

ESTIMACION DE LA DENSIDAD BAJO EL METODO
ESPECIAL KRIGING Y EL INDIVIDUO MAS CERCANO

ISMAEL HERNANDEZ BETANCOURT

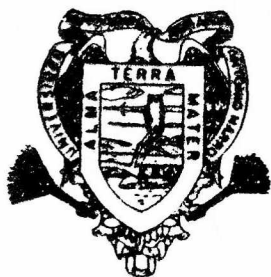
T E S I S

PRESENTADA COMO REQUISITO PARCIAL

PARA OBTENER EL GRADO DE Universidad Autónoma Agraria
MAESTRO EN CIENCIAS "ANTONIO NARRO"
EN ESTADISTICA EXPERIMENTAL



BIBLIOTECA



Universidad Autónoma Agraria
Antonio Narro

PROGRAMA DE GRADUADOS

Buenavista, Saltillo, Coah.

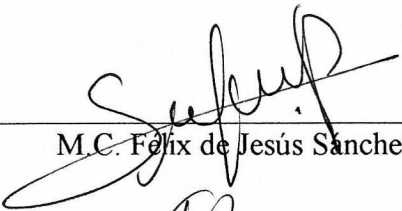
DICIEMBRE DE 1997

Tesis elaborada bajo la supervisión del comité particular de asesoría y aprobada como requisito parcial, para optar al grado de

MAESTRO EN CIENCIAS EN
ESTADÍSTICA EXPERIMENTAL

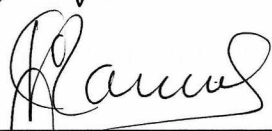
COMITÉ PARTICULAR

Asesor principal:



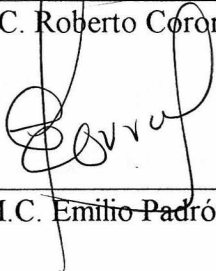
M.C. Félix de Jesús Sánchez Pérez

Asesor:



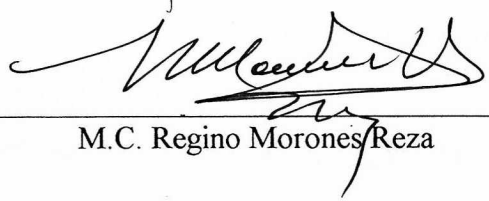
M.C. Roberto Coronado Niño

Asesor:




M.C. Emilio Padrón Corral

Asesor:



M.C. Regino Morones Reza



Dr. Jesús Manuel Fuentes Rodríguez
Subdirector de Postgrado

Buenavista, Saltillo, Coahuila. Diciembre de 1997

AGRADECIMIENTOS

A la Secretaría de Agricultura, Ganadería y Desarrollo Rural (SAGAR), por el apoyo rotundo que me otorgó durante el postgrado.

A mi Alma Mater, por darme abrigo una vez más.

Al Ing. M.C. Félix de Jesús Sánchez Pérez, mi maestro y amigo, por su denodado esfuerzo en el presente estudio.

Al Lic. M.C. Roberto Coronado Niño, por sus comentarios y consejos en el trabajo de campo, de igual manera al Lic. M.C. Emilio Padrón Corral y al Ing. M.C. Regino Morones Reza.

A los señores: Dr. Rolando Cavazos Cadena, Dr. José Antonio Díaz García y Dr. Mario Cantú Sifuentes por sus mil afanes para compartir conmigo su gran conocimiento.

A todos mis compañeros, quienes de alguna forma siempre me tendieron la mano durante nuestra convivencia académica.

DEDICATORIA

Para Martha, mi nunca bien ponderado amor.

COMPENDIO

ESTIMACIÓN DE LA DENSIDAD BAJO EL MÉTODO ESPACIAL
KRIGING Y EL INDIVIDUO MÁS CERCANO

POR

ISMAEL HERNÁNDEZ BETANCOURT

MAESTRÍA EN ESTADÍSTICA EXPERIMENTAL

UNIVERSIDAD AUTÓNOMA AGRARIA "ANTONIO NARRO"

BUENAVISTA, SALTILLO, COAHUILA. DICIEMBRE 1997

Ing. M.C. Félix de Jesús Sánchez Pérez - Asesor -

Palabras clave: Geoestadística, variabilidad espacial, umbral, semivariograma, efecto
Nugget, rango.

En agronomía que es una actividad en la que se involucran factores bióticos y abióticos, el procedimiento de muestreo clásico aplicado a casos como el presente genera errores de estimación.

Investigaciones realizadas han demostrado que la geoestadística proporciona un método de análisis de las propiedades espaciales de un conjunto de datos, dando resultados más precisos. Se puede afirmar, que básicamente es una técnica que genera un estimador insesgado de varianza mínima.

El método antes citado se basa en el variograma, que es esencial en todo estudio geoestadístico y debe ser estimado con suficientes datos, con una escala apropiada y modelado con exactitud. Éste capta las principales formas de variación de la variable, mismas que se utilizan para realizar las estimaciones de nuevos valores.

Este trabajo describe el método espacial Kriging, así como el método clásico denominado el Individuo más Cercano, siendo su objetivo comparar las estimaciones sobre densidad en Gobernadora (*Larrea tridentata*, Cov.), la cual es una planta que abunda en el desierto mexicano y que tiene una importancia potencial en medicina, agronomía, ecología, etc.

La conclusión de este estudio es que el método del Individuo más Cercano no proporciona buenos resultados en la estimación de la densidad debido a que la variable no

tiene independencia. Por el contrario el método kriging presenta una buena robustez, ya que la dependencia en la variable hace propicio el uso de esta opción técnica.

ABSTRACT

ESTIMATION OF THE DENSITY UNDER THE SPATIAL METHOD
KRIGING AND THE CLOSEST INDIVIDUAL

BY

ISMAEL HERNÁNDEZ BETANCOURT

MASTER OF SCIENCE WITH MAJOR IN EXPERIMENTAL STATISTICS

UNIVERSIDAD AUTÓNOMA AGRARIA "ANTONIO NARRO"

BUENAVISTA, SALTILLO, COAHUILA. DECEMBER 1997

Ing. M.C. Félix de Jesús Sánchez Pérez - Advisor -

Key Words: Geostatistics, spatial variability, sill, semivariogram, Nugget effect,
range.

In agronomy that it is an activity in the one which are involved factors biotic and abiotic, the classic sampling procedure applied to cases as the present generates estimate mistakes.

Accomplished investigations have demonstrated that the geostatistics provides an analysis method of the spatial properties of a set of data, paying off more accurate. It can be asserted, that basically it is a technique that it generates a unbiased estimator of minima variance.

The method before cited is based in the variogram, that it is essential in all geostatistical study and it must be estimated with sufficient data, with an appropriate scale and modeled with accuracy. This method captures the principal forms of variation of the variable, same that are used to accomplish the new values estimates.

This work describes the spatial method kriging and the classic method designated the closest individual, being its objective to compare the estimates on density in gobernadora (*Larrea tridentata*, Cov), the one wich is a plant that abounds in the mexican desert and that has a potential importance in medicine, agronomy, ecology, etc.

The conclusion of this study is that the method of the closest individual doesn't provide good results in the estimation of the density due to the fact that the variable doesn't have independence. On the contrary the kriging method presents a good

hardiness, since the independence in the variable make auspicious the use of this technical option.

ÍNDICE DE CONTENIDO

	Página
ÍNDICE DE CUADROS	xiii
ÍNDICE DE FIGURAS	xiv
INTRODUCCIÓN	1
REVISIÓN DE LITERATURA	4
Generalidades e Importancia de la Gobernadora (<i>Larrea tridentata</i> , Cov.)	4
Antecedentes del Método Kriging	6
Aplicación del Método Kriging	8
Teoría de las Variables Regionalizadas	9
La Función de Covarianza Espacial	9
El Semivariograma	10
Efecto Nugget (Co)	14
Ajuste del Semivariograma Teórico al Experimental	15
Interpolación para la Predicción	15
Aplicación del Método del Individuo más Cercano	17
Descripción del Método	19
MATERIALES Y MÉTODOS	21
Ubicación del Área de Estudio	21

Recursos Empleados	21
Uso del Método Kriging	22
Levantamiento de los Datos de Campo	22
Cálculo del Semivariograma Experimental	26
Ajuste del Semivariograma Experimental al Teórico	31
Interpolación de Valores	34
Uso del Método del Individuo más Cercano	41
Aspectos Generales	41
Levantamiento de los Datos de Campo	44
Estimación de la Población	45
RESULTADOS Y DISCUSIÓN	47
Evaluación del Método Kriging	47
Evaluación del Método del Individuo más Cercano	49
CONCLUSIONES Y RECOMENDACIONES	52
RESUMEN	53
LITERATURA CITADA	55
APÉNDICES	57
A. ESTIMADOR INSESGADO Y DE VARIANZA MÍNIMA	58
B. CONSIDERACIONES GENERALES DEL MUESTREO CLÁSICO	62

ÍNDICE DE CUADROS

Cuadro No.		Página
3.1	Observaciones obtenidas para el método kriging	25
3.1	(continuación)	26
3.2	Resumen para la variable en el método kriging	26
3.3	Valores para la construcción del semivariograma experimental ...	29
3.4	Valores ajustados del semivariograma	34
3.5	Predicción de los $\hat{Z}(x_0)$	41
3.6	Observaciones obtenidas para el cálculo de la varianza	43
3.7	Resumen para la variable en el método del individuo más cercano	44
3.8	Datos para el individuo más cercano	46
4.1	Errores reducidos	48
4.2	Resumen de los errores reducidos	48
4.3	Parámetros verdaderos de la población	50
4.4	Comparación de los resultados en el método del individuo más cercano	51

ÍNDICE DE FIGURAS

Figura No.		Página
2.1	Localización de la gobernadora	5
2.2	Semivariograma teórico	12
2.3	Dispersiones de las poblaciones ecológicas	18
2.4	Diagrama para el método del individuo más cercano	20
3.1	Localización del área de estudio en la República Mexicana	22
3.2	Localización específica del área de estudio	23
3.3	Croquis del levantamiento de los datos para el método kriging ...	24
3.4	Gráfica de la distancia v. s. semivariograma	30
3.5	Ajuste del semivariograma	32
3.6	Trayectoria del semivariograma ajustado	33

INTRODUCCIÓN

En los tiempos actuales el aprovechamiento de los recursos naturales y la preservación del medio ambiente, son dos situaciones perfectamente amalgamadas que acaparan la atención de quienes dirigen los destinos del mundo, de tal manera que ésto es discutido en todos los medios y en ocasiones llega a constituirse hasta en motivo políticc.

El hecho de que se puedan hacer predicciones o estimaciones sobre poblaciones o volúmenes de recursos es algo muy importante, pues permite en el mejor de los casos hacer una planeación adecuada y así optimizar una determinada explotación, lo que conlleva a cuidar el medio ambiente y administrar cabalmente los recursos.

En estudios geológicos con frecuencia es necesario predecir la cantidad de mineral o de combustible fósil que es factible encontrar en un sitio. La predicción puede estar basada en valores observados en otros sitios de la región y estos otros sitios pueden estar espaciados regular o irregularmente en la misma.

En estudios agronómicos y ecológicos a veces es necesario conocer la densidad de plantas, la cantidad de biomasa o el número de animales que existen en algún lugar o en su defecto los totales existentes en una región completa. Para solucionar estos

problemas se tienen los métodos clásicos de muestreo, probabilístico y no probabilístico, los cuales llegan a proporcionar una opción satisfactoria.

Las soluciones para los casos mencionados anteriormente, tienen una opción muy satisfactoria con el Método Kriging, el cual está siendo ampliamente usado en muchas disciplinas y se ha comprobado que los resultados que arroja son más exactos, que los que se obtienen por los métodos convencionales.

En general cuando el muestreo no es estrictamente probabilístico se dá por asentado que en la selección de la muestra intervienen los criterios de los seleccionadores, lo que hace imposible medir la confianza que debe tenerse en los resultados. Algunos métodos considerados no probabilísticos usados en casos como éste son: El Vecino más Cercano, el Cuadrante de Punto Central y el Individuo más Cercano, siendo este último el que se involucra en el estudio.

El presente trabajo se hace tratando de contribuir a que los investigadores o las personas que necesiten realizar una medición semejante a la presente, cuenten con un apoyo y una referencia para que puedan efectuar sus acciones con algo más de facilidad y confianza.

Se seleccionó la gobernadora (*Larrea tridentata*, Cov.) que es una especie arbustiva con gran importancia y potencial de aprovechamiento, también muy común en

el norte de nuestro país y que ha sido objeto de diversos estudios que han arrojado luz sobre todo lo que se dice en relación a ella, a la cual se le atribuyen bondades algo exageradas, sin que este hecho sea causa para dejar de reconocer lo que se ha comprobado científicamente.

REVISIÓN DE LITERATURA

Generalidades e Importancia de la Gobernadora (*Larrea tridentata*, Cov.)

Este arbusto se desarrolla ampliamente en el norte de México y parte de los Estados Unidos de América (EUA). En cuanto a la superficie que ocupa en México se sabe que son aproximadamente 48 000 000 de hectáreas (Figura 2.1), es decir casi la cuarta parte del territorio nacional, presenta una altura que va de los treinta a los ciento veinte centímetros con una cobertura algo mayor a su altura y también se encuentra provisto de un olor característico muy fuerte.

De las propiedades que se sabe que esta planta tiene, está su poder fungicida, bactericida y nematocida por lo que no es raro que en las regiones donde tiene su hábitat las personas lo usen para tratar algunos problemas de micosis. También se sabe que puede ser usada para combatir la amibeasis humana y los cálculos renales, de igual manera puede ser usada como forraje mediante un tratamiento químico.

En un trabajo de campo hecho por Huerta (1986), se encontró que la resina de la gobernadora puede desarrollar cierta acción nematocida, que puede ser canalizada para atacar problemas de nemátodos en algunos cultivos.

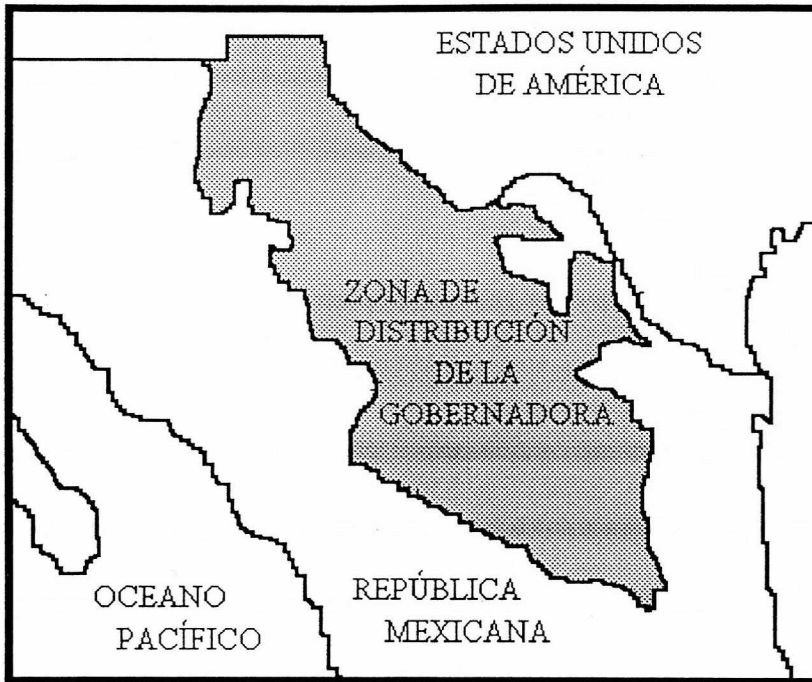


Figura 2.1. Localización de la gobernadora

González (1989), encontró que la resina de gobernadora presenta propiedades sistemáticas para controlar la bacteria *Pseudomonas solanacearum* en el cultivo de la papa (*Solanum tuberosum*, L).

La gobernadora no es palatable para los animales en su forma natural porque tiene en su constitución bioquímica resinas y ácidos que impiden su consumo directo, siendo éste un mecanismo de supervivencia, pero si recibe un tratamiento químico previo, puede serlo pues tiene un poder nutricional semejante a la alfalfa (*Medicago sativa*, L).

En una investigación realizada por García (1993), se determinó que es posible el uso de esta planta en la alimentación animal mediante un tratamiento con NaOH al 0.1 por ciento de normalidad, con lo que se favorece la digestibilidad de manera suficiente para ser consumida como forraje.

Antecedentes del Método Kriging

Los estudios que involucran situaciones de variabilidad espacial son relativamente nuevos, los primeros intentos por explicar éste fenómeno datan de principios del presente siglo. Se sabe que aproximadamente en el año de 1950 se acuñó el término kriging para nombrar éste nuevo concepto, mismo que a partir de entonces se ha extendido su aplicación a muchos campos de la ciencia y la tecnología.

Basados en el muestreo clásico se ha observado que algunas variables presentan un cierto patrón de comportamiento en función de la posición relativa de unas con otras, es decir las variables cercanas tienden a parecerse más entre ellas, mientras que con las más alejadas sucede lo contrario; es aquí donde surge el concepto de variabilidad y dependencia espacial.

A este respecto señala Cortés (1988), que si la dependencia espacial existe se puede, utilizar la metodología kriging para realizar estimaciones del valor de la variable en localidades no medidas.

Generalmente para atacar problemas de este tipo se usan dos herramientas que son el semivariograma y el krigado, el primero proporciona el modelo de dependencia espacial y el segundo permite obtener las estimaciones en los puntos o localidades que son de interés pero cuyos valores se ignoran.

Por otra parte se sabe que existen variables que no poseen dependencia espacial, es decir tienen un semivariograma constante para cualquier distancia, por lo que el tratamiento del problema deberá hacerse usando los métodos estadísticos clásicos.

En un trabajo desarrollado por Cortés (1988), se concluye que es recomendable usar la metodología kriging para todas aquellas variables que presentan una dependencia espacial, dado que la interpolación proporcionará valores insesgados y de varianza mínima.

Por otra parte Vieira *et al.* (1983), señalan que el proceso de muestreo en problemas agronómicos puede llegar a crear confusión y sesgo en el análisis y muestran que la geoestadística proporciona un método para el análisis del espacio y propiedades temporales de ciertos datos, también proporciona un método de interpolación entre puntos seleccionados.

En un estudio que realizaron Díaz y Gutiérrez (1994), para evaluar el tamaño de un banco de semillas en un área rural del Distrito Federal, se determinó que el método

geoestadístico fue el más robusto en comparación al poligonal por el hecho de existir dependencia espacial.

Por lo anterior se sabe que los métodos geoestadísticos son una herramienta potencial, misma que aún no alcanza su pleno desarrollo en muchas disciplinas que se sabe son susceptibles de aceptar el método, en nuestro país, dicha técnica no ha recibido todavía la ponderación debida, pero llegado el momento va a ayudar en forma contundente a solucionar problemas en los ámbitos agronómico y ecológico de manera rápida y confiable.

Aplicación del Método Kriging

En estudios agroecológicos las observaciones de animales o plantas en algunos sitios, pueden ser usados para predecir la abundancia de éstos organismos en un nuevo sitio o en una región completa.

Para estas situaciones se puede pensar que la cantidad de individuos de la especie de interés es una variable aleatoria $Z(x_i)$, $i = 1, 2, 3, \dots, n$ de la que se hacen observaciones en n sitios $t_1, t_2, t_3, \dots, t_n$ y se desea estimar el valor $Z(x_0)$ que es el valor de la variable aleatoria en un lugar t_0 , que es a su vez la localización del valor antes mencionado.

Las ecuaciones de predicción pueden ser escritas indistintamente con covarianzas o con el semivariograma.

Teoría de las Variables Regionalizadas

Esta teoría deja asentado que si se toman muestras cuyas distancias entre si son pequeñas, la dependencia del valor de la variable es grande sucediendo lo contrario si la toma de muestras es efectuada involucrando distancias mayores. Ramírez (1980) dice que para hacer inferencia es necesario definir el concepto de estacionaridad mismo que existe cuando los dos primeros momentos de la función de distribución son invariantes bajo traslación es decir:

(a) El valor esperado de la función Z en cualquier punto x es un valor constante $E[Z(x)] = m$.

(b) La función de Covarianza en los puntos x y $x + h$ es independiente del punto x , dependiendo unicamente del vector h es decir $E[Z(x) - Z(x + h)] - m^2 = c(h)$, siendo h la distancia entre los puntos.

La Función de Covarianza Espacial

En el contexto agroecológico los valores de la variable de interés, son generalmente no independientes, espacialmente hablando por lo que los valores cercanos

tienden a ser parecidos entre sí, sucediendo exactamente lo contrario con los valores alejados.

En síntesis la relación que guardan las variables $Z(x_1)$ y $Z(x_2)$ asociadas con los sitios t_1 y t_2 respectivamente es la covarianza:

$$\text{Cov}[Z(x_1), Z(x_2)] = E\left\{\left[Z(x_1) - E[Z(x_1)]\right]\left[Z(x_2) - E[Z(x_2)]\right]\right\}$$

Al llegar a este punto se hace necesario recalcar una situación que Thompson (1993), la señala de forma precisa escribiendo que, cuando la covarianza en dos sitios distintos depende sólo de su posición relativa y no de su localización exacta dentro del área de estudio, esta relación puede resumirse con una función de covarianza $c(h)$, donde los dos puntos están separados por una distancia h , ésto es: $c(h) = \text{Cov}[Z(x_i + h), Z(x_i)]$.

Si el valor esperado de la variable de interés es el mismo en cada sitio y la covarianza depende tanto de la distancia como de la dirección entre sitios, el proceso es llamado anisotrópico o estacionario de segundo orden. Si la covarianza depende sólo de la distancia entre sitios y no de la dirección, entonces el proceso es llamado isotrópico.

El Semivariograma

Este concepto es fundamental en geoestadística, mismo que expresa la situación cuantitativa de una variable regionalizada y se debe calcular en al menos dos

direcciones para detectar la presencia de anisotropías que es la tendencia en determinadas direcciones y de no existir tal tendencia se dice que el fenómeno es isotrópico.

Existen dos formas para el semivariograma en todo trabajo y son el experimental y el teórico, del primero se puede decir que es el que se obtiene por medio del trabajo de campo y solamente se dispone de puntos muestrales o sea no es continuo.

El semivariograma teórico (Figura 2.2) es una función continua que trata de expresar lo mejor posible al semivariograma experimental y existen básicamente cuatro modelos a usar en estas tareas ellos son:

(1) Modelo Elevado a una Potencia.

$$\hat{\gamma}(h) = ax^b \quad 0 < b < 2$$

(2) Modelo Esférico.

$$\hat{\gamma}(h) = C \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] \quad 0 \leq h \leq a$$

$$= C \quad h > a$$

Siendo:

a = rango

C = valor de umbral

(3) Modelo Exponencial.

$$\hat{\gamma}(h) = C \left[1 - e^{-\frac{h}{a}} \right]$$

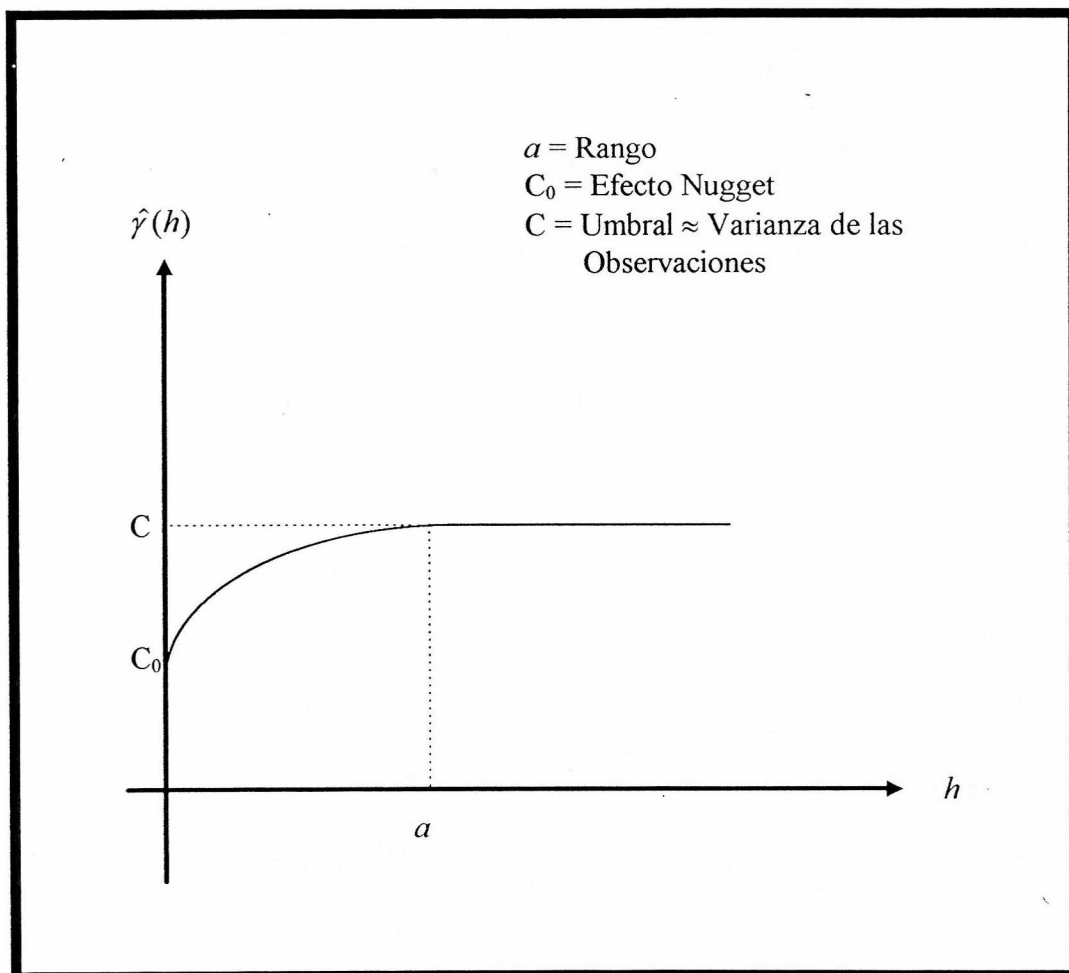


Figura 2.2. Semivariograma teórico

Siendo:

$3a \approx \text{Rango práctico (95 por ciento del valor del umbral)}$

$C = \text{valor del umbral}$

(4) Modelo Gaussiano.

$$\hat{\gamma}(h) = C \left[1 - e^{-\left(\frac{h}{a}\right)^2} \right]$$

Siendo:

$1.73a \approx$ Rango Práctico (95 por ciento del valor del umbral).

Los modelos más usados en la actualidad son los elevados a una potencia y los esféricos en primer término seguidos por los exponenciales y gaussianos en segundo.

Si se considera un campo de área S y se identifica a un conjunto de puntos denotados $\{Z(x_i), i = 1, 2, 3, \dots, n\}$, cada punto $Z(x_i)$ poseerá una determinada localización mediante un par de coordenadas, entonces el valor $Z(x_i)$ tendrá un valor particular para un sitio en particular. La variable regionalizada $Z(x_i)$ para todo x_i se considera una variable aleatoria.

Hasta entrada la década de los sesentas, los estudios de este tipo generalmente consideraban independencia en esta forma de variable, aún cuando en el mundo ya se había empezado a desarrollar el manejo del método kriging

En una investigación realizada por Vieira *et al.* (1982), se deja asentado que cuando las observaciones están distribuidas bidimensionalmente, en forma cruzada en un

campo y es necesario realizar interpolaciones, la herramienta adecuada para estimar la correlación entre los dos valores es el semivariograma, definido como:

$$\gamma(h) = \frac{1}{2} E \left\{ \left[Z(x_i) - Z(x_i + h) \right]^2 \right\}, \text{ el cual puede ser estimado mediante la}$$

expresión equivalente:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z(x_i) - Z(x_i + h) \right]^2, \text{ siendo } E \text{ el valor esperado y } N(h) \text{ el}$$

número de pares de observaciones $[Z(x_i), Z(x_i + h)]$ separados por una distancia h .

La gráfica de $\gamma(h)$ v.s. h es a lo que se le llama el semivariograma experimental y como se dejó asentado anteriormente, se espera que las observaciones cercanas entre sí, sean más parecidas que las separadas por una distancia mayor. Esto quiere decir que las diferencias entre $[Z(x_i) - Z(x_i + h)]$ crecerán al incrementarse h .

Efecto Nugget (C_0)

Analizando la ecuación $\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z(x_i) - Z(x_i + h) \right]^2$, para un proceso estacionario de segundo orden se esperaría que $\gamma(0) = 0$, sin embargo en la realidad a una distancia $h = 0$, le corresponde un cierto valor $\gamma(h)$ positivo. A este valor se le llama Efecto Nugget y generalmente se denota por C_0 .

El Efecto Nugget revela la discontinuidad que surge en el semivariograma cuando h se aproxima a cero, cerca del origen, a distancias más pequeñas que la más pequeña distancia de muestreo.

Ajuste del Semivariograma Teórico al Experimental

Existen varios programas computacionales en el mercado que pueden hacer éste trabajo como son el Geoestadística y el GS23, más sin embargo aún así es necesario desarrollar un trabajo subjetivo, ya que se sabe que no existe una metodología general para éste fin. Los parámetros del semivariograma se estiman en la forma siguiente:

(a) Pendiente al Origen.

Se calcula a partir de los tres o cuatro primeros puntos del semivariograma y esto produce el valor de la ordenada en el origen, que viene a ser el efecto Nugget.

(b) El Rango.

Este valor se estima visualmente y es aproximadamente igual a la varianza de las observaciones.

Interpolación para la Predicción

Si se hacen algunas observaciones a través de una muestra en n sitios con el fin de predecir el valor en otro sitio se considera lo siguiente:

$Z(x_i)$ = valor observado en el i -ésimo sitio de la muestra.

$c_{ij} = \text{cov}[Z(x_i), Z(x_j)]$ = covarianza entre el i -ésimo y el j -ésimo sitio.

Note que $c_{ii} = \text{var}[Z(x_i)]$

De las observaciones $Z(x_1), Z(x_2), \dots, Z(x_n)$ en los sitios t_1, t_2, \dots, t_n se puede predecir el valor de $Z(x_0)$ en el sitio t_0 , el objetivo es encontrar una función $\hat{Z}(x_0)$ de las n observaciones que sea insesgada para $Z(x_0)$ es decir $E[\hat{Z}(x_0)] = Z(x_0)$ lo cual debe minimizar a $E[Z(x_0) - \hat{Z}(x_0)]^2$ por lo que $\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$

Ahora el problema se reduce a encontrar los valores de $\lambda_1, \lambda_2, \dots, \lambda_n$ que hagan mínima la predicción del cuadrado medio del error. La solución se obtiene mediante el método de los multiplicadores de Lagrange y está dado en forma matricial por:

$f = G^{-1} h$ donde los vectores f y h son:

$$f = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ m \end{bmatrix}, \quad h = \begin{bmatrix} c_{10} \\ c_{20} \\ \vdots \\ c_{n0} \\ 1 \end{bmatrix}$$

y la matriz G es:

$$G = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} & 1 \\ c_{21} & c_{22} & \cdots & c_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}$$

La constante m , es obtenida junto con los coeficientes λ_i y es el multiplicador de Lagrange usado para calcular la predicción del cuadrado medio del error, a su vez el mejor predictor lineal $\hat{Z}(x_0)$ es llamado el predictor kriging y la predicción del cuadrado medio del error de $\hat{Z}(x_0)$ es también llamada la varianza kriging y es:

$$E[Z(x_0) - \hat{Z}(x_0)]^2 = c_{00} - \sum_{i=1}^n \lambda_i c_{i0} - m$$

Aplicación del Método del Individuo más Cercano

El presente método es uno de los llamados de distancia, es ampliamente usado para cuantificar plantas arbustivas y pastos; tiene su fundamento en el muestreo no probabilístico, y se parte del precepto de que la población está distribuida al azar.

Se sabe que existen tres tipos de dispersiones en las poblaciones ecológicas que son: Al azar, agregada y uniforme (Figura 2.3). Existe un trabajo hecho por

Coronado (1992), en donde se discute el tipo de dispersión que presentan la *Yucca carnerosana* y la Lechuguilla (*Agave lechuguilla*).

Por otra parte Salinas (1981), en un trabajo hecho con métodos de muestreo observa que el método del Individuo más Cercano es propenso a seguir tendencias subjetivas, hecho que lo atribuyó al bajo número de muestreos realizados.

A pesar de las circunstancias anteriores el presente método se usa frecuentemente por lo que se consideró interesante y de provecho incluirlo en el estudio.

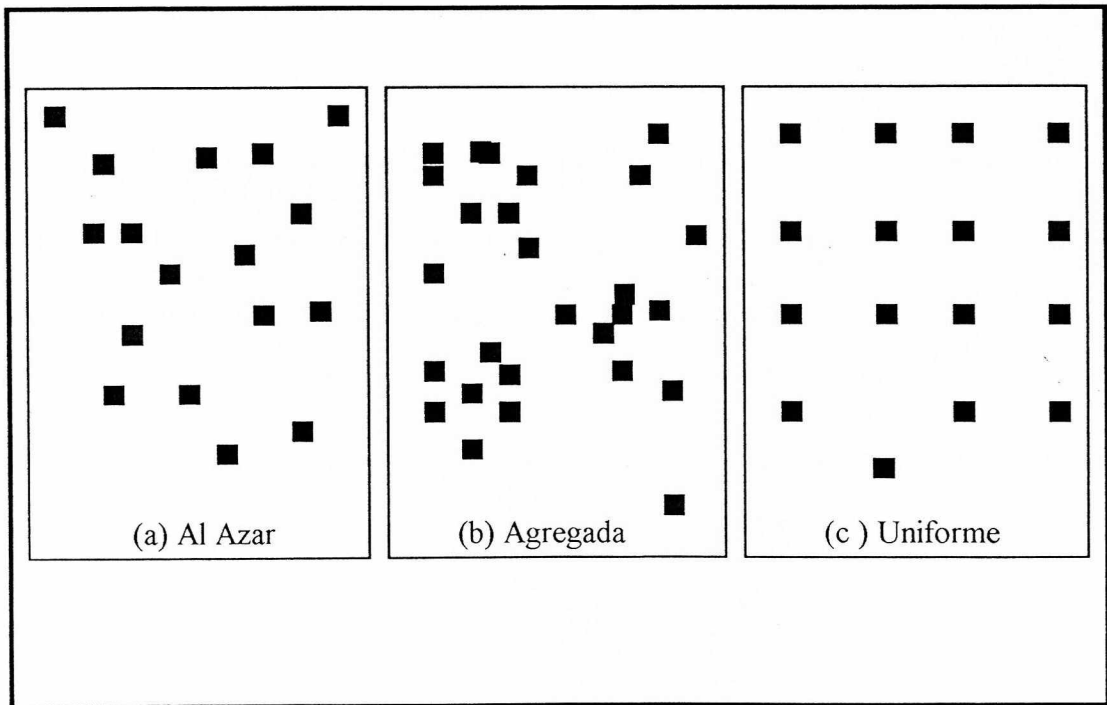


Figura 2.3. Dispersiones de las poblaciones ecológicas

Descripción del Método

Primero que todo se determina un tamaño de muestra y se consideran cuando menos ocho direcciones de la Rosa de los Vientos; Norte (N), Sur (S), Este (E), Oeste (W), Noreste (NE), Noroeste (NW), Sureste (SE) y Suroeste (SW), también cinco distancias de 3, 5, 8, 12 y 15 metros, con la finalidad de formar parejas al azar de una dirección con una distancia.

Habiendo ya determinado el tamaño de muestra n se forman así mismo n parejas de una dirección con una distancia, se toma cualquier punto del terreno y se empieza a recorrer de acuerdo a lo que indica la primer pareja de dirección y de distancia, al llegar al lugar determinado se registra la distancia existente entre el punto y el individuo más cercano lo que nos permite conocer el primer valor l_1 , después a partir del punto anteriormente determinado se sigue la indicación de la segunda pareja de dirección y de distancia para determinar el segundo sitio, tomándose el segundo valor l_2 y así sucesivamente hasta determinar l_n , según lo muestra la Figura 2.4.

Las acciones anteriores sirven para determinar la distancia media (lm) que se calcula con la fórmula:

$$lm = \frac{l_1 + l_2 + l_3 + \dots + l_n}{n}, \text{ ahora el valor } lm \text{ se multiplica por el factor de}$$

corrección 2.0, con lo cual se obtiene la distancia media corregida (lmc). El área media

ocupada por un individuo (AM) se encuentra con la ecuación $AM = (lmc)^2$ y por último la densidad D es determinada con la ecuación $D = \frac{10\,000}{AM}$, que es lo que muestra el número de individuos en una hectárea.

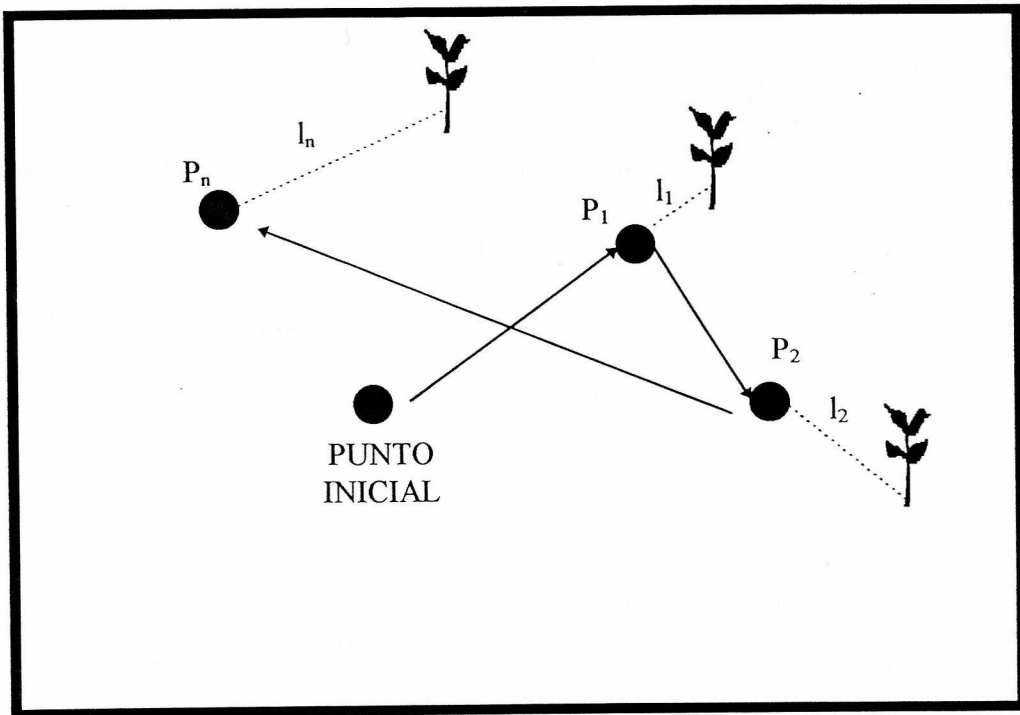


Figura 2.4. Diagrama para el método del individuo más cercano

MATERIALES Y MÉTODOS

Ubicación del Área de Estudio

El presente trabajo fue desarrollado en un área ubicada a quince kilómetros de la Universidad Autónoma Agraria “Antonio Narro” (UAAAN), rumbo al sur por la carretera a Zacatecas (Figuras 3.1 y 3.2). De acuerdo a información obtenida en una carta de la Comisión de Estudios del Territorio Nacional (CETENAL), este lugar se encuentra localizado en los 25° 15' de latitud norte y 101° 05' de longitud oeste a 1940 metros sobre el nivel del mar (msnm).

Recursos Empleados

Para las acciones inherentes al estudio fue necesario el uso de un teodolito, una brújula, balizas, estacas y cinta métrica. También se precisó del empleo de algunas herramientas computacionales como el programa Statgraphics Plus versión 6.0, Matlab para Windows y sendos programas implementados en los lenguajes Matlab y Pascal, los anteriores elementos de software se trabajaron mediante una computadora personal con procesador 486 DX y coprocesador matemático integrado.

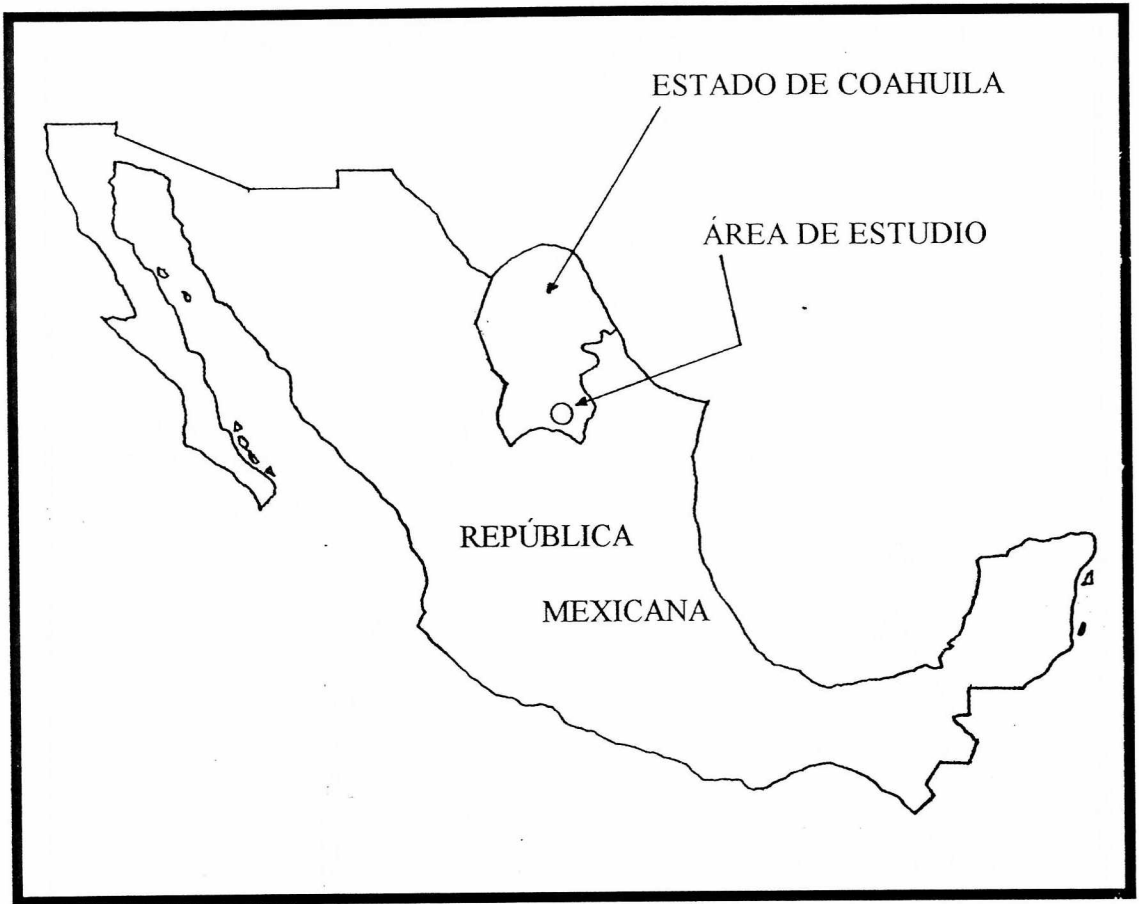


Figura 3.1. Localización del área de estudio en la República Mexicana

Uso del Método Kriging

Levantamiento de los Datos de Campo

Se diseñó el trazo de una línea de 500 metros de longitud y dos metros de anchura, con la finalidad de hacer un conteo de los individuos existentes de la especie

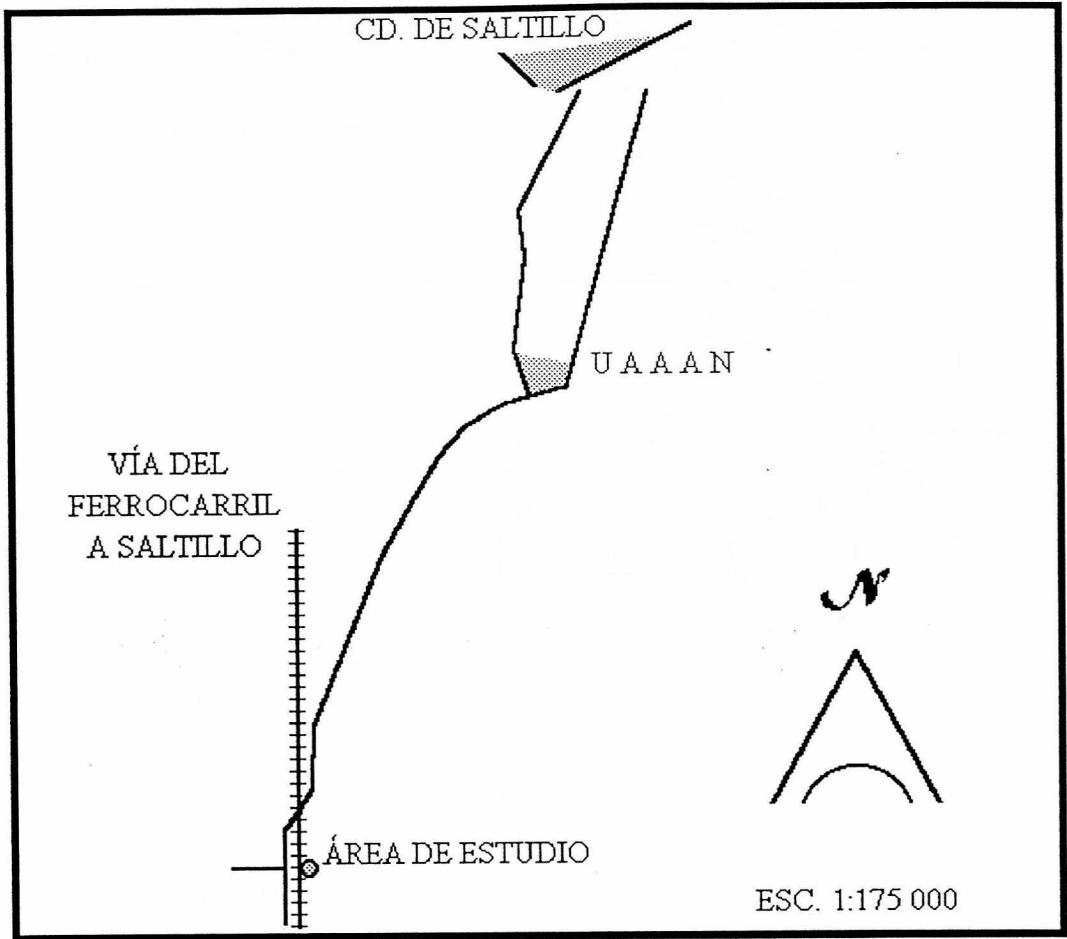


Figura 3.2. Localización específica del área de estudio

de interés. La mencionada línea fue dividida en rectángulos de 10 metros de longitud, lo que produce áreas de 20 metros cuadrados, pudiendo así determinar la densidad, es decir el número de individuos por unidad de área.

Las lecturas se hicieron como ya se indicó anteriormente, a intervalos de 10 metros (Figura 3.3) y las observaciones concentradas en el Cuadro 3.1, así mismo en el Cuadro 3.2 se especifica un resumen de la variable en cuestión.

El primer registro del conteo de individuos de gobernadora fue de tres para una superficie de 2×10 m o sea 20 metros cuadrados, con lo que resulta una densidad de $3 \text{ ind} / 20 \text{ m}^2$ igual a $0.15 \text{ ind} / \text{m}^2$.

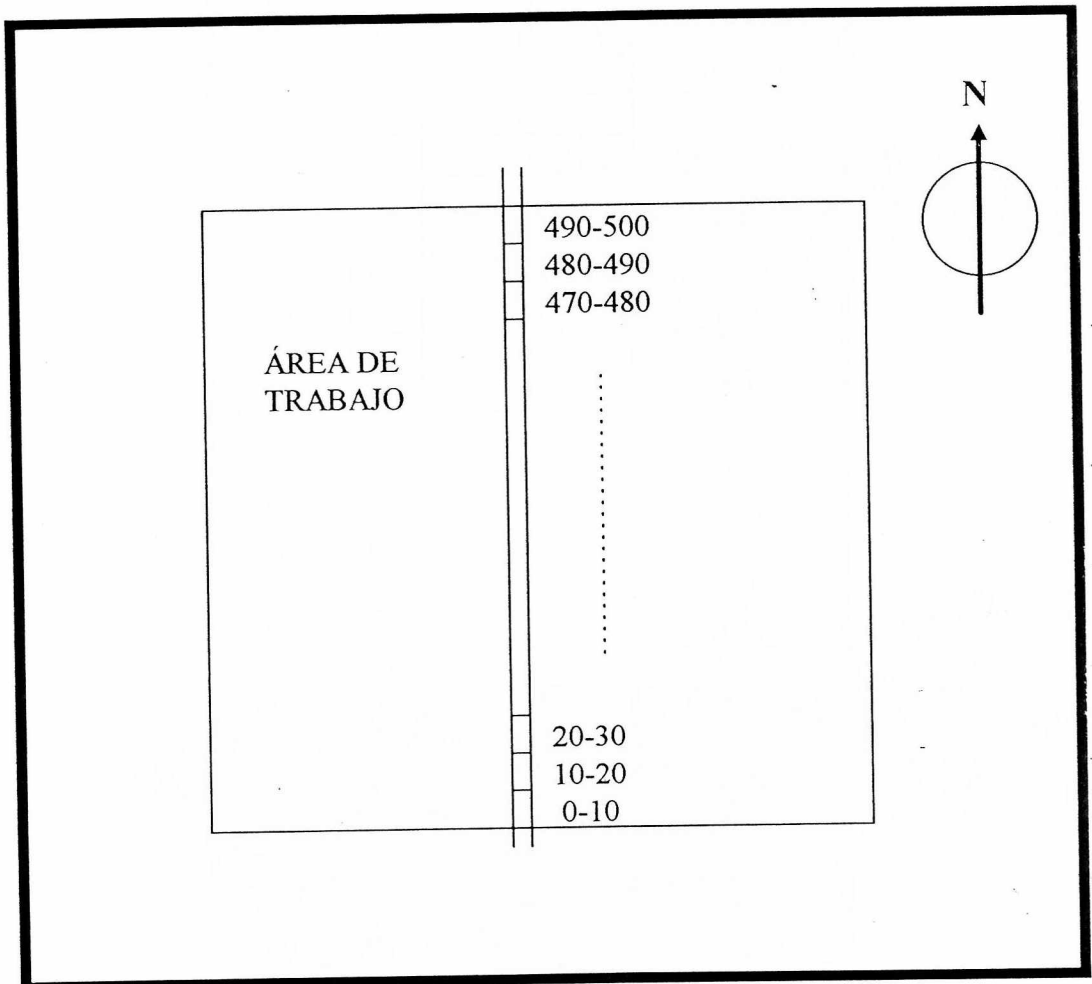


Figura 3.3. Croquis del levantamiento de los datos para el método kriging.

El segundo registro es: $6 \text{ ind} / 20 \text{ m}^2$ igual a $0.3 \text{ ind} / \text{m}^2$ y así sucesivamente; los valores obtenidos se muestran en el Cuadro 3.1.

Cuadro 3.1. Observaciones obtenidas para el método kriging

Intervalo m	No. de individuos	Densidad ind/m ²
0 10	3	0.15
10 20	6	0.3
20 30	6	0.3
30 40	8	0.4
40 50	6	0.3
50 60	15	0.75
60 70	14	0.7
70 80	19	0.95
80 90	7	0.35
90 100	11	0.55
100 110	14	0.7
110 120	5	0.25
120 130	2	0.1
130 140	25	1.25
140 150	18	0.9
150 160	7	0.35
160 170	15	0.75
170 180	6	0.3
180 190	7	0.35
190 200	3	0.15
200 210	2	0.1
210 220	3	0.15
220 230	5	0.25
230 240	6	0.3
240 250	8	0.4
250 260	12	0.6
260 270	10	0.5
270 280	4	0.2
280 290	11	0.55
290 300	10	0.5
300 310	13	0.65
310 320	7	0.35
320 330	10	0.5
330 340	16	0.8

Cuadro 3.1..... (continuación)

Intervalo m	No. de individuos	Densidad ind/m ²
340 350	16	0.8
350 360	12	0.6
360 370	10	0.5
370 380	10	0.5
380 390	18	0.9
390 400	23	1.15
400 410	7	0.35
410 420	11	0.55
420 430	18	0.9
430 440	18	0.9
440 450	11	0.55
450 460	10	0.5
460 470	9	0.45
470 480	21	1.05
480 490	20	1.0
490 500	7	0.35

Cuadro 3.2. Resumen para la variable en el método kriging

Tamaño de la Muestra	50
Promedio	10.7
Varianza	32.78
Desviación Estándar	5.72
Coficiente de Variación	53.51 %

Cálculo del Semivariograma Experimental

Para realizar esta acción se utiliza la fórmula:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

Donde $\gamma(h)$ = Expresión del semivariograma, valor para ser colocado en el eje de las ordenadas.

$N(h)$ = Número de pares de valores para una distancia h .

$Z(x_i)$ = Valor de la variable en la coordenada x_i .

h = Distancia; valor para ser colocado en el eje de las abscisas.

De tal manera que al realizar los primeros cálculos para cuando $h = 10$ y consecuentemente $N(h) = 49$, se tiene que:

$$\begin{aligned}\gamma(h) &= \frac{1}{2(49)}[(3-6)^2 + (6-6)^2 + (6-8)^2 + \dots + (21-20)^2 + (20-7)^2] \\ &= 22.91\end{aligned}$$

Para $h = 20$, implicando esto $N(h) = 48$

$$\begin{aligned}\gamma(h) &= \frac{1}{2(48)}[(3-6)^2 + (6-8)^2 + (6-6)^2 + \dots + (9-20)^2 + (21-7)^2] \\ &= 32.89\end{aligned}$$

Para $h = 30$, implicando esto $N(h) = 47$

$$\begin{aligned}\gamma(h) &= \frac{1}{2(47)}[(3-8)^2 + (6-6)^2 + (6-15)^2 + \dots + (10-20)^2 + (9-7)^2] \\ &= 25.28\end{aligned}$$

Y así sucesivamente hasta $h = 490$ lo que implica un valor para $N(h) = 1$

$$\begin{aligned}\gamma(h) &= \frac{1}{2(1)}[(3-7)^2] \\ &= 8.00\end{aligned}$$

Los cálculos anteriores fueron realizados mediante un pequeño programa hecho en Matlab, situación que ya se mencionó con anterioridad y de no haber utilizado este recurso, el trabajo aunque simple es tremendamente laborioso. Los resultados en su totalidad se muestran en el Cuadro 3.3 y con ellos mismos se procede a la construcción del semivariograma experimental. En primer lugar se grafican y en la Figura 3.4, se muestra la geometría de este arreglo, después se analizan los cuatro modelos teóricos más comunes (Lineal, Esférico, Exponencial y Gaussiano) y el que más se ajusta es el esférico. De hecho el resultado no es extraño porque según se ha apreciado en otros trabajos hechos con plantas, éste es el que proporciona mejores resultados.

La fórmula para el modelo esférico es:

$$\begin{aligned}\hat{\gamma}(h) &= C \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] & 0 \leq h \leq a \\ &= C & h > a\end{aligned}$$

Concordando con el método se sabe que $C \approx$ la varianza de las observaciones, entonces se asume que $C = 32.78$, de antemano es conocido que el valor C se descompone de la forma:

$$C = C_0 + C_1, \text{ en donde el valor } C_0 \text{ es el llamado Efecto Nugget.}$$

Para encontrar la ordenada en el origen o efecto nugget, se hace una regresión lineal simple con los primeros cuatro valores de h v.s. los primeros cuatro valores de $\gamma(h)$, lo que proporciona la ordenada en el origen $C_0 = 7.39$ y con la pendiente igual a 0.85, se determina la trayectoria que sigue la recta resultante $y = 7.39 + 0.85 x$.

Cuadro 3.3. Valores para la construcción del semivariograma experimental.

h	$N(h)$	$\gamma(h)$	h	$N(h)$	$\gamma(h)$
10	49	22.91	260	24	32.43
20	48	32.89	270	23	43.15
30	47	25.28	280	22	36.18
40	46	24.21	290	21	26.71
50	45	29.80	300	20	38.65
60	44	32.65	310	19	38.84
70	43	36.09	320	18	31.44
80	42	26.46	330	17	36.41
90	41	23.03	340	16	25.59
100	40	33.96	350	15	41.20
110	39	35.61	360	14	50.50
120	38	36.43	370	13	34.90
130	37	38.33	380	12	37.95
140	36	34.80	390	11	51.27
150	35	32.97	400	10	27.45
160	34	35.92	410	9	25.72
170	33	30.90	420	8	38.93
180	32	31.90	430	7	40.42
190	31	29.83	440	6	43.16
200	30	30.36	450	5	42.80
210	29	34.17	460	4	57.25
220	28	35.07	470	3	86.83
230	27	30.79	480	2	72.50
240	26	36.44	490	1	8.00
250	25	28.62			

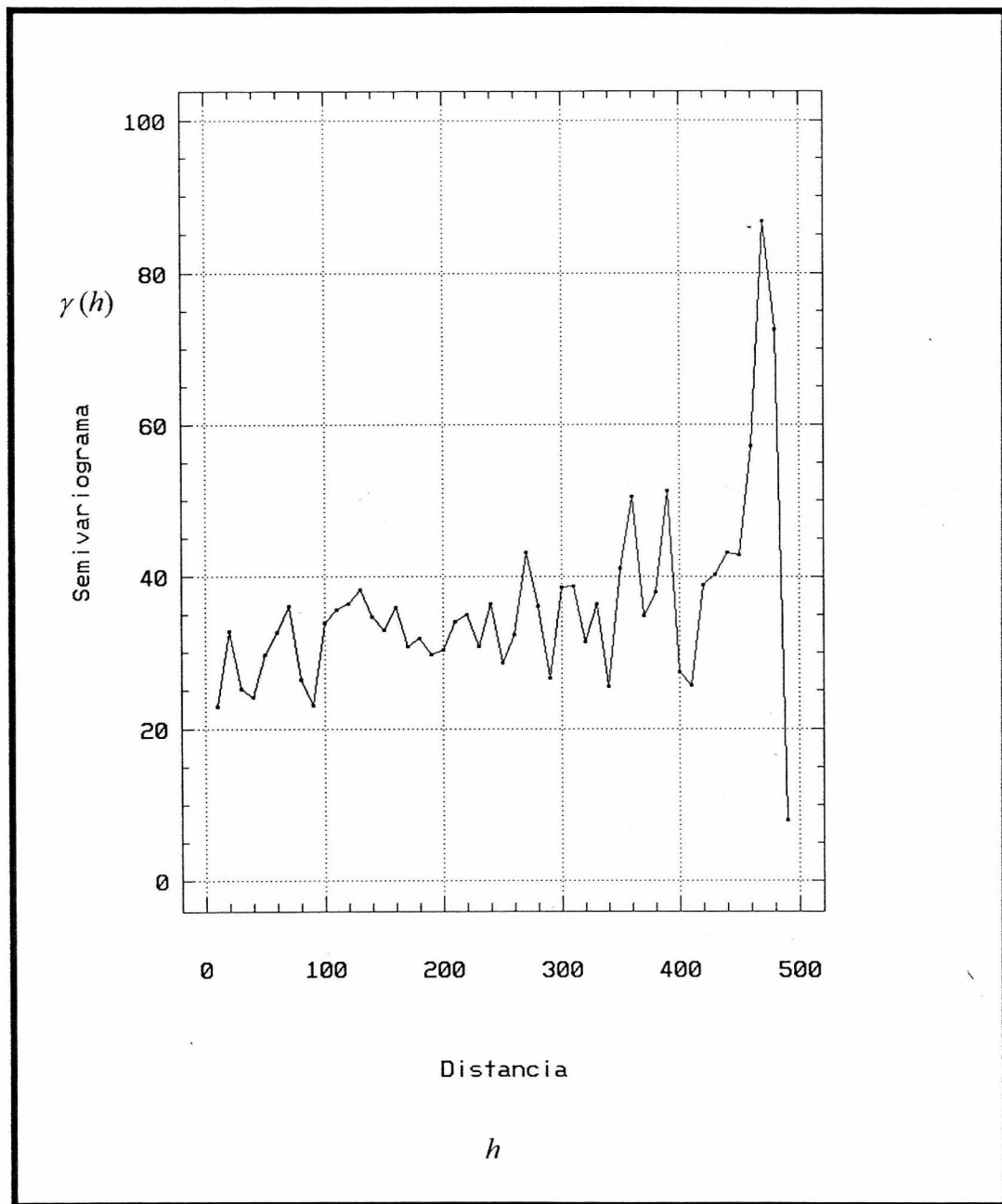


Figura 3.4. Gráfica de la distancia v. s. semivariograma

Existe un elemento que ayuda a determinar el valor de a , y es $a' = \frac{2}{3}a$. De acuerdo al procedimiento metodológico se sabe que a se encuentra en la intersección de la recta resultante de la regresión y la recta paralela al eje de las abscisas que describe el valor de la varianza de las observaciones, en decir $C = 32.78$.

Analítica o gráficamente se puede constatar que $a' = 30$ por lo que $a = 45$. Las consideraciones anteriores quedan plasmadas en la Figura 3.5.

Ajuste del Semivariograma Experimental al Teórico

Como se indicó con anterioridad el modelo que responde a las necesidades del caso, es el esférico cuya expresión es:

$$\hat{\gamma}(h) = C \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] \quad 0 \leq h \leq a$$

$$= C \quad h > a$$

Se procede a realizar las operaciones sabiendo que $C = 32.78$, $a = 45$ y el valor de h variará de 0 a 45. Los resultados se presentan en el Cuadro 3.4 y a su vez la Figura 3.6, permite apreciar plenamente la trayectoria del semivariograma ajustado. Es importante destacar que si bien es posible graficar todos los valores de h , esto no es necesario porque la función describe una recta paralela al eje de las abscisas cuando $h > a$ según se muestra en la expresión original del modelo esférico.

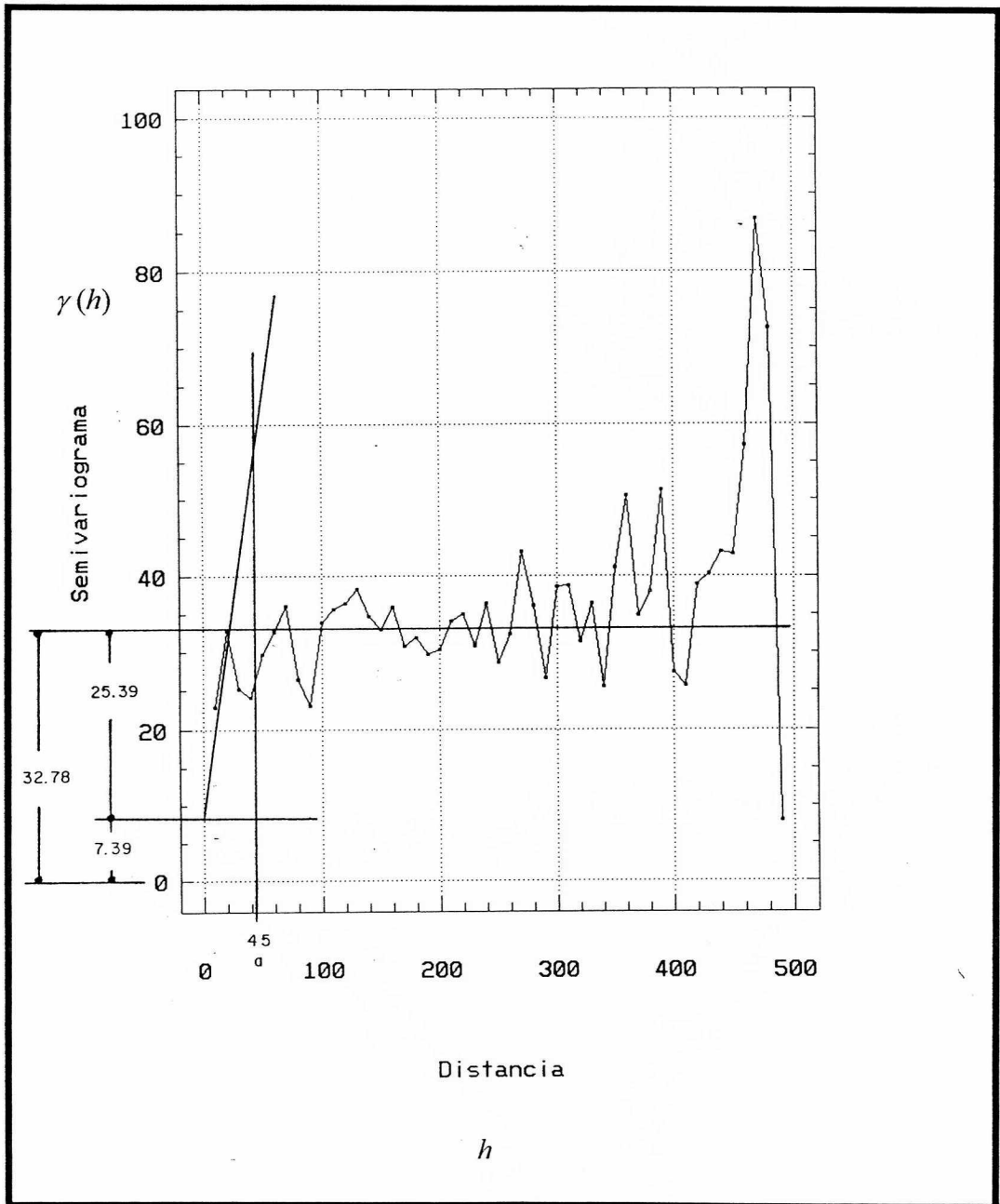


Figura 3.5. Ajuste del semivariograma

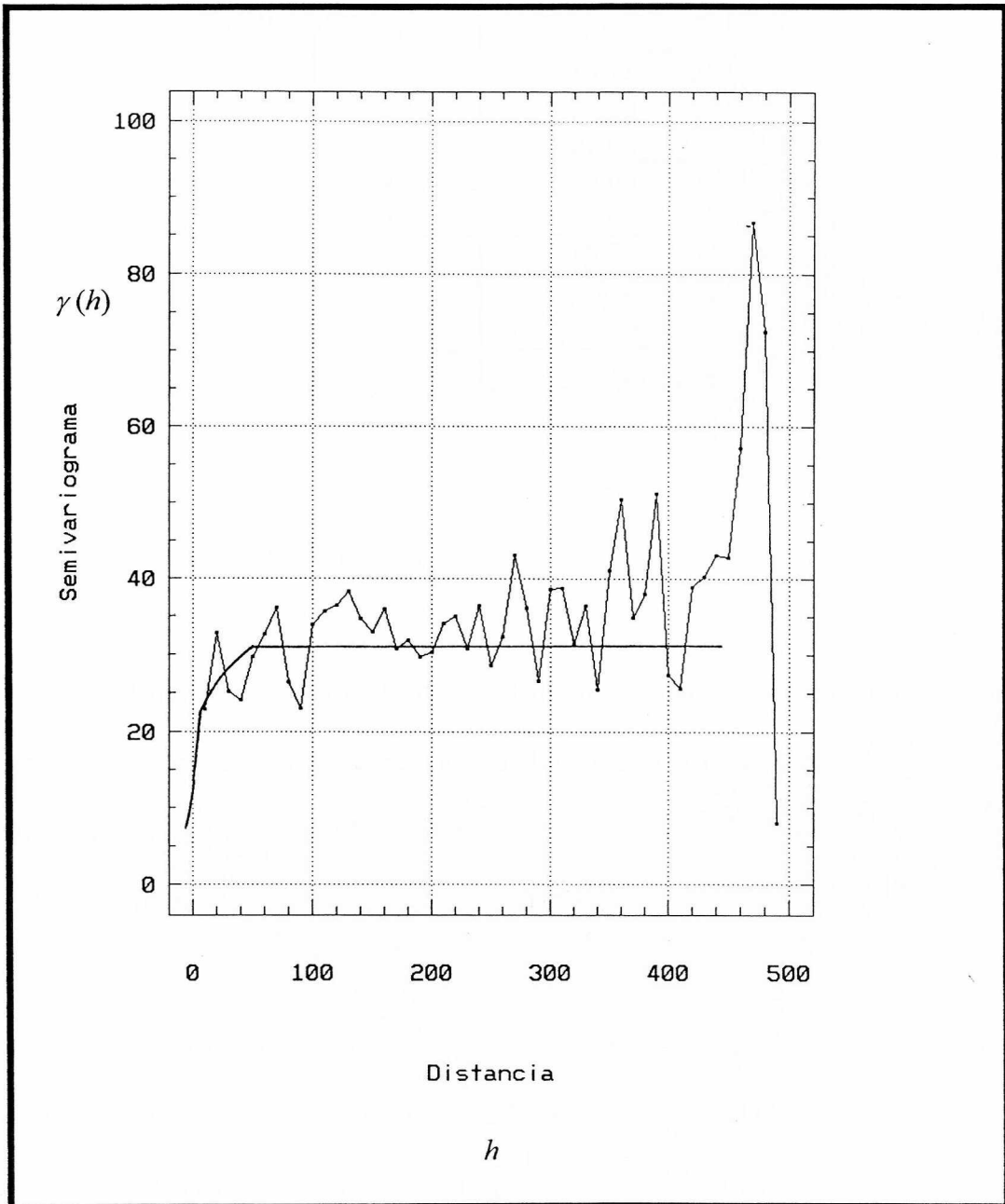


Figura 3.6. Trayectoria del semivariograma ajustado

Cuadro 3.4. Valores ajustados del semivariograma

h	$\gamma(h)$
0	7.39
10	10.72
20	20.36
30	27.84
40	32.08
45	32.64
50	32.78
⋮	⋮
⋮	⋮
500	32.78

Interpolación de Valores

Con fundamento en lo desarrollado anteriormente se forma un conjunto de cinco elementos con los primeros valores del Cuadro 3.1, es decir considerando los valores hasta una distancia de 50 metros de los cuales se ignora el primero para predecirlo mediante el método y evaluar su efectividad.

Después se repite la operación ignorando el segundo y así sucesivamente hasta llegar al quinto de tal manera que se tendrán en realidad cinco conjuntos de cinco elementos de los cuales hipotéticamente se desconoce un valor en cada caso.

El primer conjunto lo forman $Z(x_1) = 3$, $Z(x_2) = 6$, $Z(x_3) = 6$, $Z(x_4) = 8$ y $Z(x_5) = 6$, se asume que $Z(x_1)$ es un valor desconocido. Para facilitar el manejo en

notación y a la vez evidenciar la analogía con las explicaciones del método en las páginas anteriores, se optará por llamar a la $Z(x_1)$ como $Z(x_0)$, a $Z(x_2)$ como $Z(x_1)$, a $Z(x_3)$ como $Z(x_2)$, a $Z(x_4)$ como $Z(x_3)$ y a $Z(x_5)$ como $Z(x_4)$; establecidas estas convenciones se calculan las covarianzas c entre los puntos utilizados que son:

$$c_{00} = c_{11} = c_{22} = c_{33} = c_{44} = 32.78$$

$$\begin{aligned} c_{12} = c_{21} &= 32.78 - \hat{\gamma}(10) \\ &= 32.78 - [1.09(10) - 1.8 \times 10^{-4}(10)^3] \\ &= 22.06 \end{aligned}$$

$$\begin{aligned} c_{13} = c_{31} &= 32.78 - \hat{\gamma}(20) \\ &= 32.78 - [1.09(20) - 1.8 \times 10^{-4}(20)^3] \\ &= 12.42 \end{aligned}$$

$$\begin{aligned} c_{14} = c_{41} &= 32.78 - \hat{\gamma}(30) \\ &= 32.78 - [1.09(30) - 1.8 \times 10^{-4}(30)^3] \\ &= 4.94 \end{aligned}$$

$$\begin{aligned} c_{23} = c_{32} &= 32.78 - \hat{\gamma}(10) \\ &= 32.78 - [1.09(10) - 1.8 \times 10^{-4}(10)^3] \\ &= 22.06 \end{aligned}$$

$$\begin{aligned} c_{34} = c_{43} &= 32.78 - \hat{\gamma}(10) \\ &= 32.78 - [1.09(10) - 1.8 \times 10^{-4}(10)^3] \\ &= 22.06 \end{aligned}$$

$$c_{24} = c_{42} = 32.78 - \hat{\gamma}(20)$$

$$\begin{aligned}
 &= 32.78 - [1.09(20) - 1.8 \times 10^{-4}(20)^3] \\
 &= 12.42
 \end{aligned}$$

$$\begin{aligned}
 c_{10} = c_{01} &= 32.78 - \hat{\gamma}(10) \\
 &= 32.78 - [1.09(10) - 1.8 \times 10^{-4}(10)^3] \\
 &= 22.06
 \end{aligned}$$

$$\begin{aligned}
 c_{20} = c_{02} &= 32.78 - \hat{\gamma}(20) \\
 &= 32.78 - [1.09(20) - 1.8 \times 10^{-4}(20)^3] \\
 &= 12.42
 \end{aligned}$$

$$\begin{aligned}
 c_{30} = c_{03} &= 32.78 - \hat{\gamma}(30) \\
 &= 32.78 - [1.09(30) - 1.8 \times 10^{-4}(30)^3] \\
 &= 4.94
 \end{aligned}$$

$$\begin{aligned}
 c_{40} = c_{04} &= 32.78 - \hat{\gamma}(40) \\
 &= 32.78 - [1.09(40) - 1.8 \times 10^{-4}(40)^3] \\
 &= 0.70
 \end{aligned}$$

Por lo que la ecuación matricial $Gf = h$ es:

$$\begin{bmatrix} 32.78 & 22.06 & 12.42 & 4.94 & 1 \\ 22.06 & 32.78 & 22.06 & 12.42 & 1 \\ 12.42 & 22.06 & 32.78 & 22.06 & 1 \\ 4.94 & 12.42 & 22.06 & 32.78 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ m \end{bmatrix} = \begin{bmatrix} 22.06 \\ 12.42 \\ 4.94 \\ 0.70 \\ 1 \end{bmatrix}$$

Y siguiendo el procedimiento de la solución se tiene que:

$f = G^{-1}h$, por lo que:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ m \end{bmatrix} = \begin{bmatrix} 32.78 & 22.06 & 12.42 & 4.94 & 1 \\ 22.06 & 32.78 & 22.06 & 12.42 & 1 \\ 12.42 & 22.06 & 32.78 & 22.06 & 1 \\ 4.94 & 12.42 & 22.06 & 32.78 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 22.06 \\ 12.42 \\ 4.94 \\ 0.70 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ m \end{bmatrix} = \begin{bmatrix} 0.0473 & -0.446 & 0.0023 & -0.0049 & 0.4314 \\ -0.0446 & 0.0889 & -0.0465 & 0.0023 & 0.0686 \\ 0.0023 & -0.0465 & 0.0889 & -0.0446 & 0.0686 \\ -0.0049 & 0.0023 & -0.0446 & 0.0473 & 0.4314 \\ 0.4314 & 0.0686 & 0.0686 & 0.4314 & -18.6376 \end{bmatrix} \begin{bmatrix} 22.06 \\ 12.42 \\ 4.94 \\ 0.70 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ m \end{bmatrix} = \begin{bmatrix} 0.9274 \\ -0.0401 \\ -0.0514 \\ 0.1641 \\ -7.6283 \end{bmatrix}$$

Entonces $\lambda_1 = 0.9274$, $\lambda_2 = -0.0401$, $\lambda_3 = -0.0514$, $\lambda_4 = 0.1641$ y $m = -7.6283$

Usando la ecuación de predicción $\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$, se tiene que:

$$\hat{Z}(x_0) = (0.9274)(6) - (0.0401)(6) - (0.0514)(8) + (0.1641)(6) = 5.89$$

Después se obtiene la varianza kriging mediante la expresión:

$$E[Z(x_0) - \hat{Z}(x_0)]^2 = c_{00} - \sum_{i=1}^n \lambda_i c_{i0} - m, \text{ por lo que:}$$

$$\begin{aligned}
&= 32.78 - [(0.9274)(22.06) + (-0.0401)(12.42) + \\
&\quad + (-0.0514)(4.94) + (0.1641)(0.70)] - (-7.6283) \\
&= 20.58
\end{aligned}$$

Consecuentemente la desviación estándar es 4.53

El segundo conjunto a considerar estará estructurado como se indica:

$$Z(x_1) = Z(x_1), Z(x_2) = Z(x_0), Z(x_3) = Z(x_2), Z(x_4) = Z(x_3) \text{ y } Z(x_5) = Z(x_4); \text{ es}$$

decir se desconoce el valor de $Z(x_2)$.

$$\text{Entonces } c_{00} = c_{11} = c_{22} = c_{33} = c_{44} = 32.78$$

$$c_{12} = c_{21} = 32.78 - \hat{\gamma}(20)$$

$$= 12.42$$

$$c_{13} = c_{31} = 32.78 - \hat{\gamma}(30)$$

$$= 4.94$$

$$c_{14} = c_{41} = 32.78 - \hat{\gamma}(40)$$

$$= 0.70$$

$$c_{23} = c_{32} = 32.78 - \hat{\gamma}(10)$$

$$= 22.06$$

$$c_{34} = c_{43} = 32.78 - \hat{\gamma}(10)$$

$$= 22.06$$

$$c_{24} = c_{42} = 32.78 - \hat{\gamma}(20)$$

$$= 12.42$$

$$c_{10} = c_{01} = 32.78 - \hat{\gamma}(10)$$

$$= 22.06$$

$$c_{20} = c_{02} = 32.78 - \hat{\gamma}(10)$$

$$= 22.06$$

$$c_{30} = c_{03} = 32.78 - \hat{\gamma}(20)$$

$$= 12.42$$

$$c_{40} = c_{04} = 32.78 - \hat{\gamma}(30)$$

$$= 4.94$$

Por lo que la ecuación matricial $Gf = h$ es:

$$\begin{bmatrix} 32.78 & 12.42 & 4.94 & 0.70 & 1 \\ 12.42 & 32.78 & 22.06 & 12.42 & 1 \\ 4.94 & 22.06 & 32.78 & 22.06 & 1 \\ 0.70 & 12.42 & 22.06 & 32.78 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ m \end{bmatrix} = \begin{bmatrix} 22.06 \\ 22.06 \\ 12.42 \\ 4.94 \\ 1 \end{bmatrix}$$

Y siguiendo el procedimiento de la solución se tiene que:

$f = G^{-1}h$, por lo que:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ m \end{bmatrix} = \begin{bmatrix} 32.78 & 12.42 & 4.94 & 0.70 & 1 \\ 12.42 & 32.78 & 22.06 & 12.42 & 1 \\ 4.94 & 22.06 & 32.78 & 22.06 & 1 \\ 0.70 & 12.42 & 22.06 & 32.78 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 22.06 \\ 22.06 \\ 12.42 \\ 4.94 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ m \end{bmatrix} = \begin{bmatrix} 0.0258 & -0.0215 & 0.0025 & -0.0067 & 0.4849 \\ -0.0215 & 0.0648 & -0.0465 & 0.0033 & 0.1292 \\ 0.0025 & -0.0465 & 0.0890 & -0.0450 & 0.0878 \\ -0.0067 & 0.0023 & -0.0450 & 0.0485 & 0.3681 \\ 0.4149 & 0.1292 & 0.0878 & 0.3681 & -15.8973 \end{bmatrix} \begin{bmatrix} 22.06 \\ 22.06 \\ 12.42 \\ 4.94 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ m \end{bmatrix} = \begin{bmatrix} 0.5059 \\ 0.5216 \\ 0.0006 \\ -0.0281 \\ -0.9849 \end{bmatrix}$$

Entonces $\lambda_1 = 0.5059$, $\lambda_2 = 0.5216$, $\lambda_3 = 0.0006$, $\lambda_4 = -0.0281$ y $m = -0.9849$

Usando la ecuación de predicción $\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$, se tiene que:

$$\hat{Z}(x_0) = (0.5059)(3) + (0.5216)(6) + (0.0006)(8) - (0.0281)(6) = 4.48$$

Después se obtiene la varianza kriging mediante la expresión:

$$E[Z(x_0) - \hat{Z}(x_0)]^2 = c_{00} - \sum_{i=1}^n \lambda_i c_{i0} - m, \text{ por lo que:}$$

$$\begin{aligned} &= 32.78 - [(0.5059)(22.06) + (0.5216)(22.06) + \\ &+ (0.0006)(12.42) + (-0.0281)(4.94)] - (-0.9849) \\ &= 11.22 \end{aligned}$$

Consecuentemente la desviación estándar es 3.34

Continuando con el mismo manejo para los siguientes conjuntos, se conforma el Cuadro 3.5.

Cuadro 3.5. Predicción de los $\hat{Z}(x_0)$

$Z(x_1)$	$Z(x_2)$	$Z(x_3)$	$Z(x_4)$	$\hat{Z}(x_0)$	Varianza Kriging	Desviación Estándar
6	6	8	6	5.89	20.58	4.53
3	6	8	6	4.48	11.22	3.34
3	6	8	6	7.10	11.23	3.35
3	6	6	6	6.08	11.24	3.35
3	6	6	8	7.76	20.58	4.53

Uso del Método del Individuo más Cercano

Aspectos Generales

El trabajo se desarrolló siguiendo la metodología propia del caso, pero fue necesario determinar un tamaño de muestra n , de tal manera que respondiera a las exigencias de la expresión:

$$n = \frac{Ns_n^2 Z^2 \alpha/2}{d^2 N + s_n^2 Z^2 \alpha/2}$$

Se definió un cuadrado de 32 m de lado con el fin de hacer una cuadrícula que comprendiera cuadros de 2x2 m, dando un total de 256 cuadros, mismos que pasaron a ser las unidades de muestreo. Evidentemente existe en la expresión de n el valor s_n^2 que es fundamental para este cálculo, mismo que es desconocido, pero para subsanar esta deficiencia se obtuvo una muestra preliminar n' que se calcula con la expresión:

$$n' = N(10\%), \text{ que para el caso es } n' = 256(0.10) \approx 26.$$

Acto seguido se procedió a manipular la tabla de números aleatorios para determinar la ubicación de las unidades a considerar para contabilizar el número de individuos existentes en cada unidad, acción que permite calcular la densidad. La información anterior se encuentra consignada en el Cuadro 3.6, de igual manera en el Cuadro 3.7 se presenta un resumen de la variable considerada.

Para el valor de la precisión d se considera $d = |\bar{x} - \mu| = 1$, también existe el detalle de que $n' = 26$ por lo que se usa la tabla de valores para la t-student, siendo la expresión $t_{\alpha/2, (n-1)gl} \therefore t_{0.025, 25gl} = 2.060$; con lo que la expresión de n queda como se indica a continuación:

$$n = \frac{Ns_n^2 t_{\alpha/2}^2}{d^2 N + s_n^2 t_{\alpha/2}^2}, \text{ y sustituyendo valores:}$$

$$n = \frac{(256)(2.28)(2.060)^2}{(1)^2(256) + (2.28)(2.060)^2}, \text{ por lo que } n = 9.5 \approx 10$$

Cuadro 3.6. Observaciones obtenidas para el cálculo de la varianza

No. de Cuadro	Individuos	Densidad ind/m ²
6	2	0.50
16	2	0.50
33	5	1.25
37	0	0.00
41	0	0.00
42	2	0.50
51	4	1.00
59	3	0.75
85	3	0.75
97	2	0.50
101	3	0.75
102	2	0.50
117	1	0.25
120	3	0.75
129	2	0.50
131	4	1.00
135	2	0.50
160	1	0.25
168	0	0.00
173	4	1.00
174	1	0.25
210	5	1.25
211	4	1.00
219	1	0.25
248	0	0.00
255	3	0.75

Sabiendo que el tamaño de muestra n fue igual a 10, se procedió a la aplicación del método.

Es conveniente aclarar que en la cuadrícula mencionada anteriormente fue contabilizada la totalidad de la población para así obtener los parámetros verdaderos que servirán para realizar la evaluación del método.

Cuadro 3.7. Resumen para la variable en el método del individuo más cercano

Tamaño de Muestra	26
Promedio	2.26
Varianza	2.28
Desviación Estándar	1.50
Coficiente de Variación	66.37 %

Levantamiento de los Datos de Campo

Se consideraron ocho direcciones N, S, E,W, NE, NW, SE Y SW, así como cinco distancias diferentes expresadas en metros que son 3, 5, 8, 12 y 15. Diez direcciones fueron tomadas al azar, una a una con su respectiva distancia también al azar, lo cual fue indicando los lugares desde donde se hicieron las mediciones de las distancias

l_1, l_2, \dots, l_{10} , es decir la distancia del punto determinado al individuo más cercano. El Cuadro 3.8 contiene los datos obtenidos en esta acción.

Estimación de la Población

La distancia media lm se calcula en metros mediante la fórmula:

$$lm = \frac{l_1 + l_2 + l_3 + \dots + l_{10}}{n} = \frac{0.98 + 0.28 + 0.58 + \dots + 0.75}{10} = 0.61 \text{ m.}$$

La distancia media lm resultante se multiplica por un factor de corrección de 2.0, con el cual se obtiene la distancia media corregida lmc , por lo que:

$$lmc = (0.61) (2.0) = 1.22 \text{ m.}$$

El área media (AM) ocupada por un individuo se obtiene de la ecuación:

$$AM = (lmc)^2 = (1.22 \text{ m})^2 = 1.49 \text{ m}^2 .$$

La densidad por unidad de área sale de la ecuación:

$$D = \frac{10\,000 \text{ m}^2}{1.49 \text{ m}^2} = 6711 \text{ individuos en una hectárea.}$$

Pero para el estudio, únicamente interesa la estimación en una superficie de 1024 metros cuadrados, por lo que:

$$D = \frac{1\,024 \text{ m}^2}{1.49 \text{ m}^2} = 687 \text{ individuos.}$$

Cuadro 3.8. Datos para el individuo más cercano

Direcciones	Distancias m	Distancia al Individuo más Cercano I_i m
NW	8	0.98
S	15	0.28
W	15	0.58
N	8	0.70
S	12	0.32
NE	8	0.46
NE	15	0.86
NW	12	0.70
SW	5	0.52
SW	5	0.75

RESULTADOS Y DISCUSIÓN

Evaluación del Método Kriging

A través de los errores reducidos que son obtenidos mediante la expresión

$$M(x_i) = \frac{\hat{Z}(x_0) - Z(x_0)}{VK}, \quad (\text{Cuadro 4.1 y 4.2})$$

es posible evaluar el método por lo que asumiendo distribución normal de los mismos se llega al planteamiento de la hipótesis nula H_0 que asevera que la media de los errores reducidos $Mr(x_i)$ es igual a cero y su contraparte la hipótesis alternante H_a que asevera que $Mr(x_i)$ es diferente de cero, es decir:

$$H_0: Mr(x_i) = 0$$

$$H_a: Mr(x_i) \neq 0$$

Por ser una muestra pequeña se usa la distribución t-student:

$$t_c = \frac{\sqrt{n}[\bar{x} - Mr(x_i)]}{s}$$
$$t_c = \frac{\sqrt{5}[0.065 - 0]}{0.542} = 0.268$$

La regla de decisión es rechazar H_0 si $|t_c| \geq t_{\alpha/2, (n-1)g.l.}$. El nivel de significancia es $\alpha = 0.05$; por lo que:

$$t_{0.025, (4)g.l.} = 2.77$$

Cuadro 4.1. Errores reducidos

$Z(x_0)$	$\hat{Z}(x_0)$	Varianza Kriging VK	$r(x_i)$
3	5.89	4.53	0.637
6	4.48	3.34	- 0.455
6	7.10	3.35	0.328
8	6.08	3.35	- 0.573
6	7.76	4.53	0.388

Cuadro 4.2. Resumen de los errores reducidos

\bar{x}	s	n
0.065	0.542	5

Resulta como consecuencia que H_0 no se rechaza con un nivel de significancia de $\alpha = 0.05$

Más aún, un intervalo de confianza basado en la distribución t-student con un 95 por ciento de confiabilidad es:

$$\bar{x} - t_{\alpha/2(n-1)g.l.} \frac{s}{\sqrt{n}} \leq Mr(x_i) \leq \bar{x} + t_{\alpha/2(n-1)g.l.} \frac{s}{\sqrt{n}}, \text{ entonces:}$$

$$0.065 - 2.77 \frac{0.542}{\sqrt{5}} \leq Mr(x_i) \leq 0.065 + 2.77 \frac{0.542}{\sqrt{5}}$$

$$\therefore -0.606 \leq Mr(x_i) \leq 0.736$$

Por lo que el estimador por intervalo para $Mr(x_i)$ es de -0.606 a 0.736, es decir existe una probabilidad de 0.95 de que $Mr(x_i)$ se encuentre entre estos valores.

Con fundamento en lo anterior se puede afirmar que el método kriging es una magnífica opción en este tipo de trabajos, ya que se satisfacen las expectativas que convienen a la hipótesis.

Evaluación del Método del Individuo más Cercano

La evaluación del método se hará mediante una comparación directa de la cantidad de individuos estimada v.s. la cantidad real de individuos, por lo que es

conveniente tomar en cuenta el Cuadro 4.3, en el que se presentan los verdaderos parámetros de la población.

En relación a este método se puede intuir de antemano que habrá ciertas incongruencias que se reflejarán en errores por estar diseñado para trabajar sólo cuando la dependencia espacial no existe y en este caso se sabe que el concepto anterior está presente.

Aún cuando no exista la dependencia espacial parece ser que presenta en ocasiones serias diferencias, cuando es comparado con otros métodos también clásicos, de tal forma que se ha observado que llega a sobrestimar o a subestimar el parámetro buscado.

Cuadro 4.3. Parámetros verdaderos de la población

TAMAÑO DE LA POBLACIÓN N	256
PROMEDIO μ	1.85
VARIANZA σ^2	1.02
DESVIACIÓN ESTANDAR σ	1.00
COEF. DE VARIACIÓN	54.05 %

El total de la población en el área de 32 x 32 metros cuadrados de acuerdo al método fue de 687 individuos; comparado lo anterior con la población real de 474, existe una diferencia muy marcada. La situación anterior se plasma en el Cuadro No. 4.4.

Cuadro 4.4. Comparación de los resultados en el método del individuo más cercano

POBLACIÓN ESTIMADA	POBLACIÓN REAL	DIFERENCIA	DIFERENCIA EN %
687	474	213	45

Por la información anterior es evidente que se sobrestima la población real, con una diferencia por demás considerable.

Ante tal evidencia se puede concluir que el método presenta algunas deficiencias en la estimación; hecho que lo coloca en desventaja respecto al método Kriging.

CONCLUSIONES Y RECOMENDACIONES

- a) Dependiendo de la precisión utilizada el método del individuo más cercano puede aportar resultados confiables.
- b) Dada la dependencia espacial en la especie estudiada el método kriging resultó ser más confiable.
- c) Es necesario definir el patrón espacial antes de aplicar los métodos de estimación, ya que esto trae consigo una mayor precisión en los resultados.

RESUMEN

En el ámbito agroecológico es muy frecuente que se presente la necesidad de realizar un muestreo con el propósito de realizar inferencia sobre alguna población; ante tal necesidad se desarrollan métodos para la realización de estos trabajos de tal manera que ofrezcan la máxima precisión posible.

El concepto de dependencia espacial viene a cambiar la panorámica que se tenía para el tratamiento de este tipo de casos.

El método kriging presenta una robustez muy buena por lo que se consolida como una opción a considerar, sobre los métodos clásicos.

El presente trabajo se desarrolla sobre la Gobernadora, planta arbustiva ampliamente conocida y desarrollada en buena parte del territorio nacional, misma que posee muchas cualidades comprobadas y a su vez también se le atribuyen cualidades casi mágicas, obviamente no comprobadas.

Desde el levantamiento de los datos de campo es posible casi siempre, intuir la existencia de la dependencia espacial, característica imprescindible para la aplicación

del método kriging. Después de realizar el cálculo del semivariograma experimental se corrobora lo antes planteado, al observar el comportamiento de la gráfica.

La siguiente etapa que es el ajuste del semivariograma experimental al teórico, revela hasta que distancia se debe considerar que existe la dependencia espacial, ésto es el valor del rango.

La interpolación predice el valor de la variable en localidades no medidas y da la oportunidad de hacer la evaluación del método mediante los errores reducidos.

El método del Individuo más Cercano es un recurso ampliamente usado en este tipo de trabajos pero al confrontar las estimaciones con los parámetros verdaderos se aprecian diferencias significativas, debido a que la población no está distribuida al azar y consecuentemente existe la dependencia espacial. Por lo que el Individuo más Cercano no es una buena opción para casos como el presente.

LITERATURA CITADA

- Brunk, H. D. 1979. Introducción a la Estadística Matemática. Trillas. México. 579 p.
- Comisión de Estudios del Territorio Nacional (CETENAL). 1984. Agua Nueva. Carta Topográfica G14C43. Escala 1:50 000. Color: Varios. México.
- Coronado N., R. 1992. Discusión de Cuatro Métodos Estadísticos para definir el Patrón Espacial, en Poblaciones Ecológicas. Tesis de Maestría. Universidad Autónoma Agraria "Antonio Narro". Buenavista, Saltillo, Coah. 57 p.
- Cortés B., J. J. 1988. Variabilidad Espacial en Variables Geohidrológicas. Tesis de Maestría. Universidad Autónoma Agraria "Antonio Narro". Buenavista, Saltillo, Coah. 115 p.
- Díaz A., C. y R. M. Gutiérrez O. (1994). Estimación del Tamaño de un Banco de Semillas en Milpa Alta, D. F., por el Método Geoestadístico y por el Método Poligonal. Revista de Estadística. 6(8):73-82. Instituto Nacional de Estadística Geografía e Informática. Asociación Mexicana de Estadística. México, D. F.
- Dudewics, E. J. and Satya N. M. 1988. Modern Matematical Statistics. John Wiley and Sons. United States of America. 838 p.
- García S., L. P. 1993. Digestibilidad *In Vitro* de la Harina de Gobernadora (*Larrea tridentata*, Cov.), Tratada con NaOH al 0.1 % de Normalidad a Diferentes Tiempos de Agitación (0, 2, 4 y 6 Horas). Tesis de Licenciatura. Universidad Autónoma Agraria "Antonio Narro". Buenavista, Saltillo, Coah. 50 p.

- González S., F. A. 1989. Determinación de la Persistencia de la Actividad Bactericida de la Recina de *Larrea tridentata*, sobre *Pseudomonas solanacearum* en Laboratorio e Invernadero. Tesis de Licenciatura. Universidad Autónoma Agraria "Antonio Narro". Buenavista, Saltillo, Coah. 78 p.
- Huerta de la P., A. 1986. Acción Nematicida de la Resina de Gobernadora (*Larrea tridentata*, Ccv.) en el Guayule (*Parthenium argentatum*, Gray.) bajo Cultivo. Tesis de Licenciatura. Universidad Autónoma Agraria "Antonio Narro". Buenavista, Saltillo, Coah. 54 p.
- Mendenhall, W., R. L. Scheaffer and L. Ott. 1987. Elementos de Muestreo. Grupo Editorial Iberoamérica. México, D. F. 321 p.
- Raj, D. 1980. Teoría del Muestreo. Fondo de Cultura Económica. México, D. F. 305 p.
- Ramírez A., C. 1980. Interpolación Espacial de Datos de Sales en el Suelo. Tesis de Maestría. Colegio de Postgraduados. Montecillo, Mex. 158 p.
- Salinas C., S. 1981. Evaluación de Métodos de Muestreo para Estimar Densidad en Arbustos. Tesis de Licenciatura. Universidad Autónoma de Nuevo León. Facultad de Agronomía. Marín, N. L. 76 p.
- Thompson, S. K. 1993. Sampling. John Wiley and Sons, Inc. New York, U.S.A. 343 p.
- Vieira, S. R., J. L. Hatfield, D. R. Nielsen and J. W. Biggar. 1982. Geostatistical Theory and Application to Variability of Some Agronomical Properties. *Hilgardia*. 51(3):1-75. United States of America.
- Yamane, T. 1979. Estadística. Tercera Edición. HARLA. México D. F. 771 p.

A P É N D I C E S

APÉNDICE A

ESTIMADOR INSESGADO Y DE VARIANZA MÍNIMA

ESTIMADOR INSESGADO Y DE VARIANZA MÍNIMA

Estimación Puntual

Dudewics y Mishra (1988), describen con propiedad y exactitud este importante concepto.

Si se tienen variables aleatorias X_1, X_2, \dots, X_n no necesariamente independientes e idénticamente distribuidas, la función de distribución de X_1, X_2, \dots, X_n depende de un parámetro desconocido θ , por ejemplo la media de una distribución normal puede ser desconocida, entonces aquí la estimación de θ es un problema de estimación puntual.

Se dice que $t_n (X_1, X_2, \dots, X_n)$ es un estadístico y también un estimador de θ , por lo que un valor particular del estimador es llamado estimación de θ .

Estimador Insesgado y Sesgo

Se dice que $t_n (X_1, X_2, \dots, X_n)$ es un estimador insesgado de θ si:

$$E_{\theta} [t_n(x_1, x_2, \dots, x_n)] \equiv \theta \quad \text{para todo } \theta \in \Theta.$$

$$\text{El sesgo de } t_n = E_{\theta} [t_n(X_1, X_2, \dots, X_n)] - \theta \equiv b_n(\theta).$$

Con frecuencia algunos estimadores suelen ser sesgados, pero este sesgo puede desaparecer si n se incrementa, con lo que pasa a ser estimador asintóticamente insesgado, es decir:

$t_n(X_1, X_2, \dots, X_n)$ es un estimador asintóticamente insesgado de θ si:
 $\lim_{n \rightarrow \infty} b_n(\theta) = 0$ para todo $\theta \in \Theta$.

Varianza Mínima

Este concepto está fuertemente vinculado con lo que se llama eficiencia y según lo describe Yamane (1979), se define relacionándola con el estimador que tenga la varianza más pequeña, en otras palabras, si existe una situación en la cual un estimador insesgado tiene varianza al menos tan pequeña como la de cualquier otro estimador insesgado, éste es llamado eficiente.

Expresado de manera formal:

$$e_{\theta}(\varphi_n) = \frac{1}{\left\{ n E_{\theta} \left(\frac{\partial \ln f(x/\theta)}{\partial \theta} \right)^2 \right\}} \text{Var}_{\theta}(\varphi_n)$$

Si $e_\theta(\varphi_n) \equiv 1$, entonces φ_n es considerado eficiente.

Es necesario recordar también una expresión para el caso de un estimador asintóticamente eficiente $\varphi_n(X_1, X_2, \dots, X_n)$ si $\lim_{n \rightarrow \infty} e_\theta(\varphi_n)$.

Otro punto de vista para esta situación lo plantea Brunk (1979):

Si $\hat{\theta}_1$ y $\hat{\theta}_2$, son dos estimadores de un parámetro θ que van a ser comparados, la eficiencia del primero respecto al segundo es la razón de sus errores cuadrados medios

$$E[(\hat{\theta}_1 - \theta)^2] / E[(\hat{\theta}_2 - \theta)^2]$$

Se hace necesario recordar que el estimador de un parámetro θ es insesgado si $E(\hat{\theta}) = \theta$; y siendo así, su error cuadrado medio coincide entonces con su varianza.

APÉNDICE B

CONSIDERACIONES GENERALES DEL MUESTREO CLÁSICO

CONSIDERACIONES GENERALES DEL MUESTREO CLÁSICO

Según lo menciona Mendenhall *et al.* (1987), el objetivo fundamental de la estadística es hacer inferencias acerca de una población con base a la información contenida en una muestra y a sabiendas de que la obtención de ésta va invariablemente ligada a la Teoría de la Probabilidad. Tal panorama se presenta inmediatamente al empezar a seleccionar las unidades que integrarán la muestra, siendo que cada unidad tendrá una probabilidad diferente de cero de salir seleccionada.

Como es sabido, se puede conocer información de una población a través de dos maneras, una de ellas es el censo, que básicamente consiste en tomar toda y cada una de las unidades integrantes de la población, analizarlas y concluir el trabajo con el conocimiento preciso de los parámetros de interés, los inconvenientes de esta manera se pueden resumir en dos cosas, tiempo y dinero; sería mucho más laborioso y costoso usar este recurso.

Por otra parte la segunda opción, es decir el muestreo, en cuanto a tiempo y dinero produce el efecto contrario, es más rápido y menos costoso, sólo que para el empleo adecuado de tal recurso se requiere de una capacidad más técnica. La concepción

anterior se puede resumir en una frase coloquial que quizás no sea rigurosamente cierta pero que presenta un enfoque filosófico del caso y dice:

Un censo cualquier persona lo puede hacer, pero un muestreo no cualquiera.

Muestreo Simple Aleatorio (MSA)

Raj (1980) describe de forma magistral el fundamento matemático para el MSA. Si se tiene una población de N unidades y se seleccionan n unidades pero cuidando que cada unidad tenga igual probabilidad de ser seleccionada y además cada unidad es reintegrada a la población una vez que se toman los datos pertinentes, es decir la unidad vuelve a ser susceptible de seleccionarse, se tiene el caso de un Muestreo Simple Aleatorio con Reemplazo (MSAcr). Si por el contrario se ignoran las repeticiones o las unidades seleccionadas no son reintegradas a la población y se continúa hasta completar n unidades, se tiene el caso de un Muestreo Simple Aleatorio sin Reemplazo (MSAsr).

Si se denominan a las unidades de la población N como $U_1, U_2, U_3, \dots, U_N$, en el MSAcr cualquier selección resultará U_i ($i = 1, 2, 3, \dots, N$) teniendo cada una de ellas una probabilidad de $1/N$. Todas las selecciones son independientes, puesto que la unidad seleccionada es reincorporada a la población antes de hacer la siguiente elección

En el MSAsr al seleccionar la primera unidad por ejemplo U_j , ésta tiene probabilidad de $1/N$, por lo que al seleccionar la siguiente unidad U_i , la probabilidad será de $1/(N-1)$. Por lo tanto la probabilidad absoluta de que U_i sea seleccionada al hacer el

segundo sorteo es
$$\sum_{j \neq i} \frac{1}{N} \frac{1}{N-1} = \frac{1}{N}.$$

Análogamente la probabilidad absoluta de que se seleccione en el tercer sorteo una unidad específica U_i es $1/N$ y así sucesivamente, por lo que se concluye que también en el MSAsr cada selección puede producir cualquier unidad U_i con una probabilidad de $1/N$.

Más aún; la probabilidad de que U_i sea seleccionada en la n -ésima extracción es igual a la probabilidad de que sea extraída en la n -ésima elección y no sea extraída en la $n-1$ extracción anterior, es decir:

$$\frac{N-1}{N} \frac{N-2}{N-1} \frac{N-3}{N-2} \dots \frac{N-(n+1)}{N-n+2} \frac{1}{N-n+1} = \frac{1}{N}$$