

MÉTODOS PREDICTIVOS DE PRECIPITACIÓN UTILIZANDO DATOS DE LA CUBIERTA VEGETAL MEDIANTE TELEDETECCIÓN

A. Moreno (*), E. Soria (**), J. García (*), J.D. Martín (**) y F. Belda (***)

(*) *Unidad de Investigación de Teledetección, Facultad de Física, Universitat de Valencia.*

C/ Dr. Moliner 50, 46100. Burjassot, Valencia, Spain.

(**) *Departamento de Ingeniería Electrónica, Universidad de Valencia.*

C/ Dr. Moliner, 50. 46100 Burjassot, Valencia, Spain.

(***) *Delegación Territorial en Murcia. AEMET (Agencia Estatal de Meteorología).*

RESUMEN

Este estudio se enmarca en el proyecto DULCINEA, que tiene entre sus objetivos el análisis del cambio climático y su incidencia en la vegetación de la Península Ibérica. El principal objetivo consiste en evaluar el potencial predictivo de variables que recogen la evolución de la cubierta vegetal sobre la precipitación, empleando para ello modelos temporales lineales y no lineales. Como variables adicionales se consideran la temperatura y la precipitación correspondientes a los periodos previos. Para realizar este estudio se dispone de series largas de las variables climáticas así como datos de NDVI correspondientes a un periodo de 22 años (1981- 2003). La utilización de técnicas de predicción no lineales, en concreto redes neuronales artificiales, ha mostrado ser una herramienta matemática predictiva adecuada para este problema, debido a la existencia de fenómenos marcadamente no lineales. Estos modelos permiten capturar, a un mes vista, la dinámica general de las precipitaciones, con errores medios en la predicción (RMSE) del orden de 44 mm para una serie de validación de 4 años. El análisis de sensibilidad de los diferentes modelos demuestra que las observaciones del estado de la cubierta en los meses anteriores tienen un poder predictivo superior que precipitación y temperatura.

ABSTRACT

This study has been focused on the development of rainfall prediction using data obtained by Remote Sensing techniques in the context of the DULCINEA Project. The first approach for prediction models was based on classical Auto-Regressive Moving-Average (ARMA) models, being the temperature and the NVDI external inputs. Non-linear prediction models, based on an artificial neural network with same input variables, were also used. The use of these models may be justifiable by the fact that the nature of the environmental phenomena is non-linear. The use of the neural network also allowed carrying out a sensitivity analysis of the inputs, thus rating the relevance of the different inputs in the prediction model. The obtained models have captured the overall dynamic of the annual rainfall at a number of representative sites. Results suggest that temporal signatures of satellite based vegetation observations may be valuable explanatory variables for rainfall forecasting systems.

Palabras clave: teledetección, vegetación, predicción, precipitación, redes neuronales, métodos lineales.

INTRODUCCIÓN

La precipitación se caracteriza por ser un fenómeno extremadamente irregular en el tiempo y en el espacio y esto provoca numerosas dificultades para su estudio y análisis. Las variables que afectan a la precipitación son numerosas, siendo las más importantes la orografía del terreno (en el caso de la Península Ibérica complicada y compartimentada), temperatura superficial, temperatura del mar, humedad del suelo, vegetación, viento y presión.

La vegetación lejos de ser una componente pasiva del sistema climático, es un elemento clave en la forma en que la energía, agua, momento y flujos de CO₂ de la superficie terrestre se acoplan con la atmósfera. Su influencia será especialmente

importante en zonas en las que la circulación atmosférica sea pequeña (Adams 2007). Los cambios en la vegetación no sólo tienen un alcance local sino que pueden afectar a zonas vecinas, el efecto local mas inmediato es la modificación de la temperatura, ya que la cubierta vegetal actúa directamente sobre el albedo. La radiación absorbida aumenta la temperatura de las hojas y calienta el aire próximo a ellas, este incremento es menor del correspondiente a la energía absorbida debido a que gran parte de esa energía se transforma en calor latente. Esta conversión de energía por evaporación tiene lugar en los estomas de las hojas (transpiración) y su efecto inmediato de reducir las temperaturas.(Schwartz y Karl 1990, Bounoua 2000). Los cambios de humedad en el aire tienen un radio de acción mayor debido a que la humedad se

transporta a mayores distancias. En función de qué fenómeno físico predomine para el tipo de cubierta vegetal (absorción de radiación o transpiración) tendremos un aumento de temperatura o enfriamiento a nivel local. Por último, destacar que la cubierta vegetal aumenta la rugosidad del suelo incrementando la transferencia de momento y las turbulencias de las masas de aire.

En este trabajo se pretende evaluar el potencial predictivo de variables que recogen la evolución de la cubierta vegetal (índices de vegetación o variables biofísicas de la vegetación) sobre la precipitación. Estudios realizados sobre praderas de Norteamérica han constatado que anomalías positivas en la estación de crecimiento de la cubierta provocan una disminución de la precipitación y un aumento de la temperatura en verano (Wang 2006). La vegetación se acopla con el régimen de precipitación a diferentes escalas temporales, puede comportarse como una memoria biofísica que actúa sobre la precipitación a través de retro-alimentaciones retardadas (Alessandri 2008), este efecto será más relevante en las áreas que presentan mayor sensibilidad a las variables climáticas. Dada la complejidad del sistema a modelar, y al hecho de que el acoplamiento entre variables puede ser fuertemente no lineal, se recurrirá a métodos tanto lineales como no lineales (redes neuronales).

DATOS Y METODOLOGÍA

Datos

La variable usada para caracterizar la biomasa y el vigor de la vegetación desde satélite es el NDVI (*Normalized Difference Vegetation Index*). Para el estudio, hemos empleado el producto GIMMS (*Global Inventory Modelling and Mapping Studies*), que se deriva del sensor AVHRR (*Advanced Very High Resolution Radiometer*) de la serie de satélites NOAA 7,9,11,14,12 y17. La serie está corregida de las diferencias de calibrado entre satélites, de aerosoles volcánicos y otros efectos no relacionados a cambios en la vegetación. Otra ventaja fundamental es su duración, que cubre un periodo de 22 años (1981-2003) con una resolución temporal de 15 días. Los datos climáticos (precipitación y temperatura) empleados en este estudio para caracterizar la precipitación en la Península Ibérica son los registrados por todas las estaciones pluviométricas de la Agencia Estatal de Meteorología (AEMet) durante el periodo 1950-2008. Con el fin de tener un control del tipo de cubierta y de su heterogeneidad espacial, se ha utilizado la clasificación de coberturas CORINE

Land Cover 2000. En particular, las áreas próximas a zonas urbanas y zonas cultivo irrigadas han sido excluidas.

Modelos

Dentro de las opciones de elección de modelos lineales empleados para el modelado de series temporales hemos elegido los de tipo ARMAX (*Autoregressive Moving Average with Exogenous Inputs*) (Nelles 2001), estos modelos incluyen la historia de la variable a modelar (términos autorregresivos AR), otras variables explicativas (términos X) y un modelado del ruido (términos MA) cuyo objetivo es tratar de describir la varianza que las variables consideradas no han conseguido reproducir.

$$y(t) + a_1 y(t-1) + \dots + a_{n_a} y(t-n_a) = b_1 u(t-n_k) + \dots + b_{n_b} u(t-n_k-n_b+1) + c_1 e(t-1) + \dots + c_{n_c} e(t-n_c) + e(t) \quad (1)$$

donde $y(t)$ es la salida deseada del modelo en un tiempo t , n_a es el número de retardos de la variable y , n_b es el número de retardos de cada variable de entrada u (en nuestro caso, existen dos variables, NDVI y T), n_c es el número de retardos empleados para el modelado del ruido y n_k es el retardo que representa el número de observaciones que ocurren antes de que las variables de entrada afecten a la variable de salida a predecir.

En el estudio se propone el uso de redes neuronales artificiales (RNA), que son modelos especialmente apropiados para este estudio: (i) permiten capturar las complejas relaciones entre precipitación, temperatura y NDVI, (ii) se construyen a partir de los datos, sin presuponer ningún tipo de relación *a priori* entre variables, y (iii) pueden aplicarse a conjuntos de datos relativamente grandes, ya que la carga computacional es constante una vez ha sido entrenada. Se ha utilizado un perceptrón multicapa que es la red más usada en la actualidad dado su amplio abanico de aplicaciones (Haykin 1999). Cada neurona (unidad elemental de proceso de la red) efectúa una transformación no lineal de sus entradas de acuerdo a la siguiente expresión:

$$y = \phi \left(\sum_{i=1}^m w_i \cdot x_i + b \right) \quad (2)$$

donde m es el número de entradas a la neurona, w_i son los pesos sinápticos (parámetros del sistema), b es el coeficiente de sesgo (otro parámetro a determinar) y ϕ es una función no lineal (en este trabajo se emplea la más utilizada, la tangente

hiperbólica). Estos elementos están organizados en capas (Figura 1).

Criterio de selección de modelos

Se desarrollaron sobre 4000 modelos lineales con diferente número de parámetros y retardos para cada uno de los píxeles con el fin de minimizar el error de predicción y el número de parámetros. Esta selección se llevó a cabo en dos pasos. Primero se seleccionaron los 200 mejores modelos en base al criterio de información de Akaike (AIC), este método permite maximizar el coeficiente de determinación (R^2) penalizando al modelo según el número de variables incluidas, finalmente se seleccionó el modelo con menor error cuadrático medio (MSE).

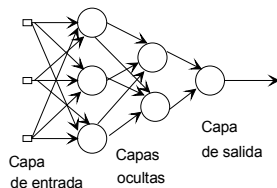


Figura 1.- Esquema de un perceptrón multicapa.

Para el caso de los modelos no lineales (RNA), se entrenaron alrededor de 6000 redes para cada píxel probando diferentes arquitecturas, realizando diferentes inicializaciones de los pesos sinápticos (debido a que el algoritmo de obtención de parámetros es un algoritmo de búsqueda local de mínimos). Para ambos tipos de modelos se reservaron los últimos 4 años de la serie (1999-2003) como conjunto de datos de validación sobre el que se estudia el valor del RMSE cometido.

Análisis de sensibilidad de variables

Con el fin de simplificar los modelos obtenidos mediante las redes neuronales se realizó un análisis de sensibilidad. Mediante este proceso se jerarquiza la influencia de las variables para desechar aquellas variables que no mejoran sustancialmente nuestro modelo predictivo. El análisis se basa en estudiar las diferencias que sufre el modelo al incluir o no una determinada variable atendiendo al error que se introduce al eliminarla.

RESULTADOS

En el desarrollo de este trabajo se ha hecho un muestreo de los principales tipos de cubierta y condiciones climáticas. Las zonas irrigadas y antropizadas fueron excluidas. Seguidamente se

muestran los resultados obtenidos para 4 zonas especialmente representativas (tabla 1).

Tabla 1.- Características de algunas de las zonas estudiadas.

Zona	Latitud	Longitud	Cubierta
1	43°52'30"N	8°57'51"W	Bosque frondoso
2	42°41'28"N	0°0'36"E	Herbácea montaña
3	43°14'59"N	9°35'21"W	Bosque frondoso
4	36°52'22"N	2°23'24"W	Matorral disperso

Los valores de error obtenidos para los conjuntos de datos de validación se muestran a continuación (tabla 2).

Tabla 2.- Características de la predicción (un mes vista) obtenida por los modelos lineales (ARMAX) y no lineales (RNA).

Zona	Num. parámetros de entrada.		RMSE(precip) (validación)(mm)	
	ARMAX	RNA	ARMAX	RNA
1	34	8	±55.4	±43.9
2	36	14	±91.7	±91.0
3	37	11	±99.0	±92.2
4	38	10	±18.9	±16.4

En las Figuras 2 y 3 se muestran a modo de ejemplo las series de precipitación predichas para el área de estudio correspondiente a la zona 1. A la vista de estos resultados se aprecia cómo el modelo no lineal es capaz de capturar en gran medida la dinámica de las precipitaciones a partir de la evolución de la vegetación, mientras que el modelo lineal ofrece peores ajustes.

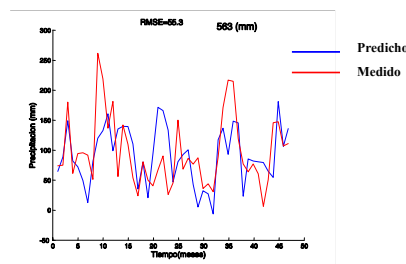


Figura 2.- Predicción a un mes vista para el modelo ARMAX ajustado en la zona 1.

En la Figura 4 se muestran los resultados de un análisis de sensibilidad. El bloque de variables correspondientes al NDVI (con sus 7 retardos) es el que más afecta a la precisión de la predicción del

modelo, en concreto NDVI (t-5) es la variable que más contribuye al modelado de la señal. Este resultado sugiere que la relación entre la vegetación y la precipitación se prolonga durante un periodo de mínimo de 7 meses. Este tipo de efectos retardados relacionados con el acoplamiento entre la vegetación y precipitación ha sido observado por otros autores (Wang 2006). Excepto para la zona 2, que corresponde a una superficie rocosa con vegetación de alta montaña muy dispersa, el NDVI muestra un peso relativo mucho mayor que el resto de variables.

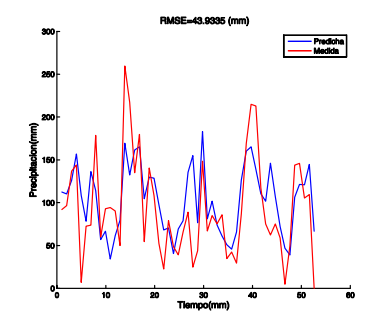


Figura 3.- Predicción a un mes vista para el modelo de redes neuronales ajustado en la zona 1.

CONCLUSIONES

Los resultados obtenidos muestran que un índice de vegetación (p.ej. NDVI) es una variable de gran interés para la modelización de la precipitación tanto a partir de modelo lineales como no lineales. En los modelos lineales la modelización del error tiene un peso demasiado elevado indicando la naturaleza no lineal de los modelos a considerar. Los modelos no lineales consiguen obtener mejores resultados y permiten cuantificar la influencia específica de las variables de entrada a través de un análisis de sensibilidad.

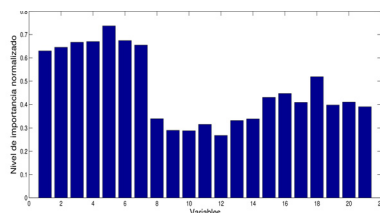


Figura 4.- Importancia normalizada de las distintas variables introducidas para la zona 1. Las variables 1-7 son retardos de NDVI, 8-14 son retardos T y 15-21 son retardos de la propia precipitación.

En general, la evolución del NDVI posee información clave para caracterizar la precipitación en periodos posteriores. El carácter de memoria biofísica de la cubierta vegetal explicaría esta característica ya que la vegetación acapara gran parte de la importancia en los modelos dada su alta sensibilidad a los factores climáticos. Esa facilidad para retener rasgos climáticos en su signature temporal le otorga un gran potencial explicativo en la modelización de la precipitación.

BIBLIOGRAFÍA

Adams, J. 2007, *Vegetation-Climate Interaction*, Springer-Praxis books in environmental sciences.

Alessandri, A., y A. Navarra 2008, On the coupling between vegetation and rainfall inter-annual anomalies: Possible contributions to seasonal rainfall predictability over land areas, *Geophys. Res. Lett.*, 35, L02718, doi:10.1029/2007GL032415.

Bishop, C.M. 1996, *Neural Networks for Pattern Recognition*, Clarendon Press.

Bounoua, L., G.J. Collatz, S.O. Los, et al., 2000: Sensitivity of climate to changes in NDVI. *J. Climate*, 13, 2277-2292.

Haykin, S. 1999. *Neural Networks: A Comprehensive Foundation*, Prentice Hall.

Nelles, O. 2001, *Nonlinear System Identification*, Springer-Verlag Berlin Heidelberg, New York.

Parker, D.E y Folland, C.K. 1988, The nature of climatic variability, *Meteorological Magazine*, 117, 201-210.

Schwartz, M.D. y T.R. Karl, 1990 Spring phenology: Nature's experiment to detect the effect of 'green up' on surface maximum temperatures. *Mon. Wea. Rev.*, 118, 883-890.

Wang, W. 2006. *Detecting and modeling large-scale interactions between vegetation, precipitation, and temperature over temperate-semiarid and boreal climate regimes*. Ph.D. dissertation, Boston University, United States -- Massachusetts.

AGRADECIMIENTOS

Este trabajo se enmarca en el proyecto DULCINEA (CGL2005-04202) y ha contado además con el soporte del proyecto ÁRTEMIS (CGL2008-00381). Los datos climáticos han sido proporcionados por AEMet.