, M. J., Sander, S., Golden, D., Hampson, R. F., Kurylo, M. J., Howard, C. J., and Ravishankara, A. R. (1987). Chemical kinetics and photochemical data for use in stratospheric modeling evaluation number 8.
- Du, Q., Fowler, J. E., and Zhu, W. (2009). On the impact of atmospheric correction on lossy compression of multispectral and hyperspectral imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(1):130–132.
- Duda, R. O., Hart, P. E., and Stork, D. G. (1999). *Pattern classification*. John Wiley & Sons,.
- Duncan, B. N., Martin, R. V., Staudt, A. C., Yevich, R., and Logan, J. A. (2003). Interannual and seasonal variability of biomass burning emissions constrained by satellite observations. *Journal of Geophysical Research: Atmospheres*, 108(D2):ACH 1–1–ACH 1–22.
- Elachi, C. and Van Zyl, J. J. (2006). *Introduction to the physics and techniques of remote sensing*, volume 28. John Wiley & Sons.
- Evensen, G. (2007). *Data assimilation*. Springer.

- Freitas, S. R., Longo, K. M., Alonso, M. F., Pirre, M., Marecal, V., Grell, G., Stockler, R., Mello, R. F., and Sánchez Gácita, M. (2011). Prep-chem-src – 1.0: a preprocessor of trace gas and aerosol emission fields for regional and global atmospheric chemistry models. *Geoscientific Model Development*, 4(2):419–433.
- Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B. (2005). Fully coupled “online” chemistry within the {WRF} model. *Atmospheric Environment*, 39(37):6957 – 6975.
- Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., McKay, W. A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., and Zimmerman, P. (1995). A global model of natural volatile organic compound emissions. *Journal of Geophysical Research: Atmospheres*, 100(D5):8873–8892.
- Guenther, A., Zimmerman, P., and Wildermuth, M. (1994). Natural volatile organic compound emission rate estimates for u.s. woodland landscapes. *Atmospheric Environment*, 28(6):1197 – 1210.
- Guenther, A. B., Zimmerman, P. R., Harley, P. C., Monson, R. K., and Fall, R. (1993). Isoprene and monoterpene emission rate variability: Model evaluations and sensitivity analyses. *Journal of Geophysical Research: Atmospheres*, 98(D7):12609–12617.
- Gutiérrez, J. M. (2004). *Redes probabilísticas y neuronales en las ciencias atmosféricas*. Ministerio de Medio Ambiente, Centro de Publicaciones.
- Hadjimitsis, D. G., Clayton, C. R. I., and Hope, V. S. (2004). An assessment of the effectiveness of atmospheric correction algorithms through the remote sensing of some reservoirs. *International Journal of Remote Sensing*, 25(18):3651–3674.
- Hadjimitsis, D. G., Papadavid, G., Agapiou, A., Themistocleous, K., Hadjimitsis, M. G., Retalis, A., Michaelides, S., Chrysoulakis, N., Toullos, L., and Clayton, C. R. I. (2010). Atmospheric correction for satellite remotely sensed data intended for agricultural applications: impact on vegetation indices. *Natural Hazards and Earth System Science*, 10(1):89–95.
- Haltiner, G. J. and Williams, R. T. (1980). *Numerical prediction and dynamic meteorology*, volume 2. Wiley New York.
- Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. (2009). *The elements of statistical learning*, volume 2. Springer.
- Helfand, H. M. and Labraga, J. C. (1988). Design of a nonsingular level 2.5 second-order closure model for the prediction of atmospheric turbulence. *Journal of the Atmospheric Sciences*, 45(2):113–132.

- Holben, B., Vermote, E., Kaufman, Y., Tanre, D., and Kalb, V. (1992). Aerosol retrieval over land from avhrr data-application for atmospheric correction. *Geoscience and Remote Sensing, IEEE Transactions on*, 30(2):212–222.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.
- Kalnay, E. (2003). *Atmospheric modeling, data assimilation, and predictability*. Cambridge university press.
- Kantz, H. and Schreiber, T. (2004). *Nonlinear time series analysis*. Cambridge university press.
- Klemp, J. B., Skamarock, W. C., and Dudhia, J. (2007). Conservative split-explicit time integration methods for the compressible nonhydrostatic equations. *Monthly Weather Review*, 135(8):2897–2913.
- Kulmala, M., Laaksonen, A., and Pirjola, L. (1998). Parameterizations for sulfuric acid/water nucleation rates. *Journal of Geophysical Research: Atmospheres*, 103(D7):8301–8307.
- Laprise, R. (1992). The euler equations of motion with hydrostatic pressure as an independent variable. *Monthly Weather Review*, 120(1):197–207.
- Lavreau, J. (1991). De-hazing landsat thematic mapper images. *Photogrammetric engineering and remote sensing*, 57(10):1297–1302.
- Lewis, J., Lakshmivarahan, S., and Dhall, S. (2006). *Dynamic Data Assimilation: A Least Squares Approach*. Number v. 13 in *Dynamic data assimilation: a least squares approach*. Cambridge University Press.
- Lillesand, T. M., Kiefer, R. W., Chipman, J. W., et al. (2004). *Remote sensing and image interpretation*. Number Ed. 5. John Wiley & Sons Ltd.
- Lin, S.-J. and Rood, R. B. (1996). Multidimensional flux-form semi-lagrangian transport schemes. *Monthly Weather Review*, 124(9):2046–2070.
- Liss, P. S. and Merlivat, L. (1986). Air-sea gas exchange rates: Introduction and synthesis. In Buat-Ménard, P., editor, *The Role of Air-Sea Exchange in Geochemical Cycling*, volume 185 of *NATO ASI Series*, pages 113–127. Springer Netherlands.
- Madronich, S. (1987). Photodissociation in the atmosphere: 1. actinic flux and the effects of ground reflections and clouds. *Journal of Geophysical Research: Atmospheres*, 92(D8):9740–9752.

- Mayer, B. and Kylling, A. (2005). Technical note: The libradtran software package for radiative transfer calculations-description and examples of use. *Atmospheric Chemistry and Physics*, 5(7):1855–1877.
- McKeen, S. A., Wotawa, G., Parrish, D. D., Holloway, J. S., Buhr, M. P., Hübler, G., Fehsenfeld, F. C., and Meagher, J. F. (2002). Ozone production from canadian wildfires during june and july of 1995. *Journal of Geophysical Research: Atmospheres*, 107(D14):ACH 7–1–ACH 7–25.
- Middleton, P., Stockwell, W. R., and Carter, W. P. L. (1990). Aggregation and analysis of volatile organic compound emissions for regional modeling. *Atmospheric Environment. Part A. General Topics*, 24(5):1107 – 1133.
- Moran, M., Bryant, R., Thome, K., Ni, W., Nouvellon, Y., Gonzalez-Dugo, M., Qi, J., and Clarke, T. (2001). A refined empirical line approach for reflectance factor retrieval from landsat-5 tm and landsat-7 etm+. *Remote Sensing of Environment*, 78(1):71–82.
- Moran, M. S., Jackson, R. D., Slater, P. N., and Teillet, P. M. (1992). Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. *Remote Sensing of Environment*, 41(2):169–184.
- Nichols, N. (2010). Mathematical concepts of data assimilation. In Lahoz, W., Khattatov, B., and Menard, R., editors, *Data Assimilation*, pages 13–39. Springer Berlin Heidelberg.
- Olivier, J. G. J., Bouwman, A., Berdowski, J., Veldt, C., Bloos, J., Visschedijk, A., Zandveld, P., Haverlag, J., et al. (1996). Description of edgar version 2.0: A set of global emission inventories of greenhouse gases and ozone-depleting substances for all anthropogenic and most natural sources on a per country basis and on 1 degree x 1 degree grid.
- Ouaidrari, H. and Vermote, E. F. (1999). Operational atmospheric correction of landsat tm data. *Remote Sensing of Environment*, 70(1):4–15.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Potter, J. (1984). The channel correlation method for estimating aerosol levels from multispectral scanner data: the errors in this method are small enough to make it useful for such applications as monitoring smoke plumes. *Photogrammetric engineering and remote sensing*, 50(1):43–52.
- Potter, J. and Mendlowitz, M. (1975). On the determination of haze levels from landsat data. *10th International Symposium on Remote Sensing of Environment, Ann Arbor, MI, Environmental Research Institute of Michigan*.

- Richter, R. (1996a). Atmospheric correction of satellite data with haze removal including a haze/clear transition region. *Comput. Geosci.*, 22(6):675–681.
- Richter, R. (1996b). A spatially adaptive fast atmospheric correction algorithm. *International Journal of Remote Sensing*, 17(6):1201–1214.
- Rozanov, A., Rozanov, V., Buchwitz, M., Kokhanovsky, A., and Burrows, J. P. (2005). Sciatran 2.0—a new radiative transfer model for geophysical applications in the 175–2400nm spectral region. *Advances in Space Research*, 36(5):1015–1019.
- Salama, M. S., Monbaliu, J., and Coppin, P. (2004). Atmospheric correction of advanced very high resolution radiometer imagery. *International Journal of Remote Sensing*, 25(7-8):1349–1355.
- Sandnes, H. and Styve, H. (1993). *Calculated Budgets for Airborne Acidifying Components in Europe, 1985, 1987, 1988, 1989, 1990 and 1991+ Corrigendum to Appendix C*. Norwegian Meteorological Institute. Meteorological Synthesizing Centre-West.
- Saxena, P., Hudischewskyj, A. B., Seigneur, C., and Seinfeld, J. H. (1986). A comparative study of equilibrium approaches to the chemical characterization of secondary aerosols. *Atmospheric Environment (1967)*, 20(7):1471 – 1483.
- Schell, B., Ackermann, I. J., Hass, H., Binkowski, F. S., and Ebel, A. (2001). Modeling the formation of secondary organic aerosol within a comprehensive air quality model system. *Journal of Geophysical Research: Atmospheres*, 106(D22):28275–28293.
- Schowengerdt, R. A. (2006). *Remote sensing: models and methods for image processing*. Academic press.
- Schubert, S. D., Rood, R. B., and Pfaendtner, J. (1993). An assimilated dataset for earth science applications. *Bulletin of the American Meteorological Society*, 74(12):2331–2342.
- Schultz, M., Rast, S., van het Bolscher, M., Pulles, T., Brand, R., Pereira, J., Mota, B., Spessa, A., Dalsoren, S., van Noije, T., and Szopa, S. (2007). Emission data sets and methodologies for estimating emissions, retro project report d1-6, hamburg.
- Simkin, T. and Siebert, L. (1994). Volcanoes of the world: A regional directory, gazetteer, and chronology of volcanism during the last 10,000 years, 349 pp. *Geoscience, Tucson, Ariz.*
- Simpson, D., Guenther, A., Hewitt, C. N., and Steinbrecher, R. (1995). Biogenic emissions in europe: 1. estimates and uncertainties. *Journal of Geophysical Research: Atmospheres*, 100(D11):22875–22890.

- Singh, S. M. and Cracknell, A. P. (1986). The estimation of atmospheric effects for spot using avhrr channel-1 data. *International Journal of Remote Sensing*, 7(3):361–377.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., and Barker, D. M. (2005). Coauthors, 2008: A description of the advanced research wrf version 3. ncar tech. Technical report, Note NCAR/TN-4751STR.
- Slinn, S. and Slinn, W. (1980). Predictions for particle deposition on natural waters. *Atmospheric Environment (1967)*, 14(9):1013 – 1016.
- Spurr, R. (2008). Lidort and vlidort: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. In Kokhanovsky, A., editor, *Light Scattering Reviews 3*, Springer Praxis Books, pages 229–275. Springer Berlin Heidelberg.
- Stockwell, W. R., Middleton, P., Chang, J. S., and Tang, X. (1990). The second generation regional acid deposition model chemical mechanism for regional air quality modeling. *Journal of Geophysical Research: Atmospheres*, 95(D10):16343–16367.
- Sundqvist, H., Berge, E., and Kristjánsson, J. E. (1989). Condensation and cloud parameterization studies with a mesoscale numerical weather prediction model. *Monthly Weather Review*, 117(8):1641–1657.
- Switzer, P., Kowalik, W., and Lyon, R. (1981). Estimation of atmospheric path-radiance by the covariance matrix method. *Photogrammetric Engineering and Remote Sensing*, 47:1469–1476.
- Teillet, P. and Fedosejevs, G. (1995). On the dark target approach to atmospheric correction of remotely sensed data. *Canadian Journal of Remote Sensing*, 21(4):374–387.
- Tourassi, G. D., Frederick, E. D., Markey, M. K., and Floyd Jr, C. E. (2001). Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical Physics*, 28(12):2394–2402.
- Van Aardenne, J. A., Dentener, F., Olivier, J. G. J., Peters, J. A. H. W., and Ganzeveld, L. N. (2005). The edgar 3.2 fast track 2000 dataset (32ft2000). *Emission Database for Global Atmospheric Research: Description of*, 32.
- Vermote, E., Tanre, D., Deuze, J. L., Herman, M., and Morcette, J. J. (1997a). Second simulation of the satellite signal in the solar spectrum, 6S: an overview. *Geoscience and Remote Sensing, IEEE Transactions on*, 35(3):675–686.
- Vermote, E., Tanre, D., Deuze, J. L., Herman, M., and Morcette, J. J. (1997b). 6S user guide version 2. *Lille: Laboratoire d’Optique Atmosphérique*.

- Walcek, C. J., Brost, R. A., Chang, J. S., and Wesely, M. L. (1986). SO₂ sulfate and HNO₃ deposition velocities computed using regional landuse and meteorological data. *Atmospheric Environment (1967)*, 20(5):949–964.
- Welch, G. and Bishop, G. (1995). An introduction to the kalman filter. Technical report, Chapel Hill, NC, USA.
- Wesely, M. (1989). Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. *Atmospheric Environment (1967)*, 23(6):1293 – 1304.
- Wesely, M. and Hicks, B. (1977). Some factors that affect the deposition rates of sulfur dioxide and similar gases on vegetation. *Journal of the Air Pollution Control Association*, 27(11):1110–1116.
- Wicker, L. J. and Skamarock, W. C. (2002). Time-splitting methods for elastic models using forward time schemes. *Monthly Weather Review*, 130(8):2088–2097.
- Wrigley, R. C., Spanner, M. A., Slye, R. E., Pueschel, R. F., and Aggarwal, H. R. (1992). Atmospheric correction of remotely sensed image data by a simplified model. *Journal of Geophysical Research: Atmospheres*, 97(D17):18797–18814.
- Xu, K.-M. and Krueger, S. K. (1991). Evaluation of cloudiness parameterizations using a cumulus ensemble model. *Monthly Weather Review*, 119(2):342–367.
- Yang, X. (2011). *Urban remote sensing: Monitoring, synthesis and modeling in the urban environment*. John Wiley & Sons.