Metodología para la creación y análisis de índices comparativos del Índice de Precipitación Estandarizado: Cuenca del Río Usumacinta, un caso de estudio

Julio Sergio Santana  
Martín Montero Martínez  
Efraín Mateos Farfán

# Introducción

El índice de precipitación estandarizado (SPI, por sus siglas en inglés, *Standarized Precipitation Index*), ha sido uno de los instrumentos numéricos predilectos para observar de manera sintética lo que ocurre con la precipitación a lo largo de períodos prolongados de tiempo, de tal modo que incluso la Organización Mundial de Meteorología (WMO, por sus siglas en inglés, *World Meteorological Organization*), lo ha adoptado como uno de los indicadores importantes para analizar lo que acontece con el asunto de la precipitación en una ubicación geográfica determinada, generando una guía de usuario para el SPI (WMO, 2012).

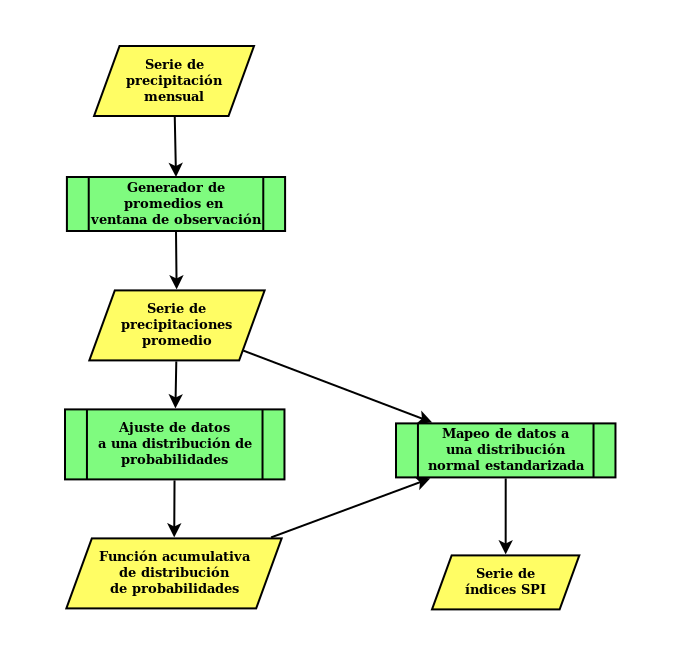
HABLAR DE LA CUENCA

HABLAR DEL ÍNDICE COMPARATIVO

En el presente trabajo se revisa la metodología para generar series de tiempo de índices de precipitación estandarizados (SPIs), tanto de puntos específicos, como de regiones geográficas completas y se introduce un índice para la comparación de los SPIs en dos o más períodos dentro de la serie de tiempo.

# Determinación del SPI

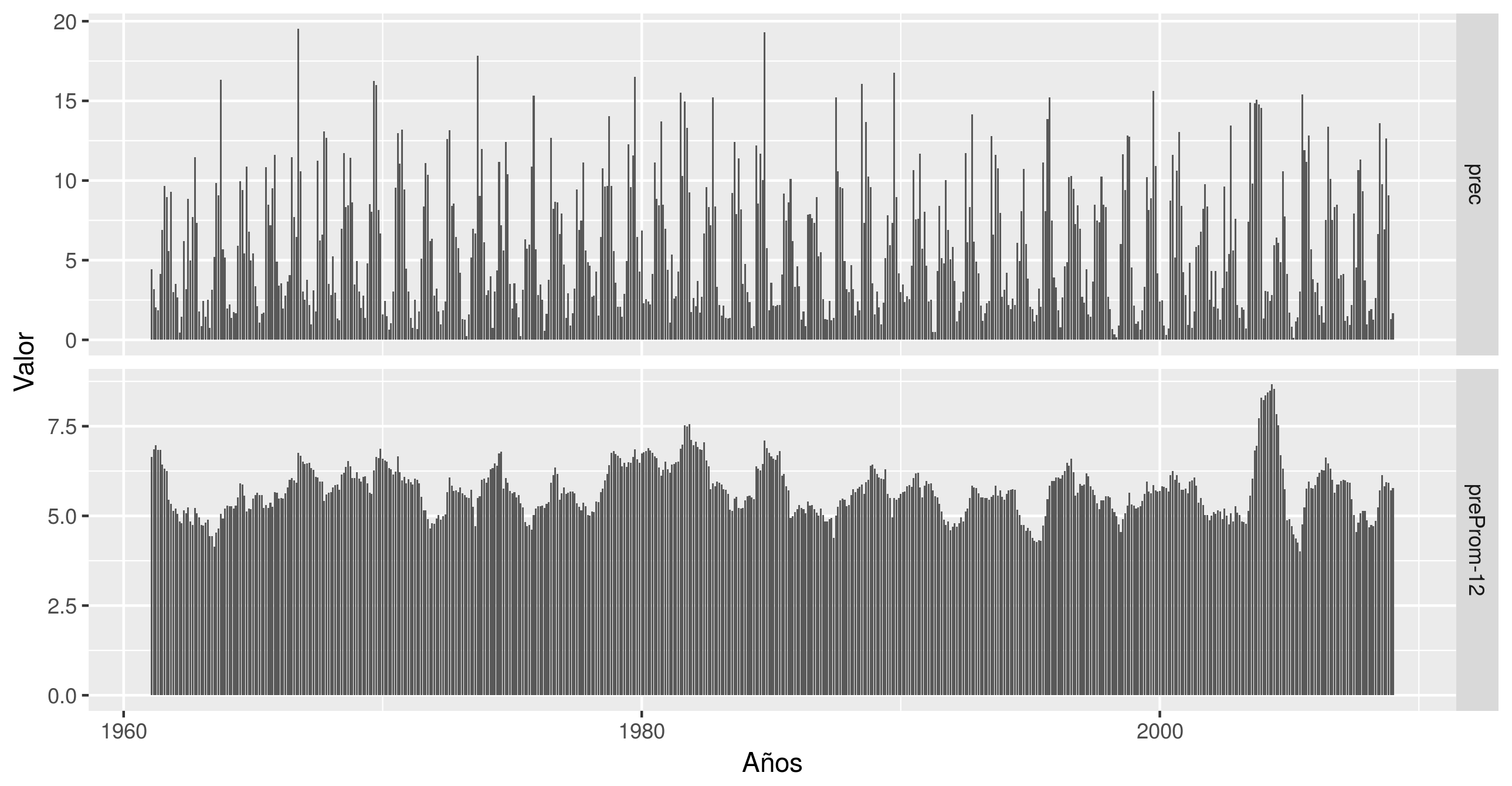
El índice de precipitación estandarizado fue introducido por McKee, Doesken y Kleist (1993). En su artículo mencionan brevemente los pasos requeridos para su generación a partir de una serie de datos de precipitación mensual. Una descripción más detallada de la metodología se puede encontrar en Edwards y McKee (1997), cuya implementación original se desarrolló en los lenguajes de programación C y FORTRAN, y que más adelante se usó como base para la implementación en el lenguaje computacional R por Wheatley (2010). Los pasos de la metodología se pueden describir como tres procesos de transformación de información, mismos que se muestran en la Fig. 1. En las secciones siguientes se describe cada uno de los procesos señalados en la figura.

  
Figura 1: Pasos para generación de serie de tiempo de SPIs. Los rectángulos representan procesos o secuencias de acciones, mientras que los romboides representan entradas o salidas a esos procesos.

## Generación de promedios en ventana de observación

El punto de partida para la generación de índices SPI, es una serie de tiempo de precipitaciones mensuales asociada a una ubicación puntual geográfica específica, o a una región determinada, como pudiera ser una cuenca, un estado, o un municipio. Para cada punto en la serie se *abre* una *ventana de observación* de la serie hacia el pasado, consistente en un número determinado de meses anteriores e incluyendo al mes correspondiente al punto en cuestión y se procede a calcular el promedio de las precipitaciones abarcadas por la ventana. El resultado es una serie de tiempo con los promedios de precipitación calculados. El tamaño de la *ventana* es arbitrario, pero típicamente se usan ventanas de 3, 6, 12, 24 o 48 meses (McKee, Doesken & Kleist, 1993). En la Fig. 2 se muestra, en su parte superior, la serie de precipitaciones medias mensuales en mm/día, promediadas para toda la parte mexicana de la cuenca del Río Usumacinta, y, en su parte inferior, la serie de promedios consecutivos para una *ventana de observación* de 12 meses, calculados por medio de la operación descrita antes en este párrafo. Los datos que se usaron para generar la serie inicial de precipitaciones, provienen de la malla desarrollada por el CICESE, a la que se denominará aquí CLICOMg, que se generó a partir de la base de datos climatológica oficial del Servicio Meteorológico Nacional (SMN) y que comprende en su totalidad más de 5000 estaciones. Para la creación de CLICOMg se aplicaron antes algunas pruebas de calidad de datos y luego, éstos se interpolaron a una malla regular mediante el método *Synographic Mapping System* (Shepard, 1984). La base de datos resultante registra precipitaciones, temperaturas superficiales mínimas y máximas, diarias, con una resolución espacial de 1/8°, para todo México, desde el año 1960 hasta el 2008 (Zhu y Lettenmier, 2007; Muñoz-Arriola et al., 2009). La plataforma fue desarrollada por el CICESE y se puede visualizar en la liga <http://clicom-mex.cicese.mx/malla>.

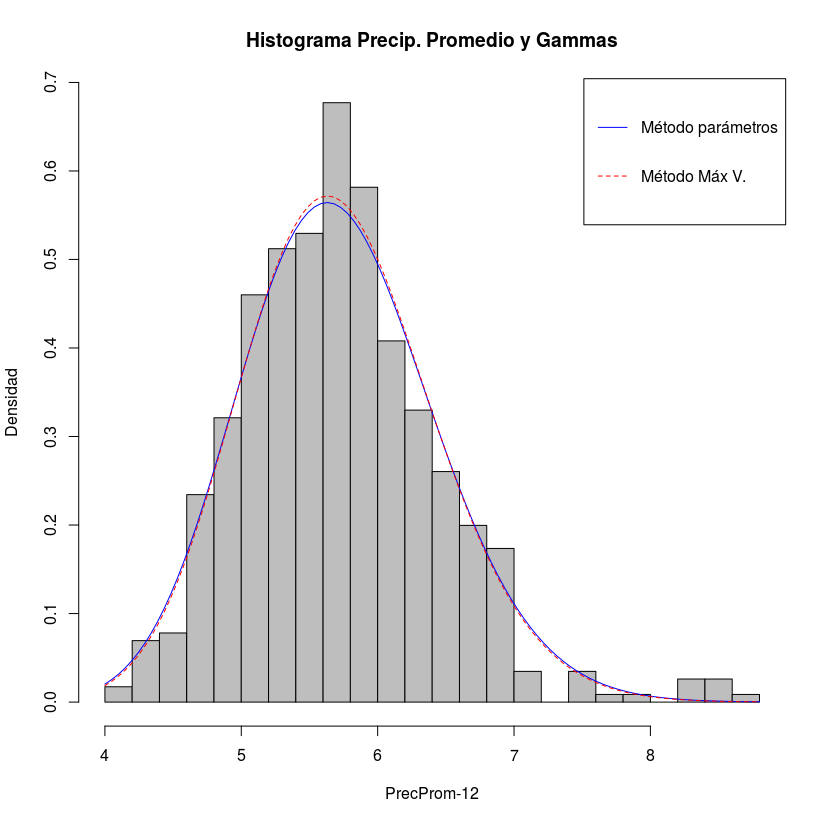
Cabe aclarar que, aunque la base de datos en malla CLICOMg, registra desde el año 1960 hasta el año 2008, las series mostradas en la Fig. 2, van desde el año 1961 hasta el año 2008, ya que los datos del año 1960 se utilizaron solamente para el cálculo de los promedios a partir del primer mes del año 1961.

  
Figura 2: Serie de precipitaciones medias mensuales (mm/día) y promedios a 12 meses, observados en la Cuenca del Río Usumacinta

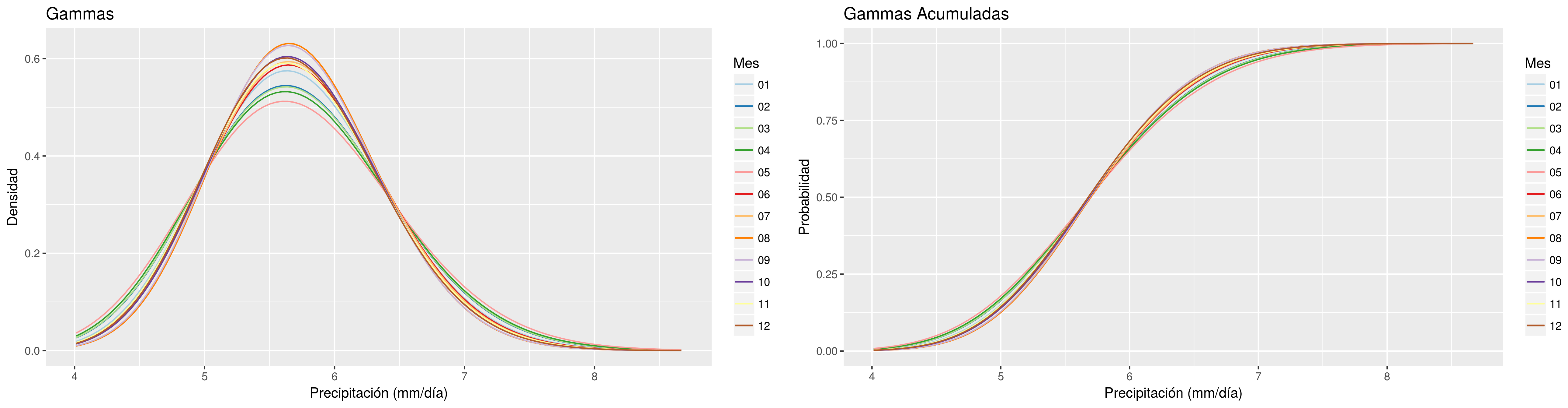
## Ajuste de datos a una distribución de probabilidades

La serie de tiempo de precipitaciones promedio calculada para una ventana de observación determinada, se debe ajustar a una distribución de probabilidades. En el artículo original de McKee, Doesken y Kleist, (1993), se sugiere la distribución de probabilidades Gamma. Aunque el tema de cuál es el tipo de distribución de probabilidades que mejor representa las precipitaciones es todavía un asunto de discusión abierta (Hanson & Vogel, 2012), la selección de la distribución continua Gamma de dos parámetros, manifiesta algunas ventajas representativas, específicamente para el cálculo de los índices SPI, que han sido señaladas en otra parte (Guttman, 1999). De este modo, para ejemplificar aquí el procedimiento se empleará dicha distribución.

El ajuste a una distribución Gamma, consiste en encontrar, a partir de los datos, los dos parámetros, *forma* y *escala*, que producen la curva de densidades o distribuciones más *cercana* a los datos. En Santana, J.S. y Mateos, E., (2014, pp. 105-123), se describen dos métodos para encontrar dichos parámetros: uno de transformación de parámetros, y el otro, conocido como método de la máxima verosimilitud (*maximum likelihood method*), cuya explicación más detallada se puede encontrar en Walpole et al. (2012, pp. 307-312), y en Venables & Ripley (2002). La Fig. 3, muestra el histograma de densidades de la serie de precipitaciones promedio de la Fig. 2, junto con dos curvas de distribución Gamma ajustadas por los dos métodos mencionados. En la figura se puede apreciar que la diferencia entre ambos ajustes es mínima.

  
Figura 3: Histograma y curvas de densidades Gamma para precipitaciones promedios de la Fig. 2.

El ajuste de los datos a una función de densidad de probabilidades, sin embargo, no se hace sobre su totalidad, sino sobre una estratificación de ellos por mes (Guttman, 1999). De esa manera el ajuste en realidad resulta en una familia de doce funciones, una por mes, cuya aplicación a una precipitación particular dependería, además, del mes en el que dicha precipitación se ha *presentado*. La Fig. 4, en su lado izquierdo, muestra el resultado de este procedimiento aplicado sobre los datos de la Fig. 2.

  
Figura 4: Ajuste de los datos de precipitación promediada de la Fig. 2 a doce funciones de densidad y de distribución de probabilidades

Las integrales de las curvas de densidad de probabilidades, mostradas en gráfico de la izquierda de la Fig. 4, corresponden a las curvas acumuladas o funciones de distribución de probabilidades, como las que se muestra en el gráfico de derecha de la misma figura, etiquetado como *Gammas Acumuladas*. Estas funciones de distribución de probabilidades o de densidades acumuladas, son las que se utilizan para el cálculo del SPI. Por ejemplo, en el gráfico de la izquierda de la Fig 5, se ha aislado la curva de distribución de probabilidades Gamma correspondiente al 6o  mes (junio) ajustada para los datos de la Cuenca del Río Usumacinta, y ahí, para una precipitación de 6.3 mm/día, con dicha función se obtiene una probabilidad de 0.797; esto es, P(X ≤ 6.3) = 0.797.

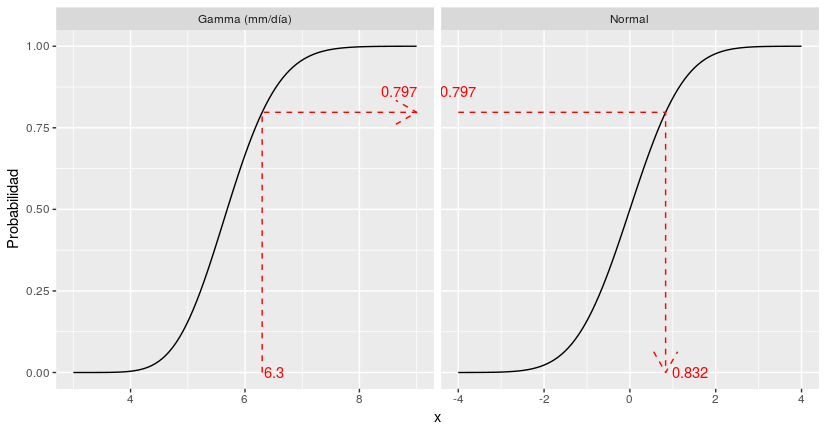
## Mapeo de datos a una distribución normal estandarizada

Para mapear una precipitación, *x*, ocurrida en el mes *m,* registrada por una de las curvas de distribución Gamma, pertenecientes a la familia de funciones ***G*** y representada por la función ***G***(), a un valor determinado por una función de distribución normal estandarizada, *N*(), con media 0 y desviación estándar 1, se compone la función ***G***(), con la inversa de la función de distribución normal estandarizada, también conocida como la función *cuantil* de dicha distribución, *QN*(), de la siguiente manera:

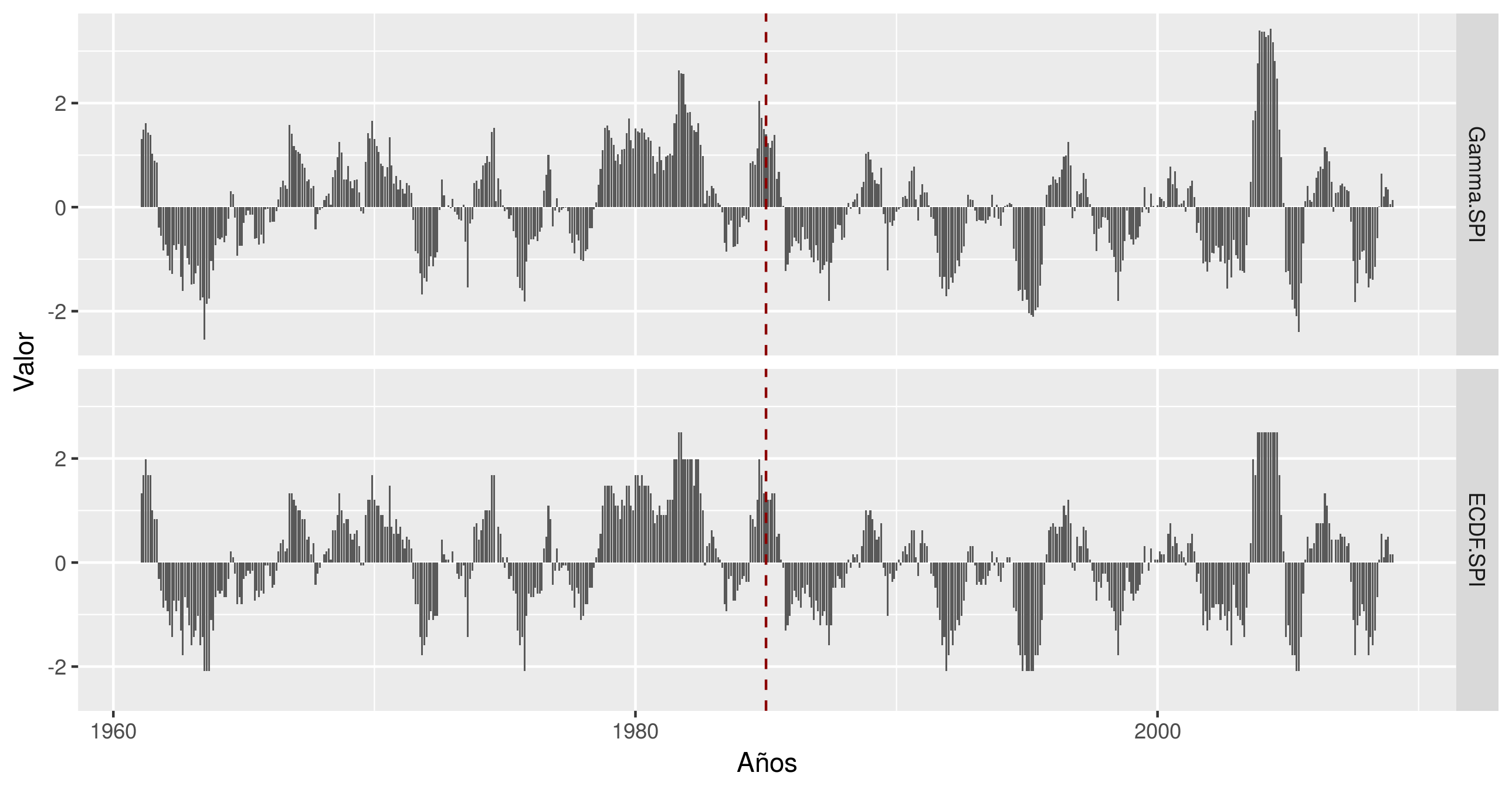


donde *x* es la precipitación, *m* es el mes en el que ocurrió dicha precipitación, *y* la composición de las funciones ***G***seguida de *QN* aplicada a dicha precipitación y mes,es ***spi****(x,m)*, que es el *índice de precipitación estandarizado correspondiente* a la precipitación *x* ocurrida en el mes *m.*

La Fig. 4 ilustra la mecánica de la operación anterior para el caso de la cuenca analizada, de tal manera que si se entra con una precipitación de 6.3 mm/día, correspondiente al 6o mes, la función Gamma arrojaría una probabilidad de 0.797, misma que se usa como entrada para la función cuantil o inversa de la función Normal, para entregar a la salida un valor del índice de precipitación estandarizado de 0.832.

Figura 5: Mapeo entre dos distribuciones de probabilidad. La distribución Gamma es la ajustada para las precipitaciones ocurridas en el 6o mes de la serie de precipitaciones promediadas de la Fig. 2.

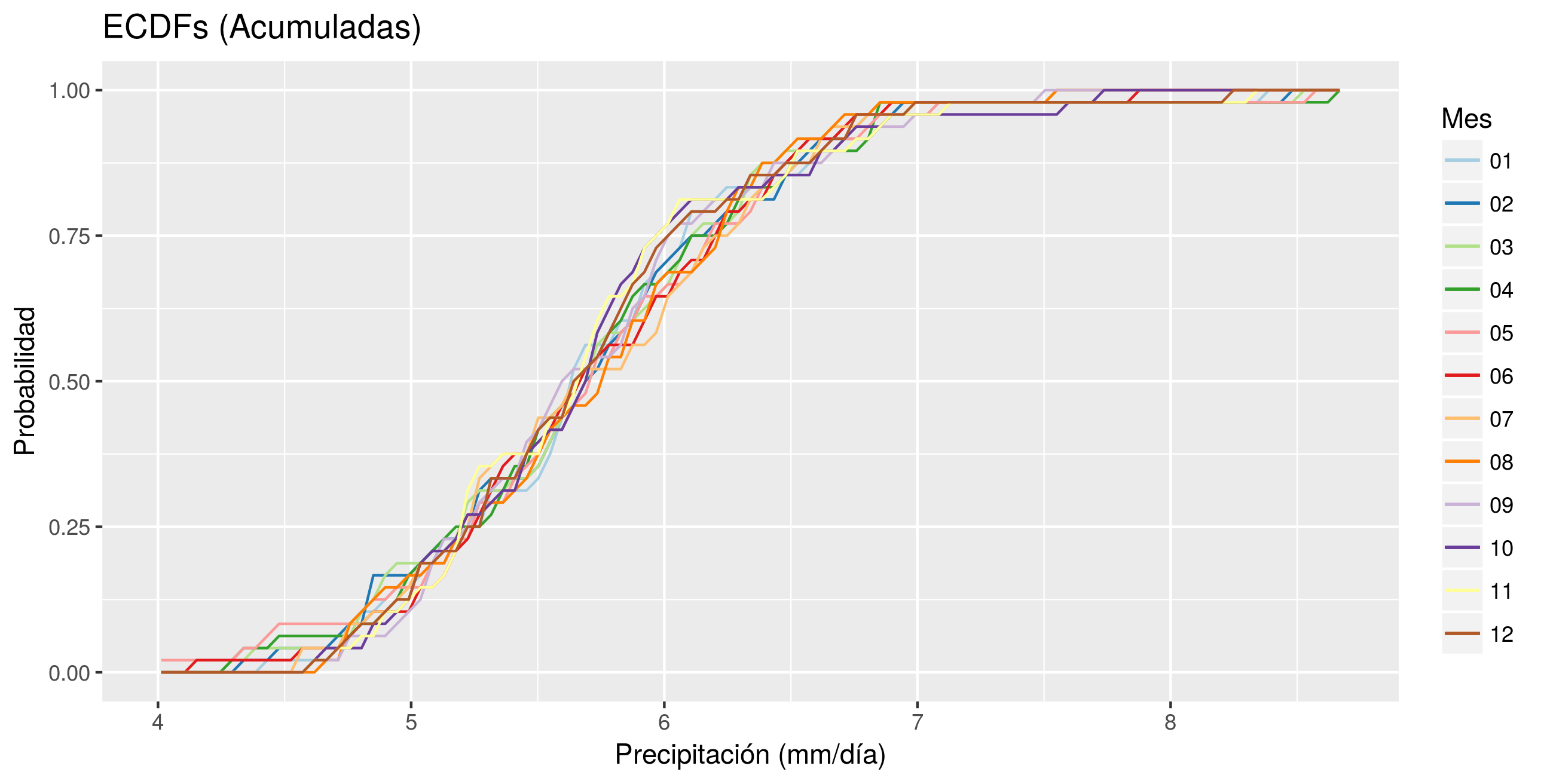
De esta manera, aplicando la función ***spi***() a cada uno de los elementos de la serie de tiempo completa de promedios de precipitaciones de la cuenca del Río Usumacinta, mostrada en la Fig. 2, se le puede transformar en una serie de tiempo de índices SPI, tal como la que se muestra en el gráfico superior de la Fig. 6.

  
Figura 6: Serie de índices SPI-12 correspondiente a la serie de promedios de la cuenca del Río Usumacinta, mostrada en la Fig. 2. En el gráfico superior, los índices se calcularon ajustando los datos a funciones de distribución Gamma; en el inferior, a funciones ECDF.

## El uso de otras funciones de distribución

Aunque la descripción original del método para el cálculo de los índices SPI señala que el ajuste de los datos de precipitación promediada se hace a una función de distribución de probabilidades Gamma (McKee, Doesken y Kleist, 1993), desde sus primeras implementaciones (Edwards & McKee, 1997), se hizo uso alternativo de otras funciones de distribución, en particular de las funciones no paramétricas conocidas como funciones acumulativas de distribución empíricas (ECDF, por sus siglas del inglés: *empirical cumulative distribution functions*). Si bien, para el efecto pudieran utilizarse cualesquiera de las funciones revisadas en otros trabajos (Hanson & Vogel, 2012, Guttman, 1999), las funciones ECDFs resultan particularmente interesantes para el presente trabajo.

Una ECDF toma una serie de datos aleatorios, y le asigna a cada uno de ellos una probabilidad de 1/*n*, donde *n* es el número de datos en la serie, ordena ascendentemente los datos y construye, a partir de ellos, una función tipo escalera en la que se cumple ***F***(*x*) = P(X ≤ *x*) , (Vaart, 1998 pp. 265-277; Gentle, 2009 pp. 59-65). Si en vez de ajustar los datos de precipitación de la cuenca del Río Usumacinta, mostrados en la Fig. 2, a una función de distribución Gamma, se hubieran empleado funciones ECDF, siguiendo el procedimiento descrito previamente, el resultado hubiera sido la familia de funciones mostrada en la Fig. 7. En esta figura, a diferencia del gráfico de la derecha de la Fig. 4, se muestran discontinuidades en cada punto donde se registra una precipitación diferente de los datos, lo que les da a las funciones una apariencia escalonada. El procedimiento, a partir de esta familia de funciones, para mapear las precipitaciones a una distribución normal estandarizada es semejante al descrito en la sección anterior. La serie de SPIs resultante de utilizar funciones ECDF en vez de funciones de distribución Gamma, se muestra en el gráfico inferior de la Fig. 6.

  
Figura 7: Ajuste de los datos de precipitación promediada de la Fig. 2 a doce funciones de acumulativas distribución empíricas (ECDFs)

Es importante destacar que la implementación en el lenguaje R, desarrollada por Wheatley (2010), deriva de la implementación alternativa de Edwards y McKee (1977) que toma como base las funciones ECDF, y que la implementación desarrollada para el presente trabajo, a su vez, se fundamenta en esas dos. Lo anterior obedece a la ventaja que tienen estas funciones al trabajar directamente con los datos, en vez de hacerlo con una aproximación de ellos. Una desventaja del uso de las funciones ECDF es que, en ellas, todas las precipitaciones máximas arrojan una probabilidad de 1 y, por consiguiente, éstos valores al ser mapeados mediante la inversa de la función normal estandarizada a sus correspondientes valores de SPI, dan un valor *infinito*. Para lidiar con este problema, se recurre a un procedimiento de optimización que reajusta, para esos valores de precipitación, valores de SPI que mantienen la distribución normal estandarizada, esto es, con media 0 y desviación estándar 1 (Wheatley, 2010).

# Interpretaciones del SPI

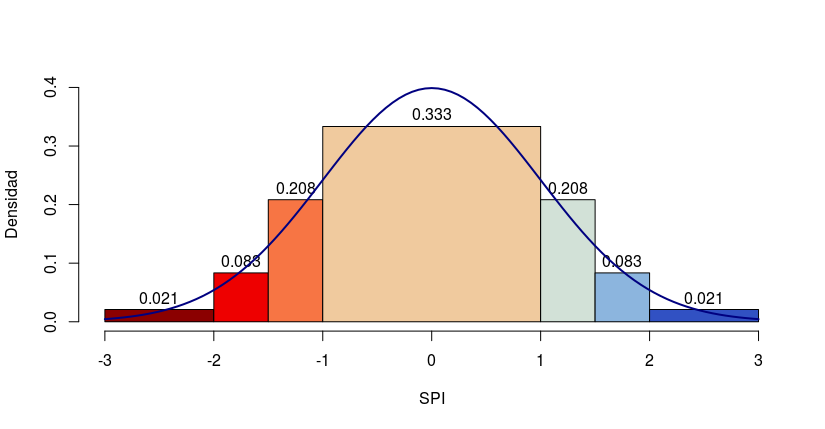
Dadas las características de la construcción del SPI, que lo hacen desembocar en una distribución normal estandarizada, típicamente sus valores variarán de manera aproximada entre -3 y 3, de tal manera que, para el período total considerado, un valor de cero para este índice estaría indicando una precipitación alrededor de la *mediana*, un valor negativo indicaría un *déficit* de precipitación y un valor positivo indicaría un *superávit*, con referencia a todo el período y a la región o punto geográfico considerados. La Organización Mundial de Meteorología (OMM) ha definido rangos de valores de los SPI para permitir la caracterización cualitativa de las precipitaciones (WMO, 2012). En el Tabla 1, se muestran los rangos definidos por la OMM.

Tabla 1. Valores y probabilidades de ocurrencia de los índices normalizados de precipitación.

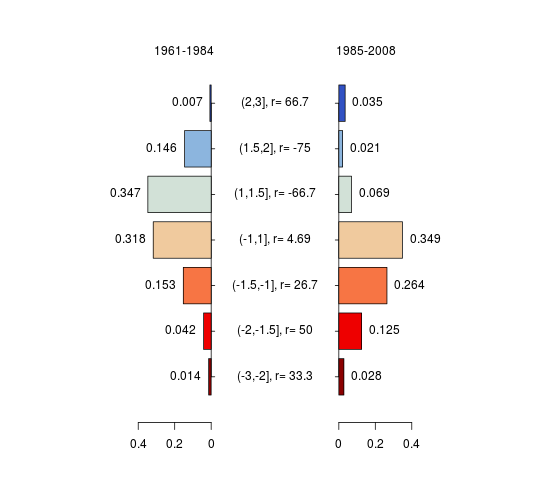
|  |  |  |
| --- | --- | --- |
| SPI | Categoría | Probabilidad |
| ≥ 2.0 | extremadamente húmedo (eh) | 0.023 |
| 1.5 a 1.99 | severamente húmedo (sh) | 0.044 |
| 1.0 a 1.49 | moderadamente húmedo (mh) | 0.092 |
| -0.99 a 0.99 | aproximadamente normal (an) | 0.682 |
| -1.0 a -1.49 | moderadamente seco (ms) | 0.092 |
| -1.5 a -1.99 | severamente seco (ss) | 0.044 |
| ≤ -2 | extremadamente seco (es) | 0.023 |

## Partición de la serie de índices SPI

Si se utilizan los rangos definidos por la OMM, mostrados en la Tabla 1, para la construcción de un histograma de densidades de probabilidad a partir de los índices SPI del gráfico inferior de la Fig. 6, se obtiene un gráfico semejante al mostrado en la Fig. 8. Debido a la forma de construir la serie de SPIs, el histograma resultante es simétrico y embona perfectamente con la curva descrita por la función normal estandarizada de densidad de probabilidades, que se muestra también superpuesta en la figura. Esto manifiesta que, lo que se ha hecho mediante la obtención de los índices SPI es clasificar los distintos promedios de precipitación ocurridos a lo largo de un período determinado, por encima o por debajo de uno de estos promedios que es la mediana de ellos. Esto, sin embargo, no tiene que ver con la distribución de los promedios a lo largo de la serie de tiempo. Por ejemplo, en el gráfico inferior de la Fig. 6, la línea punteada divide el período de la serie a la mitad, partiendolo en dos a los que se llamará subperíodos, y se puede apreciar de manera sensible que, la mayoría de los valores negativos del SPI se encuentran en el segundo subperíodo, esto es, en el más reciente, mientras que, por el contrario, la mayoría de los valores positivos se encuentran en el primero.

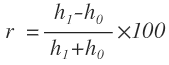
  
Figura 8: Histograma de densidad de probabilidades para la serie SPI mostrada en el gráfico inferior de la Fig. 6, con una curva de densidad normal estandarizada superpuesta.

De la observación anterior, surge la idea de construir dos histogramas: uno para cada subperíodo. Estos histogramas se muestran en la Fig. 9, a manera de espejo para hacer comparaciones entre ellos.

  
Figura 9: Histogramas comparativos de los dos subperíodos. El ancho de las barras no se muestra gráficamente; en vez de ello se han indicado los intervalos y la densidad de probabilidades se muestra en la cabeza de cada barra. Se muestra el índice comparativo, r, para cada intervalo o barras apareadas.

## Índice comparativo del SPI

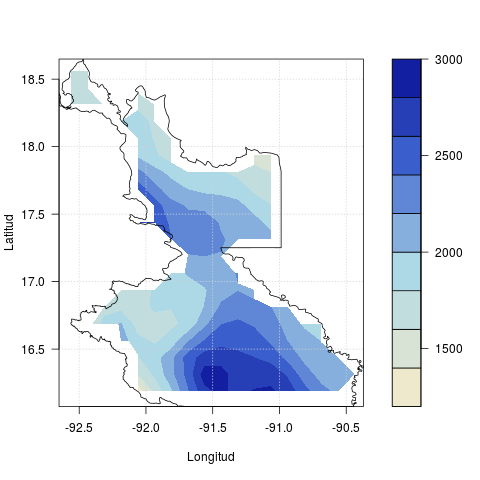
Nótese que, en la Fig. 9, al promediar la altura, o densidad, de cada dos barras apareadas por el mismo intervalo, el resultado coincide con la densidad o altura, para ese intervalo en el histograma de todo el período, presentado en la Fig. 8. Por ejemplo, para el intervalo (1, 1.5], el promedio de 0.347 y 0.069 es: (0.347+0.069)/2 = 0.208, que es la altura de la barra correspondiente a ese intervalo en la Fig. 8. Esta propiedad, derivada de la forma de construir los histogramas, indica que para cada par de histogramas de subperíodos provenientes de un período total, independientemente de la distribución que se presente en el histograma particular, la suma de las alturas de las dos barras correspondientes a un intervalo, siempre será la misma. De aquí se puede definir un índice comparativo entre los dos subperíodos que forman el período, de la siguiente manera:



donde, *r* es el índice comparativo, en porcentaje, *h*0 es la *altura* de la barra del primer subperíodo y *h*1 es la *altura* de la barra del segundo subperíodo. Las etiquetas centrales de la Fig. 9 muestran, además del intervalo, los valores calculados para el índice *r* correspondiente al intervalo. Claramente, el índice *r* varía entre -100 y 100. Un valor negativo de *r* indica la *preponderancia* del primer subperíodo sobre el segundo, un valor positivo indica que el segundo subperíodo es *preponderante* sobre el primero y un valor cero, indica un equilibrio o igualdad en ambos subperíodos. Por ejemplo, para el intervalo (1, 1.5], *r* = - 66.5, lo que indica que el cambio en los SPIs categorizados como *moderadamente húmedos* fue de un 66.5% hacia el primer subperíodo, y para el intervalo (-2, -1.5], *r* = 50.0, lo que indica que el cambio en los SPIs categorizados como *severamente secos*, fue de un 50% hacia el segundo subperíodo. Es decir, las sequías severas ocurrieron *más* en el segundo subperíodo.

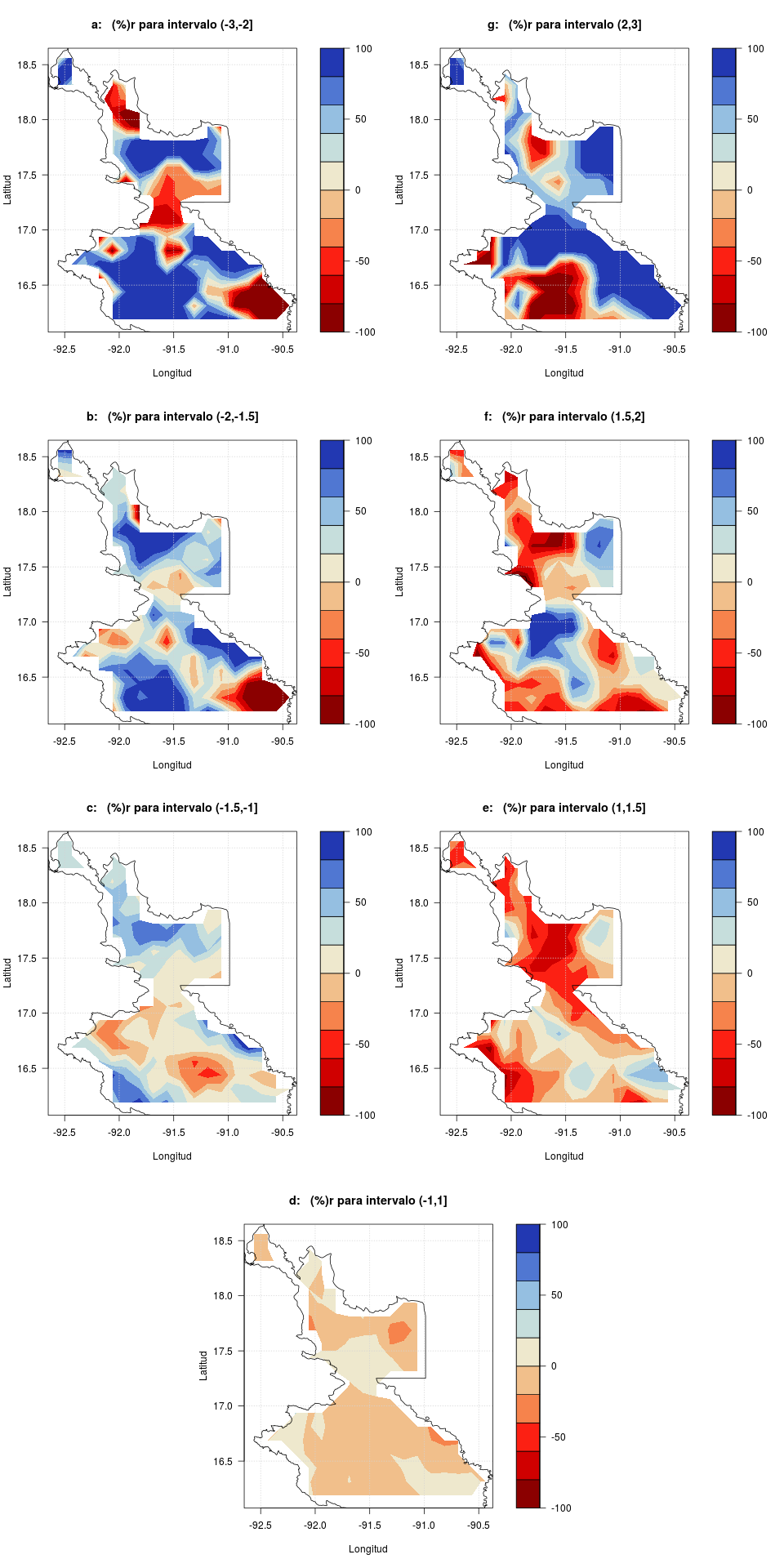
# Comparaciones geográficas de los índices SPI

Una de las ventajas del índice comparativo *r* propuesto, es que permite visualizar sobre la descripción geográfica de alguna región, el comportamiento de los índices SPI asociados a puntos definidos en ella. Para revisar este asunto, se presenta primeramente el patrón de precipitaciones promedio de la cuenca del Río Usumacinta en la Fig. 10.

  
Figura 10: Patrón anual de precipitación de la cuenca del Usumacinta (mm)

La estrategia para la determinación de un índice comparativo *visible* geográficamente, parte del hecho de que se tiene un conjunto de puntos en malla en el interior de la cuenca, cada uno de los cuales tiene asociada una serie de tiempo de precipitaciones para el período considerado. Si bien, la series, tanto de precipitaciones como de promedios hacia atrás a 12 meses mostradas en la Fig. 2 corresponden a la totalidad de la cuenca, para cada uno de los puntos de malla de la cuenca se dispone y se pueden calcular series semejantes. A partir de la serie de promedios de precipitación se desarrolla todo el proceso descrito anteriormente, pero ahora para cada uno de los puntos de malla de la cuenca, hasta llegar a la obtención de los índices comparativos *r*, para cada uno de los intervalos de histograma asociados con las condiciones o categorías climáticas mostradas en la Tabla 1. Es decir, cada punto de malla contará con siete índices comparativos *r;* uno para cada intervalo. A partir de esta información, se pueden construir siete mapas, uno para cada intervalo, en los que se muestren las variaciones del índice comparativo *r*, punto por punto.

La Fig. 11 muestra los siete mapas del índice comparativo *r* de índices SPI, con una ventana de observación de 12 meses, para la cuenca del Río Usumacinta, y para el período considerado que comprende los años de 1961 a 2008.

  
Figura 11: Mapas del índice comparativo r para los siete intervalos

# Referencias

* Edwards, D. C. and McKee T. B.,1997. Characteristics of 20th century drought in the United States at multiple time scales. Climatology report 97-2. Atmospheric science paper No. 634. Colorado State University, Department of Atmospheric Science. URI: http://hdl.handle.net/10217/170176. May, 1997. Implementation notes and code of this work can be found in URI: <http://ccc.atmos.colostate.edu/standardizedprecipitation.php>.
* Gentle, J.E., 2009. *Computational Statistics*. Statistics and Computing Series, Springer-Verlag, New York, April 2009. ISBN 978-0-387-98143-7, DOI 10.1007/978-0-387-98144-4.
* Guttman, N. B., 1999. *Accepting the Standardized Precipitation Index: a calculation algorithm*. JAWRA Journal of The American Water Resources Association 35(2), 311-322. April, 1999.
* Hanson, L.S., Vogel, R. The Probability Distribution of Daily Rainfall in the United States. *World Environmental and Water Resources Congress 2008*. American Socity of Civil Engineers. Published online: April 26, 2012. DOI: 10.1061/40976(316)585.
* McKee, T. B., Doesken, N. J., & Kleist, J. (1993, January). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology* (Vol. 17, No. 22, pp. 179-183). Boston, MA: American Meteorological Society.
* Muñoz-Arriola, F., R. Avissar, C. Zhu, D. P. Lettenmaier, (2009). Sensitivity of the water resources of Rio Yaqui Basin, Mexico, to agriculture extensification under multiscale climate conditions. *Wat. Res. Res.* Vol. 45 (11), DOI: 10.1029/2007WR006783.
* Santana, J.S., Mateos-Farfán, E. *El arte de programar en R: Un lenguaje para la estadística*. Instituto Mexicano de Tecnología del Agua. 1a Edición digital, 2014. ISBN: 978-607-9368-15-9. Disponible en: <https://www.imta.gob.mx/biblioteca/libros_html/el-arte-de-programar-en-r/>.
* Shepard, D. S., (1984). Computer mapping: The SYMAP interpolation algorithm. Spatial Statistics and Models, G. L. Gaile and C. J. Willmott, Eds., D. Reidel, 133–145.
* Vaart A. W. van der, 1998. *Asymptotic statistics*. Cambridge University Press, ISBN 0-521-78450-6. The Edinburgh Building, Cambridge CB2 2RU, UK, 1998.
* Venables, W.N., Ripley, B.D. *Modern Applied Statistics with S*. Springer, New York, NY, 4th edition, 2002. URL <http://www.stats.ox.ac.uk/pub/MASS4>. ISBN 0-387-95457-0.
* Walpole, R.E., Myers, R.H., Myers, S.H., Ye, K. *Probability & Statistics ofr Engineers & Scientists*. Prentice Hall / Pearson Education, Inc., Boston, MA, 9th edition, 2012. ISBN 10: 0-321-62911-6.
* Wheatley J., 2010. Standardized Precipitation Index Code. Descripción y código disponible en URI: https://joewheatley.net/2010/03/.
* World Meteorological Organization. *Standarized Precipitation Index: User Guide*. WMO-No. 1090. World Meteorological Organization, 2012.
* Zhu, Chunmei, Dennis P. Lettenmaier, 2007: Long-Term Climate and Derived Surface Hydrology and Energy Flux Data for Mexico: 1925–2004. *J. Climate*, 20, 1936–1946.