

METODOLOGÍA PARA LA IMPUTACIÓN DE DATOS FALTANTES EN METEOROLOGÍA.

Methodology for the Imputation of data Faltantes in the Meteorology

RESUMEN

El artículo presenta la metodología que se debe seguir para la imputación de datos en series de precipitación y/o temperatura. El procedimiento consiste en hacer uso de correlaciones parciales, modelos de regresión, ajustes de los datos por medio del método de doble masa y verificación de la tendencia a través del test de Kendal. Se pudo evaluar que para un faltante de datos de 20% esta metodología resulta apropiada, sin embargo, para faltantes de más de 20% se recomienda usar los modelos AMI bivariada para la imputación de los mismos.

PALABRAS CLAVES: Imputación de datos, Correlaciones Parciales, Modelos de regresión, Test de Kendall, Método de doble Masa.

ABSTRACT

The article presents the methodology that should be continued for the imputation of data in precipitation series and/or temperature. The procedure consists on making use of partial correlations, regression models, adjustments of the data by means of the method of double mass and verification of the tendency through the the test of Kendal. You could evaluate that for a faltante of data of 20% this methodology is appropriate, however, for faltantes of more than 20% it is recommended to use the model AMI bivariados for the imputation of the same ones

KEYWORDS: *Imputation of data, Partial Correlations, regression Models, Test of Kendall, Method of double Mass.*

1. INTRODUCCIÓN

Uno de los problema que se presenta en climatología al momento de procesar bases de datos con el fin de realizar los análisis estadísticos, es la ausencia de registros diarios o mensuales en algunas estaciones climatológicas que se mantienen en funcionamiento de forma regular, lo que genera una serie datos faltantes en dichas bases de datos. Como solución a este problema que es ocasionado por errores humanos o informáticos, se han creado una serie de técnicas, métodos, metodologías, procedimientos, modelos estadísticos y matemáticos que permiten dar solución a esta problemática, este trabajo muestra una metodología que se utilizan para la imputación de datos en series climáticas.

2. DESARROLLO TEÓRICO

[1]**Patrones de precipitación:** Como precipitación se conocen todas las formas de humedad que caen a la tierra, provenientes de las nubes, como agua, nieve y

hielo. La precipitación constituye la entrada primordial del sistema hidrológico y es el factor principal que controla la hidrología de una región.

El conocimiento de los comportamientos y patrones de la lluvia en el tiempo y en el espacio es esencial para entender procesos como la variación de la humedad del suelo recarga de acuíferos y caudal en los ríos.

El estudio de la precipitación es entonces de capital importancia para los hidrólogos, pero una investigación detallada de los mecanismos de su formación es dominio de la meteorología.

[2]**Correlación Parcial:** El coeficiente de correlación simple está basado en la suposición de la aproximación a la distribución *normal bivalente*. Si se tiene más de dos variables, el modelo básico para la correlación múltiple, sería una ampliación de esta distribución, denominada *distribución normal multivariante*.

JORGE ANDRÉS URRUTIA

M. Sc. Investigación de operaciones y Estadística
Profesor Auxiliar
Universidad Tecnológica de Pereira
jurrutia@utp.edu.co

REINER PALOMINO

M. Sc en Meteorología
Profesor Investigador
Universidad Tecnológica del Chocó
reiner@utch.edu.co

HEVER DARÍO SALAZAR

M. Sc Candidato
Profesor
Universidad Agrícola y rural de Colombia. UNISARC.
hedasa@hotmail.com

Fecha Recepción: 9 de Septiembre de 2010

Fecha aceptación: 15 de Noviembre de 2010

Si hay tres variables, habrá tres correlaciones simples entre ellas, ρ_{12} , ρ_{13} y ρ_{23} . Estos coeficientes miden la relación lineal que existen entre estas variables, dos a dos, sin tener en cuenta la posible influencia de la tercera. La correlación parcial se define como la correlación entre dos variables si las demás variable no varían, es decir, el valor de las demás variables son fijos. Por ejemplo, el coeficiente de correlación parcial $\rho_{12.3}$. Si una correlación significativa entre dos variables se convierte en correlación parcial no significativa cuando una tercera variable permanece constante, esto sugiere, aunque no prueba, que la variable que permanece constante es la causa común de la correlación de las otras dos.

La correlación parcial $r_{12.3}$, sería la correlación lineal entre la variable 1 y 2 dejando como constante la variable 3. Esto quiere decir que hay que medir la correlación entre la variable 1 y 2 que no sea un reflejo de sus relaciones con la variable 3. Por tanto, se puede obtener una estima muestral $r_{12.3}$ calculando la desviación o residuo e_{13} , de la regresión de la variable 1 sobre la variable 3, y la desviación o residuo e_{23} , de la regresión de la variable 2 sobre la variable 3. Y $r_{12.3}$ es el coeficiente de correlación simple entre e_{13} y e_{23} . El cálculo de la correlación parcial se define como sigue:

Puede demostrarse que $r_{12.3}$ satisface la siguiente fórmula

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} \tag{1}$$

Por lo que podemos probar $H_0: r_{12.3} = 0$ por medio de la t

$$t = \frac{r_{12.3} \sqrt{n-3}}{\sqrt{1-r_{12.3}^2}} \tag{2}$$

Que se contrasta con la $t_{n-3; \alpha/2}$. De la misma manera se puede hallar la regresión entre la variable 1 y la variable 3 dejando constante la variable 2; o la correlación entre la variable 2 y la variable 3 dejando constante la variable 1.

$$r_{132} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}} \tag{3}$$

$$S_{r_{132}} = \sqrt{\frac{1-r_{132}^2}{n-3}}, \tag{4}$$

$$t = \frac{r_{132} \sqrt{n-3}}{\sqrt{1-r_{132}^2}} \tag{5}$$

$$r_{231} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}} \tag{6}$$

$$S_{r_{231}} = \sqrt{\frac{1-r_{231}^2}{n-3}} \tag{7}$$

$$t = \frac{r_{231} \sqrt{n-3}}{\sqrt{1-r_{231}^2}} \tag{8}$$

[3]**Modelos de Regresión Lineal:** El análisis de regresión es una técnica estadística para la investigación de la relación entre dos o más variables, puede emplearse para construir un modelo que permita predecir el comportamiento de una variable y (dependiente, respuesta) en función de una o más variables (independientes, predictivas) x .

Los comportamientos de estas variables pueden estar definidos de antemano lo cual nos remite a un modelo teórico, o bien, se tiene el caso de que no exista una relación establecida entre estas y sea necesario establecer una primera aproximación del comportamiento de las mismas. Este modelo puede ser expresado como,

$Y_i = \beta_0 + \beta_1 X_i + \epsilon$ (9). Dado el modelo de regresión simple, si se calcula la esperanza (valor esperado) del valor Y , se obtiene

$$E(y_i) = \hat{y}_i = E(\beta_0) + E(\beta_1 x_i) + E(\epsilon_i) \\ = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{10}$$

Calculando $\hat{\beta}_0$ y $\hat{\beta}_1$. Para esto se buscan dichos parámetros que minimicen $\sum (y_i - \hat{y}_i)^2 = \sum \hat{\epsilon}_i^2$. Derivando respecto a $\hat{\beta}_0$ y $\hat{\beta}_1$ e igualando a cero, se obtiene

$$\frac{\partial \sum (y_i - \hat{y}_i)^2}{\partial \hat{\beta}_0} = 0 \tag{11}$$

$$\frac{\partial \sum (y_i - \hat{y}_i)^2}{\partial \hat{\beta}_1} = 0 \tag{12}$$

Obteniendo dos ecuaciones denominadas ecuaciones normales que generan la siguiente solución para ambos parámetros:

$$\hat{\beta}_1 = \frac{\sum x \sum y - n \sum xy}{(\sum x)^2 - n \sum x^2} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \tag{13}$$

$$\hat{\beta}_0 = \frac{\sum y - \hat{\beta}_1 \sum x}{n} = \bar{y} - \hat{\beta}_1 \bar{x} \quad (14)$$

[4],[5] **Método de doble masa:** El método de doble masa es un método que prueba la homogeneidad de los datos meteorológicos, comparando una cantidad (anual o estacional) de precipitación acumulada, con los valores promedio concomitantes (e igualmente acumuladas) de un grupo de estaciones circunvecinas. Los puntos resultantes de las correlaciones obtenidas para cada año o estación, son unidos por segmento de recta. Cuando los valores anuales acumulados de una estación pluviométrica son homogéneos, la gráfica muestra una recta continua.

Esa representación consiste en identificar la estación que queremos controlar, tomando los valores anuales de precipitación. Luego debe contarse con al menos tres (3) estaciones vecinas, cuyos registros anuales sean confiables y que llamaremos estaciones base, cuya serie de datos anuales debe coincidir con el de la estación a controlar. En cada año, a partir del primero con registro, se promedian los valores de las estaciones base y se acumulan por años sucesivos, obteniéndose una precipitación media anual acumulada.

Luego, en un sistema de ejes ortogonales, se grafica en ordenadas los valores de precipitación anual acumulada de la estación a controlar y en abscisas los de precipitación media anual acumulada de las estaciones base. Si los registros no han sufrido variaciones, los puntos se alinean en una recta de pendiente única, por lo tanto no será necesario efectuar correcciones. Si por el contrario hay variaciones en la pendiente de la recta, significa que parte de la serie contiene valores erróneos por lo cual el registro de datos debe ser corregido a partir del año en el que cambia la pendiente de la recta, se obtiene en ese caso un factor de corrección que es proporcional a la variación de la pendiente de la recta como lo muestra la Fig 1

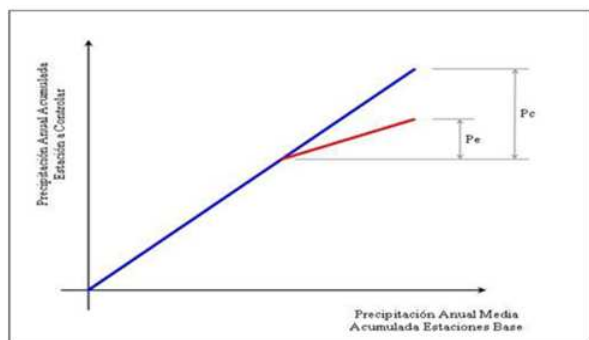


FIGURA 1 ANÁLISIS DE DOBLE MASA

El factor de corrección se obtiene haciendo Pc/Pe que en el ejemplo del gráfico será K 1, debido a que los registros anuales medidos han sido menores a los reales y deben corregirse a partir del año del error, tomando los valores anuales sin acumular y afectándolos a cada uno por el factor de corrección.

[6], [7] **Test de Mann- Kendall:** El test de Kendall, es un método no paramétrico para la detección de tendencia. La aplicación del test de Mann-Kendall para una serie de n datos (x₁, x₂, ... , x_n) independientes e idénticamente distribuidos supone la aceptación previa de la hipótesis nula H₀, que supone la no existencia de tendencia. Al ser no paramétrico detecta la presencia de cualquier forma de tendencia, en general siempre que para i < j , se tenga prob (X (i) < X(j)) > ½. Además de tendencias lineales también detecta saltos siempre que den tendencia global en el sentido anterior. Es muy usado en climatología. Tiene la ventaja sobre el test similar de Spearman de que la convergencia de la distribución del estadístico a la distribución normal es muy rápida a partir de N= 10 da buena aproximación. El estadístico t del test se obtiene comparando todas las parejas de elementos de la serie y sumando 1, ó 0 -1, según el orden creciente o decreciente, t es igual a:

$$t = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \text{signo} [x(j) - x(i)] \quad (15)$$

El estadístico t se estandariza usando las expresiones para la media y la varianza de la siguiente forma:

$$u (t) \approx \mathcal{N} (0 , 1) \quad (16)$$

3. METODOLOGÍA

Para el caso de estudio se seleccionaron las estaciones que cumplieron con el criterio de tener menos del 20% de datos faltantes, para presentar homogeneidad, continuidad y confiabilidad.

Lo primero que se realizó fue depurar aquellas estaciones con más del 20% de datos faltantes, luego se organizaron las estaciones con mayor cantidad de registros en un orden de mayor a menor, teniendo como referencia, el año de instalación y funcionamiento.

Para calcular los datos faltantes en las estaciones, se tomó un tramo de 10 años en donde todas las estaciones presentaban registros continuos y se realizaron correlaciones parciales, con el fin de establecer modelos de regresión ajustados que permitieran encontrar los datos faltantes en cada una de las estaciones.

Una vez realizado este procedimiento con cada una de las estaciones que presentaron ausencia de datos menores al

20%, se aplicaron pruebas de confiabilidad orientadas a la verificación y ajuste de los datos, aplicando el método doble masa recomendado por (OMM, 1983). Organización Meteorológica Mundial, y posteriormente el test no paramétrico de tendencia de Mann-Kendal.

4. ANÁLISIS DE RESULTADOS

La finalidad de aplicar las correlaciones parciales, es determinar cuáles estaciones presentan tendencia similares en los patrones de precipitación, para luego hacer uso de los modelos de regresión para estimar los valores faltantes. La tabla 1, muestra las estaciones de la Vuelta, Guaduas, Certegui, Managrú, Paimadó y Lloró, en donde se observan datos faltantes. La tabla 2, muestra las correlaciones parciales de las estaciones que tendremos en cuenta para determinar nuestros modelos de regresión y estimar los datos faltantes. Se observa que las estaciones que presentan mayor correlación son: Carmen de Atrato- Paimadó (0.94), Paimadó-Guadua (0.727), Pinol-Guadua (0,601), Paimadó- Pinol (0.930), Paimadó- Certegui (0.026), Apto el Caraño (0.614).

MUNICIPIO		LLORO	EL CARMEN	EL CARMEN	CERTEGUI	EL CANTON	RIO QUITO	LLORO	RIO QUITO	QUIBDO	
NOMBRE:		VUELTA LA	CARMEN DE	GUADUAS	PINON EL	CERTEGUI	MANAGRU	PAIMADO	LLORO	SAN ISIDRO	TUTUNENDO
Años	Meses										
1983	E	270	5	71	284	79	242	86	343.3	203.3	203
1983	F	261	30	102	176	38	129	30	336.7	295.8	309
1983	M	497.6	141	495	550	379	315	373	515.5	629.4	608
1983	A	626.6	231	666	474	442	379	201	548.5	247.3	682
1983	M	724.8	169	340	673	1052	737	684	691.6	628.3	773
1983	J	717.8	159	392	449	827	577	215	587.6	656.6	
1983	J	666.6	312	395	675	558	644		773.7	829	
1983	A	875.6	264	414	367	694	491.7		532.8	569.7	
1983	S	786.5	251	450	697	972	686		503.1	618.7	
1983	O	200	259	882	745	298	287		481.4	272.8	
1983	N	620	159	211	555	327	336		373.8	392.3	780
1983	D	527	207	338	687	231	456		681.4	501.6	925
1984	E	785	255	184	501	346	336	615	724.8	636.7	798
1984	F	935	232	284	483	788	810	560	1088	598.8	924
1984	M	634	96	281	534	680	785	762	627.8	1030	1051
1984	A	491	194		547	940	632	393	849.8	443.7	922
1984	M	873	195.4		789	588	616	478	902.5	686.3	931
1984	J	911	331.9		564	511	753	342	839.3		1164
1984	J	843	240		410	484	748	722	801.5		1339
1984	A	900	157		580	500	523	586	180		1811
1984	S	669	272		674	198	669	388	833.7		1335
1984	O	1179	278		794	1083	840	632	1072		1136
1984	N	840	377.9		797	630	373	311	951.1	704.7	1257
1984	D	661	144		814	674	547	588	639.6	516.2	1123
1985	E	702	133	355	649	423	396	301	1299	303.1	644

TABLA 1. MATRIZ DE DATOS FALTANTES

Recuperando hoja de trabajo del archivo: 'C:\Users\HF\Desktop\para el articulo .xlsx'
La hoja de trabajo se guardó en 06/09/2010

Resultados para: Hoja2

Correlaciones: BUEY EL CALAORRA. ALTO DEL BUE. LOMA LA-BOJA. OPOGADO. ...

	BUEY EL CALAORRA	ALTO DEL BUE	LOMA LA-BOJA	OPOGADO	VIGIA DE CURVA	RIOSUCIO	TERESITA LA	UNGUIA	HONDA LA	SAPTATA
BUEY EL CALAORRA	0.659 0.000									
ALTO DEL BUE	0.528 0.006	0.373 0.087								
LOMA LA-BOJA	0.692 0.005	0.283 0.168	0.086 0.052							
OPOGADO	0.690 0.000	0.562 0.003	0.590 0.002	0.414 0.036						
VIGIA DE CURVA	0.618 0.001	0.713 0.003	0.464 0.017	0.451 0.021						
RIOSUCIO	0.629 0.000	0.498 0.023	0.493 0.010	0.493 0.011						
TERESITA LA	0.676 0.002	0.496 0.013	0.602 0.001	0.310 0.112						
UNGUIA	0.528 0.006	0.433 0.027	0.602 0.001	0.503 0.009						
HONDA LA	0.421 0.032	0.374 0.063	0.581 0.002	0.270 0.183						
SAPTATA	0.614 0.001	0.479 0.013	0.450 0.021	0.511 0.005						
OPOGADO										
VIGIA DE CURVA	0.668 0.000									
RIOSUCIO	0.441 0.024	0.408 0.033								
TERESITA LA	0.560 0.003	0.535 0.003	0.770 0.000							
UNGUIA	0.594 0.001	0.575 0.001	0.720 0.000	0.600 0.000						
HONDA LA	0.542 0.004	0.306 0.123	0.611 0.001	0.774 0.000						
SAPTATA	0.424 0.031	0.380 0.055	0.799 0.000	0.637 0.000						

TABLA 2. CORRELACIONES PARCIALES

La tabla dos de correlaciones parciales, muestran las asociaciones entre los pares de estaciones, que parten de las observaciones por columnas, que constituyen el vector aleatorio del cual se parte para el cálculo de las correlaciones parciales. Los supuesto multivariante en que se basan las correlaciones parciales se cumplen el la construcción de los modelos de regresión, ya que un modelo de regresión es una combinación lineal $Y = a_1x_1 + \dots + a_nx_n$, y esta está normalmente distribuida; además el vector aleatorio Z de cada modelo utilizado para el cálculo de los datos ausente, son independientes.

Para este caso particular, se mostrará el proceso con dos estaciones, Certegui y Paimadó, tomando como variable de respuesta la estación de Certegui y Como variable regresora la estación de Paimadó, que en nuestro caso no presenta datos faltantes. En la tabla 3 de estadísticos de la regresión, muestran que el coeficiente de correlación múltiple es de 0.68, lo que garantiza que existe una buena relación lineal entre estas dos estaciones, el coeficiente de determinación $R^2 = 0.907$, presenta un buen ajuste, esto determina que se puede explicar el 90,7% de la variabilidad de las observaciones de la estación Certegui, a partir de las observaciones de la estación Paimadó.

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0.680006768
Coeficiente de determinación R ²	0.90762409205
R ² ajustado	0.895280937
Error típico	176.7433979
Observaciones	58

TABLA 3. ESTADÍSTICOS DE LA REGRESIÓN

En la tabla 4, de análisis de varianza, se puede observar que el valor crítico de F es 4.34235E-09, estadístico que por ser menor que $\alpha=0.05$, valida el modelo de regresión, al rechazarse $H_0 : \beta_0 = 0$, lo que indica que β_0 es estadísticamente diferente de 0.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	Valor crítico de F
Regresión	1	1504697.064	1504697.064	4.34235E-09
Residuos	56	1749340.807	31238.228	
Total	57	3254037.871		

TABLA 4. TABLA DE ANÁLISIS DE VARIANZA

El gráfico 1 de probabilidad normal, muestra el ajuste de las observaciones sobre la línea de regresión y la estimación de la ecuación de regresión, se puede observar que los datos presentan buen ajuste a lo largo de la línea de regresión.

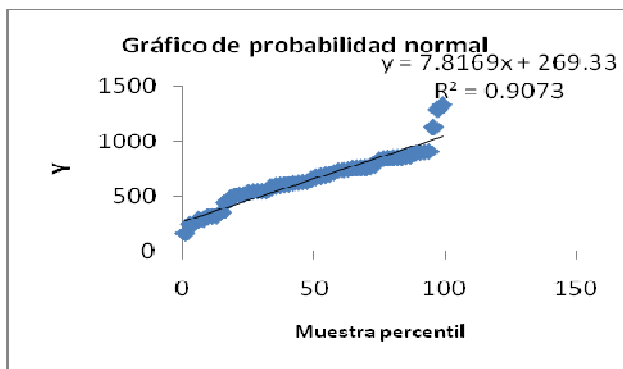


GRAFICO 1. GRAFICO DE PROBABILIDAD NORMAL

El modelo de regresión estimado $Y=269.33+7.8169X_1$, permitió estimar los datos presentados en la nueva matriz de observaciones, este procedimiento se realizó con todas las estaciones con datos faltantes presentados en la tabla 5.

MUNICIPIO	LLORO	EL CARMEN	EL CARMEN	EL CARMEN	CERTEGUI	EL CANTON	RIO QUITO	LLORO	RIO QUITO	QUIBDO
NOMBRE:	VUELTA LA	CARMEN DE	GUADUAS	PIÑON EL	CERTEGUI	MANAGRU	PAIMADO	LLORO	SAN ISIDRO	TUTUNENDU
Años	Mess									
1983	E	270	5	71	284	79	242	86	343.3	203.3
1983	F	261	30	102	176	38	129	30	336.7	295.8
1983	M	497.6	141	495	550	379	315	373	515.5	629.4
1983	A	626.4	231	666	474	442	379	201	548.5	247.3
1983	M	724.8	169	340	673	1052	737	684	691.6	628.3
1983	J	717.8	159	392	449	827	577	215	587.6	656.6
1983	J	666.6	312	395	675	558	644	875	773.7	829
1983	A	875.6	264	414	367	694	491.7	121	532.8	569.7
1983	S	786.5	251	450	697	972	686	671	503.1	618.7
1983	O	200	259	882	745	298	287	202	481.4	272.8
1983	N	620	159	211	555	327	336	518	373.8	392.3
1983	D	527	207	338	687	231	456	426	681.4	501.6
1984	E	785	255	184	501	346	336	615	724.8	636.7
1984	F	935	232	284	483	788	810	500	1088	598.8
1984	M	634	96	281	534	680	785	762	627.8	1030
1984	A	491	194	282	547	940	632	393	849.8	443.7
1984	M	873	195.4	517	789	588	616	478	902.5	686.3
1984	J	911	331.9	446	564	511	753	342	839.3	920
1984	J	843	240	437	410	484	748	722	801.5	1032
1984	A	900	157	416	580	500	523	586	180	846.8
1984	S	669	272	357	674	198	669	388	833.7	516.9
1984	O	1170	278	1120	794	1083	840	632	1072	707.2
1984	N	840	378.4	264	797	630	373	311	951.1	784.7
1984	D	661	144	280	814	674	547	588	639.6	585.2
1985	E	702	133	355	649	423	396	301	1299	303.1
1985	F	320	39	122	281	430	559	146	333.7	477.3
1985	M	637	4.34235E-09	414	673	472	57	626.7	311.7	444
1985	A	111	412	539	565	604	234	698.3	503.2	780
1985	M	937	230	601	411	319	732	404	932.3	777.7
1985	J	461	106	312	547	554.3	322	171	329	807.5
1985	J	808	160	466	488	505.4	513	167	679.4	575.1
1985	A	646	229	455	611	468	441	194	873.5	733.9
1985	S	1104	195	300	409	812	637	402	992.4	638.8
1985	O	523	205	625	793	465	446	349	594.5	519.9
1985	N	600.6	150	266	727	615	405	166	478	545.5
1985	D	755.7	151	115	605	785	663	307	274.4	314.1
1986	E	454.6	181	107	348	660	726	104	422.4	305.3

TABLA 5. MATRIZ DE ESTIMADOS

La aplicación del método de doble masa, requiere que los datos de precipitación sean estandarizados para eliminar el efecto de la estacionalidad.

La grafica presentada en la figura 2, muestra la tendencia de las series de las estaciones para las cuales se realizaron las imputaciones, para este caso como no se presentan distorsiones en las series, por tanto no se hace necesario la corrección de los datos estimados, dado que los datos estimados presentan homogeneidad en las tendencias.

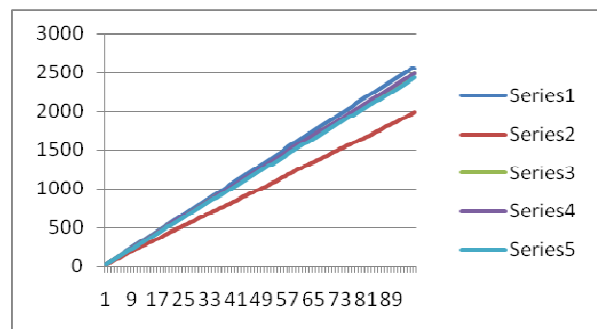


FIGURA 2. APLICACIÓN METODO DE DOBLE MASA

El objetivo de la aplicación del test de Mann-Kendall, se realiza para verificar el supuesto de la tendencia, antes de comprobar los supuestos de probabilidad normal.

Para la aplicación del test de Mann-Kendall, se construyó el estadístico de prueba t , aplicando la ecuación 15, el valor del estadístico es $t = 241.2790$, si tomamos un $\alpha = 0.05$, el estadístico de t , para 25 grados de libertad es $t = 1.7081$. Se puede ver entonces que se puede rechazar $H_0 =$ No existe tendencia. Lo que convalida que la metodología de imputación de datos ha sido eficiente.

5. CONCLUSIONES Y RECOMENDACIONES

La metodología propuesta para la imputación de datos climatológicos resulta ser una herramienta crucial para la preparación de datos antes de realizar cualquier análisis estadístico.

El uso de los modelos de regresión lineal y correlaciones parciales, deben ser aplicados cuando se tienen datos faltantes menor o igual al 20%, cuando las correlaciones parciales no resultan significativa, se recomienda hacer uso de las correlaciones cruzadas.

Para datos faltantes en más de 20%, se recomienda hacer uso de otros modelos de imputación como los modelos AMI bivariados, dado a que los modelos de regresión lineal no resultan muy robustos para estos casos.

Se deja abierta la posibilidad de explorar la bondad de los modelos de espines cúbicos para la imputación de datos climatológicos.

6. BIBLIOGRAFÍA

- [1] Tomado del libro de *HIDRÁULICA DE AGUAS SUBTERRÁNEAS* María Victoria Vélez Otálvaro Universidad Nacional de Colombia Facultad de minas Sede Medellín 1999
- [2] Tomado y Adaptado de Curso de Análisis Técnico Avanzado online, Lin <http://www.pdf-search-engine.com/analisis-de-correlacion-canonica-pdf.html>
http://www.uco.es/zootecniaygestion/img/pictorex/06_19_26_8_correlacion_multiple.pdf
- [3] Walpole, Myers Myers ye, Probabilidad y estadística para ingeniería y ciencia, 2007, p. 389-402
- [4] (OMM, 1983). Organización Meteorológica Mundial.
- [5] Tomado y adaptado de GUIA DE TRABAJOS PRACTICOS Página 5y 6 Trabajo Práctico N° 2: Procesamiento de los datos de precipitación. Facultad de Ingeniería, Depto de Hidráulica, Universidad Nacional del Nordeste.
- [6] Tomado y adaptado de Bouza Deaño, R.; Ternero Rodríguez, M.*; Fernández Espinosa, A.J. *Departamento de Química Analítica. Facultad de Química. Universidad de Sevilla.c/ Profesor García González, s/n. 41020 Sevilla (España).* Disponible en http://grupo.us.es/ciberico/archivos_word/232b.doc
- [7] Davis and Hinkley, Kendall (1976, eqn 4.4, p.55).Disponible en <http://rss.acs.unt.edu/Rdoc/library/Kendall/html/Kendall.html>.