

Adaptación y Desarrollo de un Modelo Estadístico de Predicción Estacional

Antonio Ángel Serrano de la Torre

*Departamento de Física Aplicada
Facultad de Ciencias de la Universidad de Granada*

Abstract. Se han buscado en la literatura científica, potenciales predictores para las variables temperatura y precipitación acumulada, en las estaciones de invierno y verano, y se ha desarrollado un modelo estadístico de predicción estacional basado en Regresión Lineal Múltiple que utiliza estos predictores. Además, se ha llevado a cabo una validación del modelo. Por último se han propuesto mejoras al mismo.

Keywords: Predicción estacional Modelo estadístico de predicción Regresión lineal múltiple.

INTRODUCCIÓN

Actualmente se encuentran bastante desarrollados los modelos de predicción a corto plazo (hasta 5 días) y las proyecciones climáticas para el estudio del cambio climático (IPCC, 2014). Entre estas dos escalas temporales, se encuentra la predicción estacional, que está empezando a desarrollarse, y de la que se derivarían evidentes beneficios para la sociedad.

Existen dos grandes tipos de modelos numéricos de predicción del clima: dinámicos y estadísticos. En general, los modelos dinámicos dan mejores resultados que los estadísticos, si se encuentran suficientemente desarrollados. Al ser la predicción estacional un campo de incipiente desarrollo, los modelos dinámicos estacionales no están suficientemente desarrollados, por lo que los modelos estadísticos están dando, de momento, mejores resultados que aquellos. Además, requieren de menos recursos para su desarrollo.

Por todo esto, un modelo estadístico es un excelente punto de partida para el estudio de la predicción estacional.

Dentro de los modelos estadísticos hay una gran variedad, de entre la que hemos escogido el más simple de ellos: la Regresión Lineal Múltiple (de ahora en adelante, RLM).

En toda regresión lineal hay que distinguir las variables que se quieren predecir, a las que se llama *predictandos*, y las variables con las que se pretende predecirlas, a las que se llama *predictores*. La RLM establece relaciones lineales entre las primeras y las segundas.

Las variables que se quieren predecir en el presente estudio, o sea los predictandos, son: *temperatura invernal*, *precipitación acumulada invernal*, *temperatura veraniega*

y *precipitación acumulada invernal*. Y los predictores utilizados son índices de teleconexión que, según se ha encontrado en la literatura, influyen, con cierto retraso en el tiempo, en los predictandos.

OBJETIVO

El presente trabajo tiene un doble objetivo: por un lado, la búsqueda en la literatura de potenciales predictores para las variables *temperatura* y *precipitación acumulada* en el área mediterránea en las estaciones de invierno y verano; y por otro lado, el desarrollo y validación de un modelo estadístico de predicción estacional aplicado a esta área geográfica.

DATOS

Predictandos

Son las variables que queremos predecir: temperatura media y precipitación acumulada, en las estaciones de invierno (DJF) y verano (JJA), en al área mediterránea [15.0N, 52.0N, -20.0E, 60.0E]. El invierno de un año, comienza en diciembre del año anterior.

Las podemos obtener de diversas instituciones que las ponen a disposición de la comunidad científica. Concretamente, las bases de datos utilizadas como predictandos son, para el área mediterránea:

- *Universidad de Delaware (Udel)*. Son datos observacionales de *temperatura* y *precipitación* procedentes de estaciones en tierra que se han interpolado a rejilla regular de 0,5x0,5°. Su calidad es pobre en el norte de África debido a la escasez de estaciones en esas tierras.
- *Reanálisis del NCAR*. Resolución de 2,5°x2,5°. La precipitación es un campo derivado y, por tanto, puede que poco realista.
- *GPCP precipitation*. Combina datos de observatorios terrestres con datos de satélite. Resolución de 2,5°x2,5°.
- *CMAP precipitation (standard)*. Datos de satélite. Resolución de 2,5°x2,5°.
- *CMAP precipitation (enhanced)*. Combina datos de satélite con el reanálisis del NCAR. Resolución de 2,5°x2,5°.

También, pero cubriendo sólo porciones del área mediterránea (para comparar los resultados en las zonas incluídas):

- *E-Obs*. Sólo abarcan Europa. Son datos observacionales. Resolución de 0,25°x0,25°. Contiene datos diarios que hemos convertido a mensuales.
- *Spain02* versión 2. Sólo abarcan a la península Ibérica y Baleares. Son datos observacionales. Resolución de 0,2°x0,2°.

Predictores

Se han encontrado algunos predictores con retardo de las variables a predecir, y también se han encontrado teleconexiones simultáneas de los predictandos con ciertos índices climáticos. En algunos de estos casos, se han encontrado además, predictores para dichos índices de teleconexión simultánea, por lo que sirven de predictores de las variables a predecir. El resultado de toda esta búsqueda se presenta en la siguiente tabla:

TABLA . Predictores encontrados en la literatura

Predictor	Predictando
AMO veraniego	Temperatura invernal Lluvia invernal
AMO otoñal	Temperatura invernal
AMO primaveral	Temperatura veraniega Lluvia veraniega
Tropical Pacific SST EA correlated otoñal	Temperatura invernal Lluvia invernal
Indian summer monsoon index (IM index)	Lluvia invernal
1st EOF SST Mediterráneo primaveral	Temperatura veraniega
2nd EOF SST Mediterráneo primaveral	Temperatura veraniega
NAO primaveral	Temperatura veraniega
TNA primaveral	Temperatura veraniega
SV-NAM invernal	Temperatura veraniega
EA primaveral	Temperatura veraniega
PNA primaveral	Temperatura veraniega
SCAND invernal	Temperatura veraniega

La procedencia de los predictores es la que se indica a continuación:

TABLA . Fuentes de los predictores

Predictor	Fuente
AMO	NOAA/ESRL/PSD1
Tropical Pacific SST EA correlated	Calculados para este trabajo a partir del producto ERSST de la NOAA
Indian summer monsoon index (IM index)	Monsoon Monitoring Page
1st & 2nd EOF SST Mediterráneo	Calculados para este trabajo a partir del producto ERSST de la NOAA
NAO	NOAA/NCEP/CPC
TNA	Calculados para este trabajo a partir del producto ERSST de la NOAA
SV-NAM	Mie University
EA	NOAA/NCEP/CPC
PNA	NOAA/NCEP/CPC
SCAND	NOAA/NCEP/CPC

METODOLOGÍA

El modelo lleva a cabo una regresión lineal múltiple en cada punto de malla del predictando, utilizando como predictores, el listado dado en la Sección anterior. Este listado es el mismo para todos los puntos de grid del predictando.

A efectos de este modelo, el invierno de un año, está formado por diciembre del año anterior, y por enero y febrero del año. Verano está formado por los meses de junio, julio y agosto. El periodo de calibración o regresión es aquel del que se toman los datos (predictores y predictandos) para hacer la regresión. Periodo de validación es el periodo de tiempo, disjunto al de regresión, en el que se valida el modelo. El mes de lanzamiento es el mes en el que se hace correr el modelo, y es un concepto importante porque el modelo asume que no disponemos de valores de los predictores en dicho mes; este mes siempre es el mes justo anterior a la estación a predecir.

Para que la RLM funcione bien, se deben cumplir ciertas premisas, que son, principalmente, la no multicolinealidad de los predictores, no sobreajustar el modelo, y requerimientos sobre los residuos, que a su vez son: deben ser independientes de los predictores e independientes entre sí, o sea, no autocorrelacionados; su valor medio debe ser igual a cero, deben ser homocedásticos y normalmente distribuidos. Algunas de estas premisas se comprueban mediante tests de hipótesis y otras son tenidas en cuenta a la hora de preparar los datos.

Tanto a los predictores como a los predictandos, se les elimina la tendencia lineal antes de llevar a cabo la regresión. Una vez hallada la predicción, se le vuelve a añadir la tendencia. En el caso de tener un predictando transformado (la lluvia, por la raíz cuadrada), la tendencia se elimina después de hacer la transformación, y se vuelve a añadir al resultado de la predicción antes de quitarle la transformación.

Para validar el modelo, se halla una la predicción de cada uno de los cuatro predictandos, en cada uno de los años del periodo de validación, y se halla el coeficiente de correlación de Pearson entre estos resultados y los valores realmente observados. Después se halla la Raíz del Error Cuadrático Medio (RMSE, del inglés Root of Mean Square Error). El coeficiente de correlación nos dice si el modelo *acierta*, y el RMSE nos dice cuánto *acierta*, de modo que un buen valor del RMSE no significa buena capacidad predictiva si no viene acompañado de un buen valor del coeficiente de correlación.

A continuación se presenta una tabla con los periodos de tiempo usados para calibración y validación:

TABLA . Periodos de tiempo utilizados

	Abarcado por los datos	Calibración	Validación
NCAR	Ene 1949 - Actualidad	1951 -1990 (40 años)*	1991 - 2010 (20 años)
Udel	Ene 1901 - Dic 2010	1951 -1990 (40 años)*	1991 - 2010 (20 años)
CMAP	Ene 1979 - Abr 2015	1980 -2001 (22 años)	2002 - 2011 (10 años)
GPCP	Ene 1979 - Abr 2015	1980 -2001 (22 años)	2002 - 2011 (10 años)
E-Obs	Ene 1950 - Dic 2013	1951 -1990 (40 años)*	1991 - 2010 (20 años)
Spain02 v2	Ene 1950 - Mar 2008	1951 -1990 (40 años)*	1991 - 2005 (15 años)

*Para temperatura veraniega, este periodo de calibración cambia a 1959 - 1990 (32 años)

RESULTADO DE LA VALIDACIÓN

En la siguiente tabla se presentan los resultados de la validación del modelo:

TABLA . Tabla de resultados de la validación.

Variable predicha	Coef. de determinación	Coef. de corr. validación	RMSE validación
Temperatura invernal	< 30%	No significativo	< 1.5 °C
Precipitación acumulada de invierno	< 20%	No significativo	> 80 mm/m ²
Temperatura veraniega	< 30%	Significativo en el este	< 1 °C
Precipitación acumulada de verano	< 10%	No significativo	> 80 mm/m ²

Encuanto al número de condición, indica que puede llegar a perderse el 10% de la precisión de los datos.

DISCUSIÓN Y CONCLUSIONES

Podemos identificar varias causas para estos malos resultados. Primero está la limitación del modelo de usar los mismos predictores en todos los puntos de malla, lo que introduce ruido en la solución. Por otro lado, está la posible colinealidad de los predictores, por la cual se ha estimado que este modelo puede llegar a perder un 10% de precisión. Además, es necesaria la búsqueda de nuevos predictores.

Por ello, se propone mejorar el modelo haciendo que éste busque los mejores predictores en cada punto de malla. Un método como la *introducción progresiva* eliminaría, además, la colinealidad. También se propone usar la validación por el método de *dejar uno fuera*, que permitiría ampliar el periodo de calibración y también usar otros predictores de los que, al estar basados en medidas satelitales, se dispone de una corta serie de tiempo.

AGRADECIMIENTOS

Agradezco a los respectivos proveedores de los datos.

REFERENCIAS

1. IPCC, 2014. "Climate Change 2014 – Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects." Vol. 1. *Cambridge University Press, cambridge Books Online*.
URL <http://dx.doi.org/10.1017/CBO9781107415379>