

Manual para la caracterización multivariada de eventos secos

Alessio Bocco

19 April, 2021

Contents

1	Introducción	1
2	Análisis probabilístico	2
2.1	Análisis exploratorio	4
2.2	Análisis univariado	5
	Bibliografía	7

1 Introducción

El análisis probabilístico de los eventos secos es una instancia integradora que incluye a los productos generados en los Entregables anteriores. La Figura 1 muestra el diagrama de flujo para el análisis de frecuencia de los eventos secos. Dada la escasa longitud de los registros climáticos históricos para las localidades del SISSA, es necesaria la generación de series climáticas sintéticas. Para ello se utiliza el generador estocástico presentado en el Entregable I. A partir de las series sintéticas de precipitación y de temperaturas máximas y mínimas se calculan varios índices de sequía, principalmente el Índice Estandarizado de Precipitación (SPI, por sus siglas en inglés) (McKee *et al.*, (1993)) y el Índice Estandarizado de Precipitación y Evaporación (SPEI, por sus siglas en inglés) (Vicente-Serrano *et al.*, (2010)).



Figure 1: Diagrama del procesos de análisis multivariado de métricas de eventos secos.

A continuación, los índices de sequía calculados para las series climáticas sintéticas son utilizados para identificar eventos secos sintéticos. Un evento seco comienza cuando los valores del índice de sequía perforan un determinado umbral definido por el usuario y termina cuando el índice supera nuevamente dicho umbral. Una vez definido el comienzo y fin de un evento seco, se caracteriza al evento a través de un conjunto de métricas. A partir de las series de métricas se pueden ajustar distintos tipos de modelos para analizar el riesgo de ocurrencia de distintos tipos de eventos secos. Para ello, como se observa en el diagrama de la Figura 1, se procede de dos maneras diferentes. Se pueden considerar las métricas individualmente y realizar un análisis probabilístico univariado. Este análisis implica ajustar una función de probabilidad a una métrica individual (Por ejemplo, la duración de un evento seco) y estimar el período de retorno de eventos de diferente duración. Por ejemplo, se puede calcular el período de retorno de eventos que tengan una duración igual o mayor a 10 pentadas (50 días) tanto para una localidad como para una región. No obstante, las distintas métricas no son siempre independientes entre sí, por lo que el análisis univariado puede sobresimplificar el fenómeno en estudio. Por ejemplo, dos eventos de igual duración pueden tener impactos muy diferentes si las intensidades difieren. Lo anterior hace especialmente relevante el uso de técnicas que permitan evaluar la dependencia conjunta de dos o más métricas para así caracterizar los eventos secos de la forma más completa posible.

2 Análisis probabilístico

El análisis probabilista de una amenaza natural necesita de la combinación del generador estocástico de series climáticas y del cálculo de algún tipo de indicador o índice que per-

mita caracterizar cada evento. Luego, las características de estos eventos pueden modelarse individualmente (análisis univariado) o de manera conjunta (análisis multivariado). Dada la complejidad de la temática se explicará detalladamente la metodología utilizando como ejemplo los eventos identificados en la localidad de Villa Reynolds, San Luis, Argentina. Para más detalles se sugiere ver (). Si bien este ejemplo será con eventos identificados a partir de series sintéticas, el procedimiento es válido también para series observadas.

Las series simuladas no son estacionarias, copian la tendencia observada en la serie climática original. El periodo simulado comienza en 1961 y finaliza en 2019. Para ver más detalles sobre la generación de series sintéticas se sugiere ver el siguiente link: <https://github.com/CRC-SAS/weather-generator/tree/gamwgen/workshop>

Cada serie temporal simulada de temperaturas máxima y mínima y precipitación fueron utilizadas para el cálculo de índices de sequía. El análisis probabilístico que se describirá a continuación está basado en el Índice Estandarizado de Precipitación (McKee *et al.*, (1993)) agregado a una escala de tres meses (SPI-3). Si bien aquí se mostrarán los resultados a partir del SPI 3, la misma metodología puede utilizarse con otros índices de sequía como el SPEI (Vicente-Serrano *et al.*, (2010)). También es muy interesante la utilización de distintas escalas temporales del índice ya que permite complementar el análisis y caracterizar distintos tipos de sequías. Una vez obtenidas las series temporales del índice de sequía se procede a la identificación de los eventos secos. Un evento comienza después de una cierta cantidad de periodos consecutivos con valores de los índices menores a un umbral determinado del índice especificado. El evento termina cuando el valor del índice supera nuevamente dicho umbral (Shiau, (2006)). Un ejemplo de un evento seco hipotético con las distintas métricas que lo caracterizan se muestra en la Figura 2. La línea roja corresponde a la serie temporal de SPI. En el ejemplo, el umbral que define condiciones de sequía es $SPI_n < -1$ (mostrado como una línea violeta punteada). El evento comienza (“Comienzo” en la Figura 2) cuando el índice cruza el umbral hacia condiciones secas (menores al umbral definido) y termina (“Fin” en la Figura 2) cuando el índice cruza nuevamente el umbral en el sentido contrario. Una vez definido el momento de comienzo y fin, un evento se puede caracterizar a través de una serie de métricas. Las más importantes son:

- Duración: período de tiempo durante el cual el índice de sequía se encuentra por debajo del umbral seleccionado (Shiau, 2006);
- Intensidad: promedio de todos los valores del índice durante el período en que el mismo está por debajo del umbral seleccionado (línea verde horizontal en Figura 2);
- Magnitud: acumulación del déficit por debajo del umbral seleccionado durante el período de la sequía – se estima como el producto de la duración y la intensidad (ilustrado como el área punteada en la Figura 3) (Mishra and Singh, 2010);
- Valor extremo: el menor valor del índice alcanzado durante el evento (Serinaldi et al., 2009).

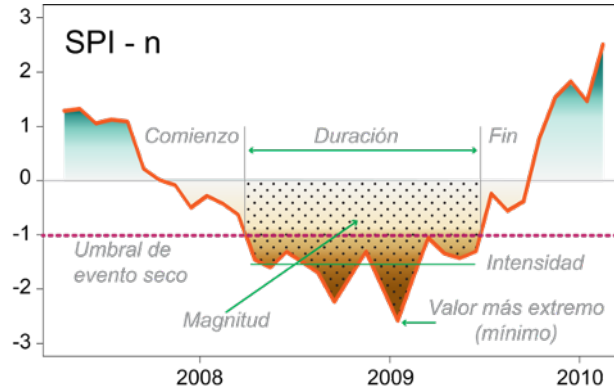


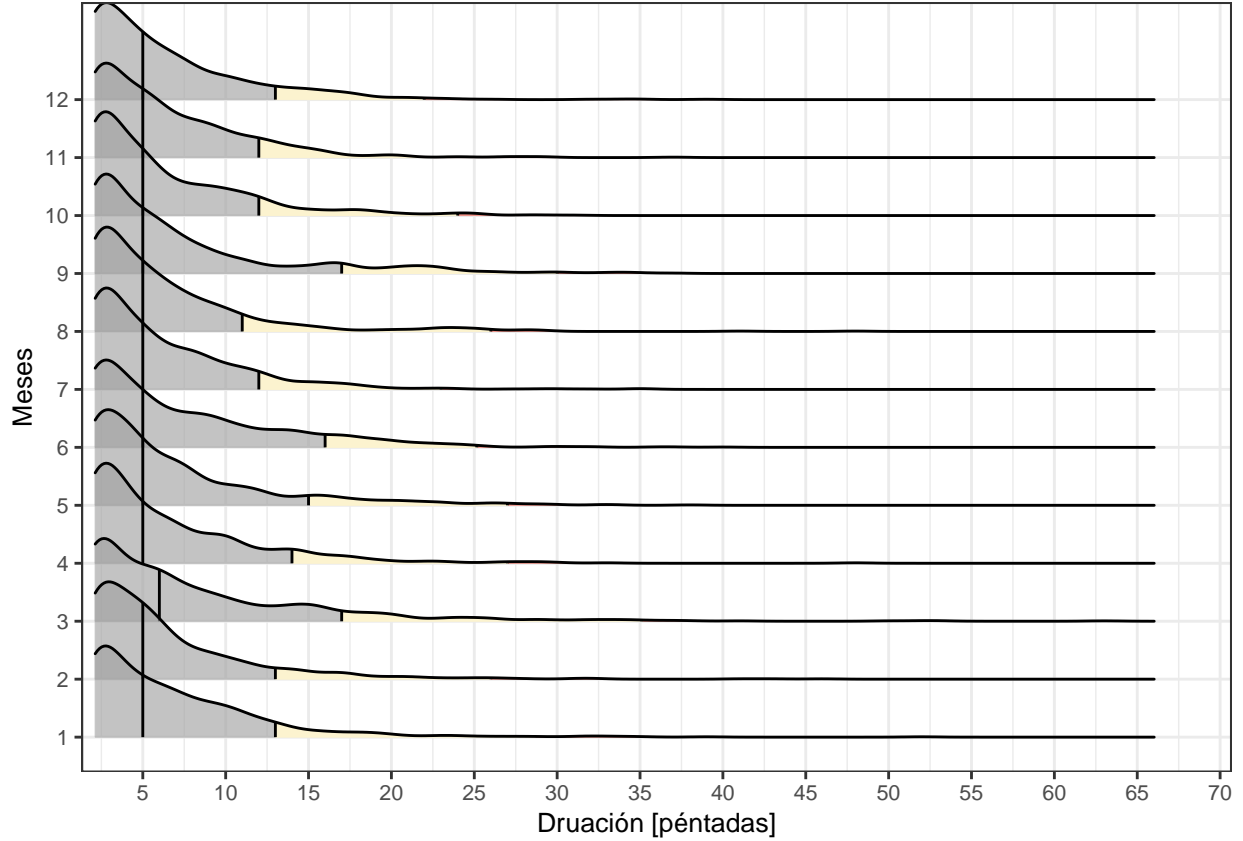
Figure 2: Diagrama del procesos de análisis multivariado de métricas de eventos secos.

Un problema frecuente en la evaluación del riesgo de la amenaza de sequías es la estimación de la probabilidad de ocurrencia de un evento seco de determinadas características expresada en términos de período de retorno (PR). El PR se define como el tiempo promedio entre dos realizaciones sucesivas de un evento tipo determinado. El concepto de PR es muy importante en términos de la amenaza, ya que permite no sólo tomar decisiones en función de la probabilidad de ocurrencia de un evento determinado sino que es la medida más comúnmente usada para la evaluación objetiva de la amenaza (Singh et al., 2007).

2.1 Análisis exploratorio

El análisis exploratorio de las series de métricas tiene como objetivo comprender su variabilidad temporal y espacial. El uso de series sintéticas permite multiplicar la cantidad de eventos secos identificados por lo que la amenaza puede ser caracterizada de una manera mucho más robusta. Cabe mencionar que en los registros históricos, que tienen una longitud de aproximadamente 60 años (1961-2020), se han identificado unos 70 eventos en esta localidad. El uso de series sintéticas permite multiplicar la cantidad de eventos y así permitir el ajuste de distintas funciones de probabilidad tanto univariadas como multivariadas a la serie de métricas para la caracterización de la amenaza de sequía.

Como se mencionó anteriormente, el análisis probabilístico implica el ajuste de una función de probabilidad a las series de métricas. Para ello es de mucha utilidad visualizar la distribución de los valores de cada una de las métricas. En la Figura 6 se muestra la distribución de la duración (panel superior) y magnitud (panel inferior) de los eventos secos en Junín y Laboulaye, ambas localidades de Argentina. La línea sólida muestra la densidad de la distribución empírica de la duración y magnitud para cada uno de los meses del año. En color gris se muestra la parte central de la distribución que comprende desde el cuantil 0 hasta el 90; en amarillo se muestra entre el 90 y 99 y en rojo desde el cuantil 99. La línea vertical central indica la ubicación de la mediana de la distribución empírica. Los eventos ubicados en la zona gris podrían considerarse relativamente frecuentes mientras que los ubicados en las zonas amarillas y rojas son extremos.



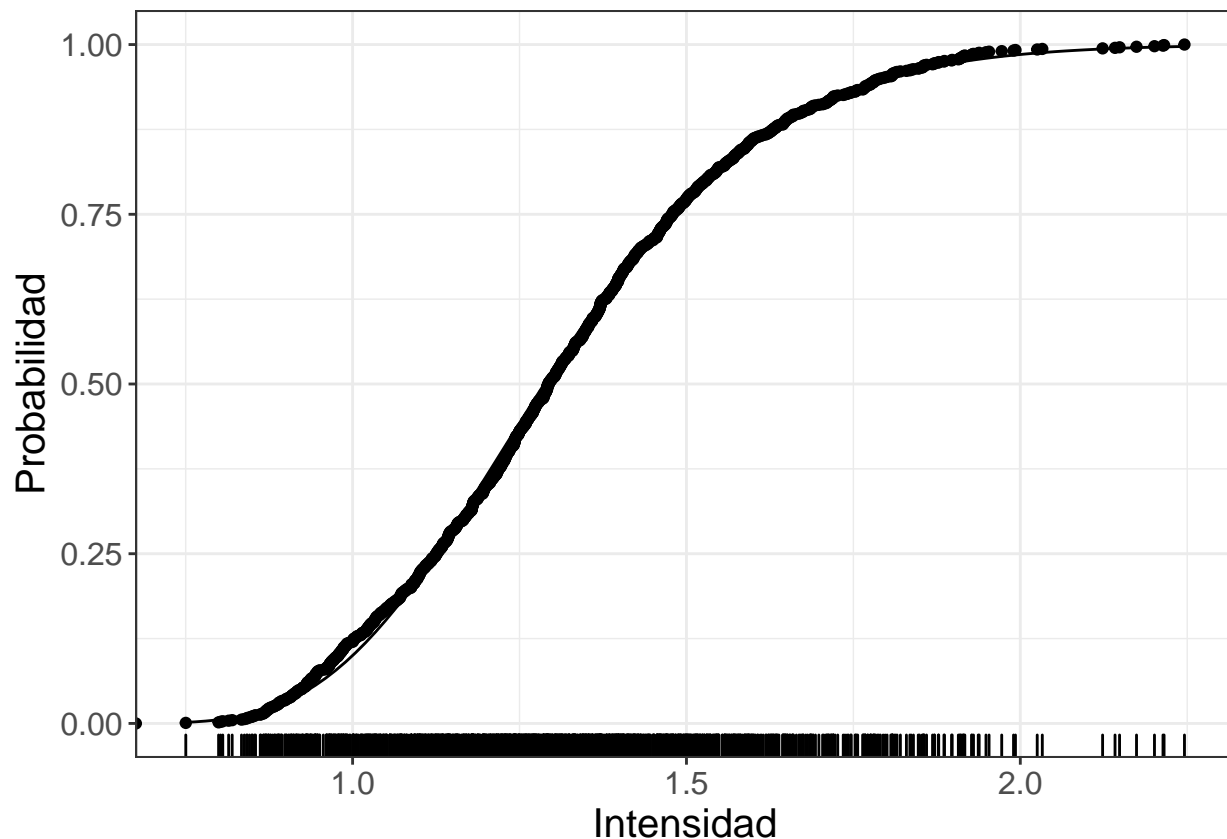
Este tipo de gráficos son muy útiles para visualizar el tipo de distribución de probabilidad que tienen los datos. Sin embargo, para sacar conclusiones estadísticamente significativas es necesario continuar con el análisis univariado.

2.2 Análisis univariado

El análisis de las métricas individuales comienza con los eventos identificados a partir de series sintéticas de índices de sequía. A cada una de las métricas se ajustan una suite de distribuciones univariadas de probabilidad. Luego de una extensa revisión bibliográfica se seleccionaron una serie de distribuciones candidatas que son ajustadas a cada uno de los vectores de métricas (De Michele et al., 2013; Gaupp et al., 2017; Madadgar et al., 2017; Shiau, 2006). Las distribuciones usadas son: Exponencial de tres parámetros, Gamma, Valor Extremo Generalizada, Log-logística Generalizada, Normal Generalizada, Pareto Generalizada, Gumbel, Laplace, Log-Normal, Normal, Pearson III, Student 3 y Weibull (Asquith, 2011). Cada una de estas distribuciones univariadas se ajustó usando los métodos de L-momentos y de máxima verosimilitud. Para aquellas distribuciones cuyo ajuste logró converger, se aplicaron tres tests de bondad de ajuste: Kolmogorov-Smirnov, Anderson Darling y Cramér von Mises (Wilks, 2011). Las tres pruebas están basadas en un test de hipótesis por lo que se estima un estadístico y un p-valor para cada uno de ellos. Si el p-valor es superior a 0.05 se considera que la distribución ajustada es consistente con las distribuciones empíricas. Las

funciones que hayan pasado los tres tests de bondad de ajuste son luego ordenadas en función de cuán bien explican el comportamiento de las variables. Las distribuciones se compararon mediante otros tres tests que fueron desarrollados específicamente para determinar cuál de las distribuciones univariadas es la mejor. La primera prueba es una variante de la raíz del error cuadrático medio entre los cuantiles observados y los teóricos. Mientras más pequeño sea el error, mejor será el ajuste (Wilks, 2011). El segundo test es una prueba de correlación basada en la concordancia de los cuantiles (Lawrence and Lin, 1989). La interpretación es idéntica a la de un coeficiente de concordancia ordinario, los valores cercanos a 1 indican la máxima concordancia mientras que valores próximos a 0 indican lo contrario. La última prueba de este tipo consiste en una comparación entre los cuantiles teóricos para una distribución determinada y los cuantiles estimados para los datos observados (Wilcox et al., 2014). Se comparan sólo 10 cuantiles que son considerados representativos de la función de probabilidad; a través de una prueba de Monte-Carlo se determina si los cuantiles teóricos son iguales a los ajustados. Para cada uno de los 10 cuantiles probados se obtiene un p-valor que debe ser superior al alpha de la prueba (0.05) para ser considerado exitoso. Si el p-valor es inferior a 0.05, se considera que los cuantiles observados y los teóricos son diferentes. Este test tiene como objetivo determinar si los cuantiles extremos de la distribución se encuentran bien ajustados ya que son los que definirán las probabilidades de ocurrencia de valores extremos de las métricas.

Por ejemplo, en la Figura ?? se muestra el resultado del ajuste de la intensidad de los eventos secos para la localidad de Villa Reynolds. La distribución que mejor ajustó la duración fue una Pearson III ajustada por máxima verosimilitud.



Los puntos corresponden a la probabilidad empírica de las métricas y la línea sólida a la probabilidad teórica. A partir del diagnóstico gráfico se puede ver que las distribuciones teóricas ajustan muy bien a los datos observados. Este mismo proceso se repite para todas las métricas.

Bibliografía

McKee, Thomas B, Nolan J Doesken, and John Kleist. 1993. "The Relationship of Drought Frequency and Duration to Time Scales." Journal Article. *AMS 8th Conference on Applied Climatology*, 179–84. <https://doi.org/citeulike-article-id:10490403>.

Shiau, Jenq-Tzong. 2006. "Fitting Drought Duration and Severity with Two-Dimensional Copulas." Journal Article. *Water Resources Management* 20 (5): 795–815. <https://doi.org/10.1007/s11269-005-9008-9>.

Vicente-Serrano, Sergio, Santiago Beguería, and Juan López-Moreno. 2010. "A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index." Journal Article. *Journal of Climate* 23: 1696–1718.