

Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Recursos Hídricos y
Ambiental



INCERTIDUMBRE Y CONFIABILIDAD DE MÉTODOS DE INTERPOLACIÓN
ESPACIAL EN EL COMPLEMENTO DE REGISTROS HIDROLÓGICOS

TESIS

Que como parte de los requisitos para obtener el grado de

Maestro en Ciencias

Presenta:

Q.A. Juan Manuel Vera Morales

Dirigido por:

Dr. Alfonso Gutiérrez López

SINODALES

Dr. Alfonso Gutiérrez López
Presidente


Firma

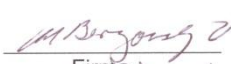
M. en C. Filiberto Luna Zúñiga
Secretario


Firma

Dr. Nabil Mobayed Kohdr
Vocal

Firma

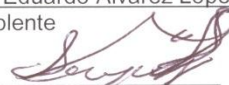
Dr. Moisés Berezowsky Verduzco
Suplente



Firma

Dr. Eduardo Álvarez López
Suplente


Firma

Dr. Aurelio Domínguez González
Director de la Facultad




Dr. Irineo Torres Pacheco
Director de Investigación y
Posgrado

Centro Universitario
Querétaro, Qro.
Noviembre de 2012
México



Hey, todos ustedes que quieren cambiar al mundo,
No se detengan.

El Otro Yo

Para todas esas personas que me apoyaron en la vida



Agradecimientos

Agradezco ante todo a mis padres que han sabido criarme y llevarme adelante toda esta vida, sin ellos no hubiera logrado nada de esto. A mis hermanos que con su lucha y superación me han motivado a hacer lo mismo siempre, a mi familia le doy las gracias hoy y siempre.

A todos los profesores que me enseñaron algo valioso, incluso sin que yo me diera cuenta de cuan necesario puede ser.

A todos mis amigos de este mundo, cercanos o lejanos físicamente, siempre puedo apoyarme en ellos.

Especialmente agradezco a la UAQ por formarme desde bachiller, como Químico Agrícola y ahora como Maestro en Ciencias en Recursos Hídricos, así mismo al CONACYT por su apoyo para el cumplimiento de esta labor de investigación y obtención de grado



RESUMEN

En este trabajo se abordó una evaluación del desempeño estadístico de varios métodos de interpolación espacial, particularmente aplicados al complemento de datos perdidos en diferentes puntos de medición. Los modelos de interpolación evaluados fueron el método de Lagrange, Distancia Inversa, Aproximación multicuadrática, redes de triángulos irregulares, Splines, Kriging y Cokriging. Se obtuvieron bases de datos de catorce estaciones climatológicas con diversas longitudes de registro y se realizó una validación cruzada para evaluar estadísticamente los resultados de los métodos de interpolación mediante análisis de varianza. Además se analizó la validez del complemento de registros mediante el análisis de la mejora de parámetros estadísticos. Los resultados obtenidos mostraron que, a excepción del método de Lagrange, todos los métodos son estadísticamente similares, aunque los métodos de Redes de triángulos irregulares y kriging mostraron el menor error cuadrático medio. Para el análisis de cokriging se presentaron dos modelos, uno que incorporó datos de elevación y otro con datos de un coeficiente k basado en el tiempo de inicio de las tormentas, determinado mediante un análisis de aglomerados, estos modelos de cokriging resultaron no presentar mejora significativa en cuanto al error cuadrático medio en validación cruzada, sin embargo, gráficamente la distribución de las tormentas resultó más adecuada. Este trabajo sugiere que la validez de los métodos de interpolación es aceptable debido a la adecuada extensión espacial y temporal de los datos analizados y que la elección del modelo a utilizar cuando se pretende rellenar datos faltantes dependerá del analista y la facilidad de aplicación de dicho método.

(Palabras clave: Interpolación, kriging, precipitación)



SUMMARY

In this work we presented a statistical performance evaluation of several interpolation methods, these methods were specifically applied to complete different time series with lost data. The evaluated interpolation models were the Lagrange method, Inverse Distance Weighting, Multiquadratic approach, Triangle irregular networks, splines, kriging and cokriging. Different data from fourteen weather stations were obtained and a cross validation model was applied to evaluate the statistical performance of the interpolation methods through analysis of variance (ANOVA). We also evaluated the effectiveness of the data complementation through a statistical analysis of diverse parameters. The results obtained showed that, except for the Lagrange method, all the evaluated methods were statistically similar, though the Triangle Irregular network and kriging methods showed the lower root mean square error. For the cokriging analysis two models were presented, one of them incorporate elevation data and the other one values of coefficient k , based on average time of rainfall events obtained from a cluster analysis, these models showed no significant improvement for the root mean square error in the cross validation compared to the standard kriging method, however, the spatial distribution of the rainfall was improved graphically. This work suggest that the effectiveness of the interpolation methods is acceptable due to the adequate distribution of the rain gauges and long enough time series and that the election of a model to interpolate data, particularly to complete lost data, depend of the analyst and the easiness of application of said method.

(Key words: Interpolation, kriging, rainfall)

ÍNDICE GENERAL

CONTENIDO	Página
AGRADECIMIENTOS	
RESUMEN	
INDICE GENERAL	i
INDICE DE CUADROS	iii
INDICE DE FIGURAS	iv
I. INTRODUCCIÓN	1
II. MÉTODOS DE CORRELACIÓN	2
II.1 Regresión y correlación	2
II.2 Análisis de varianza	5
II.3 Residuales	7
III. ANÁLISIS ESPACIAL	11
III.1 Introducción al análisis espacial	11
III.2 Semivariograma	13
III.3 Anisotropía	15
III.4 Validación y validación cruzada	16
IV. MÉTODOS DE INTERPOLACIÓN ESPACIAL	18
IV.1 Aproximación de La Grange	18
IV.2 Distancia inversa	19
IV.3 Redes de triángulos irregulares (TIN)	20
IV.4 Splines	21
IV.5 Aproximación multicuadrática	24
IV.6 Kriging y cokriging	25
V. MEJORA DE VARIANZA EN COMPLEMENTACIÓN DE SERIES	30
V.1 Criterios de mejora en complementación de series	30

VI. DESAGREGACIÓN DE TORMENTAS POR INTERPOLACIÓN	33
VII. HIPÓTESIS Y OBJETIVOS	35
VIII. METODOLOGÍA	36
VIII.1 Definición del área de estudio	36
VIII.2 Obtención de bases de datos	36
VIII.3 Interpolación espacial de precipitación	37
VIII.4 Validación cruzada	38
VIII.5 Cokriging	39
VIII.6 Mejora de la varianza	40
IX. ANÁLISIS DE RESULTADOS	41
IX.1 Información disponible	41
IX.2 Aplicación de los métodos	41
IX.3 ANOVA	44
IX.4 Mejora de varianza	45
IX.5 Consideración de los métodos de interpolación	46
IX.6 Patrones temporales de lluvia	49
IX.7 Cokriging	54
X. CONCLUSIONES	58
XI. REFERENCIAS BIBLIOGRÁFICAS	60
ANEXO I	62
ANEXO II	77

ÍNDICE DE CUADROS

Cuadro		Página
1	Ejemplos de funciones bivariadas de splines	23
2	Valores de N_1 y \hat{r}_{cm}	31
3	Nombre y ubicación de las estaciones utilizadas.	37
4	Resultados de RMSE y error absoluto para los métodos	42
5	Análisis de varianza del RMSE	44
6	Análisis de varianza del error absoluto	45
7	Análisis de mejora de varianza para las estaciones	45
8	Precipitación total para el análisis de cokriging	54
9	Malla de puntos de elevación para el análisis de cokriging	55
10	Datos valores de k para el análisis de cokriging	55
11	Datos de las estaciones para un ejemplo	63
12	Datos obtenidos para el método de distancia inversa	64
13	Valores calculados para el caso ejemplificado	69

ÍNDICE DE FIGURAS

Figura		Página
1	Representación de un modelo de regresión	3
2	Simulación de datos para diferentes tipos de correlación	5
3	Ejemplos de modelos de regresión y gráfica de residuales	9
4	Zona de estudio y estaciones utilizadas	36
5	Representación de los datos obtenidos por interpolación	43
6	Gráfica de residuales para una tormenta	44
7	Caracterización de tormentas con distribución sigmoidea	50
8	Lluvias con otras distribuciones	51
9	Representación espacial de las horas de inicio y final	53
10	Kriging simple para datos de lluvia diaria en 13 estaciones	56
11	Cokriging usando datos de elevación	56
12	Cokriging usando el coeficiente k	57
13	Puntos de las estaciones para un ejemplo	63
14	Ejemplo de kriging lineal	67
15	Evaluación de distancias a 100	67
16	Evaluación de distancias a 200	68
17	Gráfica de los puntos del variograma experimental	68
18	Variograma ajustado a diferentes modelos	69
19	Puntos teóricos de elevación	72
20	Red de triángulos irregulares	74

I. INTRODUCCIÓN

La información acerca de eventos de lluvia es requerida para el diseño, análisis y evaluación de hidrosistemas. Las características de una lluvia se pueden describir mediante su lámina total, duración y variación temporal de intensidad a lo largo del evento, estas pueden influenciar el diseño de infraestructuras de un hidrosistema. Por tanto la disponibilidad de datos de lluvia es importante para el manejo de los recursos hídricos y para varias aplicaciones hidrológicas como la modelación de erosión y transporte de sedimentos, análisis de inundaciones, modelación de calidad del agua, diseño de estructuras hidráulicas, etc.

Hidrólogos e investigadores siempre se han enfrentado a la frustrante experiencia de la falta de herramientas apropiadas para la medición directa de la precipitación en un área. Dado que el proceso de precipitación presenta un alto grado de variabilidad tanto en el tiempo como en el espacio, se ha tratado antes el problema de que las mediciones directas de la lluvia pueden ser obtenidas solo a una escala muy limitada de espacio, una escala puntual.

Las distribuciones espaciales y temporales de fenómenos físicos pueden ser aproximadas a funciones dependientes de la ubicación en un espacio multidimensional. Ejemplos típicos de estos son elevaciones, fenómenos climáticos, propiedades del suelo, densidades de población, flujos de materia, etc. Aun cuando la mayoría de estos fenómenos son caracterizados por datos medidos puntualmente e irregularmente distribuidos en el espacio y el tiempo, la visualización, análisis y modelaje normalmente necesitan representaciones continuas en un área. Sin embargo, la selección de la función apropiada para modelar la tendencia presenta un problema en análisis de tendencias de superficie ya que hay un enorme rango de funciones candidatas. Además de poder realizar representaciones continuas en áreas, el análisis espacial puede ser utilizado para complementar registros faltantes en los puntos de medición, sin embargo la funcionalidad de esta operación debe ser determinada mediante varios análisis no sólo espaciales sino de mejora de varianza e incertidumbre a la hora de complementar registros.

II. MÉTODOS DE CORRELACIÓN

II.1 Regresión y correlación

Desarrollamos modelos matemáticos para explicar la variabilidad de mediciones hechas a un resultado de interés. El objetivo es tratar de explicar los diferentes valores de este resultado en respuesta a otra información disponible. Variaciones aleatorias en los valores de la respuesta lo hace difícil, lo mejor que podemos hacer es establecer un valor típico o promedio dada una serie de circunstancias. Nuestros datos consisten de pares ordenados de coordenadas $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ con los cuales se pretende obtener un modelo de regresión

$$y = \alpha + \beta x \quad (2.1)$$

Esta ecuación es un modelo matemático para los datos, no existe una razón para que esta línea pase por alguno de los datos. Un modelo es una simplificación de la realidad. La línea provee un resumen intuitivo de la relación de nuestros datos pero no es indicativa de ellos.

Una descripción más exacta de lo que ocurre es escribir una ecuación por separado para cada par de datos (x_i, y_i) . Hay una sola línea para todos los datos, pero también hay que tomar en cuenta el hecho de que los pares de datos individuales no se encuentran justamente en la línea. Algunas observaciones se encontraran por encima y otras debajo de esta, por lo que se puede escribir

$$y_i = \alpha + \beta x_i + r_i \quad (2.2)$$

Donde r_i es la diferencia entre el valor observado y_i y donde este estaba calculado si apareciera justo sobre la línea. A estos valores se les llama residuales ya que representan el espacio que queda después de ajustar la recta. En otras palabras, hay una línea con pendiente beta e intercepción alfa, y a cada valor de x_i el valor observado y_i estará cerca de un valor estimado por la línea $\alpha + \beta x_i$. El residual es la diferencia entre el valor observado y donde se esperaba encontrar sobre la línea. La elección para el “mejor ajuste” es determinado por estos residuales.

Para determinar la línea más apropiada, necesitamos estimar tanto la intercepción alfa y la pendiente beta para esta línea y definir un criterio mediante el cual se obtengan estos estimados. El proceso de estimación es motivado por el planteamiento de los datos mostrados en la Figura 1. En esta figura imaginamos la línea de regresión pasando por tres puntos de los datos, denotados por círculos negros, la línea se muestra punteada. Esta es una ampliación de una pequeña parte de los datos. Para cada una de las observaciones podemos ver dónde estaría si se encontraran sobre la línea. Estos valores o datos estimados aparecen como círculos blancos sobre la línea punteada. Nótese que los estimados se obtienen únicamente como cambios sobre el valor y . El valor estimado no es identificado como el punto más cercano al valor observado. Las implicaciones de esta convención son que los valores de x se asumen como conocidos a un alto nivel de precisión y todas las variables estadísticas se evalúan sobre el eje y .

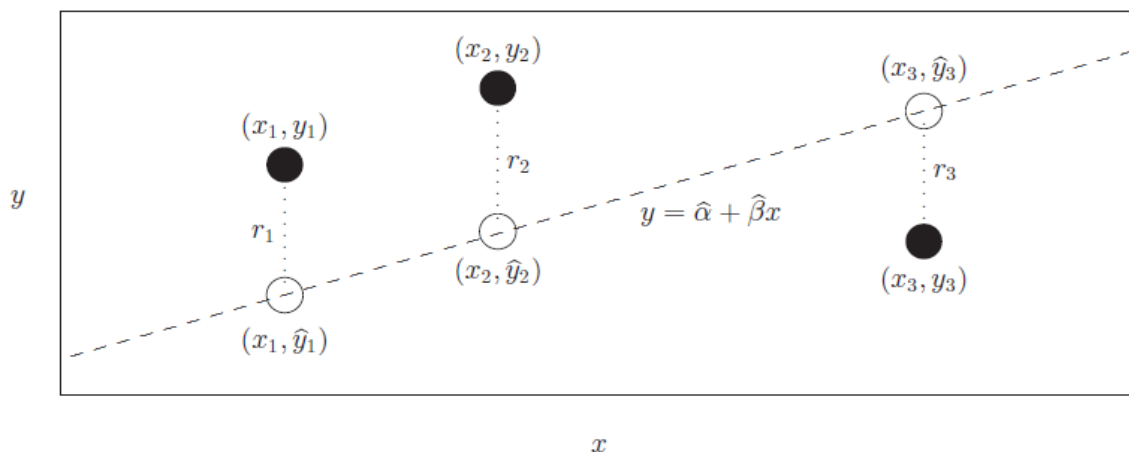


Figura 1. Representación de un modelo de regresión. Los círculos sólidos representan los datos observados mientras que los círculos sin relleno son los valores estimados, r representa el residual para cada dato.

Una de las técnicas estadísticas más comúnmente utilizada para mostrar la correlación entre dos variables, es el coeficiente de correlación, el cual es un número sencillo que resume y expresa la utilidad de una regresión lineal. El coeficiente de correlación es un número adimensional entre -1 y 1. La pendiente y la correlación tienen el mismo signo. Este número es usado para expresar la

fuerza de una relación lineal, por lo que los valores más cercanos a -1 o 1 indican una fidelidad mayor a una relación en línea recta, dándonos una idea de la efectividad de un modelo de regresión (Zelterman, 2010).

La correlación es estandarizada en el sentido de que su valor no depende de las medias o desviaciones estándar de los valores x o y si añadimos o sustraemos los mismos valores de los datos (y por tanto, cambiamos sus medias), la correlación continua siendo igual. Si multiplicamos todos los valores de x (o de y) por un valor positivo, la correlación sigue siendo la misma. Si multiplicamos los valores de x o de y por un número negativo, el signo de la correlación cambiará.

Como con cualquier simplificación de una situación compleja, el coeficiente de correlación tiene sus beneficios y sus desventajas. Una variedad de valores de correlación se ilustran en la Figura 2. Cada una de estas gráficas consta de 50 pares de datos simulados. Una correlación de 0 en la gráfica superior izquierda muestra que no hay indicios de una relación lineal entre las variables graficadas. Una correlación de 0.4 no indica mucha fuerza tampoco. Una correlación de 0.8 o de 0.9 indica una fuerte relación lineal. Nótese que el signo de la correlación es también el signo de la pendiente.

Es posible tener una alta correlación poca o nula tendencia lineal. La correlación de -0.9 en la gráfica inferior izquierda es más el resultado de un pequeño número de observaciones inusuales que la indicación de una tendencia real de los datos. El grupo mayoritario ubicado en la esquina superior izquierda es más representativo de la población y no muestra ninguna tendencia.

Así mismo, una correlación de 0 no indica que los valores medidos no se relacionen, en la gráfica inferior derecha vemos que dos variables tienen una fuerte relación pero esta no es lineal, el coeficiente de correlación es útil solo en medir la fuerza de una relación lineal.

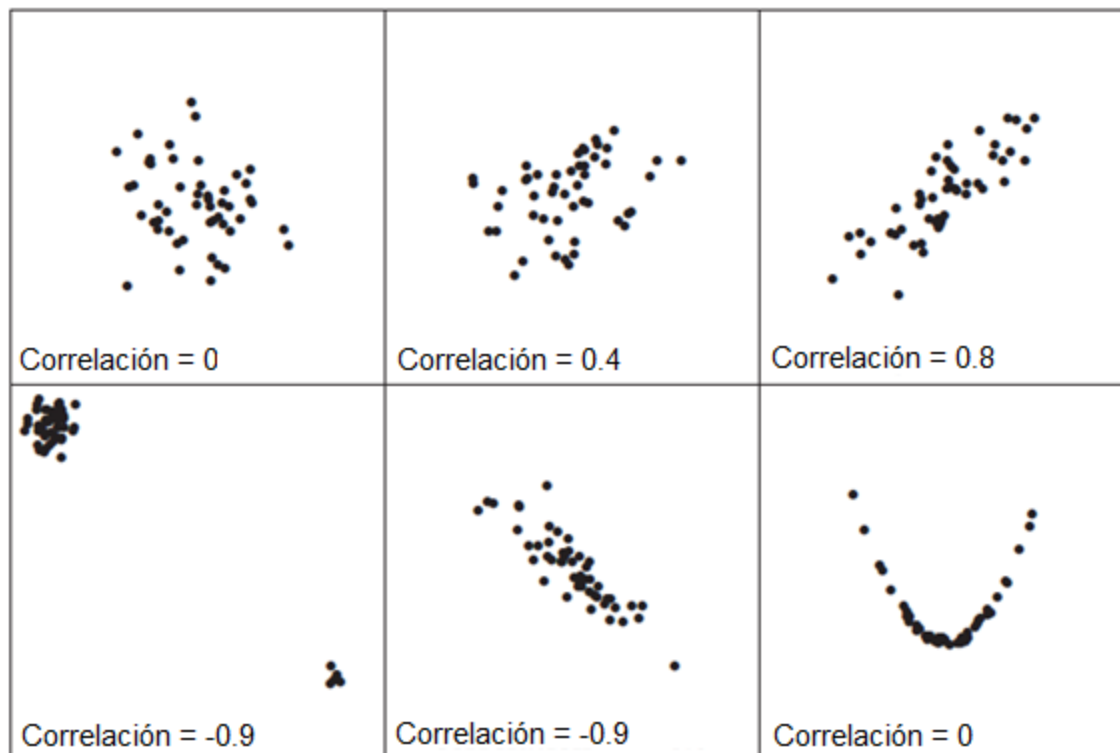


Figura 2. Simulación de datos que ilustran diferentes tipos de correlación.

El coeficiente de correlación es a menudo expresado como su cuadrado y referido como R^2 .

II.2 Análisis de varianza

El análisis de varianza (ANOVA) es la prueba estadística que sirve para demostrar si dos varianzas muestrales provienen de la misma población. Este concepto es muy útil, ya que permite discriminar cuáles variables son significativas en un modelo de regresión, también puede determinarse si dos o más poblaciones son iguales, una vez que son comparadas simultáneamente sus medias poblacionales. En estos dos casos, las poblaciones analizadas deben tener una distribución normal. Esta prueba se basa en que el estadístico de prueba sigue la distribución F .

II.2.1 Distribuciones muestrales

El estadístico F se utiliza como una prueba en varias situaciones. Las principales características de esta distribución son:

1. Existe una familia de distribuciones F . Un elemento específico de la familia está determinado por los dos parámetros que caracterizan a la distribución, los grados de libertad (g.l.) en el numerador y en el denominador
2. El valor de F no puede ser negativo.
3. La distribución F es una distribución continua.
4. La curva que representa una distribución F tiene un sesgo positivo.
5. Sus valores varían de 0 a ∞ y tiene un comportamiento asintótico sobre el eje F .

La función F con $\gamma_1=m$ y $\gamma_2=n$ grados de libertad (m y n son conocidos como numerador y denominador, respectivamente) está dada por:

$$P_F(f) = \Gamma[(\gamma_1 + \gamma_2)/2] \gamma_1^{\gamma_1/2} \gamma_2^{\gamma_2/2} f^{(\gamma_1-2)/2} (\gamma_1 + \gamma_2)^{f(\gamma_1+\gamma_2)/2} / [\Gamma(\gamma_1/2)\Gamma(\gamma_2/2)] \quad (2.3)$$

La media y la varianza de F serán:

$$E(F) = \gamma_1 / (\gamma_2 - 2) \quad (2.4)$$

$$Var(F) = \gamma_2^2 (\gamma_1 + 2) / [\gamma_1 (\gamma_2 - 2) (\gamma_2 - 4)] \quad (2.5)$$

II.2.2 Comparación de dos varianzas poblacionales

La distribución F se utiliza para demostrar la hipótesis de que la varianza de una población normal es igual a la varianza de otra población normal. Así, la prueba es útil para determinar si una población tiene o no más variación que otra. La prueba F también puede usarse para validar supuestos con respecto a ciertas pruebas estadísticas, cabe recordar que independientemente de esto, primero se plantea la hipótesis que se quiere rechazar (H_0 : hipótesis nula).

Para comparar varianzas poblacionales, la hipótesis nula es aquella en la que la varianza de una población normal, σ_1^2 es igual a la varianza de otra población normal, σ_2^2 . Para realizar la prueba, en una muestra aleatoria de N_1

observaciones, y una muestra de N_2 de una segunda población. El estadístico de prueba es S_1^2/S_2^2 , en donde S_1^2 y S_2^2 son las varianzas muestrales respectivas.

Si la hipótesis nula es verdadera ($H_0: \sigma_1^2=\sigma_2^2$), el estadístico de prueba sigue la distribución F con N_1-1 y N_2-2 grados de libertad. La varianza muestral más grande se coloca en el numerador; en consecuencia, F es mayor que 1.0 y el valor crítico del estadístico F se obtiene con el nivel de significancia (α), buscando después el número adecuado de grados de libertad, en una tabla de valores críticos de F para diferentes niveles de significación y diferentes grados de libertad del numerador.

El procedimiento ANOVA, puede ilustrarse de manera más adecuada mediante los pasos siguientes:

- **Hipótesis Nula.** H_0 expresa que no hay diferencias significativas entre las medias poblacionales; es decir $\mu_1=\mu_2=...$. La hipótesis alternativa denotada como H_1 plantea que al menos una media es diferente. En la prueba, si H_0 se rechaza, H_1 se aceptará.
- **Nivel de significancia.** Se selecciona el nivel, por ejemplo al 5% ($\alpha=0.05$)
- **Estadístico de Prueba.** El estadístico de prueba adecuado es la distribución F . Este procedimiento se basa en las consideraciones antes mencionadas. De esta manera, F es la razón de dos varianzas.

II.3 Residuales

Varios modelos matemáticos han sido desarrollados para explicar la variabilidad de mediciones de una variable de interés. El objetivo es intentar explicar los diferentes valores medidos en respuesta a otra información disponible. La variación aleatoria en los valores de la respuesta lo hace difícil, y la mejor aproximación que se puede hacer es especificar los valores típicos o promedios de la respuesta dada una circunstancia específica (Zeltermán, 2010).

Algunos diagnósticos acerca de un modelo de regresión pueden ser obtenidos usando una gráfica de los residuales del modelo de ajuste. Los residuales pueden

ser obtenidos y examinados de manera sencilla, pero el concepto crucial es que estos son obtenidos de una población más grande y no observable. Como con todas las inferencias estadísticas, una muestra (en este caso de residuales) es usada para inferir propiedades de una población más grande. Esta población se refiere a la distribución del error. Los residuales son una muestra observada de errores. Se puede pensar en el modelo de regresión como una serie de distribuciones normales centradas a lo largo de la línea de regresión. Dado un valor de x , hay un valor y en la distribución normal cuya media equivale a $\alpha + \beta x$. La desviación estándar de esta línea de regresión es estimada mediante el error cuadrático medio. Todas las suposiciones que se necesitan hacer en el modelo son concernientes al error. No se pueden observar los errores, y solo se tiene una muestra de los residuales de esa población. El mensaje es que se deben tomar los residuales seriamente. Una de las herramientas más importantes disponibles es la gráfica y el análisis de residuales (Kim y Ahn, 2009).

Esta simple técnica es, por mucho, la más poderosa y sencilla forma de diagnosticar muchos de los problemas que pueden existir en un modelo de regresión. La verdadera dificultad radica en que el análisis debe ser realizado por el usuario, una computadora no lo puede efectuar, es necesario examinar las gráficas e inferir si el modelo es apropiado o no. Los valores residuales son típicamente graficados en el eje vertical, mientras que el eje horizontal puede ser tanto el valor de x como el estimado de y .

La Figura 3 presenta una muestra de cuatro problemas en el análisis de datos que pueden surgir cuando se realiza una regresión lineal. En muchos casos, la gráfica de residuales puede ampliar el problema y hacerlo más fácil de apreciar y diagnosticar. En el inciso a de la Figura 3 se puede ver que la relación entre x y y no es una línea recta. La grafica de residuales amplía la curvatura en la relación. Estos residuales claramente no representan simplemente ruido y la tendencia curvada es más clara en la figura a la derecha. En la segunda pareja de graficas en la Figura 3, podemos ver un ejemplo de la violación del supuesto de una

varianza constante. El grafico de residuales tiene una forma de embudo, indicando una variabilidad que se incrementa a valores altos de x .

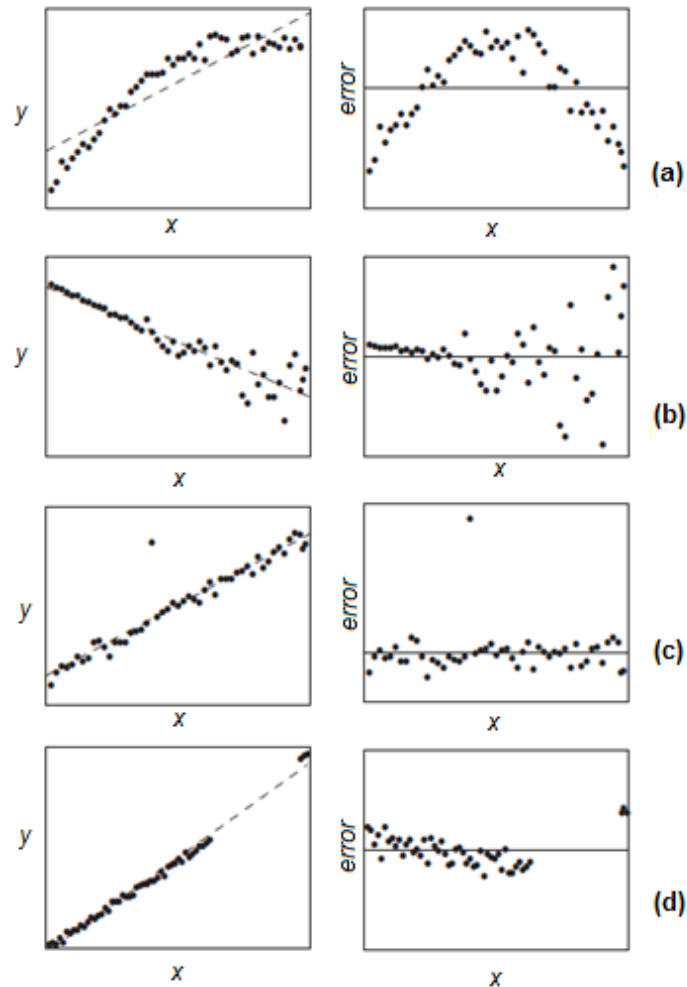


Figura 3. Ejemplos de modelos de regresión y grafica de sus residuales.

La línea de regresión estará altamente influenciada por las observaciones con alta variabilidad dado que estas observaciones tienden a aparecer más lejos de la línea. Como resultado, se puede ver que el modelo lineal tiene un ajuste muy pobre para valores pequeños de x a la izquierda donde el modelo es más apropiado. Un valor atípico es una observación que no sigue la forma del modelo por varias razones. Algunas veces estos son simples errores en los datos. Estos pueden ser reemplazados o la observación puede ser omitida. En otras situaciones estos datos pueden tener un gran valor significativo y puede proveer

una nueva perspectiva acerca del fenómeno. Es cuestión del analista decidir la naturaleza de estos datos atípicos, pero siempre deben ser identificados. En la Figura 3(d) se ilustra el efecto de observaciones influyentes. Estas no están necesariamente alejadas de la línea de regresión, pero sus valores inusuales causan que estos ejerzan una gran influencia en los valores ajustados de la pendiente y la ordenada al origen. Se dice que dichas observaciones tienen un alto apalancamiento. En la Figura 3 se aprecia una pendiente con tendencia hacia abajo en el grupo de observaciones a la izquierda en la gráfica de residuales. Los puntos con alto apalancamiento a la derecha dirigen la línea de regresión hacia ellos. Esto provoca que el resto de las observaciones no se ajusten muy bien al modelo (Zelterman, 2010).

III. ANÁLISIS ESPACIAL

III.1 Introducción al análisis espacial

La distribución espacial de las variables hidrológicas es sin duda uno de los temas de mayor importancia en el estudio de los componentes del ciclo hidrológico en la actualidad. Por ejemplo, para describir la distribución espacio-temporal de las precipitaciones es necesaria una técnica para cartografiar los campos de lluvia (Blanchet, 1981; Creutin y Obled, 1982) y un estudio detallado de la variabilidad de la lluvia (Rodríguez Iturbe et al., 1984; Huber y Iroumé, 2001). Como las precipitaciones presentan características ligadas en todo momento a la topografía y al relieve, se deben utilizar técnicas de modelación geoestadísticas, tal como la modelación con variogramas direccionales, bajo la hipótesis de que un variograma direccional puede reflejar en forma de variables espaciales la forma y comportamiento en el espacio de un evento específico (Haiden et al., 1992).

El nombre geoestadística viene de su aplicación común a la geología. El término es de alguna forma desafortunado, sin embargo, ya que la metodología es muy general, y apropiada para el esfuerzo de entender el tipo y la extensión de la correlación espacial, y para usar este conocimiento para realizar predicciones eficientes. La situación básica es la que sigue:

El proceso de interpolación espacial consiste en la estimación de los valores que alcanza una variable Z en un conjunto de puntos definidos por un par de coordenadas (x,y) , partiendo de los valores de Z medidos en una muestra de puntos situados en el mismo área de estudio (por ejemplo la interpolación de la temperatura a partir de datos de observatorios meteorológicos), la estimación de valores fuera del área de estudio se denomina extrapolación. En algunos casos pueden utilizarse otras variables de apoyo V de las que se conoce su variación espacial en el proceso de interpolación/extrapolación (en el caso de la interpolación de la temperatura puede ser la altitud ya que aquella depende de esta. El área de estudio vendría definida, aunque no de forma muy clara, por el entorno de los puntos en los que sí se dispone de datos (Mitas y Mitasova, 1993).

Un proceso de interés es observado en n lugares s_1, \dots, s_n . Estas n observaciones se denotan como Z_{s_1}, \dots, Z_{s_n} , aunque a veces pueden ser denotadas como Z_1, \dots, Z_n . Esta notación es normalmente abreviada, ya que normalmente debemos conocer las ubicaciones específicas s_1, \dots, s_n para realizar cualquier estimación o inferencia. Las principales metas son:

- a) Predecir el proceso en ubicaciones no observadas.
- b) Entender la relación entre valores en ubicaciones “vecinas”

Basado en la suposición de que las mediciones de pluviógrafos pueden medir confiablemente un “punto de precipitación” después de tomar en cuenta un cierto número de posibles errores, los estimados de precipitación en un área han sido obtenidos tradicionalmente a través de algunos métodos de interpolación y técnicas de agregación. Estos están basados en la hipótesis de que los estimados de precipitación en sitios sin medición pueden ser obtenidos como combinaciones lineales o no lineales de los valores medidos en lugares específicos usando las ponderaciones adecuadas. La mayoría de estas aproximaciones se originaron desde los métodos subjetivos de las isoyetas a las interpolaciones geométricas más objetivas basadas en los polígonos de Thiessen, o las interpolaciones matemáticas basadas en *splines*, hasta los paradigmas de la geoestadística, la nueva rama de teoría estadística introducida por Matheron en el campo de la ingeniería de minas. Por otro lado, estimados indirectos de la precipitación en un área basados en la medición de variables auxiliares ha sido usada desde finales de los 60s mediante Radares terrestres y aparatos de medición remota mediante satélites con grados de confiabilidad variados. Tales métodos indirectos, cuya significancia y confiabilidad para propósitos hidrológicos aún está siendo determinada, son frecuentemente aceptados como indicadores de patrones de lluvia a escalas amplias y utilizados en varias aplicaciones. Sin embargo, los métodos indirectos deben ser primeramente calibrados y validados usando datos históricos y entonces, ya en operación, ajustados continuamente contra mediciones directas en la superficie (como pluviómetros). Finalmente los modelos numéricos atmosféricos de base física (aunque dependientes de varias

aproximaciones teóricas) proveen predicciones de valores temporales de acumulación para áreas de precipitación en escalas espaciales muy grandes. Los datos de salida de tales modelos poco tiene que ver con las mediciones reales de precipitación; más bien proyectan en el futuro el análisis de datos del estado presente de la atmosfera, lo que permite el pronóstico de lluvia con un tiempo de hasta algunos días. La precisión de los valores pronosticados depende tanto de la calidad del modelo como de la extensión del tiempo, debido a posibles divergencias debido a la caótica naturaleza de la atmosfera. Dependiendo de las mediciones disponibles y las aproximaciones en los modelos así como de la naturaleza de la aplicación, los hidrólogos tienden a reemplazar las observaciones no disponibles de la lluvia en un área a las escalas requeridas de tiempo, con sustitutos apropiados basados en técnicas de interpolación o desagregación. En general, la interpolación es aplicada cuando hay pluviógrafos o radares disponibles, mientras que la desagregación se aplica cuando se usa una medición indirecta de otros aparatos de medición remotos, o valores predichos mediante modelos atmosféricos, los cuales están normalmente disponibles a escalas más amplias que las requeridas en la mayoría de las aplicaciones hidrológicas. Sin embargo, no se puede aplicar una clara diferenciación ya que ambas pueden ser incorporadas en una única aproximación, por ejemplo al analizar alternadamente la información obtenida mediante mediciones remotas y datos de pluviometría. (Fiorucci et al., 2001; Lanza et al., 2001; Todini, 2001)

III.2 Semivariograma

El caso del espacio es similar al del tiempo pero algo más complejo. El espacio es bidimensional y no hay una dirección preferente. Esto implica que los valores de autocorrelación no tienen por qué ser los mismos en todas las direcciones.

Otra dificultad es que mientras que en una serie temporal disponemos de todos, o casi todos, los datos medidos como valores agregados (diarios, mensuales, etc.); en un estudio espacial sólo disponemos de los valores medidos en algunos puntos de muestreo en un espacio infinito, por tanto la estimación de medias, desviaciones típicas y covarianzas no es muy fiable. Por ello se prefiere utilizar

otro estadístico alternativo al coeficiente de correlación y que no se basa en estos parámetros, es el semivariograma.

El semivariograma nos da una visión de cuál es la estructura de variabilidad espacial de una variable medida en un conjunto de puntos, para su cálculo se siguen los siguientes pasos:

1. Determinar todos los posibles pares de puntos (en total $\sum_{i=1}^{n-1} i$ donde n es el tamaño de la muestra).
2. Para cada par (i, j) , anotar la distancia $(d_{i,j})$ entre los puntos y el cuadrado de la diferencia de los valores $(dZ^2 = (Z_i - Z_j)^2)$.
3. Establecer una serie de valores críticos de distancia (h) , por ejemplo 500, 1000, 1500, 2000 m, y asignarles los pares cuya distancia d sea más similar, de manera que cada valor de h recibe n_h pares.
4. Calcular para cada valor de h la función semivariograma como:

$$\gamma(h) = \frac{\sum_{k=1}^{n_h} dZ_k^2}{2n_h} \quad (3.1)$$

donde k hace referencia a cada uno de los pares.

Calculando para diferentes valores de h obtenemos un semivariograma experimental. Lógicamente para valores de h pequeños la función $\gamma(h)$ alcanzará valores pequeños (puesto que los valores en puntos cercanos tenderán a ser más parecidos) y conforme aumenta h aumentará $\gamma(h)$. El valor de $\gamma(h)$ suele alcanzar un máximo (denominado meseta). La representación gráfica (h en abcisas y $\gamma(h)$ en ordenadas) proporciona un resumen de la estructura de variación de la variable.

Este gráfico nos proporciona información de gran utilidad para interpolar nuevas capas de información espacial a partir de datos puntuales. Nos da una indicación de como la distancia entre los puntos influye sobre las diferencias en los valores

de las variables medidas en ellos. La interpretación del semivariograma utiliza tres conceptos fundamentales:

- efecto pepita, es el valor del semivariograma en el origen. Resulta del componente aleatorio, no correlacionado espacialmente, que experimenta cualquier variable espacial. Se denomina así por las pepitas de
- meseta, es el valor máximo que adopta el semivariograma para distancias elevadas más allá de las cuales no hay autocorrelación espacial.
- rango, es la distancia a la que se alcanza la meseta. Puede asimilarse a la distancia más allá de la cual dos medidas pueden considerarse independientes

III.3 Anisotropía

Los variogramas son ejemplos de modelos isotrópicos, en los cuales la estructura de correlación (en particular el variograma) no difiere con la orientación. El variograma depende sólo de la distancia de separación h y no de la orientación del segmento lineal que conecta los dos puntos; esto es, la diferencia cuadrada promedio entre dos muestras a una distancia h es la misma ya sea en dirección horizontal y vertical. Sin embargo, hay varios casos donde la estructura de una cierta variable depende de la dirección (anisotropía). El mejor ejemplo de estructuras anisotrópicas puede ser encontrado en formaciones estratificadas. Considérese por ejemplo, una unidad aluvial formada por la deposición de capas de grava, arena y limo. Las capas y depósitos de los diversos materiales están orientadas en dirección horizontal, como resultado, la diferencia de conductividad hidráulica medida en dos puntos a corta distancia h en la horizontal es más pequeña que la diferencia medida en la vertical.

La anisotropía es normalmente el resultado de dinámicas en el sistema, en el caso de la hidrometeorología la estructura de la precipitación es frecuentemente anisotrópica como resultado de los vientos predominantes en una cierta dirección o por efecto de la orografía.

Los variogramas direccionales experimentales son una herramienta muy útil en la exploración del grado de anisotropía en los datos. Para graficarlo, consideremos el variograma directo. Para dos dimensiones, además de la distancia de separación, se computa el ángulo de dirección (dónde los ángulos que difieren 180 grados se consideran iguales) para cada par de mediciones. Los pares de datos son agrupados con respecto a la orientación. Entonces un variograma experimental direccional es graficado para cada grupo de diferentes ángulos. Esto es, se grafica un variograma experimental para el intervalo de -90° a 0° y otro para el intervalo de 0° a 90° . Comúnmente, en dos dimensiones, dos direcciones ortogonales (aquellas cuya orientación difiere 90°) son seleccionadas.

Diferencias significativas entre los variogramas direccionales pueden ser indicativas de anisotropía, y las direcciones ortogonales donde el contraste es más evidente pueden ser tomadas como las principales direcciones de la anisotropía. Sin embargo, las mismas diferencias pueden ser indicativas de una desviación. Por tanto, el variograma direccional es una técnica exploratoria que puede ser usada para sugerir modelos que requieren análisis posteriores.

III.4 Validación y validación cruzada

Para verificar la calidad de un mapa interpolado debe utilizarse un conjunto de validación formado por una serie de puntos de muestreo (de los que por tanto se conoce el valor real) en los que se va a hacer una estimación de dicho valor real (sin utilizar por supuesto el valor medido en ellos). La diferencia entre el valor medido y el estimado es el error de estimación en ese punto. De este modo a cada punto de validación se asigna un error. El conjunto de los errores debe tener las siguientes características:

1. Media de errores y media de errores al cuadrado próxima a cero
2. Los valores de error deben ser independientes de su localización en el espacio y no estar autocorrelacionados

3. La función de distribución de los errores debe aproximarse a la distribución normal

El problema es que en muchos casos se dispone de pocos puntos de muestreo, por lo que no resulta conveniente reservar algunos de ellos como puntos de validación, la alternativa es el procedimiento de validación cruzada Este consiste en la estimación del valor de la variable Z , con el procedimiento de interpolación que quiere validarse, en cada uno de los puntos de muestreo, aunque sin incluir dicho punto de muestro. De esta manera se conoce para cada punto de muestreo tanto el valor real como el valor estimado de forma que puede llevarse a cabo el análisis estadístico de errores antes mencionado.

IV. MÉTODOS DE INTERPOLACIÓN ESPACIAL

Los métodos geoestadísticos han demostrado una gran confiabilidad en la estimación de campos de lluvia (Creutin y Obled, 1982; Lebel y Laborde, 1988; Holawe y Dutter, 1999; Goovaerts, 2000). Por ejemplo, se ha demostrado que las isoyetas tradicionales calculadas con variogramas direccionales lineales, no representan la variabilidad de las lluvias producidas por fenómenos extremos como los huracanes. Sin embargo, cuando se utiliza una función que describe la distribución espacial de la precipitación (variograma) los resultados mejoran notablemente. Estos campos de precipitación asociados a las trayectorias de huracanes o simplemente la precipitación intensa de corta duración en zonas urbanas, son útiles, por ejemplo, para Protección Civil del estado y para la Comisión Nacional del Agua, pues son las instituciones encargadas de salvaguardar los bienes nacionales y la vida de las personas en situaciones de emergencia.

IV.1 Aproximación de Lagrange

Esta aproximación es referida algunas veces como "Interpolación de Lagrange" y, es una técnica de interpolación confiable para ciertos casos; fue descubierto por el gran matemático francés Joseph Lagrange (1736-1813). Para la aplicación de esta técnica, los coeficientes a_k son evaluados de modo que el proceso P pudiera pasar a través de los valores observados en las estaciones. Por lo tanto, esta aproximación requiere que el número de monomios sea igual al número de estaciones, es decir que $m = n$.

$$P_j = \sum_{k=1}^n a_k \varphi_k(x_j, y_j) \quad ; \quad j = 1, \dots, n \quad (4.1)$$

Los coeficientes a_k se estiman como sigue

$$a_k = \sum_{j=1}^n \beta_{kj} P_j \quad ; \quad k = 1, \dots, n \quad (4.2)$$

Donde β_{kj} elementos de la matriz inversa, de $n \times n$, con elementos $\varphi(X_j, Y_j)$ para $k = 1, \dots, n$ monomios (filas) y $j = 1, \dots, n$ estaciones (columnas), así, sustituyendo los valores obtenidos de a_k , la expresión resultante para estimar P_e es:

$$P_e = \sum_{j=1}^n \left[\sum_{k=1}^n \beta_{kj} \varphi_k(x_e, y_e) \right] P_j \quad (4.3)$$

y la expresión entre corchetes son los factores de peso

$$w_j = \sum_{k=1}^n \beta_{kj} \varphi_k(x_e, y_e) \quad (4.4)$$

IV.2 Distancia Inversa

Es uno de los más simples y fácilmente disponibles métodos. Está basado en el concepto de la primera ley de Tobler de la geografía, que dice que todo está relacionado con el resto pero las cosas más cercanas se relacionan más que las lejanas, por lo que existe la suposición de que el valor en un punto no muestreado puede ser aproximado como un promedio ponderado de valores en puntos dentro de una cierta distancia, o de un número dado de m puntos cercanos. En este método los pesos son inversamente proporcionales a una potencia de la distancia denominada α :

$$\hat{R}_p = \sum_{i=1}^N w_i R_i \quad (4.5)$$

$$w_i = \frac{d_i^{-\alpha}}{\sum_{i=1}^N d_i^{-\alpha}} \quad (4.6)$$

Donde \hat{R}_p representa el dato desconocido; R_i la cantidad de puntos medidos, w_i el peso de cada estación, d_i la distancia de cada punto medido al punto deseado, α es la potencia, un parámetro de control normalmente tomado como igual a dos. Varias investigaciones se han realizado variando dicho parámetro para examinar su efecto en la distribución espacial de la información, sin embargo, sigue siendo dos el número más usado para el método.

IV.3 Redes de triángulos irregulares (TIN)

Este método utiliza un mosaico triangular formado por los puntos de medición para derivar una función bivariada para cada triángulo que entonces es usado para estimar los valores de sitios no muestreados. La interpolación lineal usa facetas planas ajustadas a cada triángulo. Funciones no lineales mezcladas (como polinomios) usan condiciones de continuidad adicionales en las derivadas de primer orden o en primero y segundo orden, asegurando una suavidad en las conexiones de los triángulos y diferenciabilidad de las superficies resultantes. Debido a su naturaleza local, este método es normalmente rápido, con una fácil incorporación de discontinuidades y características estructurales. Una triangulación apropiada que respete la geometría de las superficies es crucial. Aun cuando las TIN proveen una representación efectiva de superficies útiles para varias aplicaciones, la interpolación basada en TIN, especialmente en su versión lineal, es de los métodos menos precisos.

Este método de interpolación requiere primero la creación de la red triangular o TIN. Una vez que los triángulos han sido construidos, se describe la superficie interior del triángulo por una función lineal derivada de las altitudes de cada uno de los puntos de muestreo que constituyen los vértices del triángulo. Esta función permite calcular la altitud en cada punto del triángulo de manera que el valor Z^* estimado pueda ser transcrito tanto como un atributo en la base de datos de un programa que maneja las estructuras de datos vectoriales, o como la cota "Z" en la red regular de un programa que maneja las estructuras de datos matriciales. Formalmente

$$Z^* = a + bx + cy \quad (4.7)$$

donde a, b, c son coeficientes a determinar mientras que x, y son las coordenadas cartesianas de abscisa y de ordenada del punto a interpolar, y ya que se tiene tres vértices conocidos para cada triángulo, los valores de los coeficientes a, b y c se obtiene a través de un sistema de tres ecuaciones:

$$\begin{cases} Z_1 = a + bx_1 + cy_1 \\ Z_2 = a + bx_2 + cy_2 \\ Z_3 = a + bx_3 + cy_3 \end{cases} \quad (4.8)$$

Donde x_n y y_n indican las coordenadas de los vertices del triángulo y Z_n indica el valor medido en dicho punto. La triangulación va siempre a respetar el valor Z cuando la posición de un punto de muestreo coincide con aquel de un nodo de red o píxel a interpolar. Esta propiedad es muy deseable en todo método de interpolación. Por el contrario, la interpolación por triangulación es incapaz de “predecir” la existencia de cimas o depresiones locales salvo aquellas descritas explícitamente en los datos medidos. Esta forma de “predicción” es otra propiedad deseable.

IV.4 Splines

Este método de interpolación trata de producir la superficie continua de mínima curvatura sobre la cual todos los puntos de observación están anclados. Superficies aún menos curvadas son por ejemplo las superficies de tendencia basadas en una regresión polinomial de mínimos cuadrados o bien las superficies planas producidas por las redes triangulares TIN. Sin embargo, las superficies de tendencia no tratan de reproducir los valores observados con exactitud, mientras que las superficies en TIN no son superficies con continuidad matemática.

El método de la mínima curvatura está inspirado por el principio de las reglas flexibles (“thin plates”) en las cuales la presión ejercida en un punto dado produce una reacción (deformación) que afecta un ambiente local determinado por ciertos puntos fijos o fronteras. En el espacio 2D, este ambiente local es conocido como una placa o pieza que formando parte del mosaico total de la superficie interpolada. La unión entre las diferentes placas del mosaico debe definir una superficie continua, derivable al primer y al segundo grados (pendiente y curvatura).

La aproximación de una superficie tal es una extensión de las funciones splines, equivalentes matemáticas de las reglas flexibles utilizadas por los dibujantes. Son funciones polinomiales seccionadas en las cuales aquellas de tercer grado

reproducen bastante bien las deformaciones de las reglas flexibles. Una vez extendidas al espacio 3D, estas funciones son llamadas bicubic splines o B-splines.

La aproximación por B-splines o por la mínima curvatura es una solución de interpolación que tiene como finalidad estimar el conjunto de los valores Z^* para los nodos de una red de puntos regular, de manera que la estimación para el punto de la red se acerca al valor observado Z cuando la posición del punto de observación se acerca a la posición del punto de la red. Este enfoque de interpolación no obliga a conocer de manera explícita la función continua en el espacio 3D que reproduce la superficie real. En el primer caso, la interpolación está hecha a partir de una solución numérica por diferencias finitas, mientras que en el último, se debe encontrar una solución analítica a través del cálculo diferencial.

Desde una perspectiva diferencial, el problema a resolver consiste en encontrar la deformación inducida sobre la “regla flexible” por las fuerzas que actúan sobre puntos discretos, de manera que los desplazamientos hacia estos puntos sean iguales a las observaciones Z conocidas, el trabajo sobre splines para el análisis espacial fue principalmente iniciado por Wahba en 1980.

Smith y Wessel (1990) mencionan que el método de interpolación natural para la curvatura mínima es la función spline bicubica, pero este procedimiento produce oscilaciones y extraños puntos de inflexión sobre todo en las regiones con déficit de datos. Hutchinson (1989) introdujo mejoras importantes en el procedimiento al que llamó “Thin-plate smooth spline”:

- a) agregando términos de la primera derivada en la condición de variación, lo que minimiza la sobreestimación
- b) incorporando un algoritmo considerando la red de drenaje
- c) utilizando una red de resoluciones múltiples para reducir el tiempo de cálculo.

Ya que aproximación mediante splines, que se puede denominar aproximación variacional, está basada en la suposición de que la función de interpolación debe pasar a través (o cerca de) los puntos medidos y, al mismo tiempo, debe ser lo más

suave posible, se pueden combinar ambas condiciones en una sola formulación para minimizar la suma de las desviaciones de los puntos medidos y la seminorma de suavidad para la función spline:

$$\sum_{j=1}^N |z_j - F(r_j)|^2 w_j + w_0 I(F) = \text{minimo} \quad (4.9)$$

Donde w_j y w_0 son pesos positivos y $I(F)$ denota la seminorma de suavidad, la solución de la ecuación (4.8) puede ser expresada como la suma de dos componentes:

$$F(r) = T(r) + \sum_{j=1}^N \lambda_j R(r, r_j) \quad (4.10)$$

Donde $T(r)$ es una función “de tendencia” y $R(r, r_j)$ es una función base que tiene una forma dependiente de la elección de $I(F)$, ejemplos de esta se muestran en el Cuadro 1.

Cuadro 1. Ejemplos de funciones biariadas de splines.

Método	$I(F)$
Membrana	$\int [F_x^2 + F_y^2] dr$
Mínima curvatura	$\int [F_{xx}^2 + F_{yy}^2] dr$
Thin Plate Spline	$\int [F_{xx}^2 + F_{yy}^2 + 2F_{xy}^2] dr$
Thin Plate Spline + Tensión	$\int [\varphi^2 [F_x^2 + F_y^2] + [F_{xx}^2 + \dots]] dr$
Thien Plate Spline regularizada	$\int [[F_{xx}^2 + \dots] + \tau^2 [F_{xxx}^2 + \dots]] dr$
Spline regularizada con tensión	$\sum_{mn} c_{mn}(\varphi) \int [F_x^n y^m]^2 dr$

Debido a lo complejo de los modelos de Splines, existen paquetes informáticos comerciales que realizan estos análisis, sin embargo, es importante saber qué métodos y modelos ocupan estos paquetes, por ejemplo, en el programa Surfer, el método de la mínima curvatura fue definido siguiendo el enfoque de Briggs (1974) et Smith et Wessel (1990). Por otra parte, los cálculos de las derivadas de terreno utilizan el enfoque sugerido por Mitásová et Hofierka (1993) y Mitásová et Mitás, (1993).

IV.5 Interpolación multicuadrática

Aun cuando no es obtenida por una aproximación variacional, la interpolación multicuadrática tiene una formulación y desempeño similar a los splines. Ofreciendo alta exactitud, diferenciabilidad, formulación d-dimensional, y, mediante segmentación, es aplicable a grandes bases de datos. (Hardy 1990).

En la Interpolación Multicuadrática, la influencia de cada estación de medición es representada por superficies cuadradas de conos como función de sus coordenadas. La estimación para un punto dado (X_e, Y_e) se obtiene por la suma de las contribuciones de todas las superficies. Esto se expresa matemáticamente como

$$P_e = \sum_{i=1}^n C_i d_{ei} \quad (4.11)$$

donde C_i es el coeficiente multicuadrático del punto (X_i, Y_i) , d_{ei} es la distancia entre los puntos (X_e, Y_e) y (X_i, Y_i) .

Para estimar los coeficientes C_i y expresar la ecuación (4.10) en términos de los factores de pesos, se usará esta ecuación para cada punto (X_j, Y_j) como

$$P_j = \sum_{i=1}^n C_i d_{ij} ; j = 1, 2, \dots, n \quad (4.12)$$

y los coeficientes C_i se determinan como sigue

$$C_i = \sum_{j=1}^n \delta_{ij} P_j \quad ; \quad i = 1, 2, \dots, n \quad (4.13)$$

Donde δ_{ij} es un elemento de la matriz inversa de $n \times n$, cuyos elementos son las distancias d_{ij} entre las n estaciones. Sustituyendo la ecuación (4.12) en la ecuación (4.6) tenemos

$$P_e = \sum_{j=1}^n \left[\sum_{i=1}^n \delta_{ij} d_{ei} \right] P_j \quad (4.14)$$

y nuevamente los factores de peso para cada estación son

$$W_j = \sum_{i=1}^n \delta_{ij} d_{ei} ; j = 1, 2, \dots, n \quad (4.15)$$

IV.6 Kriging y cokriging

El método de interpolación por Krigeage o “kriging” es definido como “un método de interpolación por medias ponderadas donde el conjunto de los pesos asignados a los puntos de muestreo minimiza la varianza de estimación, que es calculada en función de un modelo de variograma y de los emplazamientos relativos de los puntos, unos con respecto a otros, así como con relación al punto de estimación”

El Krigeage es en cierto sentido, un método de interpolación basado en el principio de las medias móviles, pero es capaz de responder a ciertas consideraciones inaccesibles con los otros métodos: (a) ¿Cuál es el largo del radio de búsqueda?; (b) ¿Cuál forma y orientación debe haber para una interpolación óptima –isotropía versus anisotropía-?; (c) ¿Existe una manera óptima para estimar el peso que no sea una simple función de la distancia?; (d) ¿Cuál es el error (incertidumbre) asociado al valor interpolado?.

El Krigeage es un método de interpolación local dicho óptimo, en el sentido en como utiliza la programación dinámica para escoger los pesos de interpolación de manera a obtener la mejor estimación lineal insesgada del valor de la variable en un punto cualquiera.

El método reposa sobre el principio que dice que la variación espacial de una propiedad (lo que se llama una variable regionalizada) es muy irregular para ser modelada por una simple función matemática, pero que puede ser descrita por una superficie estocástica. La interpolación debe pues, primeramente, explorar y seguidamente modelar el aspecto estocástico de la variable regionalizada. La información que resulta es entonces utilizada para estimar los pesos de la función de interpolación.

La teoría de las variables regionalizadas (Matheron, 1971) asume que la variación espacial de una variable cualquiera puede ser obtenida por la adición de tres componentes mayores:

- a) la componente estructural, expresada con un valor medio constante o con una tendencia constante
- b) una componente aleatoria espacialmente correlacionada; y
- c) un ruido aleatorio o error residual.

Los componentes b) y c) se presumen constantes en toda la zona de muestreo.

Sea x la posición en 1, 2 ó 3 dimensiones, el valor de la variable regionalizada Z en x es:

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon'' \quad (4.16)$$

donde $m(x)$ es una función *determinística* que describe la componente estructural de Z en la posición x , $\varepsilon'(x)$ es el término que denota los residuos de $m(x)$ que son de naturaleza estocástica, localmente variables y espacialmente correlacionadas y ε'' es un ruido residual, gaussiano, espacialmente independiente teniendo por media cero y varianza s^2 .

La primera etapa en el *Kriging* consiste en escoger una función adecuada para el término $m(x)$. En el caso más simple, cuando no existe tendencia en los datos, $m(x)$ es igual al valor medio de la zona de muestreo. En este caso, la *diferencia media* o diferencia esperada entre dos posiciones cualquiera x y $x+h$ separadas por un vector de distancia h , será:

$$E[Z(x) - Z(x + h)] = 0 \quad (4.17)$$

En tal caso, se puede asumir también que la varianza de las diferencias depende solamente de la distancia h , entre los sitios, sea:

$$E\{[Z(x) - Z(x + h)]^2\} = E\{[\varepsilon'(x) - \varepsilon'(x + h)]^2\} = 2\gamma(h) \quad (4.18)$$

donde $\gamma(h)$ es una función conocida bajo el nombre de *semi-varianza* (la mitad de la *varianza*).

Las condiciones establecidas por las ecuaciones corresponden a la hipótesis intrínseca de la teoría de las variables regionalizadas. De acuerdo a esta, una vez que la componente estructural de los datos es eliminada, la variación restante es homogénea y las diferencias entre estos sitios están en función de la distancia que los separa. Si las dos condiciones especificadas en la hipótesis intrínseca son respetadas, la semivarianza puede ser establecida a partir de los puntos comprendidos en la muestra:

$$\gamma(h) = \frac{1}{2n} \sum_{i+1}^n (Z_{x_i} - Z_{x_{i-h}})^2 \quad (4.19)$$

donde n es el número de pares de puntos separados por la distancia h . La distancia entre los puntos h , es también llamada el corrimiento o "*lag*". Un gráfico de $g(h)$ sobre h es nombrado *semivariograma*. La construcción de este último constituye una etapa fundamental para establecer los pesos a utilizar al momento de la interpolación.

De la misma manera que el método del inverso de la distancia, el *Kriging* requiere del cálculo de la suma ponderada de los puntos de muestreo en el vecindario del sitio a evaluar, sea:

$$Z^* = \sum_{i=1}^n w_i z_i \quad (4.20)$$

pero en el *Kriging*, la suma de los pesos w_i debe ser igual a 1. Los pesos son calculados a partir de un conjunto de $n + 1$ de ecuaciones lineales simultáneas,

donde n es el número de puntos utilizados para la estimación en una localización cualquiera. En forma matricial, las ecuaciones de los pesos son:

$$C \cdot w = d \quad (4.21)$$

que pueden ser distribuidas en:

$$\begin{bmatrix} c_{11} & c_{12} & \cdot & c_{1n} & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & \cdot & c_{nn} & 1 \\ 1 & 1 & \cdot & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ w_n \\ \mu \end{bmatrix} = \begin{bmatrix} c_{10} \\ \cdot \\ \cdot \\ c_{n0} \\ 1 \end{bmatrix} \quad (4.22)$$

donde los términos en la matriz C son los valores de las *covarianzas espaciales* entre los puntos de la muestra y los términos en d son también de las covarianzas espaciales, pero entre los puntos de la muestra y el punto a estimar. La covarianza espacial es una medida de la correlación espacial o autocorrelación.

El término μ es un parámetro conocido como el multiplicador de Lagrange, un parámetro flexible utilizado para asegurarse que la suma de los pesos sea igual a 1. La ecuación (4.17) es resuelta premultiplicando los términos de los dos lados de la igualdad por el inverso de C :

$$C^{-1} \cdot C \cdot w = C^{-1} \cdot d \quad (4.23)$$

$$w = C^{-1} \cdot d \quad (4.24)$$

Para cada nodo que debe ser estimado al momento de la creación de la red, las matrices de las covarianzas C y d deben ser calculadas y la ecuación (4.20) resuelta para obtener los pesos. Estos son después aplicados en la ecuación (4.16) para calcular el valor estimado de la variable, es decir, la “altura” de la superficie modelada en este punto en particular.

En general, el Kriging predice valores en puntos y bloques en espacios d -dimensionales y permite la incorporación de la anisotropía. Varias extensiones al método permiten una gran flexibilidad para su aplicación.

Las principales cualidades del Kriging están en la calidad estadística de sus predicciones y en la habilidad de predecir la distribución espacial de la incertidumbre.

IV.6.1 Cokriging

En el cokriging, una variable es típicamente considerada como primaria, mientras que otras variables se consideran secundarias. Esto es, son usadas para ayudar en la predicción de la primera variable pero no son de interés básico. Frecuentemente la variable primaria no tiene suficientes puntos de muestreo a diferencia de las variables secundarias, y el realizar más mediciones en la variable primaria es costoso o imposible. En el caso de los problemas hidrológicos, concretamente en la precipitación, esta posee variabilidad espacial y temporal, y las mediciones puntuales en los pluviómetros, aunque pueden asumirse de buena calidad, normalmente son muy escasas y dispersas para identificar la variabilidad espacial de la precipitación. Sin embargo, puede asumirse que otras variables de mayor disponibilidad influyen en el proceso, como las orográficas (elevación, pendiente), climáticas (vientos, presión) o las mismas de precipitación obtenidas por otros métodos (radar, pluviógrafos).

Se puede proponer una definición de un variograma cruzado que involucre dos variables como:

$$2\gamma_{1,2} = \text{var}\{Z_1(s+h) - Z_2(s)\} \quad (4.25)$$

Comparando este variograma cruzado con la ecuación (4.18), podemos observar que por definición $2\gamma(0) = 0$, mientras que $2\gamma_{1,2}(0)$ no es necesariamente igual a 0, también hay que notar que $2\gamma_{1,2}(h)$ no es necesariamente igual que $2\gamma_{2,1}(h)$.

En el Anexo 1 de esta tesis se ejemplifica el uso particular de estos métodos, desglosándose las ecuaciones para un caso específico.

V. MEJORA DE VARIANZA EN COMPLEMENTACION DE SERIES

Los hidrólogos normalmente se encuentran en la práctica con series de tiempo que presentan datos faltantes debido a malas condiciones climáticas, mal funcionamiento del equipo, contaminación de los datos, error en el procesamiento y otras causas. Los datos incompletos normalmente causan inconsistencias y resultados desviados en análisis hidrológicos. La inferencia estadística de los modelos a través de los datos que se tengan disponibles conlleva cierto grado de arbitrariedad e incertidumbre, las expresiones teóricas usadas para los valores obtenidos y la varianza del error de los estimados presentan tres inconsistencias básicas al plantearse un modelo: Una desviación en el punto de estimación de una variable multidimensional, una distribución espacial de la medición de la incertidumbre y una elección inapropiada del modelo para el variograma. Por todo esto una estimación de datos faltantes en una serie de datos de lluvia es requerida como primer paso en la práctica de análisis hidrológicos y modelaciones.

En el estudio de fenómenos distribuidos espacialmente uno de los objetivos del análisis estadístico es cuantificar la incertidumbre en los estimados obtenidos. En el análisis espacial se formulan ciertas hipótesis estadísticas para evaluar la estructura espacial de los procesos hidrológicos, tales como la existencia de una función normal, la mejora de varianza general y en el semivariograma y la estimación de ciertos parámetros. Cada uno de los modelos existentes para la obtención de datos espaciales provee una mejora en las hipótesis estadísticas dependiendo del caso concreto de aplicación, por lo que es de vital importancia el analizar la incertidumbre obtenida al aplicar diversos métodos para elegir el más adecuado para el caso de estudio, sobre todo se pueden considerar la media y la varianza como factores estadísticos significativos para analizar la eficiencia de un método de complementación de datos faltantes.

V.1 Criterios de mejora en la estimación de la media y varianza

La mejora de la estimación de la media y varianza en una muestra extendida, es medida a través del *contenido de información relativa*.

$$I = \frac{Var(\hat{y}_1)}{Var(\hat{y})} \quad (5.1)$$

Donde $Var(\hat{y}_1)$ es la varianza del parámetro y estimado a través del registro original con longitud N_1 y $Var(\hat{y})$ es la varianza del parámetro y estimado a través del registro combinado N_1+N_2 .

Para que exista superioridad al usar las dos estaciones, " I " tiene que ser mayor a la unidad. Por lo tanto, en el caso de la media, se tiene que la varianza de ésta cuando se usan ambas estaciones, es el siguiente para el registro extendido:

$$Var(\hat{y}) = \frac{\hat{\sigma}_y^2}{N_1} \left\{ 1 - \frac{N_2}{N_1 - N_2} \left[\bar{r}^2 - \frac{(1 - \bar{r}^2)}{(N_1 - 3)} \right] \right\} \quad (5.2)$$

Donde \bar{r}^2 es el coeficiente de correlación de los valores comunes del registro. La varianza de la media, tomando solo el registro original N_1 es:

$$Var(\hat{y}_1) = \frac{\hat{\sigma}_y^2}{N_1} \quad (5.3)$$

Y dado que $I > 1$ o lo que es lo mismo $Var(\bar{y}) < Var(\hat{y}_1)$ se tiene que esto ocurre cuando:

$$\hat{r} > \left(\frac{1}{N_1 - 2} \right)^{1/2} < \hat{r}_{cm} \quad (5.4)$$

\hat{r}_{cm} es función de N_1 y se obtiene por la interpolación de los valores que se muestran en el Cuadro 2:

Cuadro 2. Valores de N_1 y \hat{r}_{cm}

N_1	5	10	15	20	25	30	40	50
\hat{r}_{cm}	0.58	0.35	0.28	0.24	0.21	0.19	0.16	0.14

El estimado de la varianza puede ser mejorado por análisis de correlación, si $Var[S^2(Y)]$, la varianza del estimado después de que el registro es extendido por

correlación, es menor que $Var|S_1^2(Y)|$, la varianza del estimado de la muestra original es:

$$Var|S^2(Y)| = \frac{2\sigma_y^4}{(N_1-1)} + \frac{N_2\sigma_y^4}{(N_1+N_2-1)^2} [Ar^4 + Br^2 + C] \quad (5.10)$$

Para que $S^2(Y)$ sea un mejor estimador de σ_y^2

$$Var|S^2(Y)| < [Var|S_1^2(Y)| = \frac{2\sigma_y^4}{(N_1-1)}] \quad (5.11)$$

Esto ocurre sí:

$$|\bar{r}| > \left(\frac{-B + \sqrt{B^2 - 4AC}}{2A} \right)^{1/2} \quad (5.12)$$

Si esta desigualdad no se satisface, entonces $S_1^2(Y)$ es mejor estimador de σ_y^2 que $S^2(Y)$.

No es suficiente el determinar que existirá mejora en la media y la varianza al extender un registro, además es necesario considerar el número de años que podrán extenderse. Para esto Langbein propone lo que se denomina *el período efectivo de registro* N_e el cual indica, con base en los registros combinados de las dos estaciones, cuántos años como máximo podrán extenderse sin que se perjudiquen las características estadísticas del registro. Este criterio relaciona el número de años de la estación base, los años de la estación a la que se le va ampliar el registro y su respectivo coeficiente de determinación r^2 obtenido de la regresión lineal.

$$N_e = \frac{N_1 + N_2}{1 + \frac{N_2}{N_1 - 2}(1 - r^2)} \quad (5.13)$$

Donde

- N_e período efectivo de registro, en años
- N_1 número de años del registro común (registro corto)
- N_2 número de años del registro largo
- r^2 coeficiente de determinación

VI. DESAGREGACION DE TORMENTAS POR INTERPOLACIÓN

Usualmente un proceso hidrológico estocástico debe ser estudiado a diferentes escalas de tiempo. Por tanto, el problema surge de cómo generar series de tiempo consistentes tanto en escalas mayores, o gruesas, como en menores, o finas, de tiempo. Una solución trivial para este problema es modelar el proceso únicamente en una escala menor, y luego agregar los valores para obtener las escalas mayores. Sin embargo, hay varias razones para evitar esta solución y modelar el proceso en diferentes escalas de tiempo por separado, concentrándose cada vez en diferentes propiedades estadísticas del proceso. Por ejemplo, a una escala anual un modelo se concentra en la persistencia a largo plazo de las propiedades del proceso; a una escala mensual otro tipo de modelo debe describir periodicidad y memoria del proceso en el corto plazo; y a la escala horaria un modelo debe describir intermitencia (por ejemplo, seco o lluvioso) y estructuras finas del proceso. En otros casos, los procesos a altos niveles pueden ser la salida de un modelo especializado (por ejemplo un modelo meteorológico o climatológico) o deducidos de mediciones (por ejemplo medición diaria de lluvia); aparentemente en tales casos la idea de agregación tal vez no funcione, sino más bien desagregación (Wu y col., 2006). Específicamente cuando se trata con datos de lluvia, hay un gran número de mediciones diarias operando durante décadas. Sin embargo, el número de pluviómetros con resolución horaria o subhoraria es mucho más pequeña. Esta situación refleja una escasez general de datos a escalas de una hora o menos tanto en número de estaciones como en longitud de las series. La necesidad de datos a escala subhoraria para aplicaciones hidrológicas, especialmente en estudios de inundaciones, sugiere el uso de técnicas apropiadas para refinar la información diaria disponible y proveer al usuario con posibles datos de precipitación que agregados simulen los datos diarios. Tales técnicas pueden proveer una herramienta de simulación continua útil para el diseño y manejo de hidrosistemas. Esta clase de problemas es comúnmente abordada por los modelos de desagregación. Varios modelos se han desarrollado desde los 1970s y utilizados en numerosas aplicaciones hidrológicas, incluidas, entre otras, simulaciones de sistemas de reserva tanto para propósitos de diseño como de

operación, simulaciones de tormentas e inundaciones, e incluso extender y completar series de datos. Algunos de los modelos de desagregación desarrollados son de propósito general y no son específicos para un cierto proceso hidrológico o una aplicación exacta. Sin embargo, estos modelos no son siempre aplicables a los procesos de lluvia, especialmente a escalas muy finas, debido a ciertas particularidades del fenómeno. Esto ha provocado el desarrollo de técnicas especializadas para el proceso de lluvia, cuyo estudio constituye un campo preferencial de modelos de desagregación. Hoy en día hay un renovado interés en los métodos de desagregación debido a situaciones climatológicas que han interesado a varios investigadores (por ejemplo, detallar una situación de cambio climático). Normalmente, tales escenarios son desarrollados a una escala de tiempo amplia y existe la necesidad de transformarlas a una escala más fina. Sin embargo, desagregación no es sinónimo de reescalado, ya que este último proceso busca producir series de tiempo a menores lapsos con las estadísticas requeridas pero no necesariamente añade datos al total de la serie. El reescalado es en particular usado para aplicaciones hidrológicas de modelos de circulación general donde los valores exactos de escalas mayores no son considerados particularmente exactos. En ambos tipos de problemas, series sintéticas a escalas finas deben reproducir las variables estadísticas importantes a esa escala de tiempo en el proceso hidrológico relacionado.

VII. HIPOTESIS Y OBJETIVOS

VII.1 Hipótesis

- La distribución espacial de la precipitación puede ser representada por una función matemática denominada variograma direccional, la variabilidad en la magnitud e intensidad de los campos de lluvia producidos en zonas urbanas tiene una relación directamente proporcional con dicho variograma,
- El complemento de registros históricos de variables hidrológicas tales como la precipitación e intensidad de la lluvia se puede realizar mediante técnicas de interpolación espacial en donde se involucre el variograma y la incertidumbre en el complemento espacial de registros de precipitación.
- El complemento de registros puede ser evaluado utilizando pruebas estadísticas sobre la varianza de los datos.

VII.2. Objetivos

- Proponer y analizar diversas metodologías de análisis espacial y geoestadística para la interpolación espacial de datos hidrológicos.
- Realizar la interpolación de datos hidrológicos reales en escala de tiempo menor a una hora mediante varias técnicas de desagregación y evaluar la incertidumbre mediante una validación cruzada.
- Establecer criterios para el complemento de registros históricos de valores hidrológicos mediante las técnicas analizadas previamente.

VIII. METODOLOGÍA

VIII.1. Definición del área de estudio

La ciudad de Santiago de Querétaro, es la ciudad más importante y capital del estado mexicano de Querétaro, además de cabecera de la municipalidad del mismo nombre, se encuentra localizada al suroeste del estado, en un valle homónimo, a los $20^{\circ} 35' 34.8''$ de latitud norte y $100^{\circ} 23' 31.6''$ de longitud oeste, y tiene una altitud media sobre el nivel del mar de 1, 820 m. El valle presenta una precipitación anual promedio de 555 mm, la mayor parte ocurre en la temporada de lluvias, de junio a octubre, el resto se distribuye a lo largo del año, principalmente en los meses invernales debido a frentes fríos provenientes del norte.



Figura 4. Zona de estudio y estaciones utilizadas.

VIII.2. Obtención de bases de datos

Para el presente trabajo se emplearon las bases de datos de las EMAs (Estaciones Meteorológicas Automáticas) administradas por la CEA ubicadas dentro y alrededor de la ciudad de Querétaro (Cuadro 3), cada una de estas bases de datos se analizó en cuanto a consistencia de los datos y periodos de tiempo en

las cuales existen los registros. Posteriormente se elaboraron tablas de disponibilidad de información en las cuales se pudieron cotejar aquellos días en los cuales existan registros suficientes para evaluar los métodos de interpolación espacial.

Cuadro 3. Nombre y ubicación de las estaciones utilizadas.

Clave Estación	Nombre	Ubicación (UTM WGS1984)	
		x	y
IQUERETAR6	CEA Querétaro	352010.3614	2277417.97
IQUERETA17	Centro Cívico	358236.6482	2274043.88
IQUERETA43	El Marqués	361409.3902	2279552.59
IQUERETA19	Huimilpan	367456.2422	2252937.8
IQTSANTI2	Jurica	349139.3913	2286141.91
IQUERETA39	Obrera	352039.3313	2280738.78
IQUERETA37	Plaza Altamira	346838.6042	2281892.02
IQTSANTI3	Protección Civil	353091.0652	2281836.64
IQUERETA23	Rancho Tec Monterrey	373849.5706	2270600.12
IQUERETA35	Rochera Santa Rosa	338660.6501	2298574.38
IQUERETA33	San Antonio De La Punta	350968.0175	2277427.09
IQROQUER2	Santa Rosa Jáuregui	349031.6445	2294049.81
IQROQUER3	UTEQ	354161.7447	2285148.43
IQUERETA36	Villa Corregidora	350938.8841	2274106.29

VIII.3. Interpolación espacial de precipitación

Diversos métodos de interpolación se analizaron tanto en su método como en su fundamentación teórica, partiendo de este análisis se determinaron las posibles ventajas y desventajas de cada uno de estos.

Una vez elaboradas las tablas de disponibilidad de información se procedió a realizar la interpolación espacial de las variables hidrológicas mediante los diversos métodos propuestos (Aproximación de LaGrange, Splines, Distancia Inversa, Aproximación Multicuadrática, Redes de Triángulos Irregulares, Kriging) con la finalidad de rellenar todas las series de tiempo faltantes en las tablas de disponibilidad, completando los registros para cada estación y finalmente evaluar

la distribución y propiedades estadísticas de cada serie tales como la media aritmética, la desviación estándar, la varianza y la curtosis, entre otras

$$\bar{X} = \frac{1}{n} \sum_i^n n_i \quad (8.1)$$

$$S_x = \left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \right)^{0.5} \quad (8.2)$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \quad (8.3)$$

$$g_2 = \frac{\frac{1}{n} \sum (X_i - \bar{X})^4 * n_i}{\frac{1}{n} (\sum (X_i - \bar{X})^2 * n_i)^2} - 3 \quad (8.4)$$

Donde \bar{X} representa la media de la muestra, S_x la desviación estándar, S^2 la varianza, g_2 el coeficiente de curtosis, x_i cada uno de los valores, n_i la frecuencia de cada valor y n número total de datos (Zelterman, 2010).

VIII.4. Validación cruzada

Después de la evaluación estadística de los métodos de interpolación espacial cada uno de los métodos fue sometido a una validación cruzada entre estaciones dentro de la zona de estudio, con el fin de observar el comportamiento con relación a datos reales, estas validaciones tuvieron como base las tablas de disponibilidad anteriormente elaboradas. La principal base de las validaciones fue el análisis de residuales, con la información arrojada por este se emitirán juicios acerca del comportamiento de los valores interpolados en relación con mediciones reales. Además se realizaran análisis basados en el error cuadrático medio así como el error absoluto. Estos son calculados mediante:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\varphi}_1 - \varphi_1)^2} \quad (8.5)$$

$$AE = \sum_{i=1}^n |\hat{\varphi}_i - \varphi_i| \quad (8.6)$$

Donde RMSE indica el error cuadrático medio, AE el error absoluto, n es el número total de observaciones, $\hat{\varphi}$ es el valor estimado y φ el valor real observado (Zeltermán, 2010).

VIII.5. Cokriging

El análisis de cokriging se realizó de manera independiente para ilustrar tanto su uso como una forma concreta de obtener un coeficiente basado en el tiempo de ocurrencia de las lluvias en los diferentes puntos de muestreo, basado en la caracterización de patrones de lluvia típicos de la región. Además se comparó este con el cokriging tradicional que involucra la elevación del terreno como factor adicional. Dado el análisis complejo que requiere este modelo y la naturaleza de las mediciones de lluvia se decidió que este modelo se evaluaría para datos de lluvias diarias.

VIII.5.1 Modelo digital de elevación

Para realizar un análisis co-kriging incorporando datos de elevación se obtuvo un modelo digital de elevación (MDE) de la zona de estudio a través del INEGI, con resolución de 10 m, a partir de este MDE se determinaron 25 puntos diferentes distribuidos de manera homogénea sobre la zona, elaborándose una tabla de datos de la cual se extrajeron los datos para incorporar la elevación al semivariograma de lluvia.

VIII.6.2 Obtención del coeficiente *k*

Para seleccionar los datos en esta parte de la metodología se tomaron en cuenta únicamente eventos con una precipitación total de un milímetro o más y una duración mayor de 30 minutos con el propósito de obtener una resolución adecuada para este estudio (Powell y col., 2008). Ya que la precipitación total varía de un evento a otro, la caracterización de las tormentas se realizó usando una escala adimensional, dividiendo la fracción acumulada entre el total de la

precipitación (P/P_{24}). Con el fin de caracterizar adecuadamente los patrones de lluvias, se tomaron periodos de 24 horas fijando el inicio al medio día, contrario al tradicional de media noche, para evitar cortar las lecturas de tormentas debido al horario natural en el que estas ocurren (Engida y Esteves, 2011).

Cada estación se caracterizó individualmente para identificar las variaciones temporales mediante análisis de aglomerados. Se utilizó el paquete Statgraphics Centurion XV (StatPoint, Inc. 2007) utilizando el método jerárquico de Ward con una distancia Euclidiana, parámetros utilizados típicamente en modelos hidrológicos (Muñoz-Díaz y Rodrigo, 2004, Crétat y col., 2010).

Una vez obtenidos los aglomerados estos se graficaron como hietogramas para observar su comportamiento en el tiempo y ajustarlos a una ecuación descriptiva, realizando posteriormente un análisis de los coeficientes de esta ecuación y su relación con la hora de inicio y la duración de los eventos de cada aglomerado.

VIII.7 Mejora de la varianza

Una vez analizados los métodos se procedió a rellenar ciertos periodos de registros en estaciones faltantes y se evaluaron los cambios en la varianza mediante el método descrito previamente para determinar cuál es la extensión que se puede considerar como válida para el complemento de registros.

En el primer caso, de una serie de datos que abarcaba tres días se calculó la mejora de varianza para rellenar una estación que sólo contara con dos días de registro. Utilizando las formulas dadas en la sección V para las 12 estaciones de referencia y la estación rellenada se determina si existe mejora de varianza.

IX. ANÁLISIS DE RESULTADOS

IX.1 Información disponible

Después de una revisión de los datos disponibles de las estaciones de la red de la CEA, se elaboró una Tabla (Anexo 2) que muestra las fechas disponibles para cada una de estas estaciones. De las 15 estaciones disponibles, en ninguna fecha coinciden todas en actividad, siendo la estación de El Marqués la de menor información, seguida por Centro Cívico y Corregidora.

IX.2 Aplicación de los métodos.

En base a las estaciones disponibles se elaboraron hojas de cálculo (Microsoft Excel 2010) para el manejo de los datos, además, en este mismo programa se elaboraron algoritmos para realizar el análisis espacial mediante los métodos de LaGrange, Distancia Inversa y Aproximación multicuadrática. Mediante el software especializado MATLAB (The Mathworks, Inc. 2010) se elaboraron los algoritmos para realizar el análisis espacial mediante los métodos de Splines (Thin-Plate Spline) y TIN.

Debido a que los métodos de LaGrange, Multicuadrático y Splines pueden arrojar valores negativos, cuando ocurría este caso estos se redondearon a un valor de 0.

Cada uno de los métodos fue realizado para diferente número de estaciones con base en la disponibilidad de los datos, realizándose una validación cruzada con las estaciones que se ubican dentro del polígono delimitado por las estaciones más alejadas del centro de la ciudad.

Para el co-kriging se ocupó el software ArcGIS mediante el módulo de análisis espacial.

Después de obtener los valores interpolados por medio de los métodos, los resultados obtenidos por medio de la validación cruzada se resumen en el Cuadro 4.

Cuadro 4. Resultados de RMSE y error absoluto para los métodos y número de estaciones ocupadas en el análisis. Los cálculos son para valores ajustados a 0 en los métodos que así lo requieren.

No. De Estaciones	Multi-						
	D. Inversa	cuadrática	LaGrange	Spline	TIN	Kriging	
13	0.6900	0.6829	2.4191	0.7147	0.6867	0.6829	RMSE
12	0.4318	0.4222	5.1419	0.5412	0.4150	0.4221	
11	0.4313	0.3970	1.0806	0.6665	0.3900	0.3970	
10	0.4717	0.4744	9.5529	0.5877	0.5050	0.4740	
Media	0.5062	0.4944	4.5487	0.6275	0.4992	0.4940	
13	223.77	223.30	693.54	267.90	267.19	223.30	Error ABS
12	95.87	91.43	891.58	122.69	87.92	91.43	
11	68.70	68.70	146.12	108.69	65.40	68.74	
10	163.36	161.38	2339.45	196.51	173.01	161.50	
Media	137.93	136.21	1017.68	173.95	148.38	136.25	

Se puede observar que el método de Kriging presenta el Error Cuadrático Medio (RMSE) más bajo, seguido de los Triángulos irregulares, la Aproximación multicuadrática y la Distancia Inversa. La Aproximación multicuadrática, aun cuando es uno de los métodos propensos a presentar valores negativos, no varió mucho en su desempeño una vez que dichos valores fueron ajustados a 0. De igual forma el método de Thin-Plate Spline mejoró muy poco al ajustar los valores negativos. El Método de LaGrange presenta el mayor RMSE que es mejorado ampliamente con el ajuste de valores negativos. Un aspecto interesante es el hecho de que para los métodos con menor RMSE la disminución del número de estaciones disminuyó el RMSE, esto puede deberse al hecho de que las estaciones removidas pudieran ejercer una influencia no real en estos métodos. Las estaciones removidas en orden fueron: El Marqués, Huimilpan y Centro Cívico. Esto sugiere dos posibilidades, la primera que estas estaciones se encuentren en áreas con un comportamiento espacial muy particular que distorsiona el comportamiento general de la lluvia en la zona (particularmente la estación El Marqués que se encuentra en una cañada); y la segunda es que estas estaciones presenten datos no reales o distorsionados, lo que afecta el resultado de los métodos.

La Figura 5 muestra la gráfica de una tormenta en particular (10 de agosto de 2010) y la comparación de los resultados de cada método.

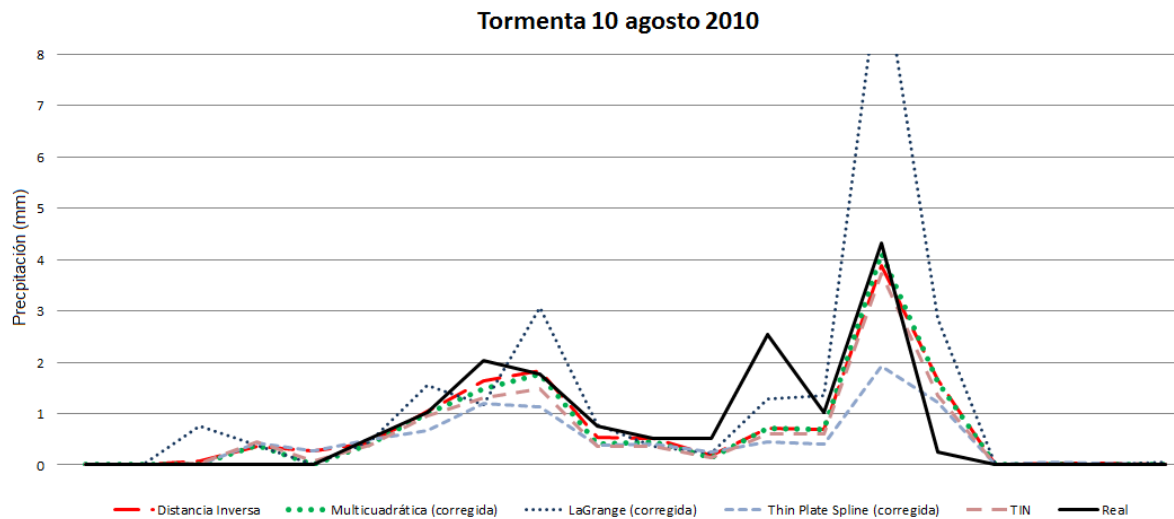


Figura 5. Representación gráfica de los datos obtenidos por interpolación en relación a los datos reales de precipitación.

El error principal del Método de LaGrange radica en el efecto Runge, de ahí que es el método que presenta más números negativos, y también presenta el mayor RMSE aun cuando se corrigen estos valores. Debido a la característica principal del método de Splines (la “suavización” de las curvas) este presenta los mayores errores por subestimación en los valores extremos de las tormentas. Estos dos últimos métodos presentan errores contrarios en los valores extremos, LaGrange por encima y Splines por debajo de los observados.

Una característica particular de los métodos de interpolación espacial es su deficiencia para representar valores extremos. Debido a la naturaleza de los métodos, estos no son capaces de representar el valor máximo en un punto interpolado, pues siempre son calculados con base en el resto de los datos. En caso de que el sitio interpolado presente un valor extremo, los métodos aquí analizados no son capaces de representarlo adecuadamente, este fenómeno se ejemplifica en la gráfica de residuales para la tormenta del 21 de septiembre de 2010 (Figura 6), en la cual se encuentran mayormente ubicados por debajo de la

línea de 0. En dicha figura no se grafican los residuales del método de LaGrange por ser estos demasiado grandes para ser observados adecuadamente.

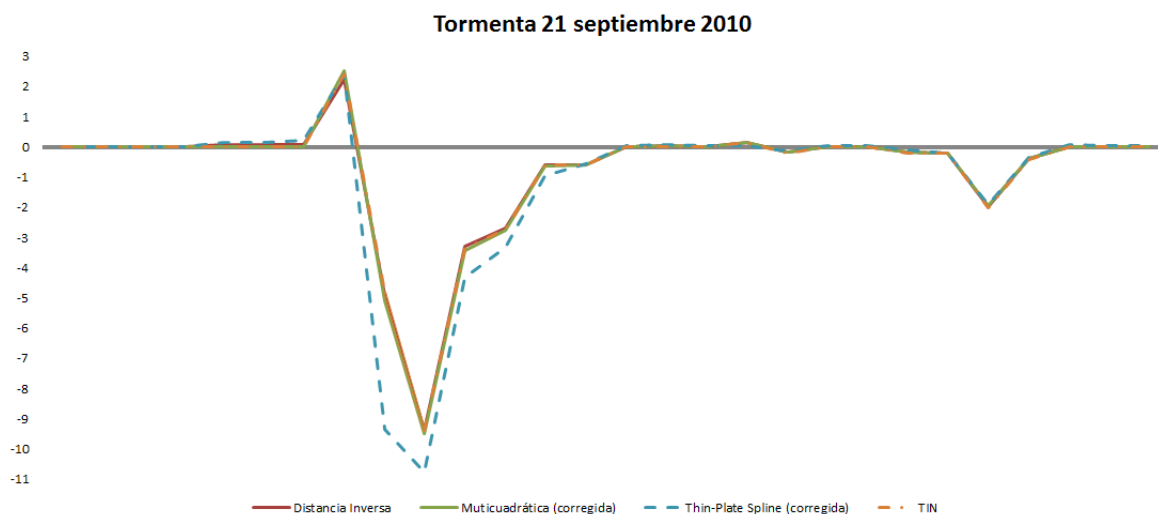


Figura 6. Gráfica de residuales para una tormenta, se muestran los residuales para cuatro métodos únicamente.

IX.3 ANOVA

Al realizar un Análisis de Varianza de los métodos mediante el software Statgraphics se encontró que únicamente la variación en el método de La Grange es estadísticamente significativo para considerarse, el resto de los métodos, con un intervalo de confianza de 99% se puede considerar sin diferencias significativas, como puede observarse en los Cuadros 5 y 6.

Cuadro 5. Análisis de varianza del RMSE para métodos de interpolación.

<i>Fuente</i>	<i>Suma de cuadrados</i>	<i>G.L.</i>	<i>Cuadrado de la media</i>	<i>F</i>	<i>Valor P</i>
Entre grupos	54.0403	5	10.8081	4.61	0.0070
Dentro de los grupos	42.1766	18	2.34314		
Total	96.2169	23			
<i>Método</i>	<i>Media</i>	<i>Grupos homogéneos</i>			
Kriging	0.494	X			
Multicuadrática	0.494125	X			
TIN	0.499175	X			
Distancia Inversa	0.5062	X			
Spline	0.627525	X			
La Grange	4.54862	X			

Cuadro 6. Análisis de varianza del error absoluto para métodos de interpolación.

Fuente	Suma de cuadrados	G.L.	Cuadrado de la media	F	Valor P
Entre grupos	2.53374x10 ⁶	5	506748.	3.36	0.0256
Dentro de los grupos	2.71329x10 ⁶	18	150738.		
Total	5.24703x10 ⁶	23			
Método	Media	Grupos homogéneos			
Kriging	136.202	X			
Multicuadrática	136.243	X			
TIN	137.925	X			
Distancia Inversa	148.38	X			
Spline	173.947	X			
La Grange	1017.67	X			

IX.4 Mejora de varianza

Ya que se considera a los métodos estadísticamente iguales el método de interpolación usado en este apartado fue el de Kriging tomando un variograma lineal. Para el análisis de la mejora de varianza se evaluaron doce estaciones con respecto a la estación a rellenar (La Obrera) en el periodo del 5 al 8 de julio de 2010, la estación complementada carece de datos para el 8 de julio, los resultados se ilustran en el Cuadro 7.

Cuadro 7. Análisis de mejora de varianza para las estaciones usadas en el complemento de registros para tres días

Estación	Media	Desv Est	Var	coef reg	Var($\hat{\gamma}$)	Var($\hat{\gamma}_1$)	I
Altamira	0.1468	0.5483	0.3006	0.6740	0.0082	0.0027	3.0514
Centro Cívico	0.1616	0.3999	0.1599	0.4049	0.0060	0.0027	2.2281
Corregidora	0.2685	0.9345	0.8732	0.2340	0.0046	0.0027	1.7054
UTEQ	0.1764	0.4477	0.2005	0.2456	0.0047	0.0027	1.7407
Huimilpan	0.1613	0.3883	0.1507	0.3014	0.0052	0.0027	1.9116
Jurica	0.3372	1.4484	2.0979	0.4311	0.0062	0.0027	2.3082
La Rochera	0.0542	0.2467	0.0609	0.0904	0.0034	0.0027	1.2662
CEA Qro	0.2083	0.5917	0.3501	0.3889	0.0059	0.0027	2.1791
P Civil	0.2848	1.0235	1.0475	0.8313	0.0095	0.0027	3.5326
P Escobedo	0.1870	0.3010	0.0906	0.4984	0.0068	0.0027	2.5143
Rancho Tec	0.1324	0.4323	0.1869	0.0961	0.0035	0.0027	1.2834
San Antonio	0.1907	0.7744	0.5997	0.4994	0.0068	0.0027	2.5172
Santa Rosa	0.1565	0.4809	0.2313	0.0657	0.0032	0.0027	1.1907

Todas las estaciones cumplen el criterio de $I > 1$ para que exista superioridad en el complemento de registros, además, la estación de Protección Civil, con $I = 3.5326$ es también la estación con mayor peso w en la interpolación.

Al variar el número de datos posibles a interpolar se encontró que $I < 1$ en tres estaciones cuando sólo se tienen dos días de registro, pero incluso en estos niveles se podría considerar un análisis espacial que no incluya estas.

Al calcular el N_e se encontró que para estas estaciones con $I < 1$ el periodo HJYKGH abarcaría sólo entre 560 y 569 de los 576 registros necesarios para completar los datos faltantes.

Esto indica que, si el comportamiento de las estaciones de la región de estudio muestra una correlación adecuada como se muestra en la Tabla IUL, es posible completar los registros faltantes si se cuenta con poco más de la mitad de los otros registros, principalmente de los sitios que el mismo análisis espacial indique son los que tienen un mayor peso en la interpolación.

IX.5 Consideraciones de los métodos de interpolación

Los modelos de regresión y de series de tiempo han sido usados anteriormente para la estimación de datos faltantes de lluvia. Los métodos de interpolación global que usan análisis de tendencias y regresión proporcionan varias ventajas comparados con técnicas determinísticas de ponderación (Sullivan y Unwin 2003, Teegavarapu, 2009). Varios métodos determinísticos de ponderación y de interpolación estocástica han sido usados en el pasado para la construcción espacial de campos de precipitación o estimación de datos faltantes de lluvia en un punto en el espacio. Los métodos de ponderación tradicionales y los basados en datos son usados generalmente para estimar precipitación no registrada. Los métodos de ponderación pertenecen a una clase de técnicas de interpolación espacial tales como la distancia inversa, el método determinístico no lineal y la interpolación estocástica (como el Kriging). Los métodos de regresión y análisis de series de tiempo pertenecen a modelos basados en datos. El Manual de Hidrología (ASCE 1996) recomienda dos métodos para la estimación de datos

perdidos. Estos métodos son determinísticos basados en promedios normalizados y en la distancia inversa.

Aproximaciones polinómicas como las de LaGrange presentan la problemática del fenómeno de Runge, este es un problema que sucede cuando se usa una interpolación con polinomios de alto grado. Fue descubierto mientras se exploraba el comportamiento de los errores al usar interpolación polinómica para aproximar determinadas funciones. El fenómeno demuestra que los polinomios de grado alto no son, en general, aptos para la interpolación.

Los métodos de interpolación local tales como los *thin-plate Splines* tienden a generar gradientes amplios en áreas con escasos datos y los errores en los procesos de estimación se agravan (Chang 2004). Otras limitaciones de los métodos de interpolación espacial han sido reportadas por varios investigadores en recientes estudios. Vieux (2001) señaló varias limitaciones del método de la distancia inversa, incluido el "*tent pole effect*" que produce estimaciones muy altas cerca del punto de interés en el espacio. Grayson y Bloschl (2001) mencionaron varias limitaciones de los métodos de polígonos de Thiessen y de la distancia inversa. Ellos sugirieron que estos métodos no son recomendables para interpolación espacial, considerando sus limitaciones. Sin embargo, recomiendan el método de *thin-splines* y el Kriging para la interpolación de variables hidrológicas. La aproximación por polígonos de Thiessen tiene la gran limitación de no proveer campos continuos de estimados si se usa para interpolación espacial. Brimicombie (2003) indicó que el punto de contención principal para la aplicación de distancias inversas para interpolación espacial es la selección del número y la relevancia de los puntos de observación para la interpolación espacial en un punto. Todos estos problemas asociados con las técnicas de interpolación espacial pueden llevar a una sub o sobre estimación de magnitudes de lluvia en un punto basadas en observaciones en otros sitios. Eischeid et al. (2000) reportaron varios métodos de interpolación para estimar datos faltantes en temperatura y precipitación diaria y discutieron sus limitaciones.

Singh y Chowdhury (1986) compararon trece métodos de estimación de lluvia y encontraron que el método de las isoyetas producía estimados más altos que la media diaria y mensual. Tung (1983) comparó cinco métodos usados para estimar datos puntuales de lluvia y encontró que los métodos de la media aritmética y la distancia inversa no producían resultados adecuados para regiones montañosas. Métodos de interpolación de superficie dependientes de la varianza pertenecientes a la familia del Kriging han sido aplicados para múltiples problemas de interpolación geofísica en hidrología (Grayson y Bloschl 2001; Vieux 2001). Estos métodos estocásticos de interpolación están basados en el principio de minimizar la varianza de los estimados en puntos donde no hay mediciones registradas. El Kriging en varias de sus formas es aplicado para la estimación de datos de precipitación perdidos y precipitación aérea de varios puntos de medición. Ashraf et al. (1997) compararon métodos de interpolación (Kriging, distancia inversa, cokriging) para estimar valores perdidos de precipitación. Ellos indicaron que el método de interpolación de Kriging producía el Error Cuadrático Medio más bajo. Sin embargo, los métodos de interpolación de Kriging poseen varias limitaciones. La selección de un modelo de semivariograma, la asignación de valores arbitrarios para parámetros *sill* y *nugget* y los intervalos de distancia son algunas de las dificultades asociadas a este método.

La subestimación de precipitación causada por redondear valores bajos a un valor de cero provoca errores significativos en balances hídricos regionales. Los valores bajos de precipitación pueden constituir el 10% de la precipitación promedio en verano y alcanzar hasta la tercera parte del total en algunos años, por lo que se deben de efectuar correcciones para evitar subestimaciones asociadas con valores pequeños de precipitación. La sobreestimación de precipitación por técnicas de interpolación espacial no es inusual. Las técnicas determinísticas tradicionales de interpolación espacial (por ejemplo, las basadas en la distancia) y las técnicas estocásticas tales como el Kriging isotrópico no consideran la variabilidad espacial de los patrones de precipitación. En general, los métodos de ponderación basados en la distancia padecen de una gran limitación conceptual basada en el hecho de que la distancia Euclidiana no es siempre una medición

definitiva de la correlación entre mediciones de puntos en el espacio. Esto también niega la primera ley de la geografía de Tobler: “todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las lejanas”, lo que forma la base de muchas técnicas de interpolación. También los métodos de interpolación fallan al estimar valores perdidos correctamente si se introducen errores en el proceso de medición de la lluvia en una o más estaciones (Teegavarapu, 2009). Finalmente, como ya se mencionó antes, debido que los métodos basan sus cálculos en los valores medidos en puntos específicos es muy difícil que alguno pueda predecir de manera acertada un pico máximo o valores más altos que los ya medidos.

IX.6. Patrones temporales de lluvia

Dada la naturaleza exploratoria del proceso de análisis de aglomerados y el hecho de que el número apropiado de patrones no está predeterminado, éste se eligió al examinar de manera visual tanto los dendogramas como las gráficas obtenidas de aglomerados a diferentes niveles, tomándose finalmente la decisión de trabajar con seis grupos.

Dentro del análisis para cada estación, de los siete grupos obtenidos, cuatro o cinco se identificaban con una forma sigmoidea de duración menor a tres horas, distribuyéndose esta forma en diferentes horarios para cada aglomerado, los cuales fueron clasificados de la siguiente forma: lluvias vespertinas (V) (5-8 pm) nocturnas (N) (8 pm - 11 pm), de media noche (MD) (11 – 2 am) y de madrugada (M) (2 – 6 am), las curvas características de estos aglomerados pueden verse en la Figura 7 para tres diferentes estaciones. Es interesante notar que mientras la tormenta sea más tardía, mayor es su duración, este comportamiento vario ligeramente para cada estación.

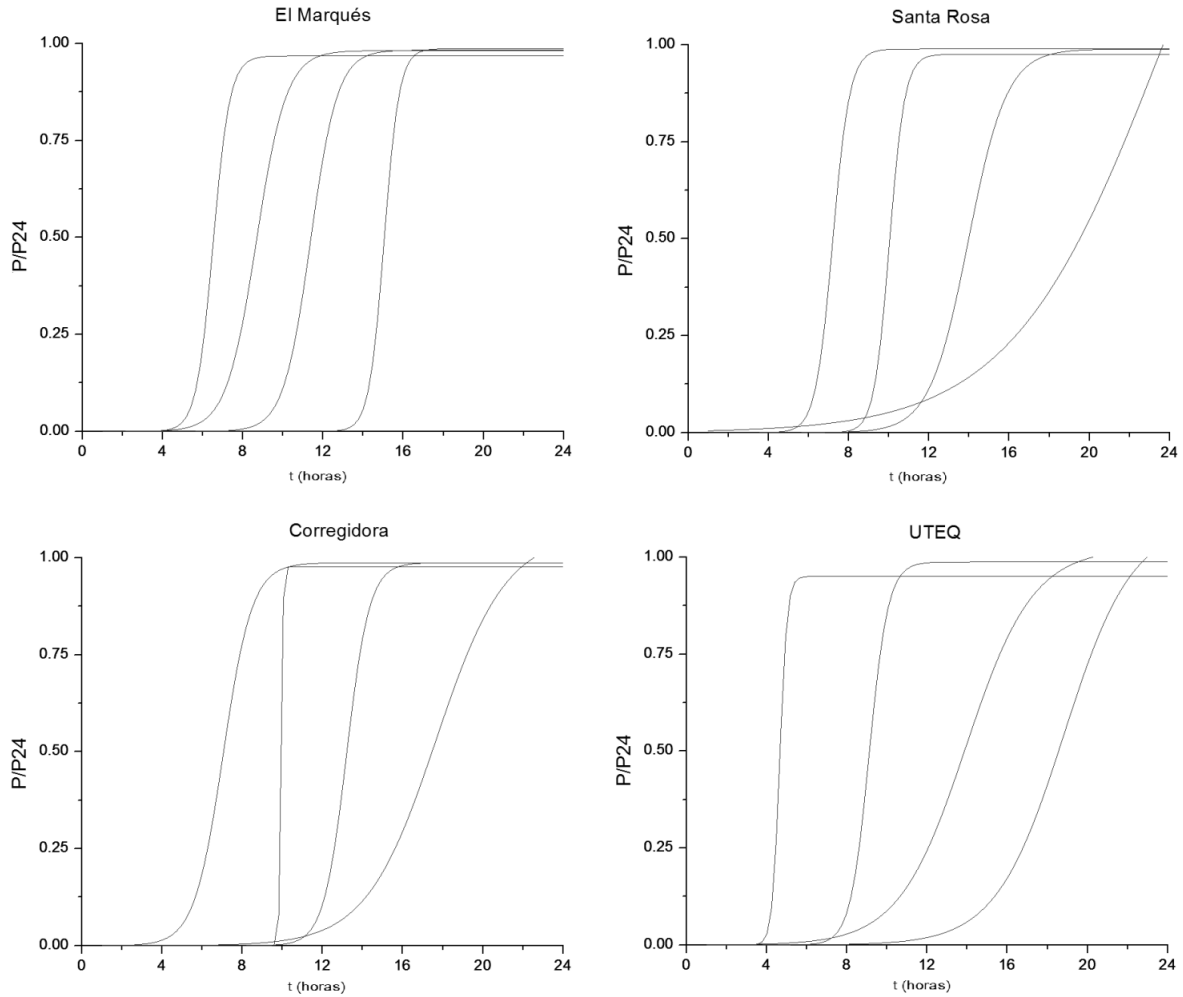


Figura 7. Caracterización de tormentas con distribución de tipo sigmoidea.

La ecuación característica que mejor se ajustó a este tipo de tormentas corresponde a una del tipo sigmoidea:

$$y = \frac{a}{1 + be^{-kt}} \quad (9.1)$$

Dado que el coeficiente a determina la altitud de la curva, y siendo el eje de las ordenadas una escala adimensional, este se considera como 1 para todos los casos, por lo que la ecuación puede expresarse como:

$$P/P_{24} = \frac{1}{1 + be^{-kt}} \quad (9.2)$$

Tomando en cuenta que el valor de t expresado en horas va de 0 a 24 donde 0 equivale al medio día y 12 a la media noche. El coeficiente b es un buen indicador de la duración de la tormenta, siendo más breve este en cuanto mayor sea el valor de b , mientras que las horas de inicio y final aproximado de las tormentas puede obtenerse expresando la ecuación en función de t , siendo para el inicio:

$$t_i = \frac{\ln(\frac{1}{y_i b} - \frac{1}{b})}{-k} \quad (9.3)$$

Mientras que para el final:

$$t_f = \frac{\ln(\frac{1}{y_f b} - \frac{1}{b})}{-k} \quad (9.4)$$

Considerando t_i y t_f como los tiempos de inicio y final de la tormenta, mientras que y_i y y_f son las fracciones a las cuales se considera una tormenta como iniciada y finalizada. Aquí se utilizaron valores de 0.05 y 0.95 respectivamente. Sin embargo, algunos aglomerados para cada estación consistían en grupos de tormentas que no tomaban la forma presentada antes. Dentro de estos grupos la característica principal era una duración mayor a 3 horas, y estas se clasificaron en vespertinas de larga duración (VLD) (5 pm – 2 am), matutinas de larga duración (DLD) (1 – 9 am) y constantes (C) (24 hrs). La Figura 8 muestra estas formas, aunque las dos primeras se ajustaron a una ecuación sigmoidea similar, la forma de la curva es diferente, mientras que la última se puede considerar prácticamente lineal.

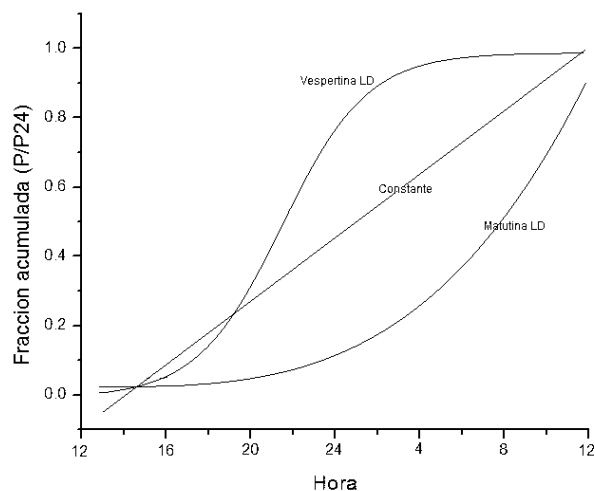


Figura 8. Lluvias con otras distribuciones.

Para estos casos, aun cuando las ecuaciones anteriormente descritas pueden utilizarse, las fracciones para y_i y y_f adecuadas fueron estimadas en 0.1 y 0.9 respectivamente, pues la forma alargada prolongaba en exceso la duración de la tormenta.

Analizando las tormentas por fecha de ocurrencia se observó que la mayor parte de las tormentas del tipo sigmoideo (N, MD, M y D) ocurren indistintamente a lo largo del año, pero son más abundantes en la estación de lluvias (junio-octubre) al igual que las tormentas VLD, aunque estas son más frecuentes al final de la estación (septiembre-octubre). Las tormentas DLD ocurren frecuentemente al final de la temporada (octubre-diciembre) mientras que las tormentas C están asociadas a frentes fríos (diciembre-abril).

Tomando en cuenta los valores de k y b obtenidos para cada estación es interesante calcular el tiempo promedio de inicio y final de cada tipo de lluvia y proyectarlo en un plano, de esta forma se obtiene una imagen que muestra el comportamiento típico espacial de la tormenta, y ya que es conocido, por la situación orográfica del valle y los vientos dominantes registrados, que las tormentas provenientes del Pacífico llegan por el sur mientras que las provenientes del Golfo llegan por el este con variaciones provenientes del sureste y noreste, la Figura 9 coinciden con la dirección típica de una tormenta (Groovaerts, 2000).

Cabe destacar que ciertas lluvias que provienen del sur no son generales para el valle y sólo afectan la zona sur-suroeste, mientras que lluvias que provienen del noreste típicamente afectan sólo la parte norte del municipio, debido a esto la imagen de la dirección típica pareciera indicar que las tormentas también tienen una dirección norte-sur, cuando esto no es así. Finalmente, las tormentas causadas por un frente frío y que suelen afectar a lo largo del día suelen ocurrir en todo el municipio sin variación espacial identificable.

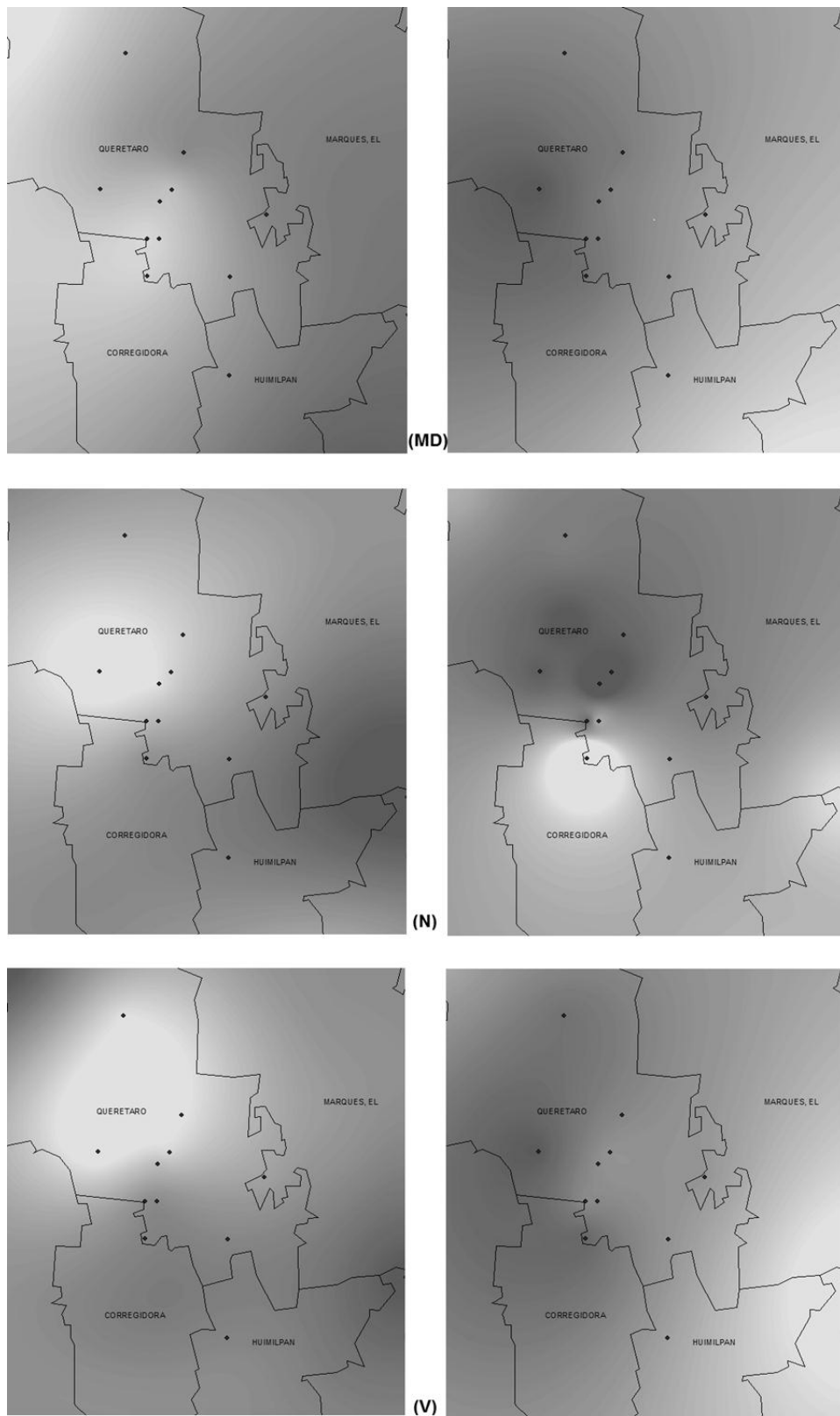


Figura 9. Representación de las horas de inicio (izquierda) y final (derecha) para tres tipos de tormenta, la escala de grises indica la dirección de los fenómenos.

IX.7 Cokriging

El método de cokriging para la interpolación de lluvias toma comúnmente datos de elevación dado que este se considera uno de los factores con mayor influencia en la distribución espacial de dicho fenómeno. La malla formada para representar los datos de elevación tomó en cuenta un área cuadrada que contuviera la mayor parte de las estaciones utilizadas en este estudio, se consideraron 25 puntos de elevación distribuidos de manera homogénea en el área donde las estaciones se encuentran. Para los datos del coeficiente k se utilizaron 10 estaciones con los valores obtenidos en la sección anterior, particularmente para el patrón de lluvia de media noche. Para este caso en particular se interpolaron datos de precipitación diaria, ya que la interpolación de datos de precipitación a 10 minutos mostraba una variabilidad muy amplia. Para ejemplificar el cambio en la distribución espacial se muestran los mapas obtenidos para el kriging normal, cokriging con puntos de elevación y cokriging con datos del coeficiente k de la ecuación descriptiva discutida anteriormente (Figuras 10, 11 y 12), los datos usados se muestran en los Cuadros 8, 9, y 10.

Cuadro 8. Precipitación total para el día 4 de julio de 2010 para las estaciones usadas en el análisis de cokriging.

Estación	x	y	P total
Plaza Altamira	-100.47	20.63	37.4
Centro Cívico	-100.36	20.56	50.6
Corregidor	-100.43	20.56	84.32
UTEQ	-100.4	20.66	63.8
Huimilpan	-100.27	20.37	44.12
Jurica	-100.448	20.668	110.43
La Rochera	-100.55	20.78	20.6
CEA	-100.42	20.59	59.4
Prot civil	-100.41	20.63	83
Pedro Escobedo	-100.14	20.5	55.6
Rancho Tec	-100.21	20.53	33
S. Antonio	-100.43	20.59	56.6
S. Rosa Jáuregui	-100.45	20.74	29.8
La Obrera	-100.42	20.62	78.6

Cuadro 9. Malla de puntos de elevación para el análisis de cokriging.

Elevacion	x	y	Elevación	x	y
1893	-100.461585	20.7172342	2006	-100.374601	20.6444565
1980	-100.435196	20.7172342	1986	-100.343529	20.6444565
2048	-100.407418	20.7172342	1800	-100.461862	20.6108453
2165	-100.376307	20.7172342	1804	-100.434918	20.6108453
1952	-100.344918	20.7172342	1816	-100.408251	20.6108453
1878	-100.46214	20.6819565	1901	-100.37464	20.6108453
1852	-100.435196	20.6819565	1990	-100.343529	20.6108453
2007	-100.407696	20.6819565	1799	-100.461862	20.5791787
2062	-100.37464	20.6819565	1802	-100.434918	20.5791787
2162	-100.343529	20.6819565	1813	-100.408529	20.5791787
1880	-100.461307	20.6444565	1841	-100.374918	20.5791787
1813	-100.434362	20.6444565	2004	-100.344085	20.5791787
1951	-100.407696	20.6444565			

Cuadro 10. Datos de puntos con valores de coeficiente k para el análisis de cokriging.

Estación	x	y	Coef. k
Altamira	346838.604	2281892.02	2.3538
CEA	352010.361	2277417.97	1.74859
Corregidora	350938.884	2274106.29	20.79021
El Marques	361409.39	2279552.59	1.33831
Huimilpan	367456.242	2252937.8	3.38657
Jurica	349139.391	2286141.91	1.19371
La Rochera	338660.65	2298574.38	1.5397
Obrera	352039.331	2280738.78	1.40456
Prot. Civil	353091.065	2281836.64	1.06024
Rancho TEC	373849.571	2270600.12	1.35139
San Antonio	350968.017	2277427.09	1.29552
Santa Rosa			
Jauregui	349031.645	2294049.81	1.36808

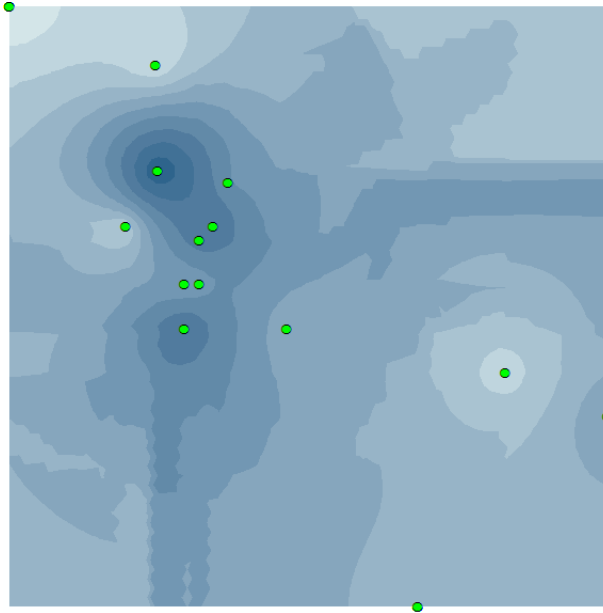


Figura 10. Kriging simple para datos de lluvia diaria en 13 estaciones.

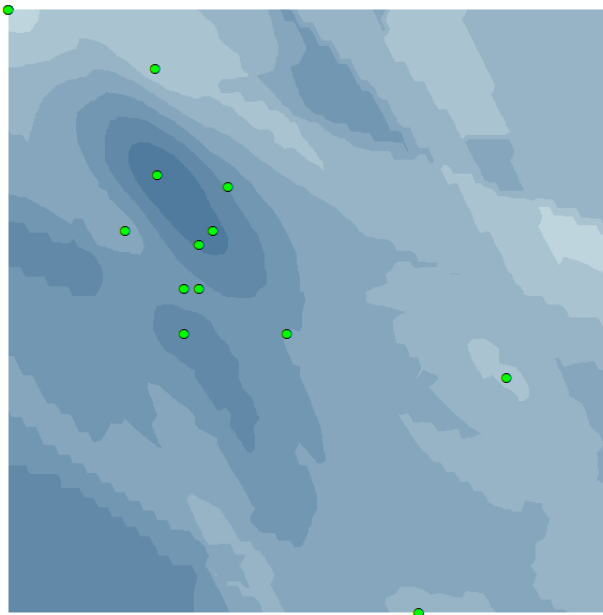


Figura 11. Cokriging usando datos de elevación.

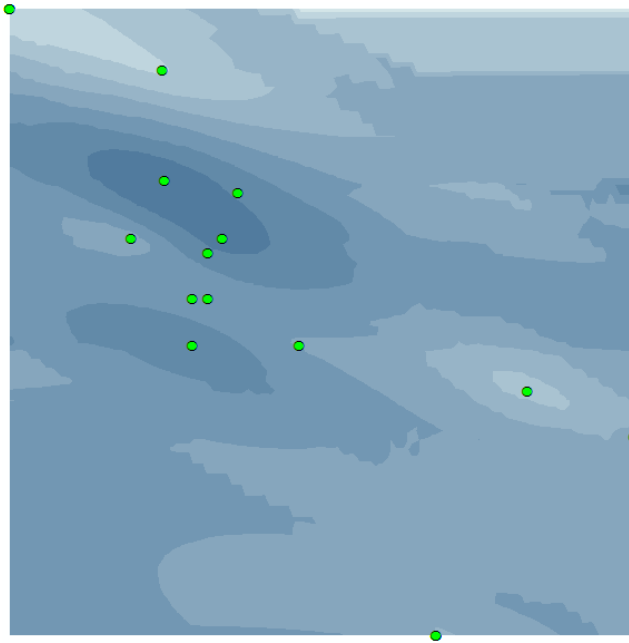


Figura 12. Cokriging usando el coeficiente k.

Podemos notar que, como en varios métodos de interpolación, en el kriging simple aparecen áreas muy concentradas alrededor de valores atípicos, formando una especie de “isla” de valores extremos.

Al usar la malla formada por los puntos de elevación se puede observar una distribución de los valores de precipitación que no presenta estas zonas de valores extremos tan pronunciadas, igualmente, al usar la malla con los valores de k estas islas se ven mejor distribuidas.

Si comparamos los resultados de los mapas interpolados a través del cokriging con los datos de elevación y del coeficiente k podemos apreciar un comportamiento aparente de la lluvia que concuerda con el descrito a través de los patrones de precipitación.

Es resaltable notar que el método del cokriging en este caso tampoco refleja una diferencia notoria con respecto a la validación cruzada de los valores de cada punto de medición de la precipitación, por lo que matemáticamente no muestran mejora en la interpolación, sólo al proyectarlos espacialmente se puede considerar el efecto del método como útil.

X. CONCLUSIONES

Debido a que la calidad de cualquier análisis hidrológico depende de la base de datos disponible, es importante siempre realizar un análisis previo de la información con la que se cuenta, clasificándola y agrupándola de manera adecuada para descartar datos erróneos o seleccionar el mejor conjunto de estos.

El análisis espacial es una herramienta muy útil para realizar complementos de series de datos meteorológicos, sin embargo, la existencia de diversos métodos, cada uno con sus características propias, nos obliga a seleccionar con cuidado cuál conviene más a nuestros objetivos, ya que cada uno de estos métodos tiene características específicas y dependerá del caso único de aplicación la elección del mejor modelo.

La región de Querétaro cuenta con una red de pluviógrafos suficientemente amplia para proporcionar datos confiables aplicables al análisis espacial, sin embargo, muchas de las estaciones no presentan una medición continua y es típico encontrarse con amplios faltantes en diversas estaciones, por lo que se debe considerar una revisión al funcionamiento de cada una de estas.

En el caso particular que nos ocupa, considerando que dentro de la región de estudio estadísticamente no existe diferencia entre la mayoría de los métodos analizados, la elección del método a usar dependerá de la facilidad del manejo de cada uno de estos en lo que respecta a la eficiencia cuando se incorpora a un sistema computacional y a la comodidad del usuario.

Los métodos espaciales vistos en este estudio, aun presentando un desempeño muy satisfactorio, siguen presentando la limitante de carecer de una función para determinar máximos en la interpolación, esto afecta principalmente a las áreas donde existe menor número de estaciones, pues esto lleva a subestimar la precipitación total.

El complemento de registros hidrológicos mediante análisis espacial también requiere un estudio de mejora de varianza, ya que se trabaja con una serie de

tiempo que, no obstante tener un método de interpolación adecuado, siempre tendrá que comprobarse su eficiencia en las tendencias estadísticas necesarias, en este estudio se evaluó el desempeño del complemento de registros y se pudo observar que, dada la alta correlación entre estaciones cercanas, es posible complementar series de casi el doble de longitud de los registros reales sin que existan cambios significativos en la varianza, esto también se debe a la buena distribución de la mayoría de las estaciones.

Los patrones espaciales y temporales de lluvia son útiles para evaluar modelos cuando se cuenta solo con datos puntuales de precipitación diaria. En este estudio se presentó un modelo para identificar estos modelos basados en registros pluviográficos recientes a través del análisis de aglomerados, una vez identificados los patrones típicos, estos pueden ser utilizados para desagregar registros pluviométricos o producir pronósticos. Además, al proponer una ecuación descriptiva esta se puede aprovechar para, mediante el análisis espacial, mostrar el comportamiento de estos patrones en el espacio. Con pocos años de operación, una red de pluviógrafos puede producir información muy útil y de gran ayuda en el estudio de diversas áreas de la hidrología.

El análisis particular mediante co-kriging debe incorporar datos adecuados que se puedan sustentar con una base física que indique el porqué de la utilización de ciertas variables. Aun cuando el presente caso de estudio no mostró mejora significativa al usar este método, no deja de ser importante el conocer estas herramientas de análisis aplicables en otros casos semejantes.

XI. REFERENCIAS BIBLIOGRÁFICAS

- ASCE 1996. Hydrology Handbook, 2nd edition. American Society of Civil Engineers ASCE, New York.
- Ashraf, M., Loftis, J. C. & Hubbard, K. G. 1997. Application of geostatistics to evaluate partial weather station network. *Agricultural and Forest Meteorology* 84, 255–271.
- Blachet, G. 1981. Analyse cartographique de la pluviométrie annuelle dans la région Rhône-Alpes. *Eaux et climat*, CNRS. 109-115.
- Briggs, I. C. 1974. Machine contouring using minimum curvature. *Geophysics*, 39, 39–48.
- Brimicombe, A. 2003. GIS, Environmental Modeling and Engineering. Taylor and Francis, London.
- Clark, I. 1979. Practical geostatistics. Applied Science Publishers LTD, Essex.
- Chang, K. T. 2004. Introduction to Geographic Information Systems. McGraw-Hill, New York.
- Creutin, J. y Obled, C. 1982. Objective analyses and mapping techniques for rainfall fields: an objective comparison. *Journal, Water Resources Research*, 18(2), 413-431.
- Eischeid, J. K., Pasteris, P. A., Diaz, H. F., Plantico, M. S. & Lott, N. J. 2000. Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *Journal of Applied Meteorology* 39, 1580–1591.
- Fiorucci, P., La Barbera, P., Lanza, L.G. and Minciardi, R., 2001. A geostatistical approach to multisensor rain field reconstruction and downscaling, *Hydrology and Earth System Science*, 5, 201–213.
- Goovaerts, P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, *Journal of Hydrology*, 288 (1-2), 113-129.
- Grayson, R. & Blöschl, G. 2001. *Spatial Patterns in Catchment Hydrology: Observations and Modeling*. Cambridge University Press, Cambridge.
- Haiden, T., Kerschbaum, M., Kahlig, P y Nobilis, F. 1992. A refined model of the influence of orography on the mesoscale distribution of extreme precipitation. *Journal des Sciences Hydrologiques*, 37 (5), 417-427.
- Hardy, R. L. 1990. Theory and applications of the multiquadricbiharmonic method. *Computers and Mathematics with Applications* 19, 163–208.

- Holawe, F. y Dutter, R. 1999. Geoestatistical study of precipitation series in Austria: time and space, *Journal of Hydrology*, 219 (1-2), 70-82.
- Huber , A. y Iroume, A. 2001. Variability of annual rainfall partitioning for different sites and forest covers in Chile, *Journal of Hydrology*, 248 (1-4) 78-92.
- Kim, T. W. & Ahn H. 2009. Spatial rainfall model using a pattern classifier for estimating missing daily rainfall data. *Stochastic Environmental Research and Risk Assess* 23, 367–376.
- Lanza, L.G., Ramírez J.A. and Todini E. 2001. Stochastic rainfall interpolation and downscaling. *Hydrology and Earth System Sciences*, 5(2), 139–143.
- Lebel, T. y Laborde, J. 1988. A geostatistical approach for areal rainfall statistics assessment. *Stochastic Hydrology and Hydraulics*. 2, 245-261.
- Mitasova. H. y Hofierka, J. 1993. Interpolation by regularized spline with tension: II. Application to terrain modeling and surface geometry analysis. *Mathematical Geology* 25, 657–69.
- Mitasova, H. y Mitas, L. 1993. Interpolation by regularised spline with tension: I. Theory and implementation. *Mathematical Geology* 25, 641–55.
- Rodriguez, I, Gupta, V. K. y Waymire, E. 1984. Scale considerations in the modeling of temporal rainfall. *Water Resou. Res.* 20, 1611-1619.
- Singh, V. P. & Chowdhury, K. 1986. Comparing some methods of estimating mean areal rainfall. *Water Resources Bulletin* 22 (2), 275–282.
- Sullivan, D. O. & Unwin, D. J. 2003. *Geographical Information Analysis*. John Wiley & Sons, New York.
- Teegarapu, R. S. V. 2009. Estimation of missing precipitation records integrating surface interpolation techniques and spatio-temporal association rules. *Journal of Hydroinformatics* 11 (2), 133-146.
- Todini, E., 2001. Influence of parameter estimation uncertainty in Kriging. Part 1 – Theoretical Development, *Hydrology and Earth System Science*, 5, 215–223.
- Tung, Y. K. 1983 Point rainfall estimation for a mountainous region. *Journal of Hydraulic Engineering ASCE* 109 (10), 1386–1393.
- Vieux, B. E. 2001. *Distributed Hydrologic Modeling Using GIS*. Water Science and Technology Library. Kluwer. Amsterdam.
- Zelterman, D. 2010. *Applied Linear Models with SAS*. Cambridge University Press. New York.

Anexo I

Ejemplificación de algunos métodos de interpolación.

Supongamos un caso práctico de aplicación de los métodos de interpolación en el cual tenemos 5 estaciones de medición y queremos determinar el valor de lluvia para un punto en específico donde no se cuenta con estación.

La Figura 13 nos muestra la ubicación de las estaciones y el punto a interpolar en un plano, el Cuadro 11 nos da un resumen de estos datos, para la ejemplificación de los métodos se utilizarán las formulas presentadas en el capítulo IV de este documento.

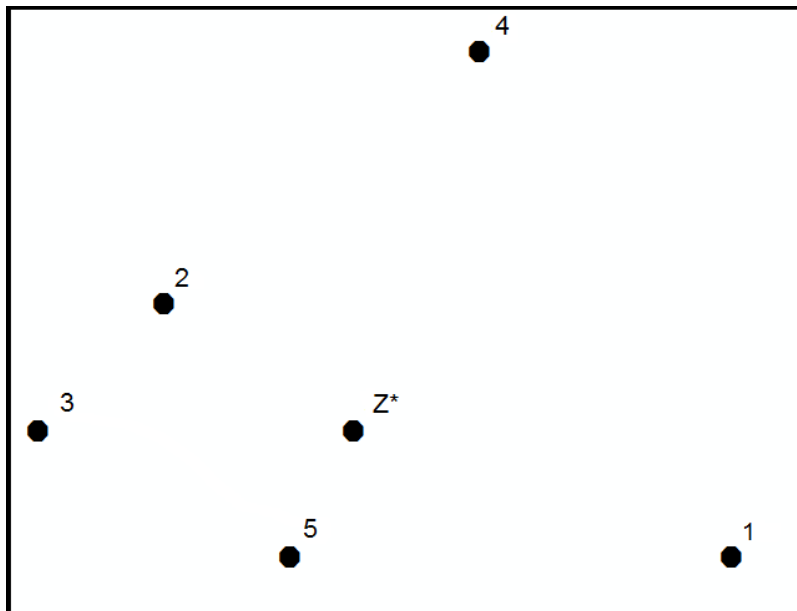


Figura 13. Puntos de las estaciones para este ejemplo.

Cuadro 11. Datos de las estaciones para este ejemplo.

Estación	Ubicación		Valor
	x	y	
Z*	70	30	R_p
1	130	10	100
2	40	50	60
3	20	30	40
4	90	90	95
5	60	10	80

Distancia Inversa

La formulación básica de este método es:

$$\hat{R}_p = \sum_{i=1}^N w_i R_i \quad ; \quad w_i = \frac{d_i^{-\alpha}}{\sum_{i=1}^N d_i^{-\alpha}}$$

Para obtener w para cada estación cercana se considera primeramente el valor de alfa, siendo el más común y el utilizado para este ejemplo el de $\alpha=2$ por lo que la distancia de cualquier punto al punto que se desea interpolar está dado por:

$$d_i = \sqrt{(x_i - x_p)^2 + (y_i - y_p)^2}$$

Y se tiene entonces que:

$$d_i^{-\alpha} = \left(\frac{1}{\sqrt{(x_i - x_p)^2 + (y_i - y_p)^2}} \right)^2$$

El siguiente paso es generar una tabla de distancias (Cuadro 12) de cada una de las estaciones con respecto al punto que se desea interpolar, posteriormente se obtiene la suma de estas distancias y se calcula w para cada estación, dividiendo cada distancia entre la suma total de estas.

Cuadro 12. Datos obtenidos para el método de distancia inversa.

Estación	Ubicación		d_i	$d_i^{-\alpha}$ ($\alpha=2$)	w
	x	y			
Z*	70	30			
1	130	10	63.245	0.00025	0.068
2	40	50	36.055	0.000769	0.209
3	20	30	50	0.0004	0.109
4	90	90	63.245	0.00025	0.068
5	60	10	22.360	0.002	0.545
			Suma	0.00366923	

Finalmente se interpola multiplicando cada w por el valor de tal punto:

$$R_p = (0.068*100) + (0.209*60) + (0.109*40) + (0.068*95) + (0.545*80)$$

$$R_p = 73.8312$$

Aproximación Multicuadrática

Para la interpolación multicuadrática podemos partir de la ecuación 4.15:

$$W_j = \sum_{i=1}^n \delta_{ij} d_{ei} ; j = 1, 2, \dots, n$$

Por lo que se requiere una matriz de distancias D_{ij} , estas se calculan entre los cinco puntos de referencia, la distancia calculada es la geométrica, quedando su formulación de la siguiente manera:

$$D_{ij} = \begin{bmatrix} 0 & d_{12} & d_{13} & d_{14} & d_{15} \\ d_{21} & 0 & d_{23} & d_{24} & d_{25} \\ d_{31} & d_{32} & 0 & d_{34} & d_{35} \\ d_{41} & d_{42} & d_{43} & 0 & d_{45} \\ d_{51} & d_{52} & d_{53} & d_{54} & 0 \end{bmatrix}$$

Y ya que $\delta_{ij} = D_{ij}^{-1}$, falta obtener d_{ei} , que es la matriz de distancias al punto a interpolar, semejante a la obtenida para el método de la distancia inversa.

$$d_{ei} = \begin{bmatrix} d_{e1} \\ d_{e2} \\ d_{e3} \\ d_{e4} \\ d_{e5} \end{bmatrix}$$

Y los coeficientes W_j se obtienen multiplicando δ_{ij} por d_{ei} . En este ejemplo, la matriz de distancias D_{ij} quedaría con los siguientes valores:

0	98.488	111.803	89.442	70
98.488	0	28.284	64.031	44.721
111.803	28.284	0	92.195	44.721
89.442	64.031	92.195	0	85.440
70	44.721	44.721	85.440	0

Luego se forma la matriz de distancias respecto al punto a interpolar, d_{ei}

63.245
 36.055
 50
 63.245
 22.360

Se calcula la matriz inversa de D_{ij} que es δ_{ij} :

-0.0069 -0.0009 0.0012 0.0054 0.0065
 -0.0009 -0.0268 0.0149 0.0069 0.0049
 0.0012 0.0149 -0.0188 0.0009 0.0077
 0.0054 0.0069 0.0009 -0.0086 -0.0002
 0.0065 0.0049 0.0077 -0.0002 -0.0189

Y se multiplica δ_{ij} por d_{ei} para obtener los pesos W para cada estación:

1	0.08646
2	0.27376
3	-0.08622
4	0.09270
5	0.53759

Por lo que se calcula:

$$R_p = (0.0865 \cdot 100) + (0.2737 \cdot 60) + (-0.0862 \cdot 40) + (0.0927 \cdot 95) + (0.5376 \cdot 80)$$

$$R_p = 73.4383$$

Kriging

El modelo de Kriging en su versión más sencilla incorpora un variograma lineal, esta concepción dependerá siempre del modelo de variograma óptimo para cada caso en específico, por lo que el primer paso es precisamente obtener el variograma, usando la ecuación 4.19:

$$\gamma(h) = \frac{1}{2n} \sum_{i+1}^n (Z_{x_i} - Z_{x_{i-h}})^2$$

A efectos prácticos los cálculos son realizados por programas como Surfer, sin embargo, estos requieren que el usuario sepa elaborar un variograma y el ejemplo aquí presentado tiene pocos puntos para realizarlo de manera adecuada, por lo que se presenta un ejemplo más adecuado, tenemos una cuadrícula con diferentes valores para ilustrar la elaboración más simple (Clark, 1979):

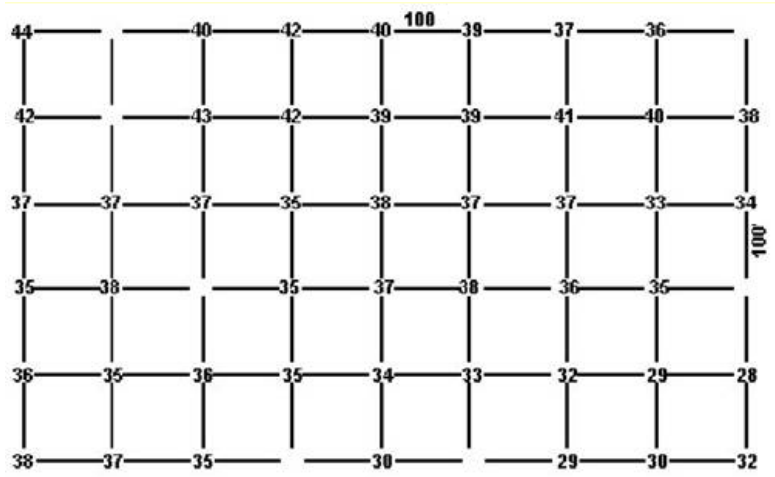


Figura 14. Ejemplo de kriging lineal.

La distancia entre cada punto es de 100 y existen puntos sin valores. El primer paso es determinar una distancia inicial h para evaluar la función del semivariograma, normalmente se utiliza la distancia menor entre dos puntos, en este caso, 100. Como se vio en el apartado III.2 el semivariograma posee una dirección específica, primeramente evaluaremos la dirección horizontal:

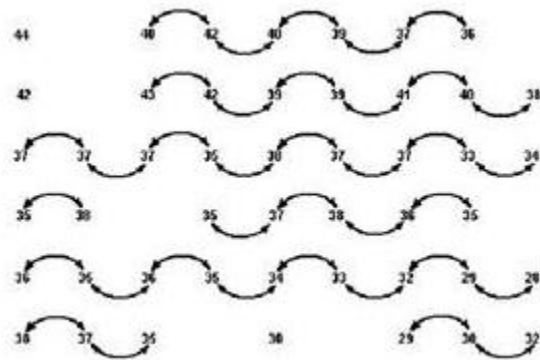


Figura 15. Evaluación de distancias entre puntos en dirección horizontal.

Como podemos ver, únicamente se evalúan los pares de puntos $(Z_{x_i} - Z_{x_i-h})$ que están a una distancia de 100, quedando algunos sin evaluar debido a los valores ausentes, el total de estos pares es el valor de n que se utiliza en la ecuación 4.19, en este caso $n=36$; y el valor calculado de $\gamma(100)$ es de 1.46. El siguiente paso es aumentar la distancia a 200 y volver a realizar el cálculo:

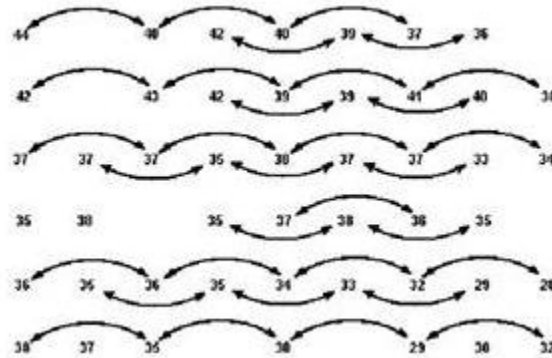


Figura 16. Evaluación de distancias entre puntos en dirección horizontal y distancia de 200.

El valor obtenido en este caso es $\gamma(200) = 3.3$, y podemos hacer esto sucesivamente hasta una distancia de 800, la mayor posible, aunque en la mayoría de los casos normalmente se llega sólo a la mitad de la distancia máxima. El siguiente paso es graficar los valores obtenidos, en este caso, con los dos puntos podemos esbozar el variograma:

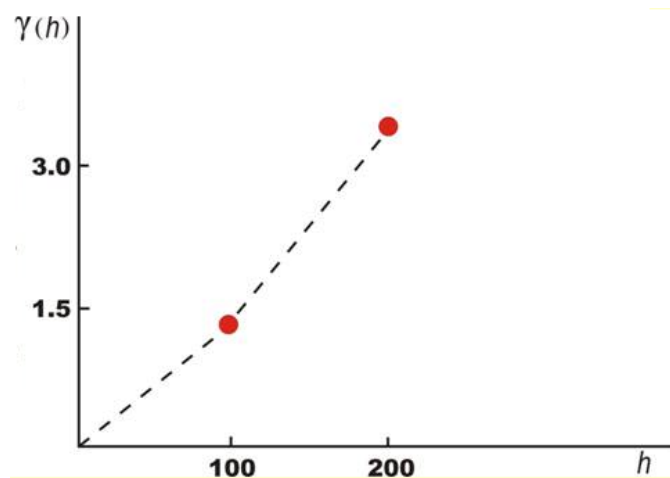


Figura 17. Grafica de los puntos del variograma experimental.

Una vez completado el variograma experimental en dirección horizontal lo adecuado sería realizarlo en otras direcciones: verticalmente, en diagonal y con ángulos diferentes, para ver si hay una variación notable y por lo tanto anisotropía, ya que esta labor resulta tediosa, como ya se mencionó antes, existen programas que pueden ayudarnos a elaborar estos variogramas.

El variograma obtenido se llama *variograma experimental*, y el siguiente paso es ajustarlo a una ecuación para ser utilizada en el procedimiento de kriging, las ecuaciones más típicas son la lineal, esférica, exponencial, logarítmica y gaussiana, además es en este paso cuando hay que elegir el valor de otros parámetros como el *sill* y el *nugget*:

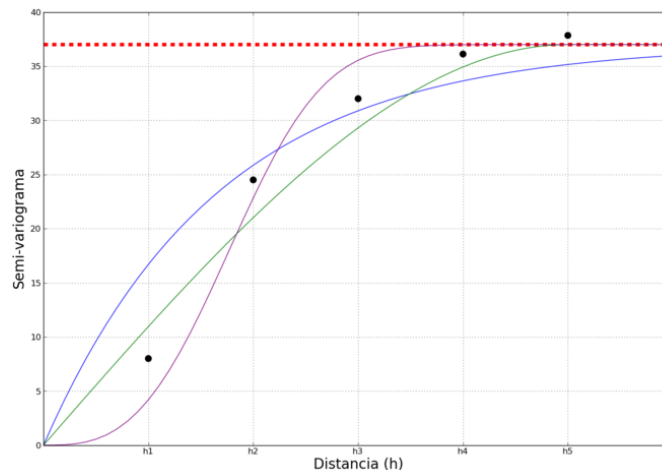


Figura 18. Variograma ajustado a diferentes modelos.

Este proceso es en sí exploratorio, aunque normalmente el programa Surfer y otros nos dan un modelo sugerido ya con valores ajustados, no quiere decir que los aceptemos de manera automática, pues estas herramientas también nos facilitan el probar diferentes tipos de ajuste y valores de parámetros. Finalmente se obtiene una ecuación descriptiva del variograma, llamada *variograma teórico* que será usada en el procedimiento del kriging que se describe a continuación.

Ya que el ejemplo de este anexo no tiene suficientes datos para elaborar un variograma, este se considerará como lineal, por lo que el proceso a seguir queda simplificado de la siguiente forma, tomando como base la ecuación 4.22:

$$\begin{bmatrix} c_{11} & c_{12} & \cdot & c_{1n} & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & \cdot & c_{nn} & 1 \\ 1 & 1 & \cdot & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ w_n \\ \mu \end{bmatrix} = \begin{bmatrix} c_{10} \\ \cdot \\ \cdot \\ c_{n0} \\ 1 \end{bmatrix}$$

Se forma una matriz de distancias semejante a la del modelo multicuadrático, que llamaremos C:

0	98.488	111.803	89.442	63.245
98.488	0	28.284	64.031	36.055
111.803	28.284	0	92.195	50
89.442	64.031	92.195	0	63.245
70	44.721	44.721	85.440	0

Igualmente una matriz de distancias respecto al punto a interpolar, que llamaremos F:

63.245
36.055
50
63.245
22.360

Si se tuviera un modelo de variograma es aquí cuando se tiene que formar una matriz D multiplicando cada uno de los valores de C por el modelo de variograma, en este caso considerando un variograma lineal la matriz C queda igual.

A continuación, para efectuar los cálculos descritos en la ecuación 4.22, a la matriz C se le agrega una columna y una fila cuyos valores sean =1 a excepción del elemento inferior derecho cuyo valor es siempre 0.

0	98.488	111.803	89.442	63.245	1
98.488	0	28.284	64.031	36.055	1
111.803	28.284	0	92.195	50	1
89.442	64.031	92.195	0	63.245	1
70	44.721	44.721	85.440	0	1
1	1	1	1	1	0

Así mismo a la matriz F se le agrega una fila con un valor de 1:

63.245
 36.055
 50
 63.245
 22.360
 1

A continuación, para obtener la matriz de valores w se resuelve 4.22, para lo que necesitamos la matriz inversa de C :

-0.008 -0.0004 -0.0007 0.003 0.006 0.359
 -0.001 -0.027 0.014 0.007 0.008 -0.058
 0.0005 0.016 -0.019 -0.0009 0.003 0.403
 0.003 0.006 -0.001 -0.009 0.002 0.293
 0.006 0.005 0.008 -0.0003 -0.019 0.001
 0.402 -0.026 0.454 0.292 -0.123 -65.72

Y se multiplica $C^{-1} \cdot F$ para obtener la matriz de pesos W .

0.1357
 0.1798
 0.0756
 0.0493
 0.5598
 -2.7527

Para interpolar el valor buscado se multiplican los pesos obtenidos por los valores medidos, descartándose el último peso, debido a que este proviene del renglón añadido por el método. Por tanto:

$$R_p = (0.1352 \cdot 100) + (0.1798 \cdot 60) + (0.0757 \cdot 40) + (0.0494 \cdot 95) + (0.5598 \cdot 80)$$

$$R_p = 76.8214685$$

Es en estos casos donde se tienen pocos datos disponibles que se puede usar el método del cokriging, ya que teniendo una variable auxiliar se puede obtener un variograma adecuado a nuestro conjunto de datos, en este caso podríamos usar datos de elevación (normalmente disponibles detalladamente) y formar una malla de puntos:

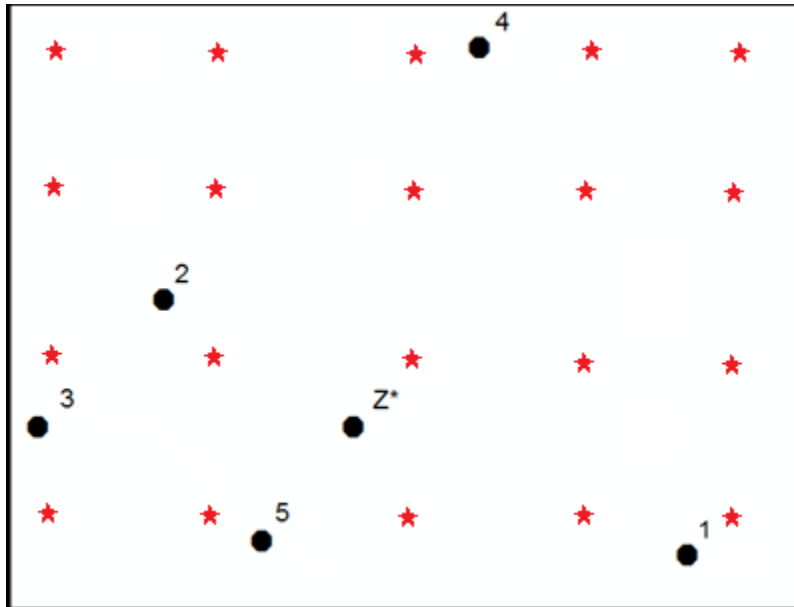


Figura 19. Puntos teóricos de elevación para un análisis de cokriging.

Así, teniendo más datos disponibles se puede aplicar la ecuación 4.25:

$$2\gamma_{1,2} = \text{var}\{Z_1(s+h) - Z_2(s)\}$$

Que se puede escribir de manera desglosada como:

$$\gamma_{v_1v_2}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} \{Z_{v_1}(x+h) - Z_{v_1}(x)\} \{Z_{v_2}(x+h) - Z_{v_2}(x)\}$$

Mediante esta fórmula podemos obtener un variograma alternativo que relacione dos series de datos distintas, la estimación de los parámetros w para calcular valores se puede obtener con una ecuación como la 4.22, sin embargo, para diferenciarla del kriging ordinario la reescribimos:

$$Y_{ij}\lambda = Y_{i0}$$

Donde:

$$Y_{ij} = \begin{bmatrix} \gamma_{v_1}(1,1) & \dots & \gamma_{v_1}(n,1) & \gamma_{v_1v_2}(1,1) & \dots & \gamma_{v_1v_2}(m,1) & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{v_1}(1,n) & \dots & \gamma_{v_1}(n,n) & \gamma_{v_1v_2}(1,n) & \dots & \gamma_{v_1v_2}(m,n) & 1 & 0 \\ \gamma_{v_1v_2}(1,1) & \dots & \gamma_{v_1v_2}(n,1) & \gamma_{v_2}(1,1) & \dots & \gamma_{v_2}(m,1) & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{v_1v_2}(1,m) & \dots & \gamma_{v_1v_2}(n,m) & \gamma_{v_2}(1,m) & \dots & \gamma_{v_2}(m,m) & 0 & 1 \\ 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0 \end{bmatrix}$$

$$\lambda = \begin{bmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \\ \mu_1 \\ \mu_2 \end{bmatrix} \quad y \quad Y_{i0} = \begin{bmatrix} \gamma_{v_1}(0,1) \\ \vdots \\ \gamma_{v_1}(0,n) \\ \gamma_{v_1v_2}(0,1) \\ \vdots \\ \gamma_{v_1v_2}(0,m) \\ 1 \\ 0 \end{bmatrix}$$

Se debe elaborar un variograma para cada variable además del variograma que las contenga a ambas, estos pueden ser lineales según el caso.

La solución es similar al proceso de kriging, nótese que para la formulación de Y_{ij} ya se consideran los valores transformados mediante la función del variograma teórico y los valores agregados ahora constan de dos filas y columnas que alternan unos y ceros. Sin embargo en este procedimiento no existe una matriz de pesos w sino que se obtienen dos matrices de pesos; a y b las cuales, para obtener los valores calculados, se emplean en la formulación:

$$Rp = \sum_{i=1}^{n1} a_i Z_{v_1}(x_i) + \sum_{j=1}^{n2} b_j Z_{v_2}(x_j)$$

Es por eso que los datos complementarios deben estar altamente correlacionados con la variable de nuestro interés, además de tener valores en una escala semejante, por lo que en caso de utilizar datos de elevación y datos de precipitación lo más común es someterlos a un proceso de transformación para adecuarlos a este análisis.

TIN

Para generar la red de triángulos irregulares simplemente se unen los puntos de referencia a los puntos más cercanos, formando el conjunto de triángulos mostrados en la Figura 14:

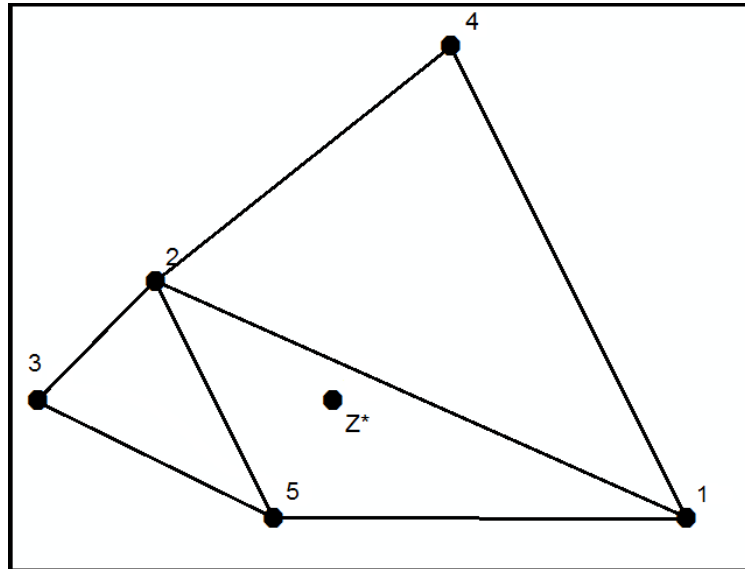


Figura 20. Red de triángulos irregulares.

Y para calcular el valor que se desea interpolar se toman en cuenta los vértices del triángulo en el cual se encuentra el punto buscado, formando un sistema de tres ecuaciones con tres incógnitas en el cual se toman en cuenta las coordenadas de dichos vértices y su valor medido:

$$\begin{cases} Z_1 = a + bx_1 + cy_1 \\ Z_2 = a + bx_2 + cy_2 \\ Z_3 = a + bx_3 + cy_3 \end{cases}$$

Para este caso en particular, en la figura 14 podemos observar que el punto a interpolar se encuentra en el triángulo delimitado por los puntos 1, 2 y 5 por lo que podemos formar las siguientes ecuaciones:

$$\begin{cases} 100 = a + 130b + 10c \\ 60 = a + 40b + 50c \\ 80 = a + 60b + 10c \end{cases}$$

Resolviendo el sistema de tres ecuaciones obtenemos que:

$$a= 66.4286 ; b= 0.2857 ; y c= -0.3571$$

por lo que para encontrar el valor buscado, sustituimos en la ecuación:

$$Z^* = a + b(x) + c(y)$$

$$66.4286 + (0.2857*70) + (-0.3571*30) = 75.7141429$$

Así como se observó en las conclusiones de esta tesis, los métodos proporcionan valores muy semejantes, y dada una red adecuadamente distribuida y extensa, no existe diferencia significativa entre los métodos y dependerá del usuario final la decisión de usar cada método dependiendo también de las consideraciones citadas también en este trabajo.

Tabla 13. Valores calculados para el caso ejemplificado.

Método	Valor para Z*
Distancia Inversa	73.8312
Aproximación multicuadrática	73.4383
Kriging	76.8214
TIN	75.7141

Anexo II.

Tabla de disponibilidad de datos de precipitación por estación

