

Método de Análisis de Consistencia Computacional, aplicado al estudio y comparación de registros de precipitaciones

Jorge Bonilla, Patricia López y Sabrina Miranda

Instituto Nacional del Agua – Centro Regional Andino – Belgrano Oeste 210, Ciudad, Mendoza. Tel: +54 261 4288005

E-mail: jbonilla@ina.gob.ar

RESUMEN: El objetivo del trabajo es establecer un método racional, estructurado y normalizado para el análisis de consistencia de precipitaciones. Se trata de minimizar o eliminar los errores asociados al análisis de doble masa tradicional, debido a apreciaciones subjetivas del profesional interviniente, y su evidente inconveniente al momento de decidir si un cambio visualizado en el diagrama es significativo o no.

Basándose en el trazado de la curva de doble masa, se reemplaza la determinación manual de la recta de alineación de registros anuales, por la búsqueda iterativa de la mejor combinación de tramos y se considera la alineación óptima por el método de los mínimos cuadrados, asociado a verificaciones estadísticas de distribución de residuos. El mejor trazado se considera aquel que presenta el menor error cuadrático medio. La implementación se hace en una combinación de planilla de cálculo y lenguaje de programación C#.

El método presenta una excesiva exactitud, que lleva a detectar en algunos casos un mayor número de variaciones que las realmente significativas, dada la limitada precisión de las mediciones utilizadas como base. Se utiliza una fórmula empírica que sugiere una solución alternativa, pero sin obviar la alternativa original. Finalmente se aplica un test estadístico para determinar la significación de los cambios detectados.

Los resultados obtenidos permiten inferir objetivamente si los cambios detectados entre registros acumulados de estaciones pluviométricas, corresponden a variaciones propias de las mediciones de fenómenos naturales, o por el contrario establecen alteraciones producidas por un factor de origen antrópico, como cambios de ubicación, condiciones de trabajo o del personal a cargo de las lecturas.

OBJETIVO

El objetivo del trabajo es establecer un método racional, estructurado y normalizado para el análisis de consistencia de precipitaciones. Se trata de minimizar o eliminar los errores asociados al análisis de doble masa tradicional.

INTRODUCCIÓN

El Modelo MACC (Metodología de Análisis de Consistencia Computacional) surge por la necesidad de disminuir los efectos negativos en el trazado de las curvas de doble masa debido al:

- Trazado de la línea de tendencia

- Precisión del trazado

Trabajando con los mismos puntos (precipitación acumulada), dos profesionales que analizan el caso pueden tener diferente criterio subjetivo para el trazado de la curva de doble masa y proponer soluciones diferentes (Figura 1).

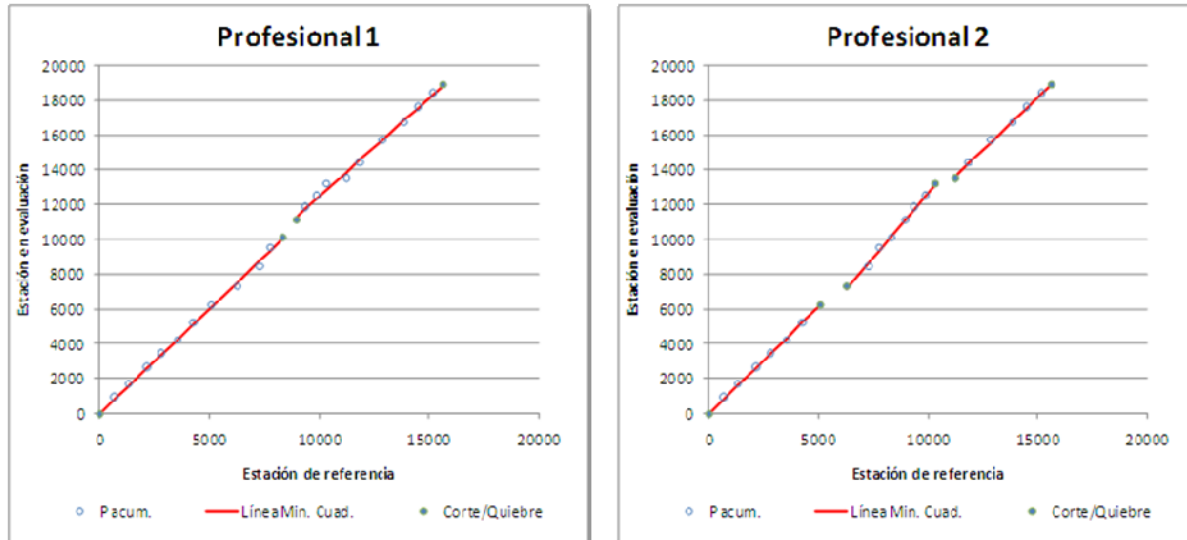


Figura 1.- Trazado ambiguo de la curva de doble masa

Esto puede deberse a dos factores: insuficiente precisión en el trazado manual y falta intrínseca de información en los datos disponibles. En cualquier caso se genera la duda acerca de que gráfico representa en forma más fiel al fenómeno en estudio.

El primero de los factores que influyen se soluciona eliminando la componente manual del proceso, responsable de la falta de precisión. La segunda no es posible eliminarla, por lo que se intenta es brindar herramientas de decisión que ayuden a comprender el comportamiento estudiado.

FUNDAMENTOS DEL MÉTODO

El método se basa en comprobar en forma iterativa todas las combinaciones posibles de rectas de alineación y determinar la óptima. Es evidente que este proceso no es factible realizarlo en forma manual por lo que se recurre a las herramientas informáticas.

El esquema detallado del método se observa en la Figura 2.

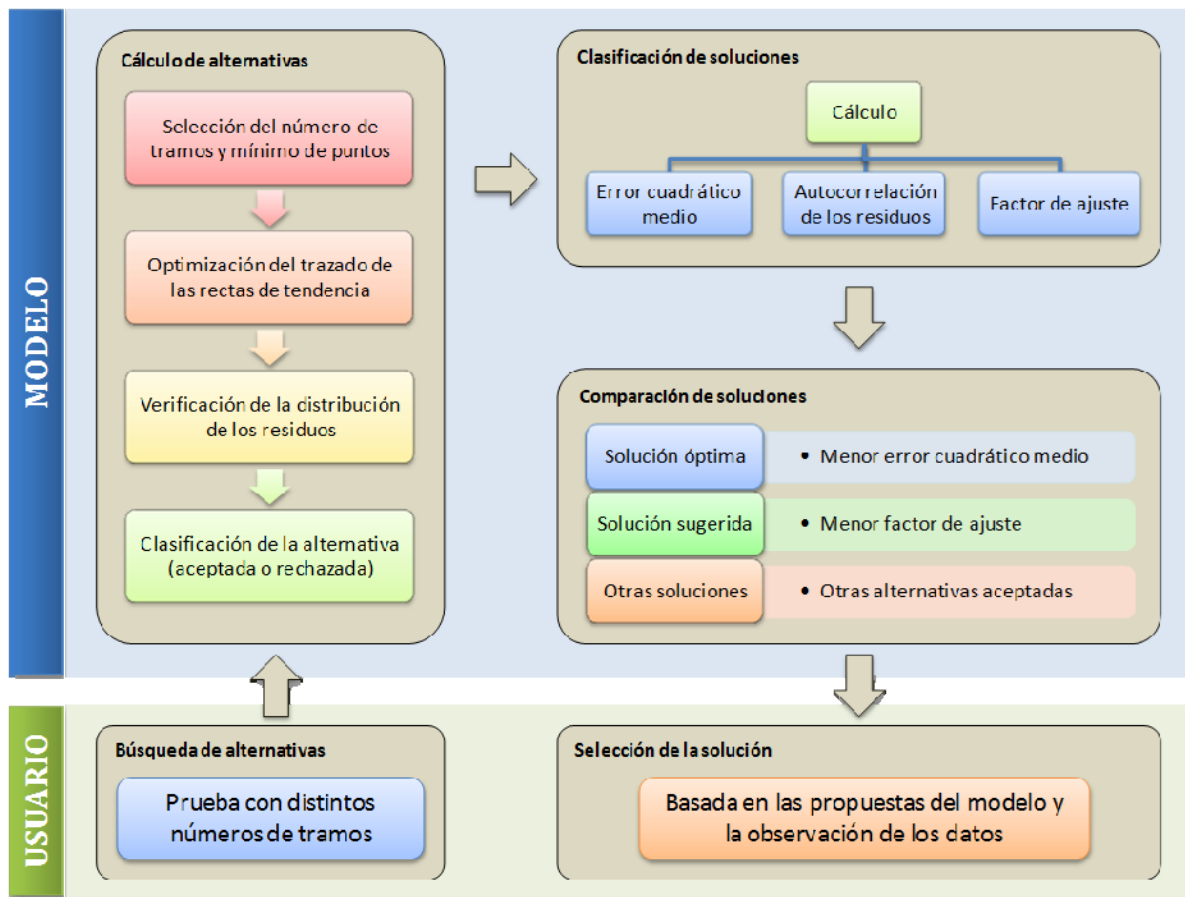


Figura 2.- Metodología general de trabajo con el Modelo MACC

DESCRIPCIÓN GENERAL DEL PROCESO

Introducción

Antes de comenzar con la descripción es conveniente establecer el uso que se le da a dos términos muy comunes en este documento: alternativa y solución.

El proceso se basa en la proposición de una serie de alternativas a evaluar. Cuando una alternativa es aceptada según los criterios establecidos se llama solución.

Búsqueda de alternativas

La búsqueda de alternativas es la prueba sistemática de curvas de doble masa considerando distintos número de tramos. Estas son validadas desde el punto de vista estadístico y son comparadas a los efectos de encontrar la mejor solución posible.

Se debe comenzar con 1 tramo e ir incrementando el número hasta un valor razonable en función del número de puntos y la respuesta obtenida en cada propuesta.

Cálculo de alternativas

Determinado el número de tramos por parte del usuario, el modelo encuentra el mejor ajuste de rectas basándose en la optimización por mínimos cuadrados, y buscando la combinación que produzca el menor error cuadrático medio.

Además verifica que los residuos sean independientes. Si no se cumple esta condición se estaría tomando una recta donde los residuos son sistemáticamente positivos o negativos y denotaría un ajuste ineficiente. Esta comprobación determina si la propuesta es aceptada o rechazada.

Clasificación de soluciones

Cuando se tiene una o más alternativas aceptadas, pasan a la categoría de soluciones y se clasifican las mismas mediante tres parámetros: el error cuadrático medio, la autocorrelación de los residuos y el factor de ajuste.

Comparación de soluciones

Este punto está íntimamente relacionado con el anterior. Se basa en la clasificación realizada para dar básicamente tres soluciones:

- Óptima: es la que presenta el menor error cuadrático medio corregido por grados de libertad. Representa la mejor solución desde el punto de vista estrictamente matemático.
- Sugerida: es la que tiene el mejor factor de ajuste. Representa una solución basada en un criterio empírico que toma en cuenta la ganancia relativa de precisión con respecto al aumento en el número de tramos.
- Otras: las soluciones que cumplen con la condición de independencia de los residuos, y que no están incluidas en las anteriores.

Selección de la solución

La disponibilidad de propuestas para seleccionar es variable. En algunos casos se tendrá la óptima y la sugerida, o en otro ambas coincidirán. Pueden quedar o no otras soluciones.

En cualquier caso observando los datos y la curva de doble masa se obtendrán una serie de herramientas que permitirán determinar la solución que mejor representa el fenómeno real. Es posible que en algún caso se elija una que no sea la óptima o la sugerida. Esto es perfectamente válido, recordando que el criterio del

profesional interviniente es fundamental para el trabajo. El modelo no toma decisiones, sólo sugiere alternativas y ayuda a través de las herramientas de decisión.

LA FUNCIÓN DEL USUARIO

El usuario básicamente realiza dos tareas:

- Búsqueda de alternativas: probando distintos números de tramos
- Selección de la solución: eligiendo la más adecuada del caso

Esto representa la clave y el mayor avance en el tema. El usuario no realiza cálculos, procesos complejos, gráficos, ni nada que se le parezca. Su función es simple: prueba y observa. Esto permite disminuir las posibilidades de error y eliminar el componente subjetivo inherente al proceso gráfico tradicional de la metodología de doble masa.

Con los mismos datos, dos profesionales que utilicen el modelo llegarán a la misma solución óptima y sugerida. Si por ejemplo uno de ellos prefiere la óptima y el otro la sugerida, podrán discutir el criterio tomando como referencia los índices calculados (error cuadrático medio, autocorrelación de residuos y factor de ajuste). Se evita totalmente la discusión acerca de la alineación visual de puntos, común en el método tradicional.

ANÁLISIS DE AJUSTES

Se realizó un estudio detallado de los casos posibles de ajustes obtenidos, para poder modelarlos y determinar la validez de cada uno de ellos. Los datos presentados son valores sin significado y se usan únicamente para ejemplificar los casos posibles.

Ajuste eficiente

Si un ajuste es eficiente se espera que la distribución de residuos sea aleatoria y que responda a los factores de variación normal de precipitación entre una estación y otra. Esto se puede expresar matemáticamente diciendo que los residuos deben ser independientes uno de otros. En la Figura 3 se considera la función $y = x + r$, donde r es un residuo aleatorio.

En este caso los residuos son positivos o negativos de acuerdo a un patrón desconocido, pero se van alternando en signo. En el gráfico hay como máximo 2 puntos seguidos por arriba o por debajo de la línea de tendencia.



Figura 3.- Ajuste eficiente por mínimos cuadrados

Un caso interesante se presenta cuando se considera la siguiente ecuación $y = ax \pm r$, siendo r un valor constante y el signo se alterna de un punto a otro. Esto se conoce como autocorrelación negativa, es decir mientras que en un punto el residuo es positivo en el siguiente es de similar valor pero negativo y viceversa. Aquí no se puede hablar de independencia de los residuos ya que el valor de uno depende del anterior, pero aplicado a las precipitaciones representaría una variación cíclica anual que mantiene la relación de las precipitaciones medias. El ajuste de una recta a este caso se considera satisfactorio y se ve en la Figura 4.

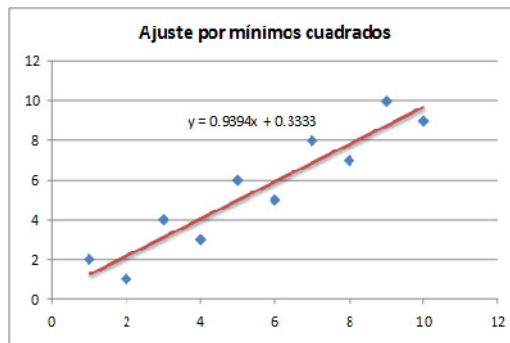


Figura 4.- Ajuste eficiente con autocorrelación negativa

Ajuste perfecto

Se denomina ajuste perfecto a aquel modelo matemático que pasa por todos los puntos de datos conocidos.

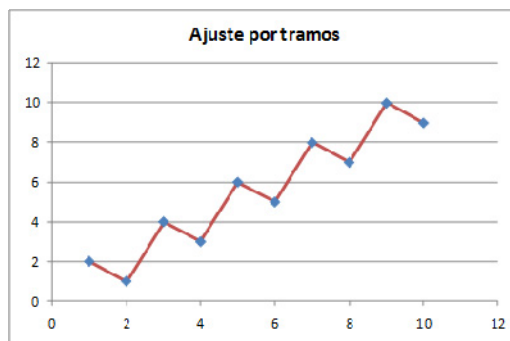


Figura 5.- Ajuste por tramos no válido en análisis de doble masa

Si se pensara en ajustar los puntos con un modelo como el de la Figura 5, no se cumpliría con la premisa que un cambio de pendiente significativo debe abarcar por lo menos 5 registros.

En el caso de la Figura 6, se estaría considerando un ajuste perfecto, que se puede aplicar a cualquier grupo de puntos, tomando la recta que los une de a dos.

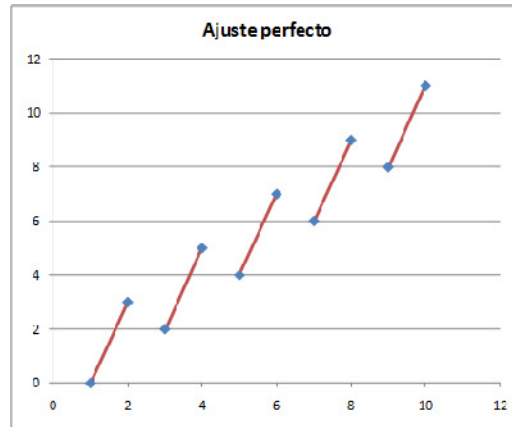


Figura 6.- Ajuste perfecto sin valor para el método

Es interesante desde el punto de vista matemático considerar que si el cálculo del error cuadrático medio, no contiene la corrección por grados de libertad, ésta solución es la que menor error cuadrático medio presenta. En la Figura 6 para 10 datos ($n=10$), se toman 5 tramos ($t=10/2=5$), quedando según la (2):

$$e = \frac{\sum(\Delta y)^2}{n} = \frac{0}{10} = 0 \quad (1)$$

Esta solución da siempre $e=0$, por lo que se convierte en la solución óptima. Esto evidencia la falta de utilidad de este cálculo.

Ahora con la consideración de los grados de libertad (3):

$$e = \frac{\sum(\Delta y)^2}{n-2t} = \frac{0}{0} \quad (2)$$

Desde el punto de vista exclusivamente matemático el error cometido por esta solución es indeterminado. Es evidente que al unir los puntos de a dos no tiene sentido el cálculo del error.

Una segunda ventaja que nos trae la corrección es que llegado a un nivel de ajuste determinado, un incremento de los tramos no trae aparejado una disminución apreciable de $\sum(\Delta y)^2$, pero disminuye sensiblemente el denominador de la ecuación de cálculo ($n - 2t$), y por lo tanto el error cuadrático medio aumenta en vez de disminuir.

En definitiva existirá una solución óptima desde el punto de vista matemático que será la que presente el menor error cuadrático medio corregido por grados de libertad.

Ajuste deficiente

Cuando el ajuste es deficiente quiere decir que hay algún fenómeno implícito en los datos que no lo representa la línea de tendencia obtenida. Para ejemplificar se supone que los puntos responden a la ecuación $y = mx$, y que el valor m cambia a partir del séptimo punto (Figura 7).

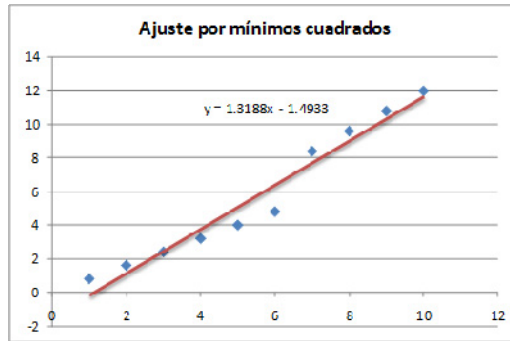


Figura 7.- Ajuste deficiente por mínimos cuadrados

La recta oculta el cambio de pendiente, pero los residuos lo evidencian al ser sistemáticamente positivos o negativos. Un efecto similar se obtiene si se toma como ecuación que rige la distribución de puntos a $y = x^2$. En este caso el ajuste lineal de la recta de mínimos cuadrados oculta la relación real que es cuadrática, como se observa en la Figura 8.

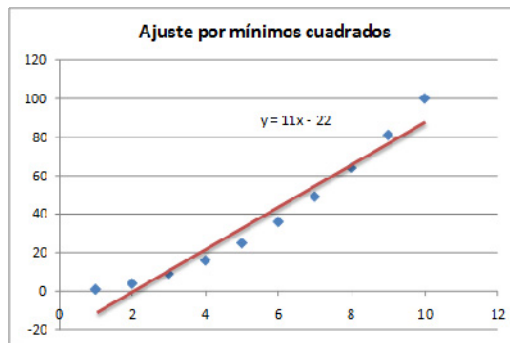


Figura 8.- Ajuste deficiente en una distribución no lineal

Conclusiones del análisis del ajuste

El valor del error cuadrático medio corregido por grados de libertad es un buen indicador del ajuste obtenido.

Se debe valorar la autocorrelación de los residuos, la cual es favorable mientras menor valor tenga, incluso un valor de autocorrelación negativo es un muy buen indicador. Se calcula un coeficiente numérico mediante la siguiente ecuación (Martínez, Seoane, y López, 1990):

$$r = \frac{n}{n-1} \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})(x_{i+1} - \bar{x})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2] [\sum_{i=2}^n (x_i - \bar{x})^2]}} \quad (3)$$

FACTOR DE AJUSTE

Las pruebas realizadas sugieren que la precisión matemática obtenida puede sobrevaluar los tramos necesarios para el ajuste de doble masa, por lo que más adelante se retoma la discusión del tema.

Cuando el número de tramos se toma tal que el error cuadrático medio sea mínimo, provee una precisión matemática importante, pero al mismo tiempo un nivel de detalle en el trazado de la curva de doble masa que en algunos casos sobreestima el número de tramos necesarios para la interpretación del fenómeno físico que se estudia. En forma empírica se toma como factor de ajuste 'f' según (5):

$$f = t^2 \cdot (\alpha + 1) \quad (4)$$

Analizando su comportamiento en función de lo deseable para un buen ajuste se tiene:

- Disminuye significativamente al tomar menor cantidad de tramos. Siempre es preferible tomar t lo más chico posible porque facilita el análisis posterior.
- Disminuye proporcionalmente con el error cuadrático medio. Mientras menor es este valor más acertado es el ajuste elegido.
- Disminuye en la medida que la autocorrelación sea negativa o no exista. Hay que evitar tener autocorrelación positiva porque es un signo de ajuste deficiente. Al valor α se le suma la unidad para que sea siempre positivo y varíe entre 0 y 2.

El factor 'f' equilibra el aumento de tramos con respecto a la ganancia de precisión y toma en cuenta la distribución de los residuos.

Al ser empírico, y por lo tanto sin demostración matemática, la alternativa que menor 'f' tenga será la sugerida por el modelo. Ésta puede ser o no coincidente con la óptima.

SELECCIÓN DE LA SOLUCIÓN

Una vez clasificadas las soluciones se debe decidir cuál es la que mejor representa la situación real.

Para la selección se recomienda seguir los siguientes pasos:

1. Observar cuidadosamente todas las soluciones.
2. Criticar el ajuste de cada una resaltando los puntos favorables y desfavorables.
3. Comparar los parámetros: error cuadrático medio, autocorrelación y factor de ajuste.
4. Observar los datos, en especial donde se producen los quiebres de tramos.
5. En base a la información recolectada elegir la solución más adecuada.
6. Documentar comentarios y observaciones que ayudaron a tomar la decisión.

IMPLEMENTACIÓN DEL MODELO

La implementación del modelo se hace en una combinación de herramientas de planilla de cálculo Excel y un software generado íntegramente en lenguaje C#, llamado MACC.

Los resultados de las distintas alternativas se observa en la Figura 9, donde el usuario puede determinar según su criterio la solución más adecuada.

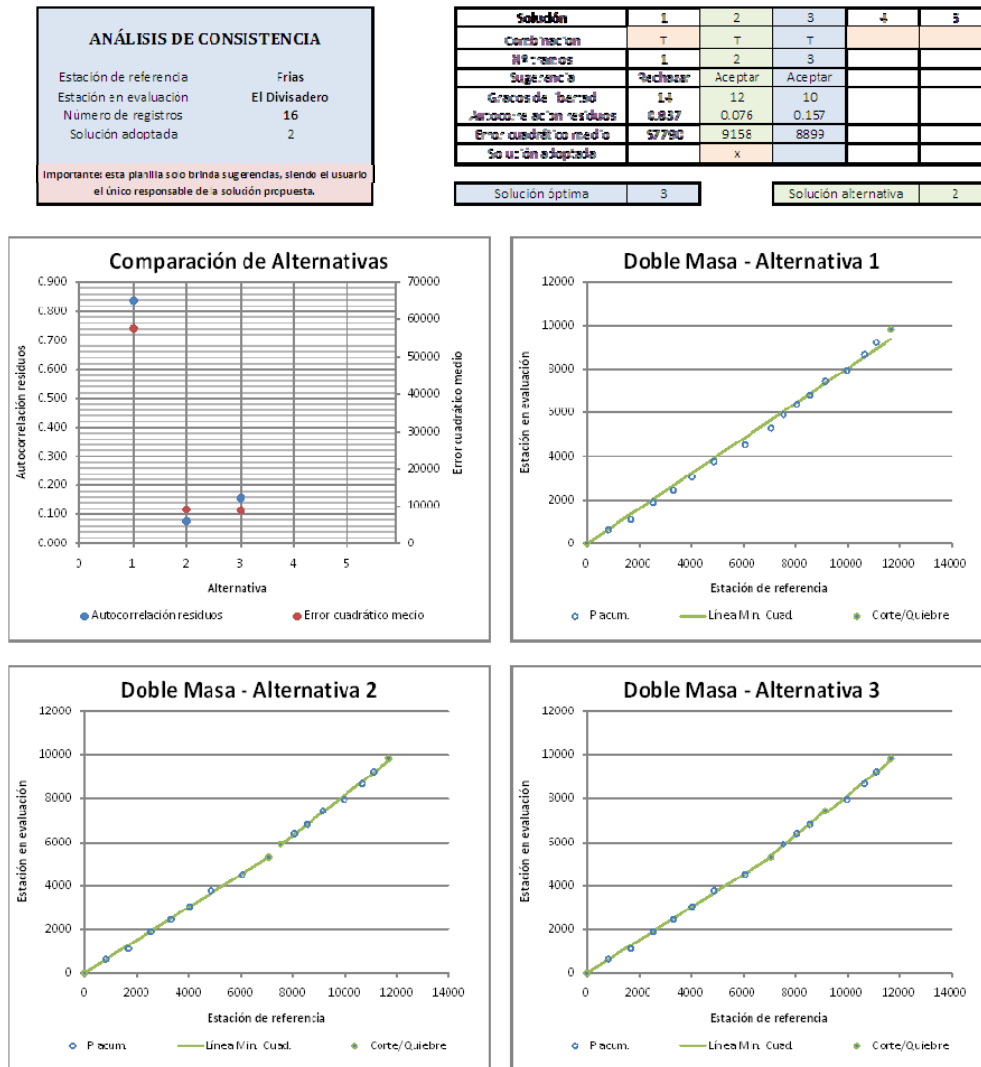


Figura 9.- Implementación del modelo

La introducción de datos se hace en una de las hojas de la planilla de cálculo (Figura 10), y se pasan al programa MACC mediante el uso del portapapeles de Windows.

Luego de procesar los datos (Figura 11), se llevan nuevamente a Excel, donde se utiliza su capacidad de graficación y análisis para proveer las herramientas de decisión descriptas.

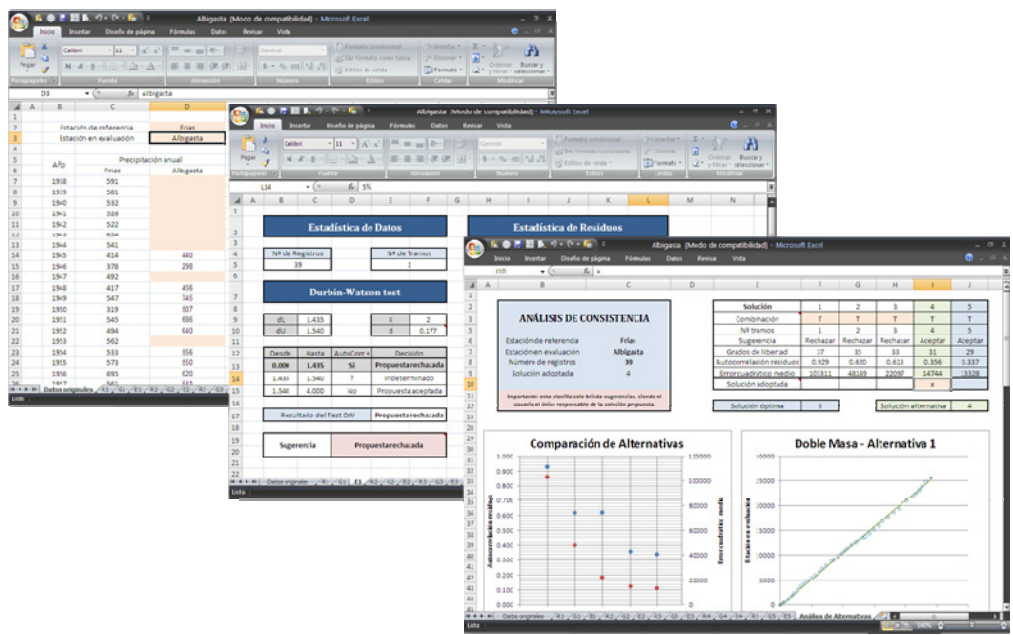


Figura 10.- Implementación del Modelo MACC (Excel)

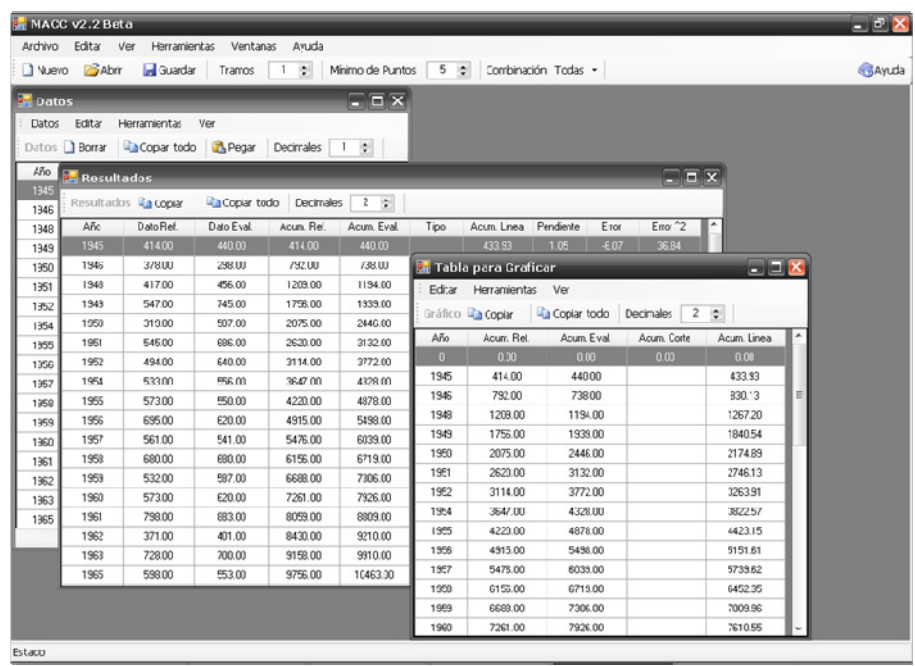


Figura 11.- Implementación del Modelo MACC (C#)

APLICACIÓN PRÁCTICA

El método fue sistemáticamente utilizado en un proyecto de asistencia tecnológica a terceros del INA-CRA (Vargas Aranibar, Maza, López, Bonilla y Burgos, 2010), donde los resultados más importantes se muestran a continuación. Los gráficos representan en todos los casos precipitación acumulada por las estaciones pluviométricas a comparar.

Estaciones con posibles problemas de consistencia de datos (cambios de pendiente, Figura 12)

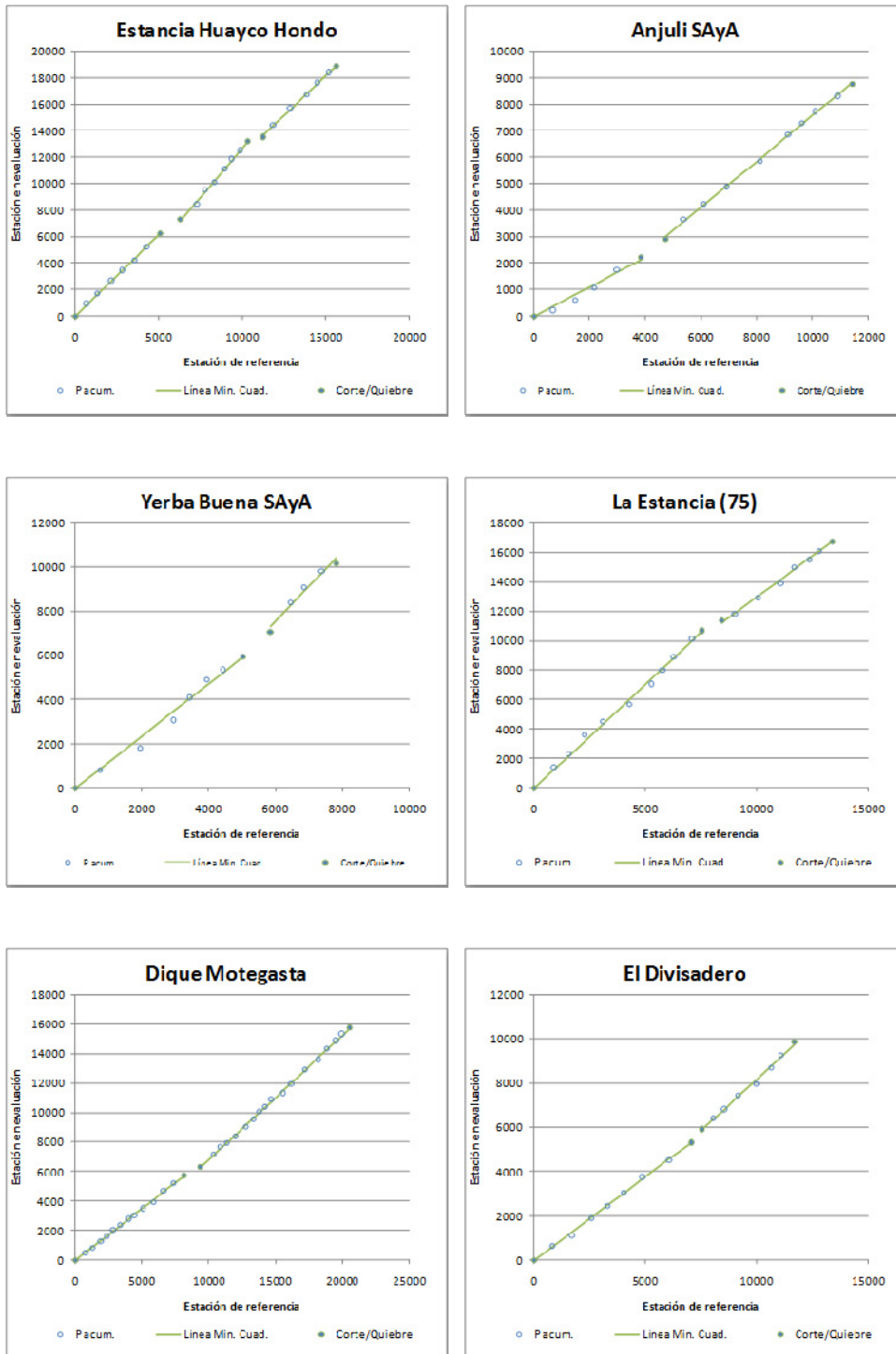


Figura 12.- Ejemplos de aplicación con cambios de pendiente detectados

Estaciones con posibles errores de medición y/o registro (desplazamiento paralelo, Figura 13)

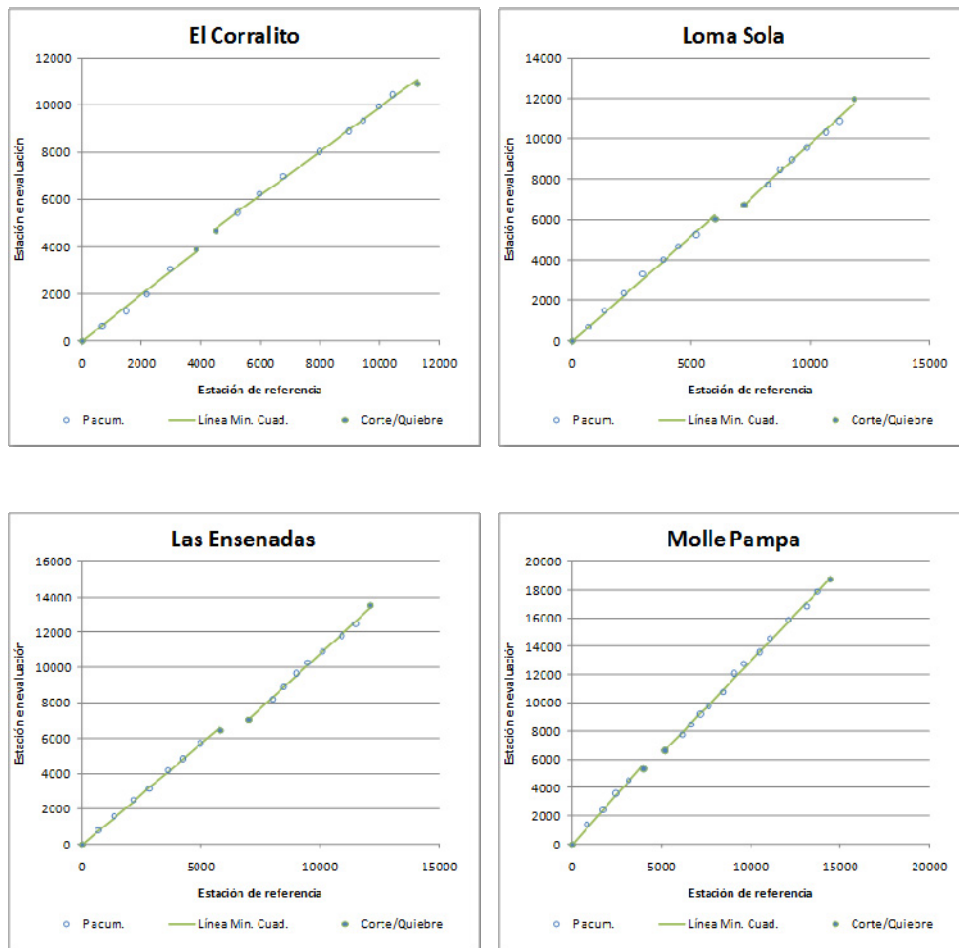


Figura 13.- Ejemplos de aplicación con desplazamientos detectados

Estaciones sin problema en la consistencia de datos (ubicados sobre una misma recta, Figura 14)

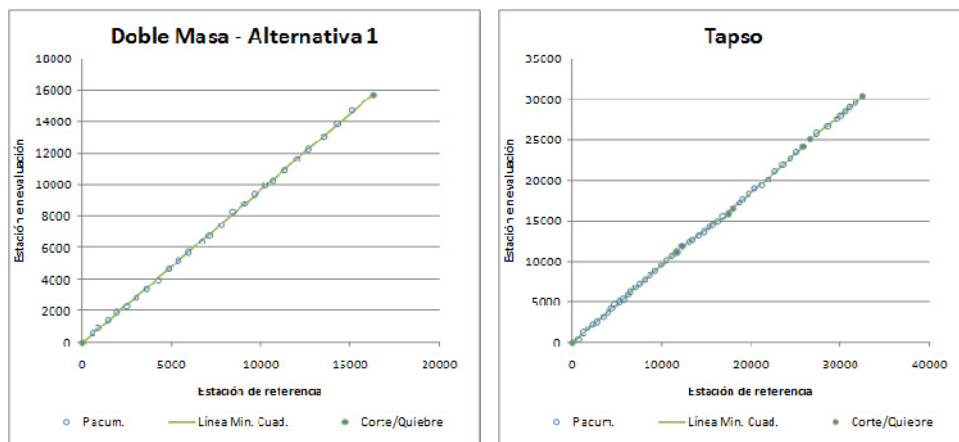


Figura 14.- Ejemplos de aplicación con datos alineados a una recta

CONDICIONES DE FUNCIONAMIENTO

Tener una estación pluviométrica de referencia, que preferentemente tenga un largo registro de datos, que haya sido operada en forma continua y con los controles adecuados para poder utilizarla como patrón.

Disponer de un mínimo de 15 años de registro completos, para poder calcular la precipitación anual acumulada, necesaria para el ingreso al modelo, y que exista la posibilidad de formar tramos de por lo menos 5 puntos.

No es recomendable utilizar precipitaciones mensuales, debido a que la variación estacional puede producir falsos resultados de cambios de pendiente o desplazamientos.

CONCLUSIONES

El método permite efectivamente detectar cambios de pendiente o desplazamientos en la curva de doble masa, de manera objetiva y sin necesidad de intervención del usuario, lo que permite analizar a posteriori los resultados sin la preocupación de la calidad del trazado.

Este análisis puede ser el estándar que plantea la bibliografía especializada, es decir observar el gráfico y de acuerdo al criterio del profesional determinar si la variación mostrada en el gráfico es significativa o no. Otra alternativa es utilizar un método estadístico para determinar si son significativos los cambios de pendiente mostrados por los gráficos.

En cualquier caso a partir de los resultados obtenidos por el modelo, se tiene un punto de partida para analizar la consistencia de los datos de base y determinar si son válidos para aplicarles procedimientos estadísticos más complejos. La decisión final dependerá de una investigación de la historia de cómo se obtuvo y registró la información original, así también de los cambios que pudieron sufrir las estaciones a lo largo de su vida.

Este trabajo fue presentado en el Congreso Nacional del Agua en 2011, donde se puede consultar con más detalle los procedimientos internos del método (Bonilla, López y Miranda, 2011).

REFERENCIAS

- Vargas Aranibar, A., Maza, J., López, P., Bonilla, J. y Burgos, V., 2010. Relevamiento de la Disponibilidad de Recursos Hídricos Superficiales de la Cuenca Salar de Pipanaco, Cuenca Abaucán – Colorado – Salado y Cuencas Faldeo Oriental de Ancasti, Tomo I y III, Instituto Nacional del Agua, Centro Regional Andino, Mendoza, Argentina.
- Bonilla, J., López, P. y Miranda, S., 2011. Método de Análisis de Consistencia Computacional, aplicado al estudio y comparación de registros de precipitaciones, Congreso Nacional del Agua 2011, Argentina.
- Martínez, E., Seoane, R. y López, P., 1990. Pruebas de Aleatoriedad en Series de Caudales Anuales. XIV Congreso Latinoamericano de Hidráulica.