MODELIZACIÓN NUMÉRICO-ESTOCÁSTICA APLICADA A LA GENERACIÓN DE SERIES DE PRECIPITACIONES DIARIAS

Eusebi JARAUTA. Dept. Matemática Aplicada III - ETSECCPB (UPC).
C/. Gran Capitán, s/n; edificio C-2. 08034-Barcelona.

M. Ángeles COLOMER. Dept. Matemática - ETSEAL (UdL).
Av. A. Rovira Roure, 177. 25006-Lleida.

RESUMEN
Se presenta en este trabajo la formulación básica de un modelo estocástico para generar series de datos de precipitación diaria en un punto cualquiera de una región, a partir de datos pluviométricos del entorno. El objetivo perseguido es la estimación del régimen de humedad de los suelos en puntos donde no hay información pluviométrica. Se utilizan cadenas de Markov de primer orden y las distribuciones lognormal y gamma.

ABSTRACT
We present a basic stochastic model for series of daily rainfall generation, in any point of a region, by knowing rainfall information of neighbour points. This is made for soil moisture regime estimation in points without rainfall data. We use first order Markov chains and lognormal and gamma distributions.

1. INTRODUCCIÓN

El régimen de humedad de un suelo es una característica del mismo que describe la evolución a lo largo del tiempo de su contenido de agua disponible para las plantas. El conocimiento del régimen de humedad de los suelos de una comarca es muy importante tanto desde el punto de vista de la genética y la evaluación de suelos, como para conocer las disponibilidades hídricas de los cultivos, la planificación de riegos y la evaluación de los recursos hídricos de dicha comarca ([Ja2]). Su definición precisa se debe al United States Department of Agriculture (USDA), en su obra "Soil Taxonomy" [So]. En virtud de su propia definición, el régimen de humedad de un suelo no puede conocerse más que por medidas directas en campo del contenido de humedad (perfiles hídricos) a lo largo del tiempo en puntos diversos; a falta de tales medidas debe recurrirse a su estimación mediante un modelo matemático de simulación del balance hídrico del suelo. Hasta la fecha se han desarrollado dos modelos para la estimación del régimen de humedad: el de F. Newhall [Ne] y el de E. Jarauta [Ja1]; ambos utilizan distribuciones deterministas de la precipitación y la temperatura. Dada la naturaleza aleatoria de la precipitación, parece razonable plantear una modificación de la simulación de la entrada del agua en el suelo utilizando un modelo estocástico. Cabe decir que el modelo se ha elaborado tomando como campo experimental la provincia de Lleida (noroeste de Catalunya).

2. DESCRIPCIÓN DEL MODELO

Se modeliza la precipitación mediante una variable aleatoria estocástica tridimensional:

$$P_n(t) = \Big(X_n(t), Y_n(t), L_n(t) \Big)$$

siendo $X_n(t)$ una variable aleatoria discreta que representa la ocurrencia o no de precipitación, $Y_n(t)$ una variable aleatoria discreta que representa el intervalo en que se halla el valor de la precipitación y,

por último, $L_n(t)$ es una variable aleatoria contínua que representa la cantidad de precipitación en el dia correspondiente. Vamos a describir a continuación la metodologia utilizada para cada una de estas variables.

2.1. Ocurrencia de precipitación

Diversos antecedentes estudiados ([Ro1], [Ro2], [Wo], [Al]) se referían mayoritariamente a zonas con una precipitación anual media bastante superior a la correspondiente a la zona estudiada, que se sitúa en torno a los 300 mm por año; en consecuencia, parece razonable cuestionar si en tal caso se mantiene o no el caracter markoviano y, en todo caso, comprobarlo. El proceso tiene únicamente dos estados posibles $\{E_0, E_1\}$, siendo E_0 = "no hay precipitación" y E_1 = "hay precipitación" ; se realizó un test χ^2 (véase [An]) para estudiar la dependencia de la ocurrencia de precipitación con la existencia de la misma en días anteriores, tomando como hipótesis nula la distribución binomial y como hipótesis alternativa la cadena de Markov de primer orden con dos estados; el estadístico utilizado ha sido

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} n_i \frac{(\widehat{p}_{ij} - p_{ij})^2}{p_{ij}},$$

que sigue una distribución χ^2 con 2 grados de libertad; p_{ij} es el valor esperado de la probabilidad de transición del estado E_i al estado E_j estimado por el método de máxima verosimilitud aceptando la distribución multinomial de la población; $\widehat{p}_{ij}$ es el valor de la misma probabilidad estimado por el mismo método aceptando que la variable sigue un proceso de Markov de primer orden y se calcula mediante $\widehat{p}_{ij} = \dfrac{n_{ij}}{n_i}$, siendo n_{ij} el número de pasos del estado E_i al estado E_j y n_i el número de veces que se ha observado el estado E_i. El resultado fue que en todos los casos se rechazaba la hipótesis nula y, por tanto, se mantiene el carácter markoviano del fenómeno. El siguiente paso era comprobar si era suficiente una cadena de Markov de primer orden o debía ser de orden superior; el criterio de información de Akaike, AIC (véase [Ei]), dio como resultado mantener la cadena de primer orden en todos los casos. En definitiva, la ocurrencia de precipitación $X_n(t)$ se modeliza mediante una cadena de Markov de primer orden con dos estados $\{E_0, E_1\}$, descrita por unas probabilidades de transición:

$$p\Big(X_n(t) = E_j \Big| X_{n-1}(t) = E_i\Big) = p_{ij}^X, \quad 0 \le i, j \le 1, \tag{1}$$

y por unas probabilidades iniciales:

$$p\Big(X_0(t) = E_i\Big) = p_i^X, \quad i = 0, 1. \tag{2}$$

2.2. Intervalos de precipitación

La variable aleatoria $Y_n(t)$ se considera estacionaria en cada mes (esto es, de comportamiento homogéneo en el mismo mes de cada uno de los años de la serie); se ha descrito mediante una distribución multinomial con seis estados correspondientes a los siguientes intervalos de precipitación:

$$E_0 = 0,\ E_1 =]0, 1],\ E_2 =]1, 4],\ E_3 =]4, 8],\ E_4 =]8, 15],\ E_5 =]15, +\infty[.$$

Las probabilidades de transición vienen dadas por:

$$p\Big(Y_n(t) = E_i \Big| X_n(t) = E_j\Big) = \begin{cases} 1, & \text{si } \ j = 0, i = 0 \\ p_i^Y, & \text{si } \ j = 1, 1 \le i \le 5 \end{cases} \tag{3}$$

(véase [Co1]).

2.3. Cantidad de precipitación

Sea α un número real cualquiera; la variable aleatoria contínua $L_n(t)$ (cantidad de precipitación) se describe por la siguiente función de distribución:

$$p\left(L_n(t) \le \alpha \;\middle|\; Y_n(t) = E_0\right) = \begin{cases} 0, \text{ si } \alpha < 0 \\ 1, \text{ si } \alpha \ge 0 \end{cases}$$

$$p\left(L_n(t) \le \alpha \;\middle|\; \omega_{i-1} < L_n(t) \le \omega_i\right) = \begin{cases} 0, & \text{si } \alpha \le \omega_{i-1} \\ \dfrac{\Phi(\alpha) - \Phi(\omega_{i-1})}{\Phi(\omega_i) - \Phi(\omega_{i-1})}, & \text{si } \alpha \in E_i =]\omega_{i-1}, \omega_i], \ 1 \le i \le 5; \\ 1, & \text{si } \alpha > \omega_i; \end{cases} \tag{4}$$

$$p\left(L_n(t) \le \alpha \;\middle|\; Y_n(t) = E_6\right) = \begin{cases} 0, & \text{si } \alpha \le \omega_6 \\ \dfrac{\Gamma(\alpha) - \Gamma(\omega_6)}{1 - \Gamma(\omega_6)}, & \text{si } \alpha \in E_6 =]\omega_6, +\infty[. \end{cases}$$

En estas expresiones, $\Phi(t)$ es la función de distribución de la distribución lognormal y $\Gamma(t)$ la función de distribución de la distribución gamma (véase [Co2]).

2.4. Estimación de parámetros

Por todo lo anterior, puede verse que el modelo utiliza los parámetros siguientes para cada punto u observatorio y para cada mes del año:

Variable $X_n(t)$: $\begin{cases} p_{ij}^X & \text{4 parametros (2 parametros independientes),} \\ p_i^X & \text{2 parametros (1 parametro independiente);} \end{cases}$
Variable $Y_n(t)$: p_i^Y 5 parámetros (4 parámetros independientes);
Variable $L_n(t)$: $\widehat{\mu}, \widehat{\sigma}$ de los datos (2 parámetros independientes).

Así pues, son necesarios 9 parámetros independientes por punto y mes para poder aplicar el modelo. Si se dispone de series de datos de precipitación, se estiman los parámetros directamente a partir de las mismas. Si no se dispone de datos (que de hecho constituye la situación de mayor interés para aplicar el modelo) la estimación de los parámetros del modelo se puede realizar empleando dos técnicas distintas:

1. A partir de los diez observatorios más cercanos geográficamente al punto en que se desea aplicar el modelo, se calculan dos coeficientes de ponderación: uno inversamente proporcional a la distancia y otro inversamente proporcional a la diferencia de altitud; a partir de estos dos coeficientes se calcula otro global para cada uno de los diez observatorios, dando un peso de 70 % a la distancia y 30 % a la altitud.

2. Se realiza una partición de la comarca en zonas climáticas homogéneas, utilizando la técnica de agregación alrededor de centros móviles o "clusters" de medias (en nuestro caso han resultado cinco zonas climáticas homogéneas). A continuación se identifica la zona a la que pertenece el punto y se estiman los parámetros a partir de todos los datos de observatorios de la zona.

3. VALIDACIÓN DEL MODELO

Para validar el modelo se han estimado los parámetros de 20 puntos distribuidos uniformemente por la provincia de Lleida de los cuales se tenia información completa (precipitación diaria) de una serie de 20 años. Se han comparado las matrices de transición real y simulada de la variable $X_n(t)$ mediante una prueba χ^2 aceptándose en todos los casos la hipótesis nula: matriz de transición igual a la real. En cuanto a la variable $Y_n(t)$ se ha comprobado que se ajustaba en todos los casos a la multinomial real.

4. REFERENCIAS BIBLIOGRÁFICAS

[Al] ALCAIDE, M., (1986). "Análisis estacional y regional de la precipitación en el valle del Guadalquivir". Tesis doctoral. Universidad de Córdoba.

[An] ANDERSON, T.W. y L.A. GOODMAN (1957). "Statistical inference about Markov chains". Ann. Math. Stat. 28, 89-110.

[Co1] COLOMER, M.A. y E. JARAUTA (1994). "Modelo estocástico de precipitación y temperatura diaria aplicado a la provincia de Lleida". XXI congrès nacional d'estadística i investigació operativa. Calella, abril 1994.

[Co2] COLOMER, M.A. y E. JARAUTA (1994). "Markov chain model for daily rainfall applied to the northeast of Spain". Fifth international conference on precipitation (Creta, junio 1995).

[Da] DAS, P.D. (1989). "Stochastic models for rainfall prediction". American assoc. for adv. of science, 311-319.

[Ei] EIDSVIK, K.J. (1980). "Identification of models for some time series of atmosferic origin with Akaike's information criterion". Journal of applied meteorology, 19(4):357-369.

[Ja1] JARAUTA, E. (1989). "Modelos matemáticos del régimen de humedad de los suelos. Aplicación a la determinación del régimen de humedad de los suelos del área meridional de Lleida". Tesis doctoral, UPC.

[Ja2] JARAUTA, E., J. PORTA y J.BOIXADERA (1993). "Règims d'humitat dels sòls: interès i problemàtica en l'aplicació als sòls de Catalunya". Butlletí de la institució Catalana d'Història Natural, 61: 87-96).

[Ne] NEWHALL, F. (1976). "Calculation of soil moisture regimes from the climatic record". Soil Conservation Service - USDA.

[Ro1] ROLDÁN, J; A. GARCIA-GUZMÁN y A. LOSADA (1980). "Modelo estocástico de precipitación diaria". ITEA, 40: 11-21.

[Ro2] ROLDÁN, J y D.A. WOOLHISER (1982). "Stochastic daily precipitation models. 1: A comparison of occurence processes". Water resources research, 18(5): 1451-1459.

[So] SOIL SURVEY STAFF (1975). "Soil Taxonomy. A basic system of soil classification for making and interpreting soil surveys". (Soil Conservation Service - USDA). Agriculture handbook, 436.

[Wo] WOOLHISER, D.A. y J. ROLDÁN (1982). "Stochastic daily precipitation models. 1: A comparison of distribution of amounts". Water resources research, 18(5): 1461-1468.