



“MANUAL BREVE DE PREPARACION DE DATOS ”

Informe Intermedio

Organizado por:
Centro del Agua para Zonas Áridas y Semiáridas de ALC – CAZALAC (Chile)
En el marco del proyecto CAZALAC-IWR: Atlas de Sequías mediante Análisis Regional de Frecuencias en áreas piloto de Argentina, Chile y Perú

Introducción

La disponibilidad de datos confiables es un requisito indispensable para cualquier tipo de análisis hidrológico, en particular, aquel relacionado con determinación de probabilidad ante eventos extremos, en la que el error o presencia o ausencia de un dato en particular, puede tener fuertes consecuencias sobre las estimaciones de probabilidad con altos periodos de retorno. Por lo mismo, los esfuerzos destinados a la recopilación y análisis de los datos, así como su corrección en la etapa previa a su incorporación en los procedimientos de análisis de frecuencias, se justifican plenamente al alcanzar estimaciones más razonables de los parámetros a utilizar como de las estimaciones y productos finales.

El propósito de este Manual Breve, en el contexto de un Análisis Regional de Frecuencias para la elaboración de mapas probabilísticas de sequía, es presentar al analista una guía de las etapas involucradas en el proceso de preparación de los datos, algunos test recomendados en el análisis de cada una de las etapas, así como referencias bibliográficas y acceso a herramientas informáticas que uniformen criterios y procedimientos que faciliten el procesamiento de la información.

No se presenta el detalle ni ejemplos de cálculo de los procedimientos estadísticos, en el supuesto que el usuario final deberá tener la preparación suficiente y un manejo adecuado de los conceptos de análisis estadístico de datos hidrológicos. El objetivo de este Manual es la unificación de criterios, en cuanto a la secuencia de etapas analíticas, así como de las técnicas estadísticas utilizadas.

La organización de la información en términos de un capítulo 3, y subtítulos correspondientes, se justifica al ser el presente documento un componente de un informe más amplio destinado a facilitar el procedimiento de elaboración del Atlas de Sequía.

3

Preparación de los datos

Se debe desarrollar un gran esfuerzo en la revisión y control de calidad de los datos a ser utilizados en el análisis regional de frecuencias. Las justificaciones de este proceder están de sobra y pueden ser revisadas por el lector en los capítulos II (Análisis exploratorio de datos) y III (Estimación y completación de datos de precipitación faltantes), ambos contenidos en la “Guía metodológica para la elaboración del mapa de zonas áridas, semiáridas y subhúmedas secas de América Latina y el Caribe”.

Sin perjuicio de lo anterior, los principales objetivos de esta etapa, previa al análisis regional de frecuencias son (UNESCO, 2006):

1. Maximizar nuestro entendimiento de los datos
2. Describir la estructura subyacente que ellos tienen (normalidad, multimodalidad, asimetría, curtosis, linealidad, homogeneidad entre grupos, homocedasticidad, etc.)
3. Detectar las variables importantes
4. Detectar valores atípicos y anomalías
5. Comprobar los supuestos implícitos en los análisis
6. Determinar los ajustes óptimos de los factores
7. Proporcionar al analista información relevante en cuanto al modelo que mejor ajusta los datos, una lista de valores atípicos, sentido de certeza en las conclusiones, estimaciones de los parámetros, grado de incertidumbre sobre las estimaciones, lista de las variables importantes, conclusiones sobre las significancias estadísticas de variables individuales.

3.1. Preparación de los datos para el análisis regional de frecuencias

La etapa de revisión de la calidad de los datos, de acuerdo a los estudios de Análisis Regional de Frecuencias, tiene como finalidad eliminar falsos valores, asociados a una amplia variedad de mediciones de datos, su registro, presencia de errores de transcripción y registros incompletos durante algunos años (Schaefer *et al*, 2006, Schaefer *et al*, 2007). Los registros para todos los sitios y años calendario deben ser revisados completamente. Además, todos los registros deben ser revisados para detectar montos de precipitación anómalas, tanto altas como bajas. En este caso, es posible utilizar la medida de discordancia de Hosking y Wallis, para identificar aquellas estaciones cuyas estadísticas muestrales difieren notablemente del resto de las estaciones ubicadas en una misma región.

Las estaciones sospechosas, así como sus datos, deben ser revisados a fin de verificar la validez de los datos. También se puede utilizar información de estaciones cercanas, para corroborar la magnitud y momento en que ocurren los datos anómalos, tanto en los valores altos como en los valores bajos de la variable en estudio. Entonces, aquellos datos que son claramente erróneos son eliminados de los registros.

Esta etapa del proceso de preparación de los datos puede ser abordada y se encuentra documentada en el Capítulo II. Análisis Exploratorio de Datos (AED) según UNESCO (2006) e incluye técnicas como:

1. Gráficos de datos crudos
2. Gráficos de estadísticos simples
3. Posicionar los gráficos anteriores
4. Gráficos de posicionamiento de los datos

Se recomienda al analista que siga las pautas desarrolladas en la referencia bibliográfica señalada. Como complemento, puede el lector recurrir a la revisión de otra fuente de información, sin que esto signifique apartarse o contradecir la recomendación previa. En particular, se insta al analista a revisar el material sobre métodos estadísticos en recursos hídricos de Helsel y Hirsch (2002), el cual se encuentra disponible en internet en la siguiente ubicación:

<http://pubs.usgs.gov/twri/twri4a3/>

3.2. Análisis de consistencia

Tal como lo señala UNESCO-ROSTLAC (1982) y UNESCO (2006), “el primer paso que debe realizarse al efectuar la evaluación espacial de la precipitación es verificar que el periodo de la estadística pluviométrica que se va a analizar es consistente, o sea, que la estación haya sido observada durante dicho periodo, de la misma forma, con el mismo criterio y que su instalación no haya sufrido variaciones de ningún tipo. Para esta verificación se recomienda el método de las curvas doble acumuladas, o de doble acumulación (CDA) o método de doble masa (MDM).

Para una descripción detallada del procedimiento, el analista debe recurrir a alguna de las siguientes fuentes bibliográficas disponibles en internet:

UNESCO-ROSTLAC. 1982. Guía metodológica para la elaboración del balance hídrico de América del Sur.

<http://www.unesco.org/phi/bibli.fr.html>

<http://www.unesco.org/phi/libros/guiametodolbalhidAS.pdf>

UNESCO. 2006. Guía metodológica para la elaboración del mapa de zonas áridas, semiáridas y subhúmedas secas de América Latina y El Caribe. CAZALAC. Documentos Técnicos del PHI_LAC, N°3

[www.cazalac.org/documentos/Guia Mapas ZA ALC.pdf](http://www.cazalac.org/documentos/Guia_Mapas_ZA_ALC.pdf)

Para facilitar el análisis de consistencia, por el método de doble masa, se encuentra disponible en internet la herramienta informática HIDROBAS v. 3.0, desarrollada por el Instituto Geológico y Minero de España. La herramienta, entre sus funciones, permite el análisis de doble masa para un total simultáneo de 9 estaciones. Se encuentra disponible, así como una guía de usuario de la misma, en forma gratuita en la siguiente ubicación (Ortiz *et al*, 2001) :

<http://aguas.igme.es/igme/aplicaciones/Hidrobas.htm>

[http://aguas.igme.es/igme/publica/sim_hidro Murcia/tomo%20XXIII/27.pdf](http://aguas.igme.es/igme/publica/sim_hidro_Murcia/tomo%20XXIII/27.pdf)

3.3. Análisis de estacionariedad e independencia serial

Dos supuestos básicos inherentes al análisis de frecuencias son la estacionariedad en el periodo de observación de los datos, y que los datos en una estación específica son serialmente independientes (autocorrelación). Solamente cumpliendo los requisitos de estacionariedad, consistencia y homogeneidad es posible efectuar análisis de frecuencias para la simulación de procesos hidrológicos (Dahmen y Hall, 1990). Como parte de la etapa de preparación de los datos, se deben aplicar procedimientos estándar para evaluar la estacionariedad y la autocorrelación.

En estudios de Análisis Regional de Frecuencias, desarrollados en Estados Unidos (Schaefer *et al*, 2006; Schaefer *et al*, 2007), no se detallan en profundidad los procedimientos estadísticos para abordar la verificación de estos criterios. Sin embargo, con base a algunos antecedentes recabados en estos estudios y referencias bibliográficas anexas, se proponen al analista las alternativas de abordaje estadístico que se presentan a continuación.

3.3.1. Análisis de estacionariedad:

El criterio de estacionariedad implica que los datos deben estar libres de tendencia durante el periodo de observación. Una serie de tiempo de datos hidrológicos, se dice estrictamente estacionaria, solo si sus propiedades estadísticas (media, varianza y momentos de orden superior) no son afectados por la elección del momento de origen de la serie de datos (Dahmen y Hall, 1990). Para abordar este criterio el analista puede utilizar alguna de las siguientes alternativas, las cuales, presentadas en orden de prioridad en cuanto a su utilización, en función de sus propiedades estadísticas, se aplican a una sola serie de datos a la vez:

3.3.1.1. Test de Mann-Kendall

El test de Mann-Kendall es un test no paramétrico, estadísticamente basado en el ranking (o número de orden) de dos variables y puede ser utilizado para tendencias en series de tiempo de datos hidrológicos si una de esas variables es el tiempo y tiene la particularidad de presentar una mayor potencia estadística en la medida que la serie de datos presenta un mayor sesgo que la aleje de una distribución normal (USGS, 2004; USGS, 2006).

El cálculo del estadístico de prueba del test de Mann-Kendall no es difícil de implementar en una planilla de cálculo. Se da al analista, sin embargo, la posibilidad de utilizar dos paquetes de cómputo disponibles para tal efecto, ambos de distribución libre. La ubicación de las herramientas de cálculo, con sus respectivos respaldos de información base para su desarrollo son las siguientes:

1. Planilla Excel: MAKESENS – For detecting and estimating trends
FINNISH METEOROLOGICAL INSTITUTE
http://www.fmi.fi/organization/contacts_25.html
2. Helsel, Mueller y Slack. 2006. Computer Program for the Kendall Family of Trend Tests. USGS Scientific Investigations Report 2005-5275
<http://pubs.usgs.gov/sir/2005/5275/>

3.2.1.2. Test de Correlación de Rango de Spearman

El test es simple y está bien documentado en Dahmen y Hall (1990). Su procedimiento consiste en:

1. Ingresar la serie de datos de precipitación en par con el correspondiente año de medición
2. Ordenar los datos de precipitación de menor a mayor
3. Calcular los valores de K_x (número de orden del año) y K_y (número de orden del valor de la precipitación que le corresponde en los datos no ordenados)
4. Obtener el coeficiente de correlación de Spearman
5. Obtener el valor del estadístico de prueba a evaluar frente a la distribución t-Student
6. Probar la hipótesis nula (no existe tendencia) frente a la hipótesis alternativa (existe tendencia), con un 5% de significancia
7. En caso de existir tendencia en la serie de datos, esta no puede ser utilizada en el análisis de frecuencia

3.2.1.3. Regresión lineal:

El procedimiento es simple y es el utilizado en estudios de análisis regional de frecuencias y está documentado en textos de análisis estadístico de recursos hídricos y

básicamente consiste en (Wallis *et al*, 2007; Schaefer *et al*, 2007; Schaefer *et al*, 2006, Helsel y Hirsch, 2002):

1. Reescalar los datos de cada estación, dividiendo cada uno de los valores de la estación por la media de dicha estación.
2. Hacer una regresión lineal entre cada valor (x_i) y el año de ocurrencia menos 1900
3. Efectuar un test de significancia estadística sobre el valor de la pendiente de la recta. Si la hipótesis nula no puede ser rechazada con un 5% de significancia, entonces los datos se consideran estacionarios.
4. En caso de existir tendencia en la serie de datos, esta no puede ser utilizada en el análisis de frecuencia

3.3.2. Análisis de correlación serial (autocorrelación):

Este criterio se refiere a la independencia de los datos de la serie de tiempo (Dahmen y Hall, 1990). Para probar el criterio de independencia serial, es posible utilizar las siguientes alternativas:

3.3.2.1. Coeficiente de correlación lag 1:

Este procedimiento busca medir la correlación que existe entre un valor de la serie temporal medido en el tiempo t y otro valor medido en el tiempo $t-1$. (De ahí deriva su nombre). El test es simple y está bien documentado en Dahmen y Hall (1990). Es posible encontrar un test de autorrelación basa en el coeficiente de correlación lag 1 en el paquete estadístico DATAPLOT.

1. Calcule el coeficiente de correlación serial lag 1 (r_1)
2. Determine los límites de confianza superior e inferior para r_1
3. Acepte o rechace la hipótesis nula ($r_1=0$) basado en si el valor de r_1 quedó dentro de los límites de confianza para un nivel de significancia de 5%

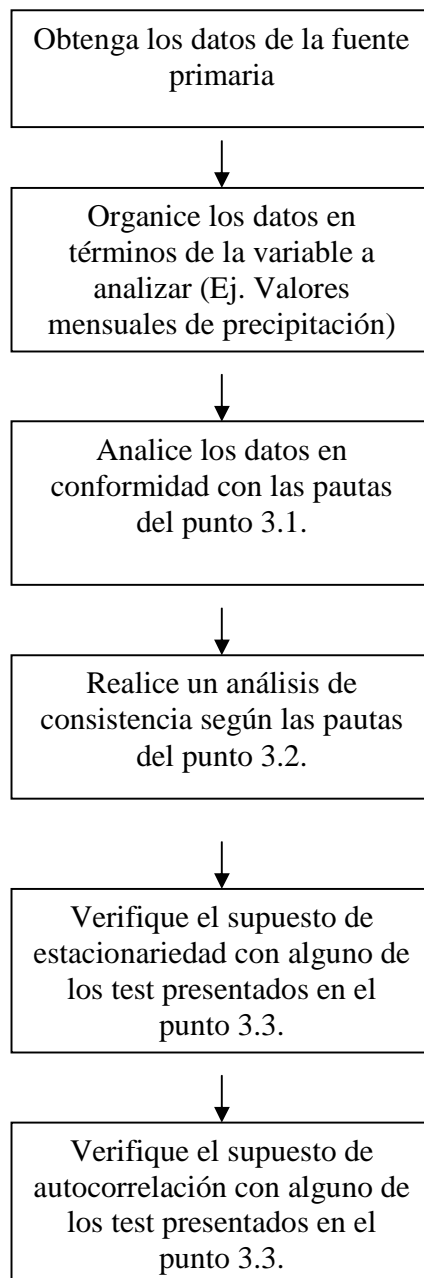
3.3.2.2. Test de Durbin-Watson:

El test de Durban Watson es un test estándar para evaluar la dependencia de los residuales. Se encuentra muy relacionado con lo que podemos denominar un coeficiente de correlación serial. Se encuentra bien documentado en Helsel y Hirsch (2002). El test calcula un estadístico de prueba "d". Un valor bajo de "d" indica dependencia serial. La hipótesis nula de que los residuales son independientes se rechaza a favor de de la alternativa, que indica una correlación serial cuando $d < d_L$, cuyo valor se encuentra tabulado en textos de series de tiempo. El valor de d_L depende del tamaño de la serie de datos, el número de variables explicatorios y del nivel de significancia, alfa. Sin embargo, un bajo valor de "d" no explica las posibles causas, por ello, el análisis gráfico de los datos (residuales) es vital. El test requiere que los datos estén regularmente espaciados y que contengan muy pocos datos faltantes.

El test se encuentra disponible como opción de cálculo en los principales paquetes estadísticos como SPSS, Statgraphics, Minitab, S-Plus, R.

3.3. Etapas en la preparación de los datos para el ARF

A continuación se presenta, en términos de diagrama de flujo, las etapas sucesivas en el análisis y procesamiento de los registros de series hidrológicas para adecuarlos a los requerimientos del Análisis Regional de Frecuencias



Sus datos, al finalizar esta etapa, debieran estar en condiciones de ser utilizados en la etapa siguiente del procedimiento de Análisis Regional de Frecuencias.

Referencias bibliográficas

Dahmen, E. y Hall, M. 1990. Screening of Hydrological Data: Tests for Stationarity and Relative Consistency. Publication 49, ILRI.

Helsel, D. y Hirsch, R. 2002. Statistical Methods in Water Resources. USGS, Book 4, Hydrologic Analysis and Interpretation, Chapter 3.

Ortiz, J.; Iglesias, D.; Hernandez, R.; De Mera, A.; Gómez, M.; López, J. y Ballester, A. 2001. Desarrollo de programas para el tratamiento de series hidrológicas y evaluación de aportaciones hídricas subterráneas (HIDROBAS v. 3.0). VII Simposio de hidrogeología. Asociación Española de Hidrología Subterránea. Murcia, 2 de Mayo-1 de Junio de 2001. Tomo 23.

UNESCO. 2006. Guía metodológica para la elaboración del mapa de zonas áridas, semiáridas y subhúmedas secas de América Latina y El Caribe. CAZALAC. Documentos Técnicos del PHI_LAC, N°3

UNESCO-ROSTLAC. 1982. Guía metodológica para la elaboración del balance hídrico de América del Sur. Estudios e Informes en Hidrología. UNESCO-ROSTLAC, Montevideo.

USGS. 2004. Trends in Streamflow Characteristics at Long-Term Gaging Stations, Hawai. USGS, Scientific Investigations Report 2004-5080

USGS. 2006. Computer Program for the Kendall Family of Trend Tests. USGS, Scientific Investigations Report 2005-5275

Schaefer, M.; Barker, B.; Taylor, G. y Wallis, J. 2006. Regional precipitation-frequency analysis and spatial mapping for 24-hour precipitation for Oregon. Oregon Climate Service.

Schaefer, . 2007. Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations in eastern Washington. Oregon Climate Service

Wallis, J., Schaefer, M., Barrer, B. y Taylor, G. 2007. Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington States. Hydrol. Earth Syst. Sci., 11(1), 415-442