

*Abstract submitted to CONGREGMET IX, 9th Argentine Congress of Meteorology,
Buenos Aires, Argentina, 3-7 October 2005*

UN GENERADOR ESTOCASTICO HIBRIDO PARA ANALISIS DE RIESGO CLIMATICO

Guillermo P. Podestá

University of Miami, Rosenstiel School of Marine and Atmospheric Science
gpodesta@rsmas.miami.edu

Balaji Rajagopalan

University of Colorado
Rajagopalan.Balaji@Colorado.edu

James Brown

jbrown@rsmas.miami.edu

María de los Milagros Skansi, Liliana Núñez, Silvia E. Núñez

Servicio Meteorológico Nacional, Argentina
mms@meteofa.mil.ar

RESUMEN

Stochastic weather generators can produce synthetic series of daily weather variables with statistical properties consistent with those of the historical record. We introduce here a hybrid generator, in which precipitation occurrence is modeled parametrically whereas values of relevant weather variables (minimum and maximum temperatures, precipitation and solar radiation) are generated non-parametrically, through resampling of historical series based on selection of “nearest neighbors” in multivariate space. As an example, we produce 25 synthetic series of length 73 years for Pergamino, an important agricultural region in the Argentine Pampas. Graphic diagnostics confirm that the historical series is consistent with the envelope of variability of the synthetic series. The generator can produce output conditioned on a given seasonal forecast. Therefore, synthetic series can be combined with relevant process models to assess climate-related risk in socioeconomic sectors sensitive to climate variability.

Los generadores estocásticos de clima producen series diarias de variables climáticas con propiedades estadísticas similares a las de los datos históricos. Presentamos un generador híbrido, en el que la ocurrencia de precipitación se modela paramétricamente mientras que los valores de temperatura máxima y mínima, precipitación y radiación solar total se generan en forma no paramétrica, remuestreando las series históricas en base a la identificación de “vecinos más cercanos” en el espacio multivariado. Como ejemplo, producimos 25 series sintéticas de 73 años de largo para Pergamino, una importante región agrícola en la Pampa argentina. Varios diagnósticos gráficos confirman que la serie histórica es consistente con el rango de variabilidad de las series sintéticas. El generador puede producir resultados condicionados a un pronóstico estacional determinado, por lo que puede combinarse con modelos de procesos para la evaluación del riesgo en sectores socioeconómicos sensibles a la variabilidad climática.

INTRODUCCION

Las interacciones entre los cultivos agrícolas y las condiciones climáticas generalmente son complejas y no lineales, por lo que algunos métodos estadísticos simples generalmente no describen adecuadamente los efectos de la variabilidad climática sobre el riesgo para la producción agrícola. Los modelos biofísicos que simulan el crecimiento y rendimiento de cultivos de importancia económica son herramientas alternativas para estimar los impactos del clima sobre los sistemas agrícolas y para evaluar los resultados de un amplio espectro de decisiones de manejo (Boote et al. 1996). Estos modelos típicamente requieren datos climáticos diarios. Aunque se pueden utilizar datos climáticos históricos como entrada para estos modelos, generalmente es difícil y costoso disponer de series históricas largas a resolución diaria. Otra limitación fundamental es que el uso de series históricas proporciona resultados para *una única realización* de los procesos climáticos. Una evaluación rigurosa del riesgo climático, por el contrario, requiere la exploración del rango de resultados del proceso de interés (por ej., rendimientos de un cultivo) en respuesta a un rango amplio de series climáticas igualmente probables que la realización histórica, y con propiedades estadísticas similares (Richardson 1981). Una posible solución es el uso de generadores estocásticos de clima que producen series diarias de variables climáticas con propiedades estadísticas similares a aquellas de la serie histórica.

Se han propuesto muchos métodos alternativos para la generación de precipitación, temperatura máxima y mínima, y radiación solar a nivel diario, las variables requeridas por la mayoría de los modelos de simulación agronómica. Revisiones de los generadores más comunes han sido publicadas por Wallis y Griffiths (1995), Johnson et al. (1996), y Wilks y Wilby (1999). La mayoría de los generadores propuestos se basan en métodos paramétricos (o sea, se asumen distribuciones que se describen mediante parámetros estimados a partir de datos históricos o publicados en la bibliografía). Sin embargo, últimamente se han desarrollado generadores basados en métodos no paramétricos (Rajagopalan y Lall 1999; Yates et al. 2003; Clark et al. 2003; Gangopadhyay et al. 2005), que tienen algunas ventajas sobre las alternativas paramétricas en algunos aspectos. Los métodos paramétricos evitan juicios subjetivos sobre las formas apropiadas de modelos o distribuciones subyacentes. Puesto que están basados en los datos mismos, los métodos no paramétricos pueden capturar desviaciones respecto de las distribuciones teóricas y asociaciones no lineales entre variables (Wilks y Wilby 1999; Rajagopalan y Lall 1999).

En este trabajo presentamos un generador estocástico de clima de naturaleza híbrida, en el cual el proceso de ocurrencia de precipitación está basado en un esquema paramétrico (una cadena de Markov de orden 1), en tanto que la generación de valores de las variables climáticas sigue un esquema no-paramétrico basado en la selección de “vecinos más cercanos” (nearest neighbors). Primero describimos brevemente el funcionamiento del generador y luego presentamos diagnósticos gráficos comparando un conjunto de series sintéticas con datos históricos para la localidad de Pergamino, Provincia de Buenos Aires, una de las zonas agrícolas más productivas de la Pampa Húmeda argentina.

METODOLOGIA Y MATERIAL

Datos utilizados

Para “entrenar” al generador estocástico de clima utilizamos una serie histórica de datos para la localidad de Pergamino, Provincia de Buenos Aires, que incluyó el período desde el 1 de enero de 1931 al 31 de diciembre de 2003. Las temperaturas máximas y mínimas y la precipitación total diaria estaban incluidas en el registro original, pero la radiación solar se estimó en base a heliofanía o, cuando esta variable faltaba, en base a amplitud térmica diaria y precipitación (ver Podestá et al. 2004 para mayores detalles).

Descripción del generador estocástico de clima

El generador que presentamos aquí está basado en la metodología desarrollada originalmente por Rajagopalan y Lall (1999). Al generador no paramétrico original se le agregó un componente paramétrico para modelar mejor el proceso de ocurrencia de precipitación, por lo que denominamos a la nueva versión como un generador “híbrido.” El resto del proceso de generación sigue la metodología descrita por Rajagopalan y Lall (1999). En esta sección presentamos una breve descripción de los componentes paramétrico y no paramétrico.

Modelación paramétrica de la ocurrencia de precipitación

Un modelo simple para la ocurrencia de precipitación es una cadena de Markov de primer orden con dos posibles estados: día lluvioso o día seco (Wilks 1992; Katz y Parlange 1993). En un proceso de este tipo, la probabilidad de un estado determinado (día lluvioso o seco) en un día t depende sólo del estado del proceso en el día $t-1$. Definimos a J como la variable estocástica que indica el estado del proceso para el día t : $J_t = 0$ si el día t es no lluvioso; $J_t = 1$ si el día t es lluvioso. Este modelo está caracterizado por cuatro probabilidades de transición:

$$P_{ij} = \Pr \{J_t = j | J_{t-1} = i\}, \quad i, j = 0, 1 .$$

Puesto que las probabilidades de las transiciones $P_{00} + P_{01} = P_{10} + P_{11} = 1$, solamente necesitamos dos parámetros para definir el proceso de Markov de primer orden con dos estados. Los parámetros fueron estimados a partir de los datos históricos utilizando una ventana móvil de 15 días centrada en el día de interés. Por ejemplo, los parámetros para el día 8 de enero se estimaron en base a todos los datos entre el 1 de enero y el 15 de enero para todos los años en el registro. Para definir un día lluvioso, la precipitación total diaria debe ser mayor o igual que 0.3 mm.

Generación no paramétrica de variables climáticas

El proceso de remuestreo por el cual se generaron las variables climáticas diarias de interés (temperatura máxima y mínima, precipitación total, y radiación solar total) está descrito con detalle en Rajagopalan y Lall (1999). Brevemente, el proceso de generación de valores para el día t involucra cinco pasos principales:

1. Se identifica un vector de condicionamiento, que en este caso contiene las variables climáticas para el día anterior ($t-1$). Ya que la radiación solar no fue observada sino

derivada de otras variables, sólo consideramos como componentes del vector de condicionamiento a las temperaturas máxima y mínima y la precipitación total.

2. Se determina un número k de “vecinos más cercanos” al vector de condicionamiento en el espacio multivariado. En el esquema híbrido, sólo se consideran como sucesores posibles a aquellos valores históricos que tengan un estado de precipitación (seco o lluvioso) igual al del día ya generado ($t-1$). Para seleccionar los vecinos más cercanos se utiliza como medida la Distancia Euclidiana.
3. Se determina, para cada uno de los “vecinos más cercanos,” un vector sucesor, que incluye los valores de las variables climáticas para el día siguiente al de cada miembro del conjunto de vecinos.
4. Se selecciona aleatoriamente uno de los vectores como el valor de las variables climáticas para el día t .
5. Se repite el proceso.

Para seleccionar los “vecinos más cercanos”, se utilizó una ventana móvil de 15 días centrada alrededor del día anterior al cual se estaban generando valores. Por ejemplo, supongamos que deseamos generar valores para un 8 de enero, y ya se han generado los valores para el 7 de enero, que en este caso provienen (como ejemplo arbitrario) del año 1973 y corresponden a un día lluvioso. Un sucesor lógico sería seleccionar los valores para el 8 de enero de 1973. Sin embargo, consideramos como candidatos a ser “vecinos más cercanos” a los valores para todas las fechas entre el 1 de enero y el 15 de enero de todos los años en el registro histórico (con la limitación de que los vecinos deben ser lluviosos). Calculamos la distancia euclidiana multivariada entre los valores para el 8 de enero de 1973 y el resto de los candidatos, y retenemos solamente un número determinado de días (unos 30), considerados “vecinos más cercanos” en base a su distancia euclidiana. Finalmente, seleccionamos aleatoriamente uno de los vecinos más cercanos. Por ejemplo, los datos seleccionados pueden corresponder al 3 de enero de 1941, pero se asignan al 8 de enero quedando así generado. El proceso se repite nuevamente hasta completar el largo deseado para una serie sintética.

La generación de las series sintéticas además fue condicionada por la fase del fenómeno ENOS (El Niño-Oscilación del Sur). Un “año ENOS” se consideró como el período entre el 1 de julio de un año determinado y el 30 de junio del año calendario siguiente. A cada uno de los “años ENOS” en la serie histórica se le asignó una fase del ENOS (o sea, años “Niño”, años “Niña”, o años “neutrales”). La asignación de fases se basó en anomalías de temperatura superficial del Océano Pacífico tropical persistentes por seis o más meses, incluyendo el trimestre octubre-diciembre (para más detalles sobre la clasificación de fases, ver Podestá et al. 1999). Para todas las series sintéticas, el arreglo temporal de fases del ENOS fue idéntico al del registro histórico. En una etapa sucesiva, planeamos tornar aleatoria también la secuencia de fases.

RESULTADOS

En la forma descrita en la sección anterior, generamos 25 series de 73 años de longitud (equivalentes al largo de la serie histórica original). Desarrollamos luego una serie de diagnósticos gráficos para investigar si la serie histórica podía ser considerada como una realización posible dentro de la envolvente definida para cada variable por el conjunto de las

series sintéticas. Los resultados se muestran en las figuras incluidas a continuación. Ponemos énfasis en diagnósticos referidos a la precipitación, ya que esta es la variable más difícil de reproducir. En especial, investigamos estadísticas como el largo de secuencias de días lluviosos o secos, que tienen dificultad en ser reproducidas por modelos simples.

Figura 1. Probabilidad de días lluviosos (precipitación diaria ≥ 0.3 mm) en Pergamino. Los puntos unidos por las líneas indican los valores estimados para la serie histórica 1931-2003. Para cada mes se muestran también la dispersión de valores para las 25 series sintéticas. La línea horizontal en el centro de cada rectángulo corresponde a la mediana de las 25 series. Los límites superior e inferior de cada rectángulo definen el 50% central de la distribución de valores en la serie. La diferencia entre estos dos valores (o sea, el alto del rectángulo) se denomina “rango inter-cuartil” (o RIC). Las líneas verticales definen el rango de valores menores o iguales que 1.5 veces el RIC (“rango contiguo”). Las líneas horizontales aisladas indican valores más allá del rango contiguo.

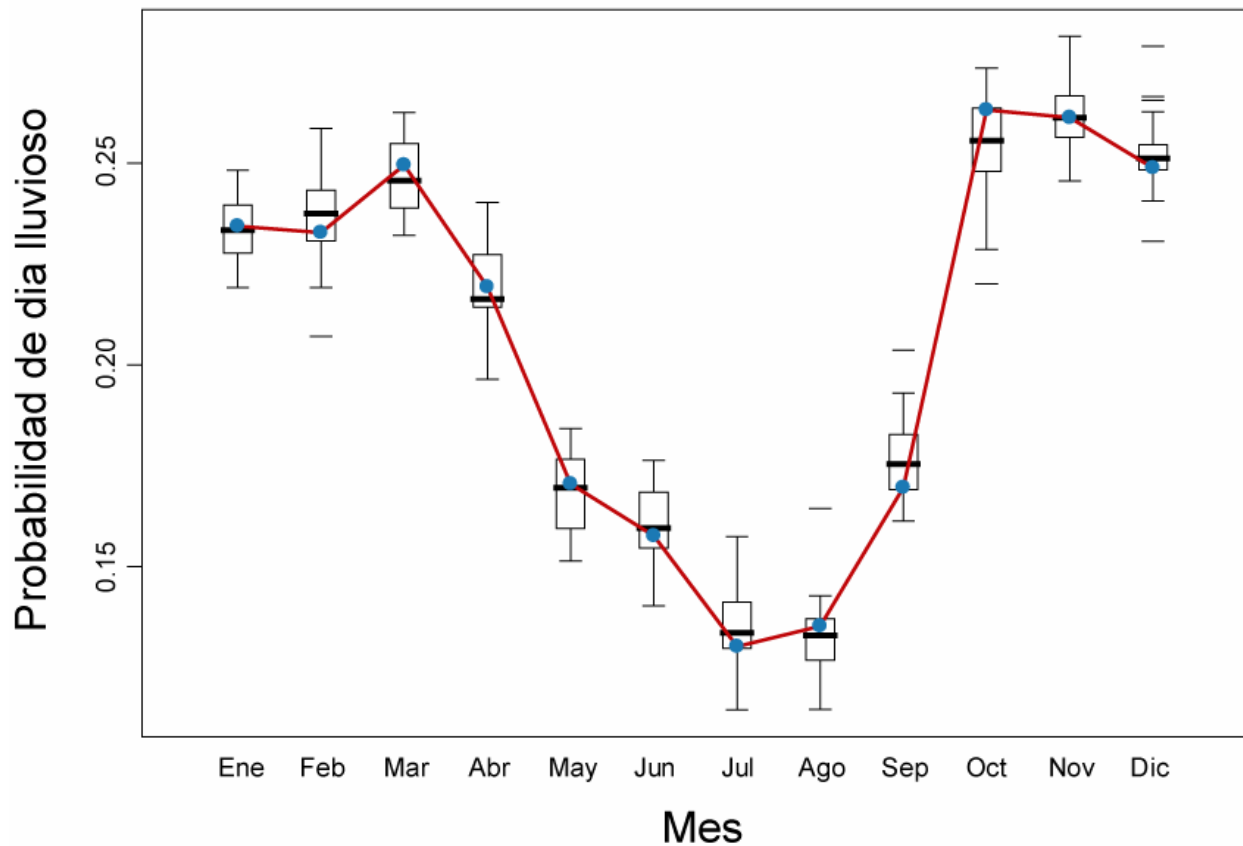


Figura 2. Largo promedio de secuencias de días lluviosos (panel izquierdo) y secos (panel derecho) por trimestre. Los puntos unidos por las líneas indican los valores estimados a partir de la serie histórica 1931-2003. Para cada mes se muestran también la dispersión de valores para las 25 series sintéticas, graficados de la misma manera que en la Figura 1.

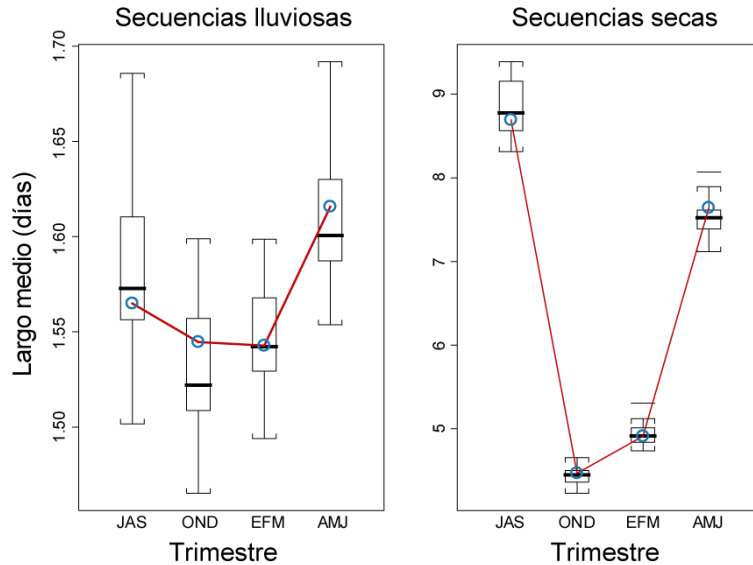


Figura 3. Mediana de precipitación total por trimestre, para cada fase del fenómeno ENOS. Las líneas gruesas unen los valores para cada fase estimados a partir de la serie histórica 1931-2003. Para cada trimestre y fase se muestran también la dispersión de valores para las 25 series sintéticas, graficados de la misma manera que en las figuras anteriores (en este caso, la mediana de las 25 series está indicada por un punto negro en el centro de las cajas).

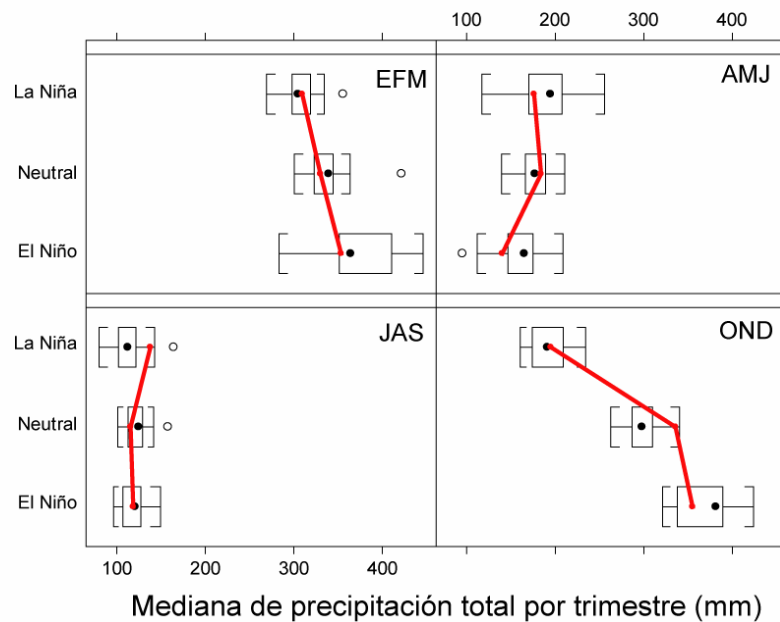
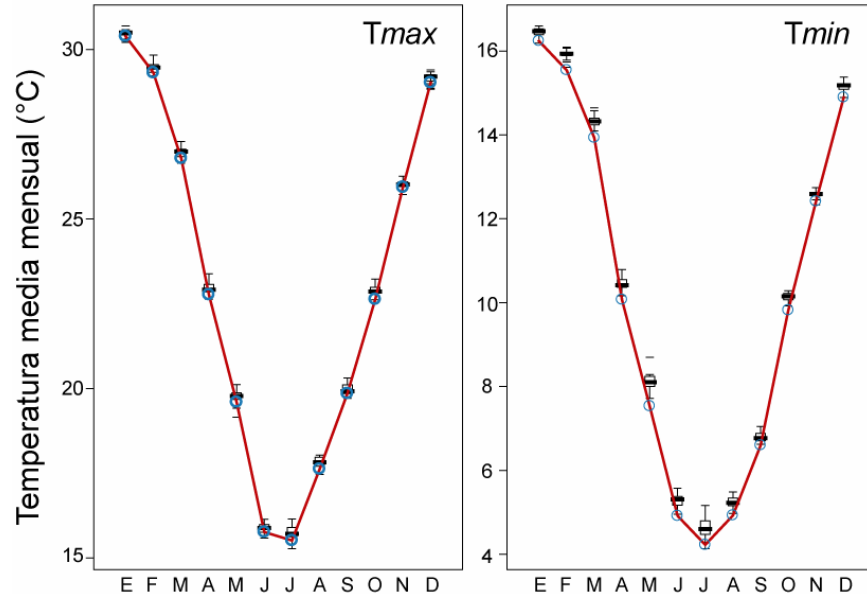


Figura 4. Temperaturas medias mensuales máxima y mínima. Los puntos unidos por las líneas indican los valores estimados para la serie histórica 1931-2003. Para cada mes se muestran también la dispersión de valores para las 25 series sintéticas, graficados de la misma manera que en la Figura 1.



DISCUSION

En este trabajo presentamos un generador estocástico que puede producir secuencias sintéticas de variables climáticas (temperaturas máxima y mínima, precipitación, radiación solar) a escala diaria. El esquema híbrido propuesto incluye una modelación paramétrica de la ocurrencia de precipitación, y un enfoque no paramétrico para la generación de las variables climáticas de interés.

Los diagnósticos presentados en la sección anterior sugieren que las estadísticas de las series sintéticas producidas son consistentes con las de la serie histórica. Es decir, la serie histórica puede ser considerada como una realización más del conjunto definido por las simulaciones. En especial, es interesante verificar que se reproducen propiedades de orden superior de las series climáticas históricas. Por ejemplo, el generador reproduce apropiadamente la distribución de secuencias de días lluviosos y secos (Figura 2).

En los últimos años ha habido avances considerables en el pronóstico de escenarios climáticos estacionales. Los generadores estocásticos de clima pueden ser utilizados para traducir escenarios climáticos pronosticados en resultados para sectores de interés socioeconómico. Por ejemplo, el generador presentado en este trabajo utiliza una variable de condicionamiento que permite remuestrear selectivamente los datos históricos de modo de producir secuencias sintéticas consistentes con un escenario determinado. A modo de ejemplo, generamos separadamente series sintéticas para las diferentes fases del fenómeno ENOS. Como sugiere la

Figura 3, la generación condicionada permite capturar diferencias entre condiciones asociadas con diferentes fases (ver, por ejemplo, el panel correspondiente al trimestre octubre-diciembre).

Las series sintéticas, a su vez, pueden alimentar modelos de procesos de interés, como modelos de simulación de crecimiento de cultivos, lo que permite obtener una distribución de frecuencias de resultados (ej., rendimientos de cultivos). Este enfoque produce una evaluación realista de los resultados posibles de la variabilidad climática. Cada resultado tendría asociada una probabilidad de ocurrencia, lo cual es un resultado sumamente útil para análisis de riesgo.

REFERENCIAS

- Boote K, J.W. Jones, y N.B. Pickering. 1996. Potential uses and limitations of crop models. *Agronomy Journal* 88: 704-716.
- Clark, M.S., S. Gangopadhyay, L. Hay, B. Rajagopalan y R. Wilby. 2004. The Schaake Shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology* 5: 243-262.
- Gangopadhyay, S., M. Clark, y B. Rajagopalan. 2005: Statistical downscaling using K-nearest neighbors. *Water Resources Research* 41.
- Johnson, G.L., C.L. Hanson, S.P. Hardegree, y E.B. Ballard. 1996. Stochastic weather simulation: Overview and analysis of two commonly used models. *Journal of Applied Meteorology* 35: 1878–1896.
- Katz, R. y M. Parlange. 1993. Effects of an index of atmospheric circulation on stochastic properties of precipitation. *Water Resources Research* 29: 2335-2344.
- Podestá, G. P., C. D. Messina, M. O. Grondona, y G. O. Magrin. 1999. Associations between grain crop yields in central-eastern Argentina and El Niño-Southern Oscillation. *Journal of Applied Meteorology* 38: 1488-1498.
- Podestá, G.P., L. Nuñez, C.A. Villanueva, y M.A. Skansi. 2004. Estimating daily solar radiation in the Argentine Pampas. *Agricultural and Forest Meteorology*.
- Rajagopalan, B. y U. Lall. 1999. A k-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resources Research* 35: 3089-3101.
- Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research* 17: 182-190.
- Wallis, T.W.R., y J.F. Griffiths, 1995. An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC). *Agricultural and Forest Meteorology* 73: 115-133.
- Wilks, D. S. 1992. Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change* 22: 67-84.
- Wilks, D.S., y R.L. Wilby. 1999. The weather generation game: A review of stochastic weather models. *Progress in Physical Geography* 23: 329-357.
- Yates, D., S. Gangopadhyay, B. Rajagopalan, y K. Strzepek. 2003. A technique for generating regional climate scenarios using a nearest-neighbor algorithm. *Water Resources Research*, 39: 1199-1214.