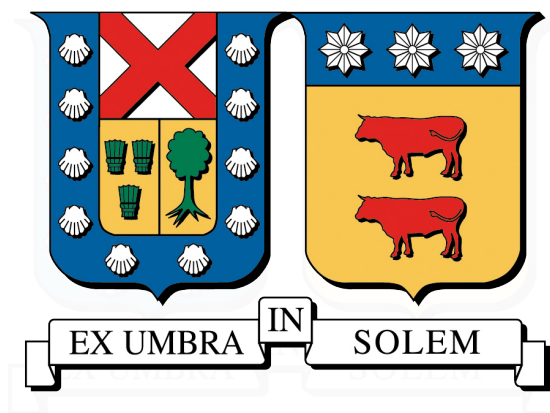


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS
SANTIAGO - CHILE



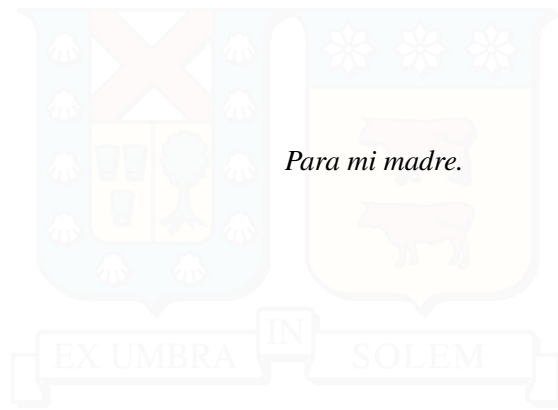
**OPTIMIZACIÓN DE MUESTREO GEORREFERENCIADO CON FINES DE
MAPEO**

CATALINA REYNOLDS CAVALLIERI

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

PROFESOR GUÍA	:	SR. RODRIGO ORTEGA BLU
PROFESOR CORREFERENTE	:	SR. VICTOR ALBORNOZ
PROFESOR CORREFERENTE EXTERNO	:	SR. LUIS ACOSTA

ENERO 2017



Resumen Ejecutivo

En la presente memoria se propone y se desarrolla una estrategia de aproximación de muestreo sistemático con fines de mapeo para ser utilizados en agricultura de precisión.

La agricultura de precisión se basa en la variabilidad a nivel de predio para la gestión agrícola de estos. Utilizando diversas tecnologías permite optimizar los recursos utilizados en los predios agrícolas en un nivel técnico-productivo, económico y ambiental.

La posibilidad de identificar, cuantificar y mapear la variabilidad de un predio agrícola ofrece la oportunidad de georreferenciar y aplicar diversos insumos con dosis variables en áreas con coordenadas geográficas específicas (TDV).

La primera etapa de la TDV consiste en la recolección de datos, estos datos luego son procesados y asociados para la elaboración de mapas digitales que son posteriormente analizados y en base a este análisis es que se hace la recomendación de la aplicación de insumos.

Para la recolección de datos existen diversas técnicas de muestreo. Una de estas técnicas corresponde al muestreo sistemático, en el cual el predio se divide en cuadrantes dentro de los cuales se toma la muestra.

La aproximación al muestreo presentada en esta memoria se divide en 4 etapas sucesivas. Primero se determinó el tamaño de la muestra asegurando que esta sea representativa. Luego, se determinaron los cuadrantes en los cuales se divide el predio, considerando la posibilidad de movilidad del número total de muestras por diferentes cuadrantes. Posteriormente se compararon dos técnicas de predicción, Kriging Ordinario y Distancia Inversa Ponderada. Para finalmente terminar con la optimización mediante la técnica elegida con su respectivo mapa georreferenciado.

Palabras Clave: Agricultura de precisión, Muestreo sistemático, Optimización de Muestreo, Mapeo Georreferenciado

Índice de Contenidos

1. Antecedentes Generales	1
1.1. Problema de Investigación	1
1.1.1. Contexto	1
1.1.2. Problema	2
1.1.3. Agentes involucrados	2
1.2. Objetivos	3
1.2.1. General	3
1.2.2. Específicos	3
2. Marco Teórico	4
2.1. Por qué muestrear	5
2.2. Objetivo del muestreo	6
2.3. Impacto sobre costos	6
2.4. Métodos de muestreo	6
2.4.1. Muestreo simple al azar	7
2.4.2. Muestreo estratificado al azar	8
2.4.3. Muestreo compuesto	8
2.4.4. Muestreo sistemático	9
2.5. Fuentes de error	10
2.6. Variabilidad espacial de las propiedades químicas del suelo	11
2.7. Dependencia o autocorrelación espacial	12
2.8. Conceptos generales sobre el modelamiento geoestadístico de datos	12
2.8.1. Momentos	13
2.8.2. Hipótesis de estacionaridad	15
2.9. Análisis variográfico	17
2.9.1. Anisotropía	18
2.10. Kriging: predicción e interpolación	24
2.10.1. Kriging de media conocida (Kriging Simple) (SK)	25
2.10.2. Kriging de media desconocida (Kriging Ordinario) (OK)	27
2.10.3. Otros	28
2.11. Medidas de autocorrelación	29
2.11.1. I de Moran	31
2.11.2. C de Geary	32
2.12. Distancia Inversa Ponderada (IDW): predicción e interpolación	32
2.13. Validación del modelo de predicción	33

3. Materiales y métodos	34
3.1. Área de estudio	34
3.2. Recopilación y gestión de datos	35
3.2.1. Definición del tamaño de muestra	36
3.2.2. Definición de los cuadrantes	36
3.2.3. Definición de método a utilizar	38
3.2.4. Algoritmo de optimización	39
4. Resultados y discusión	41
4.1. Tamaño de la muestra y diseño de cuadrantes	41
4.2. Método de predicción	42
4.2.1. Kriging Ordinario	42
4.2.2. IDW	44
4.3. Algoritmo de optimización	45
4.4. Comparación Kriging y muestreo aleatorio	49
5. Conclusiones	58
Bibliografía	60
A. Marco Teórico	62
B. Valores Muestreo	63
C. Códigos de Programación	68
D. Mapas muestra aleatoria sistemática	75
E. Mapas muestreo aleatorio	77

Índice de Figuras

2.1. Muestreo compuesto	9
2.2. Tipos de Muestreo. a) Simple al azar; b)Estraificado al azar; c) Sistemático rejilla rectangular; d) Sistemático rejilla polar	10
2.3. Variabilidad espacial del Nitrógeno total (izq.) y P Olsen (der.) en un potrero de dos hectáreas de un suelo de la serie Diguillin	11
2.4. Variograma típico sobre el cual se ajustó un modelo esférico	18
2.5. Ejemplo (2D) de anisotropía geométrica de razón a/b y ángulo θ . Izquierda: elipse de alcances. Derecha: variograma en las direcciones principales de anisotropía	19
2.6. Ejemplo (3D) de anisotropía zonal en dirección vertical. Derecha: variogramas en las direcciones horizontales (D_1, D_2), vertical (D_3) y una dirección oblicua (D_θ)	20
2.7. Variograma pepítico (derecha) y ejemplo de variable regionalizada asociada (izquierda)	21
2.8. Variograma esférico (derecha) y ejemplo de variable regionalizada asociada (izquierda)	21
2.9. Variograma exponencial (derecha) y ejemplo de variable regionalizada asociada (izquierda)	22
2.10. Variograma Gaussiano (derecha) y ejemplo de variable regionalizada asociada (izquierda)	22
2.11. Variograma seno cardinal (derecha) y ejemplo de variable regionalizada asociada (izquierda)	23
2.12. Variograma potencia (derecha) y ejemplo de variable regionalizada asociada a un variograma lineal (izquierda)	24
2.13. Contigüidad en un entramado regular	30
2.14. Matriz de interacciones para nueve datos situados en una cuadrícula	30
3.1. Área de estudio	34
3.2. Definición de los cuadrantes para muestreo sistemático, las flechas indican posibilidad de movimiento del cuadrante	37
3.3. Grilla de interpolación con valores de toda la población, círculos rojos representan a los valores muestrales	38
3.4. Configuración inicial de muestreo (en negro)	39
4.1. Cuadrantes optimización finales, 15 columnas y 19 filas, con un total de 285 cuadrantes	42
4.2. Variograma experimental y teórico para los valores de toda la población	43
4.3. Mapa mediante K.O realizado con valores de toda la población	44
4.4. Mapa mediante IDW realizado con valores de toda la población	45

4.5. Mapa mediante IDW realizado con la muestra inicial	46
4.6. Mapa mediante IDW realizado con la muestra inicial con puntos de muestreo en negro	46
4.7. Mapa MSE mediante IDW realizado con la muestra inicial, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población	47
4.8. Mapa mediante IDW realizado con la muestra optimizada	48
4.9. Mapa mediante IDW realizado con la muestra optimizada con puntos de muestreo en negro	48
4.10. Mapa MSE mediante IDW realizado con la muestra optimizada, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población	49
4.11. Variograma experimental y teórico para los valores de la muestra inicial	50
4.12. Variograma experimental y teórico para los valores de la muestra optimizada	50
4.13. Mapa mediante K.O. realizado con la muestra inicial	51
4.14. Mapa mediante K.O. realizado con la muestra inicial con puntos de muestreo en negro	51
4.15. Mapa mediante K.O. realizado con la muestra optimizada	52
4.16. Mapa mediante K.O. realizado con la muestra optimizada con puntos de muestreo en negro	52
4.17. Mapa MSE mediante K.O. realizado con la muestra inicial, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población	53
4.18. Mapa MSE mediante K.O. realizado con la muestra optimizada, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población	54
4.19. Mapa MSE mediante K.O. realizado con la muestra inicial, donde el MSE se calcula con los valores obtenidos mediante K.O. para toda la población	55
4.20. Mapa MSE mediante K.O. realizado con la muestra optimizada, donde el MSE se calcula con los valores obtenidos mediante K.O. para toda la población	55
5.1. Recomendación aproximación problema de muestreo con fines de mapeo	59
D.1. Mapa mediante IDW realizado con muestra aleatoria dentro de los cuadrantes definidos	75
D.2. Mapa mediante IDW realizado con muestra aleatoria dentro de los cuadrantes definidos con puntos de muestreo en negro	76
D.3. Mapa MSE con muestra aleatoria dentro de los cuadrantes definidos, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población	76
E.1. Mapa mediante IDW realizado con muestreo aleatorio	77
E.2. Mapa mediante IDW realizado con muestreo aleatorio con puntos de muestreo en negro	78
E.3. Mapa MSE realizado con muestreo aleatorio, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población	78

Índice de Tablas

3.1. Resumen estadístico variable N _{dvi}	35
4.1. Comparación MSE de intra y extrapolación	56
A.1. Prueba T de dos colas	62
B.1. Puntos Originales	64
B.2. Puntos aleatorios	65
B.3. Puntos optimización	66
B.4. Puntos aleatorios de toda la población	67
C.1. Estimación tamaño de muestra	68
C.2. Prueba modelo de predicción e interpolación	69
C.3. Algoritmo de optimización	70
C.4. Algoritmo de optimización (continuación)	71
C.5. Algoritmo de optimización (continuación)	72
C.6. Algoritmo de optimización (continuación)	73
C.8. Estimación tamaño de muestra	73
C.7. Comparación Kriging datos originales vs. datos optimizados	74

Antecedentes Generales

Problema de Investigación

Contexto

La Agricultura de Precisión (AP) es un concepto utilizado para la gestión de predios agrícolas basado en la variabilidad espacial a nivel de predio. Utiliza diversas tecnologías para estimar, evaluar y entender las variaciones inherentes de los predios agrícolas con el fin de optimizar los recursos en un nivel técnico-productivo, económico y ambiental.

La posibilidad de identificar, cuantificar y mapear la variabilidad de un predio agrícola ofrece la oportunidad de georreferenciar y aplicar diversos insumos con dosis variables en áreas con coordenadas geográficas específicas. Esta práctica es conocida como Manejo de Nutrientes Sitio-Específico, o también, como Tecnología de Dosis Variables (TDV). La TDV contempla diversas áreas como la aplicación variable de dosis de fertilizante, plaguicida y semillas y control sitio-específico de malezas, insectos y enfermedades entre otras.

La primera etapa para desarrollar y adoptar las prácticas de la TDV corresponde al diagnóstico o recolección de datos. En una segunda etapa estos datos son luego procesados y asociados para la elaboración de mapas digitales que son posteriormente analizados y en base a este análisis es que se hace la recomendación de la aplicación de insumos.

Para el caso del suelo, la recolección se realiza mediante un muestreo del predio. Actualmente la técnica de muestreo más utilizado corresponde al muestreo sistemático o de grilla. Las muestras son tomadas a intervalos regulares en todas las direcciones. Este tipo de muestreo ha sido extensamente aceptado debido al potencial de incremento en la exactitud de los análisis de suelos ([Roberts y Henry, 2000](#)).

Problema

Sin ir más lejos en los problemas actuales de la AP en el país, en específico la TDV, como falta de personal capacitado o conocimiento de la misma, un problema inherente a la práctica misma corresponde al muestreo.

Primeramente, en la actualidad no existe ningún modelo que permita identificar la cantidad adecuada de muestras ni dónde ubicarlas. Hoy en día se utiliza el muestreo por grillas fijas, en donde se puede determinar la cantidad de cuadrillas que contendrá la grilla y la ubicación de la muestra (al centro o desplazada) mas no se sabe si corresponde a un diseño óptimo de muestreo. No existe aún ningún software que permita optimizar la cantidad de muestras para que estas mantengan su dependencia espacial (lo que se traduce en un mejor mapeo).

Además existe un tema económico relacionado al muestreo, actualmente un análisis completo de una muestra bordea los \$40.000¹.

Es por esto que la construcción de un modelo de muestreo que permita mantener la dependencia espacial de las muestras utilizando la menor cantidad posible de estas contribuiría a resolver una de las problemáticas de la TDV.

Agentes involucrados

La resolución del problema antes planteado tendría un potencial impacto en diversos agentes como:

- Productores agrícolas: Les permitirá realizar un mejor muestro de sus predios.
- Ingenieros agrónomos: Les permitirá realizar un mejor muestreo para diversas aplicaciones.
- Empresas de servicios: Un buen muestreo les permite mejorar su servicio de mapeo.
- Académicos: El muestreo de suelos no solo se utiliza en la AP, si no en diversas áreas relacionadas con la geoestadística.

¹Laboratorio de nutrición y suelos forestales, Universidad Austral de Chile

Objetivos

General

Diseñar y evaluar un modelo de optimización para la ubicación espacial de muestras a partir de un diseño sistemático.

Específicos

- Diseñar algoritmo de optimización de muestreo utilizando muestreo sistemático.
- Evaluar diferentes métodos de interpolación para una misma muestra.
- Utilizar el algoritmo de optimización diseñado para generar mapas georreferenciados mediante interpolación basada en el variograma (Kriging) y la interpolación geométrica (Distancia inversa Ponderada) para representar el sitio evaluado.

Marco Teórico

La AP es un concepto agronómico de gestión de predios agrícolas, basado en la existencia de una importante variabilidad espacial a nivel de predio. Utiliza diversas tecnologías para estimar, evaluar y entender dichas variaciones. El uso de estas tecnologías contribuye a una adecuada toma de decisiones desde el punto del manejo técnico-productivo, económico y ambiental del predio ya que permite la recolección de información que puede ser utilizada para evaluar diversos ámbitos tales como la densidad óptima de siembra, estimar el uso de fertilizantes, optimizar la gestión del recurso hídrico, predecir con más exactitud la producción de los cultivos y la calidad del producto final, así como también optimizar los rendimientos a través de prácticas de cosechas automatizadas (Villalobos et al., 2009).

En Chile la AP ha sido incorporada crecientemente como una herramienta de gestión en la fruticultura y vitivinicultura chilena desde mediados de los noventa. Estas técnicas se han propagado a diversos cultivos industriales y de cereales tales como trigo, maíz y remolacha. La contribución de la AP ha sido demostrada en términos de disminución de costos de producción, ahorro energético e hídrico y gestión sitio-específico en los huertos, entre otros aspectos por diversas investigaciones realizadas por Centros tecnológicos y de investigación (Villalobos et al., 2009).

La TDV se basa en el supuesto de que los distintos sectores del predio responden de manera diferente a una misma técnica de cultivo aplicada de manera uniforme en toda la extensión del terreno. Por lo tanto una de las aplicaciones de la TDV sería el aumento de la productividad de la explotación agrícola mediante la división del predio en sub-zonas en donde se ajuste la técnica de cultivo y la dosis de insumos como fertilizantes según las características particulares de cada sub-zona, maximizando la producción y calidad del producto.

Está comprobado que la variabilidad de los factores de suelos induce a una variabilidad de los rendimientos y la calidad de los productos agrícolas (Kitchen et al., 1999). En Chile, esta variabilidad es altamente conocida (Ortega B. y Flores M., 2000). La estimación de esta variabilidad

se puede hacer por medio del muestreo del suelo y registro de datos, con el fin de establecer un modelo de dispersión y por tanto un plan de TDV.

Actualmente, los procedimientos oficiales para el muestreo de suelos con fines de diagnóstico en Chile están contenidos en la NCh 2060 de 1999. Pero estos procedimientos no aseguran la representatividad del suelo ni existen evidencias de que correspondan a los procedimientos más óptimos de muestreo. Un plan de muestreo óptimo debe permitir aprovechar los múltiples beneficios de la TDV buscando también una minimización en los costos.

Por qué muestrear

El muestreo de suelos es un componente esencial e integral en la evaluación de la fertilidad del suelo, recomendación de nutrientes y manejo de la fertilidad. La efectividad del muestreo es un prerequisite para que el análisis de suelo logre su objetivo de un uso eficiente y racional de nutrientes para un mejor cultivo y calidad en la agricultura práctica. Sin un muestreo preciso, es poco probable que la evaluación del suelo permita determinar los nutrientes que se encuentran disponible y por tanto el requerimiento de nutrientes faltantes ([Sahrawat et al., 2008](#)).

En efecto, la mayor dificultad radica en describir adecuadamente el nivel de fertilidad de un lote ya que ésta (fertilidad) varía espacialmente (en X, Y, y Z) y temporalmente (época del año). Un muestreo será considerado correcto en la medida que describa y represente con una exactitud aceptable (lo más cerca a la media verdadera), la característica deseada del lote. Una gran proporción de los errores en la recomendación de fertilizantes se debe al inadecuado muestreo de suelos y manejo de las muestras previo a su ingreso al laboratorio ([Ortega, 2013](#)).

Los procedimientos oficiales para el muestreo de suelos con fines de diagnóstico en Chile están contenidos NCh 2060 de 1999 El alcance y campo de aplicación de la norma es el siguiente:

1. Establece el procedimiento de muestreo de suelos para análisis de fertilidad.
2. Es aplicable al muestreo de suelos para cultivos y praderas, mientras que para el sector forestal y frutícola la norma debería aplicarse sólo con fines de diagnóstico preliminar.

Al realizar un buen muestreo es posible obtener información de calidad con el fin de aplicar un buen plan de TDV que permita obtener una mayor rentabilidad versus un predio manejado sin las tecnologías de la AP. Por lo tanto, una optimización de este permitiría obtener todos los beneficios de la TDV al menor costo posible.

Objetivo del muestreo

El principal objetivo del muestreo es obtener información sobre un suelo particular. Estadísticamente corresponde a estimar los parámetros de la población (valor verdadero de cada característica en la población) es llamado parámetro. Estadísticamente los parámetros más usados son el promedio y la desviación estándar. Para el caso del promedio este se puede obtener mediante el análisis de la muestra compuesta, no así para la desviación estándar, para la cual es necesario analizar todas las sub muestras.

La optimización de un plan de muestreo debe tener como objetivo también proporcionar la mayor cantidad de información posible sin dejar de lado la minimización de los costos asociados al muestreo.

Impacto sobre costos

El muestreo del suelo conlleva diversos costos en el manejo de un predio. Existen no solo impactos positivos en los costos, sino también, impactos negativos.

Como fue mencionado anteriormente, la TDV puede ser utilizada para determinar los niveles de fertilizante necesarios en el predio. La inversión en fertilizantes representa una parte importante de los costos de producción. Mientras más cercana a las necesidades reales de las plantas sea la fertilización, permitirá que la inversión en fertilizantes sea sólo la necesaria para obtener un retorno adecuado(Bernier, 1999).

Es importante mencionar que representar a toda la población consume tiempo, esfuerzo y costos inherentes al análisis de las muestras ya que como se menciona anteriormente, actualmente un análisis completo de una muestra bordea los \$40.000. Además existen ciertos análisis que contemplan un muestreo destructivo, lo que implica un costo en la pérdida de la muestra.

Métodos de muestreo

Dentro del diseño de muestreo probabilístico se tienen cuatro tipos distintos de muestreo, los cuales serán explicados a continuación:

Muestreo simple al azar

Este tipo de muestreo (Figura 2.2 a) se utiliza en casos en los que se dispone de poca información acerca de las características de la población a medir; se basa en la teoría de probabilidades y siempre requiere de un análisis estadístico (Volke et al., 2015). Una vez definida la población de estudio (ej. población de volúmenes de suelos) se extrae un número determinado de muestras desde el lote, cada una de las cuales tiene la misma probabilidad de ser elegida. La posición de cada muestra dentro del lote puede hacerse a través de la generación de coordenadas x,y aleatorias, o bien numerando todas las posibles muestras y extrayendo el número deseado de ellas (Ortega, 2013).

De un muestreo simple al azar, pueden estimarse algunos parámetros de la población a través del cálculo de algunos estadísticos. Los más comunes son el promedio y la varianza:

$$\bar{y} = \frac{\sum_i^n y_i}{n} \quad \text{Promedio}$$

$$s^2 = \frac{\sum_i^n (y_i - \bar{y})^2}{n - 1} \quad \text{Varianza}$$

La desviación estándar (s) corresponde a la raíz cuadrada de la varianza. El tamaño de la varianza (valor absoluto) es afectado por el tamaño de las unidades de medición de la propiedad de interés.

Para comparar la variación de distintas propiedades en una misma base se calcula el coeficiente de variación que corresponde al porcentaje de la desviación estándar con respecto al promedio de cada propiedad medida, de acuerdo a la siguiente fórmula:

$$CV = \frac{s}{\bar{y}} \times 100 \quad \text{Coeficiente de variación}$$

De la varianza también puede calcularse el error estándar ($s_{\bar{y}}$) que indica el grado de error en la estimación de la media de la población:

$$s_{\bar{y}} = \sqrt{\frac{s^2}{n}} \quad \text{Error estándar}$$

El error estándar más un valor de t obtenido desde una tabla para un cierto nivel de confianza, permite calcular intervalos de confianza alrededor de la media:

$$l_s = \bar{y} + t_{0,025} \times s_{\bar{y}} \quad \text{Límite superior}$$

$$l_i = \bar{y} - t_{0,025} \times s_{\bar{y}} \quad \text{Límite superior}$$

De acuerdo a esta misma fórmula rearrreglada, es posible calcular cuántas muestras se requerirían para estimar el promedio de la población con un cierto nivel de error a un nivel de probabilidad dado:

$$n = \frac{t^2 s^2}{D^2} \quad \text{Número de observaciones}$$

Donde:

t = valor de la tabla de t elegido arbitrariamente en primera instancia y modificado luego por reiteración (Anexo A)

s = variancia de la muestra estimada de experiencias anteriores o calculada de acuerdo al rango de valores esperado $s^2 = \left(\frac{R}{4}\right)^2$

D = nivel de precisión permitido (error) en la estimación de la media

Muestreo estratificado al azar

En este tipo de muestreo (Figura 2.2b), la población en estudio se subdivide en estratos o subgrupos que tienen cierta homogeneidad en el terreno y en cada estrato se realiza un muestreo aleatorio simple. (Volke et al., 2015). Por ejemplo, el lote puede ser dividido por topografía, manejo, textura, etc. Las razones técnicas para muestrear el suelo de esta manera incluyen (Ortega, 2013):

1. Hacer inferencias acerca de cada una de las sub-poblaciones (sectores) separadamente
2. Incrementar la precisión de las estimaciones (media, desviación estándar, etc.) sobre la población entera.

Muestreo compuesto

El muestreo compuesto es similar al muestreo simple al azar (o estratificado al azar), la diferencia radica en que toda la unidad de muestreo es cubierta y no se realizan repeticiones, debido a

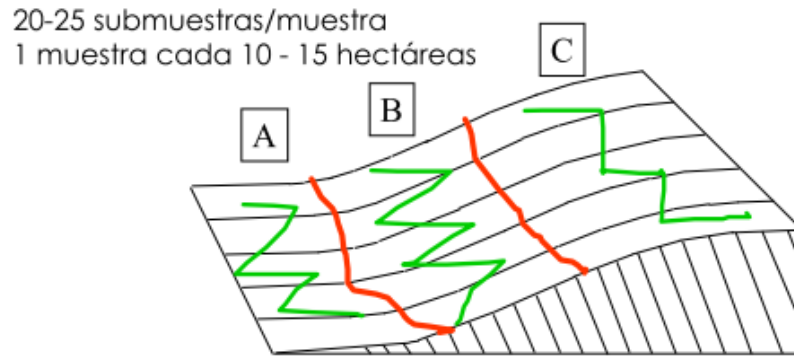


Figura 2.1: Muestreo compuesto
(Fuente: [Ortega, 2015](#))

que el suelo obtenido de todos los puntos de muestreo es mezclado en una sola muestra. La localización de las muestras no es al azar, sino, sigue un patrón, cuyo inicio y distanciamiento queda a juicio quien realice el muestreo ([Ortega, 2013](#)). Este muestreo tiene la ventaja que es de fácil implementación y la desventaja que solo permite la estimación del promedio ([Ortega, 2015](#)).

En la práctica, deberían extraerse 25 muestras, por cada unidad homogénea de muestreo la cual no debería exceder las 10 hectáreas. Estas 25 muestras deben ser compuestas en una sola muestra que se envía al laboratorio ([INN, 1999](#)).

Poco se gana en precisión cuando se excede 25 puntos de muestreo. Asimismo, el coeficiente de variación no se afecta demasiado sobre un mínimo de 8 a 12 hectáreas. Si se necesitara una mejor estimación del promedio del lote, el área debería ser subdividida en más sectores homogéneos. La ubicación dentro del lote se hace a juicio personal, siguiendo una recorrido en X o Zig Zag, que han arrojado menores coeficientes de variación que el muestreo simple al azar (Figura 2.1) ([Ortega, 2013](#)).

Muestreo sistemático

El muestreo sistemático es una herramienta que puede utilizarse para reducir la variabilidad de las muestras. Las muestras se ubican en un patrón regular en toda la zona de estudio; puede realizarse a partir de un punto determinado al azar, a partir del cual se establece cierta distancia para ubicar los demás puntos (a distancias uniformes entre sí). Este tipo de muestreo puede realizarse por rejilla rectangular o polar (Figura 2.2 c y d) ([Mason, 1992](#)).

Puede llevarse a cabo en superficies de cualquier tamaño. La distancia equidistante entre los puntos de muestreo pueden ser de unos centímetros, metros o hasta kilómetros, lo cual depende del

tipo de estudio que se esté realizando, lo que permite que las muestras puedan ubicarse de acuerdo con las dimensiones y forma del terreno (Volke et al., 2015).

La intensidad de muestreo depende del objetivo. Una alta intensidad se utiliza para el mapeo de propiedades, en cambio, una baja intensidad se utiliza cuando se desea una estimación de los parámetros de la población (promedio y desviación estándar) (Ortega, 2015).

Las muestras pueden ser analizadas separadamente y de manera compuesta según la necesidad del estudio, este tipo de muestreo es ideal para sistemas de muestreo automáticos georreferenciados

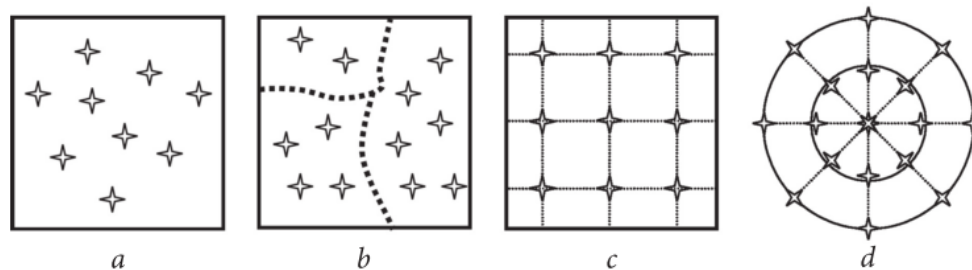


Figura 2.2: Tipos de Muestreo. a) Simple al azar; b)Estratificado al azar; c) Sistemático rejilla rectangular; d) Sistemático rejilla polar
(Fuente: Mason, 1992)

Debido a que el muestreo sistemático es el más utilizado gracias a su potencial de incremento en la exactitud de los análisis de suelos (Roberts y Henry, 2000), resulta lógico que al desarrollar un modelo de optimización de muestreo georreferenciado se utilice ese modelo.

Fuentes de error

Siempre se ha reconocido que la principal causa de errores en el análisis de suelos es el muestreo, antes que errores propios de los procedimientos analíticos (Roberts y Henry, 2000). Los errores asociados al muestreo de suelos pueden caer en tres categorías (Ortega, 2015):

1. Error de muestreo: causado por el hecho de que la muestra incluye solo los puntos muestreados en vez de toda la población de puntos posibles de muestrear. La única posibilidad de eliminarlo es incluir en el muestreo a toda la población.
2. Error de selección: causado por la tendencia a seleccionar algunas unidades muestrales dentro de la población con mayor o menor frecuencia. Por ejemplo, evitar los sitios pedregosos del lote o sobremuestrear los bordes de un lote. El muestreo simple al azar puede eliminar parte de este sesgo.

3. Error de medición: la medición observada no corresponde con el valor verdadero de la unidad muestreada. Normalmente el error de muestreo es mayor que el error de medición.

Variabilidad espacial de las propiedades químicas del suelo

Las propiedades químicas del suelo son altamente variables dentro de un lote de producción considerado visualmente homogéneo. A modo de ilustración, se caracteriza la variabilidad espacial del nitrógeno disponible y el P Olsen²

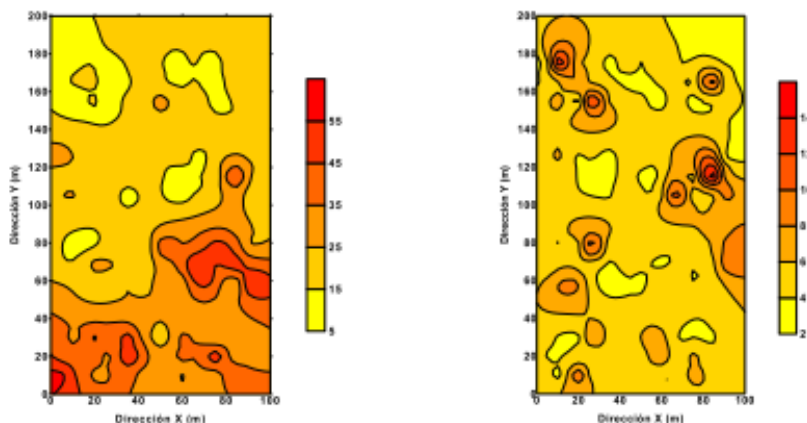


Figura 2.3: Variabilidad espacial del Nitrógeno total (izq.) y P Olsen (der.) en un potrero de dos hectáreas de un suelo de la serie Diguillin
(Fuente: [Ortega, 2015](#))

De la figura se puede apreciar que incluso en superficies tan pequeñas (dos hectáreas), tanto el Nitrógeno disponible ($\text{NO}_3\text{-N} + \text{NH}_4\text{-N}$) como el P Olsen varían ampliamente. Es importante mencionar que la variabilidad aumenta a medida que aumenta el tamaño de la superficie a muestrear, aunque este aumento pareciera ser no muy significativo sobre las 12 hectáreas ([Ortega, 2013](#)):

La norma Chilena indica que la mayor variabilidad está representada por aquel elemento nutritivo más móvil en el suelo cuyas concentraciones y solubilidad son susceptibles de sufrir rápidas oscilaciones. Por esta razón la norma recomienda muestrear una superficie no superior a las 10 hectáreas ([INN, 1999](#)).

La variabilidad del predio es un factor muy importante a considerar en el modelamiento de la optimización de un muestreo georreferenciado, un buen plan de muestreo debe ser capaz de reflejar la variabilidad del terreno a analizar.

²En Chile, la estimación del suministro del P del suelo se realiza por el método denominado P Olsen (extracción con NaHCO_3 0,5 M a pH 8,5)

Dependencia o autocorrelación espacial

Los procedimientos estadísticos clásicos en los cuales se basa el muestreo de suelos asumen que las observaciones en el campo son independientes unas de otras no importando su ubicación en el terreno. Sin embargo, debido a la naturaleza continua del paisaje, procesos de formación de suelos, y zonas climáticas, es razonable esperar que exista autocorrelación (dependencia) espacial, es decir que muestras tomadas cerca unas de otras sean más parecidas que muestras tomadas a mayor distancia entre ellas (Ortega, 2015). Se dice que una variable está espacialmente autocorrelacionada cuando los valores que se observan en un lugar específico, no solo depende de factores externos, sino que también depende de los valores de la variable observados en las regiones vecinas (Pantoja, 2013).

La cuantificación de la dependencia espacial de propiedades del suelo es crítica para definir estrategias de muestreo adecuadas, debido a que, por un lado, aumentando el número de observaciones no necesariamente se aumenta la exactitud del estudio, pues se puede producir información redundante, mientras que por otro si se desea interpolar lugares no muestreados dentro del lote, se requiere dependencia espacial (Ortega, 2015).

Un plan óptimo de muestreo por tanto debe representar la dependencia espacial con el fin de poder interpolar los lugares no muestreados utilizando la menor cantidad de muestras posibles.

Conceptos generales sobre el modelamiento geoestadístico de datos

Como fue mencionado anteriormente, la existencia de dependencia espacial entre las observaciones ubicadas en el espacio geográfico conlleva a que las hipótesis de la estadística clásica, donde se consideran los datos como realizaciones independientes, sean poco realistas. Por lo tanto, en general, no puede considerarse modelar las variables regionalizadas por medio de funciones determinísticas, debido a su complejidad, ni como variables aleatorias independientes.

La geoestadística entrega una representación intermedia, a la vez aleatoria y funcional, al basarse en el concepto de función aleatoria. Los modelos geoestadísticos consideran el valor $z(x)$ de la variable regionalizada en un sitio x del campo D como una representación de una variable aleatoria $Z(x)$ ³. Cuando x recorre D , se obtiene un conjunto de variables aleatorias $Z = \{Z(x), x \in D\}$ que constituye función aleatoria. Así pues, la variable regionalizada $z = \{z(x), x \in D\}$ es una realización

³Las variables determinísticas serán denotadas con minúscula y las aleatorias con mayúscula

de la función aleatoria Z (Emery, 2007).

El recurrir al concepto de función aleatoria es una decisión, ni verdadera ni falsa, pues dicha función aleatoria es un objeto teórico que no existe en la realidad. Es importante mencionar que pueden existir varios modelos aceptables (mientras sean compatibles con la información disponible sobre la variable regionalizada), lo que hace que determinar una función aleatoria a partir de una variable regionalizada no sea una operación única o unívoca (Emery, 2007).

Si se considera una función aleatoria $Z = \{Z(x), x \in D\}$ y una serie de sitios $\{x_1, \dots, x_k\}$ en D , el grupo de variables aleatorias $\{z(x_1), \dots, z(x_k)\}$ se puede caracterizar por una función de distribución aleatoria que depende de k argumentos:

$$F_{x_1, \dots, x_k}(z_1, \dots, z_k) = \text{Prob} \{Z(x_1) < z_1, Z(x_k) < z_k\} \forall z_1, z_k \in \mathbb{R}$$

El conjunto de funciones de distribución, para todos los enteros k y todas las elecciones posibles de $\{x_1, \dots, x_k\}$ en D , constituye distribución espacial de la función aleatoria. En general se utilizan las simplificaciones donde $k = 1$ y $k = 2$.

Donde distribución univariable:

$$F_{x_1}(z_1) = \text{Prob} \{Z(x_1) < z_1\}$$

Y distribución bivariable:

$$F_{x_1, x_2}(z_1, z_2) = \text{Prob} \{Z(x_1) < z_1, Z(x_2) < z_2\}$$

Momentos

Existe la posibilidad de simplificar aún más la caracterización de la función aleatoria al considerar sólo algunos parámetros descriptivos o momentos de las distribuciones univariables y bivriables que reflejan la información esperada. (Gutierrez-Lopez et al., 2011). Estos son:

- El valor esperado (esperanza, o momento de primer orden):

$$m(x) = E[Z(x)]$$

En cada sitio x dado, $m(x)$ representa la media alrededor de la cual se distribuyen los valores tomados por las realizaciones de la función aleatoria

- La varianza:

$$\begin{aligned}\sigma^2(x) &= \text{var}[Z(x)] \\ &= E\{[Z(x) - m(x)]^2\} \\ &= E[Z(x)^2] - m(x)^2\end{aligned}$$

La varianza corresponde a una cantidad positiva, su raíz cuadrada se llama desviación estándar y ambas constituyen medidas de la dispersión de $Z(x)$ en torno a su valor medio $m(x)$ y cuantifican su carácter aleatorio.

- La covarianza centrada entre dos variables aleatorias $Z(x_1)$ y $Z(x_2)$:

$$\begin{aligned}C(x_1, x_2) &= \text{cov}[Z(x_1), Z(x_2)] \\ &= E\{[Z(x_1) - m(x_1)][Z(x_2) - m(x_2)]\} \\ &= E[Z(x_1), Z(x_2)] - m(x_1)m(x_2)\end{aligned}$$

La covarianza da una visión elemental del vínculo que existe entre $Z(x_1)$ y $Z(x_2)$. La relación entre varianza y covarianza se expresa en la desigualdad de Cauchy-Schwarz:

$$|\text{cov}[Z(x_1), Z(x_2)]| \leq \sqrt{\text{var}[Z(x_1)]\text{var}[Z(x_2)]}$$

- El correlograma (coeficiente de correlación lineal) entre dos variables aleatorias $Z(x_1)$ y $Z(x_2)$:

$$\begin{aligned}\rho(x_1, x_2) &= \text{corr}[Z(x_1), Z(x_2)] \\ &= \frac{\text{cov}[Z(x_1), Z(x_2)]}{\sqrt{\text{var}[Z(x_1)]\text{var}[Z(x_2)]}}\end{aligned}$$

El correlograma es adimensional y toma sus valores en el intervalo $[-1, 1]$. Un coeficiente nulo indica independencia y un coeficiente igual a 1 o -1 indica que las variables son proporcionales.

- El semi-variograma o variograma entre dos variables aleatorias $Z(x_1)$ y $Z(x_2)$:

$$\gamma(x_1, x_2) = \frac{1}{2}\text{var}[Z(x_1) - Z(x_2)]$$

Hipótesis de estacionaridad

Existen dos razones por las cuales no es posible determinar, por lo menos parcialmente, la distribución espacial de la función aleatoria a través de los datos disponibles (etapa de inferencia estadística). La primera razón es que la variable regionalizada es sólo una realización de la función aleatoria y la segunda es que esta realización solo se conoce de manera fragmentaria en algunos sitios de muestreo (Emery, 2007).

La noción de estacionaridad permite sortear estos problemas. Esta consiste en permitir la inferencia estadística reemplazando la repetición sobre las realizaciones de la función aleatoria, por una repetición en el espacio. Esta repetición supone que los valores que se encuentran en las diferentes regiones del campo presentan las mismas características y por lo tanto pueden considerarse como distintas realizaciones del mismo proceso aleatorio. Matemáticamente, la distribución espacial de la función aleatoria es invariante por traslación, es decir, las propiedades de un conjunto de datos no dependen de su posición absoluta en el espacio sino que solamente de sus posiciones relativas (Emery, 2007).

La hipótesis de estacionaridad permite hacer las siguientes simplificaciones:

- La distribución univariable no depende del sitio considerado:

$$F(z_1) = Prob \{Z(x_1) < z_1\}$$

En particular la esperanza y la varianza son constantes en el espacio:

$$m = E[Z(x)]$$

$$\sigma^2 = var[Z(x)]$$

- La distribución bivariable solo depende de la separación de los sitios considerados:

$$F_h(z_1, z_2) = Prob \{Z(x+h) < z_1, Z(x) < z_2\}$$

Lo mismo ocurre para los momentos de esta distribución (covarianza, correlograma y vario-grama):

$$C(h) = cov[Z(x+h), Z(x)]$$

$$\rho(h) = \text{corr}[Z(x+h), Z(x)]$$

$$\gamma(h) = \frac{1}{2} \text{var}[Z(x+h) - Z(x)]$$

Estos parámetros muestran la interacción existente entre dos valores, dando una descripción de la continuidad espacial de la variable regionalizada. La covarianza y el correlograma indican qué tan semejantes son los valores entre dos sitios y el variograma a su vez indica que tan desemejantes son. Si la esperanza es constante, se tiene también:

$$\gamma(h) = \frac{1}{2} E \{ [z(x+h) - z(x)]^2 \}$$

Bajo la hipótesis de estacionaridad, se tienen las siguientes relaciones:

- La varianza es igual a la función de covarianza evaluada para el vector $h = 0$:

$$\sigma^2 = C(0)$$

- El correlograma es igual a la covarianza dividida por la varianza:

$$\rho(h) = C(h)/C(0)$$

- El variograma es igual a la varianza menos la covarianza::

$$\gamma(h) = C(0) - C(h)$$

- Cuando la norma del vector de separación K se vuelve infinita, la covarianza tiende a 0 y el variograma es igual a la varianza:

$$\gamma(\infty) = C(0) = \sigma^2$$

Resumiendo, una variable regionalizada se interpreta a través de una función aleatoria, la cual es caracterizada mediante una distribución espacial, esta distribución es resumida mediante sus momentos, los cuales se simplifican bajo la hipótesis de estacionaridad. Los conceptos mencionados anteriormente son conceptos claves y necesarios a entender y considerar si se desea modelar variables dependientes en el espacio geográfico y realizar una optimización del modelo de muestreo.

Análisis variográfico

La dependencia espacial de un predio es usualmente cuantificada a través del uso de variogramas. Esta es una relación entre variancia y distancia entre muestras. La optimización del muestreo espacial en un contexto geoestadístico primero requiere la estimación de un modelo para expresar la dependencia espacial a diferentes pares de distancias. Esto es resumido en la función del variograma (o covariograma) ([Santacruz Delgado, 2010](#)).

La continuidad espacial de las observaciones queda reflejada por medio del variograma, en el cual se observan los cambios de la tasa de crecimiento de la variancia ($\hat{\gamma}$) de acuerdo al aumento de la distancia de muestreo (h).

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} [z(x_\alpha) - z(x_\beta)]^2$$

Donde $N(h) = \{(\alpha, \beta) \text{ tal que } x_\alpha - x_\beta = h\}$

$|N(h)|$ es el número de pares contenidos en el conjunto $N(h)$

Un variograma típico tiene tres componentes (Figura 2.4):

- Sill: Frecuentemente, el variograma crece a partir del origen y se estabiliza, a partir de una distancia h , en torno a una meseta (sill) ([Emery, 2007](#))
- Rango: También conocido como alcance, corresponde a la noción intuitiva de zona de influencia. Más allá de esta distancia, el variograma es constante e igual a su meseta y sus muestras son independientes (en rigor, no están correlacionadas) ([Emery, 2007](#))
- Nugget (efecto pepita): corresponde a la variancia entre muestras cuando la distancia entre ellas es cero y que correspondería al error de medición ([Ortega, 2013](#))

Si los datos están irregularmente distribuidos en el campo D , el número de pares $|N(h)|$ que intervienen para calcular el variograma experimental es muy pequeño. Por lo tanto, para evitar un variograma experimental con un aspecto errático e imposible de modelar e interpretar se suele permitir algunas tolerancias, esto con el fin de hacerlo más robusto. En la dirección horizontal, para cada h se define una tolerancia Δh y se utilizan solo las observaciones que están a una distancia $[h - \Delta h, h + \Delta h]$.

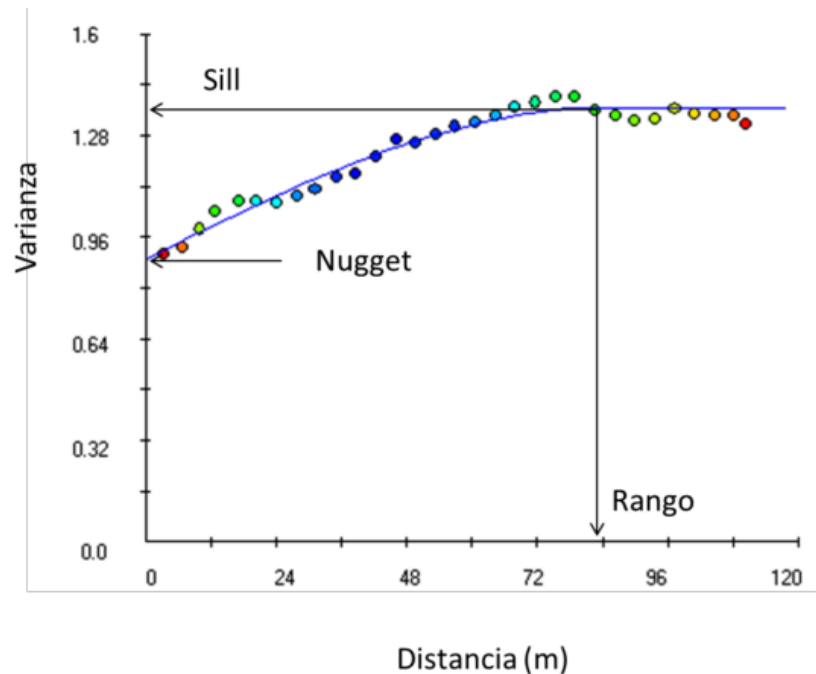


Figura 2.4: Variograma típico sobre el cual se ajustó un modelo esférico
(Fuente: [Ortega, 2013](#))

Anisotropía

A menudo la correlación espacial entre puntos del espacio no es igual en todas las direcciones. Cuando ocurre lo anterior, se tiene un proceso anisotrópico. Como los modelos elementales de variograma están definidos para el caso isótropo, se examinar las transformaciones que permiten el paso a modelos anisótropos. Se distingue varios tipos de anisotropía, en especial, la anisotropía geométrica y la anisotropía zonal ([Hernández y Corvalán, 2000](#)).

Una primera aproximación para distinguir si existe o no anisotropía en el variograma experimental correspondería a la realización de un mapa variográfico. Este consiste en visualizar el variograma experimental en todas las direcciones del espacio, bajo la forma de un mapa con escala de colores o de grises. Cuando hay isotropía, los variogramas experimentales en las diferentes direcciones se superponen y el mapa variográfico dibuja círculos (caso bidimensional) o esferas (caso tridimensional). En el caso contrario, se está en presencia de una anisotropía. Se distingue varios tipos de anisotropía, en especial, la anisotropía geométrica y la anisotropía zona. ([Emery, 2007](#)).

Anisotropía geométrica

En este caso se dibujarán elipses concéntricas (caso 2D) (ver Figura 2.5) o elipsoides (caso 3D) en el mapa variográfico. Para el caso bidimensional, la anisotropía se define por un coeficiente de anisotropía a/b , razón entre el eje menor a y el eje mayor b de cualquiera de las elipses de isovalores (refleja la intensidad de la anisotropía), y el ángulo de anisotropía θ , formado por el eje mayor y el eje de las abscisas (indica orientación de la anisotropía).

Este tipo de anisotropía se puede escribir:

$$\gamma(h) = \gamma_0(|Ah|)$$

donde $\gamma_0(|Ah|)$ es un variograma isotrópico y A una matriz de dimensión $d \times d$ representando una transformación lineal de R^d . La idea subyacente es esta situación es que el proceso no es isotrópico en el espacio original, pero sí en algún espacio transformado linealmente, el cual puede, por ejemplo, corresponder a una transformación de las coordenadas (Ginzo Villamayor, 2009).

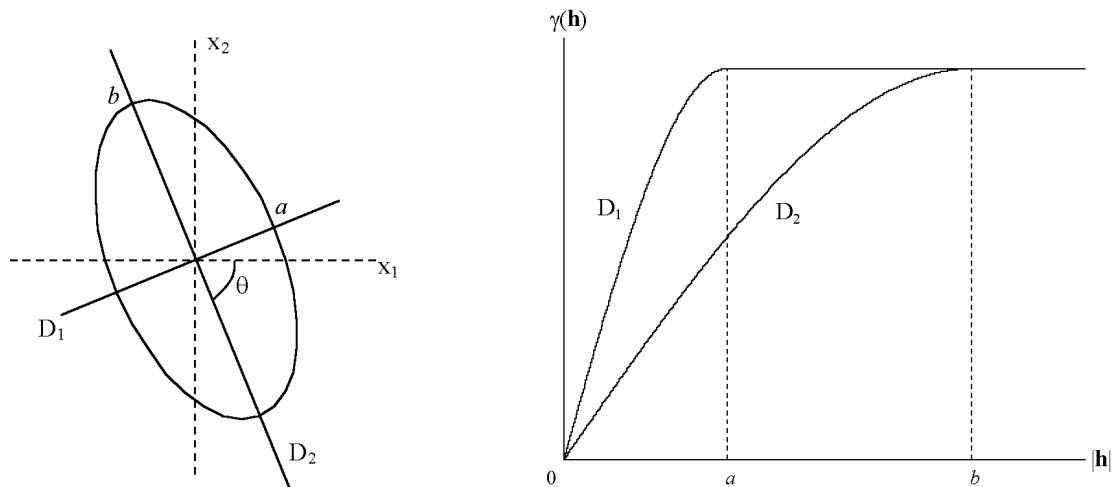


Figura 2.5: Ejemplo (2D) de anisotropía geométrica de razón a/b y ángulo θ . Izquierda: elipse de alcances. Derecha: variograma en las direcciones principales de anisotropía (Fuente: Emery, 2007)

Anisotropía zonal

La anisotropía zonal es un modelo en el cual la función aleatoria no depende de una (o varias) coordenada(s). Esto sucede a veces cuando se trabaja en el espacio de tres dimensiones. Como una primera aproximación, la función aleatoria es constante en el plano horizontal, de manera que su

variograma $\gamma(h)$ sólo depende de la componente vertical de h . Denotando como θ el ángulo entre h y la dirección vertical y u un vector unitario orientado verticalmente (ver Figura 2.6), se puede escribir (Emery, 2007):

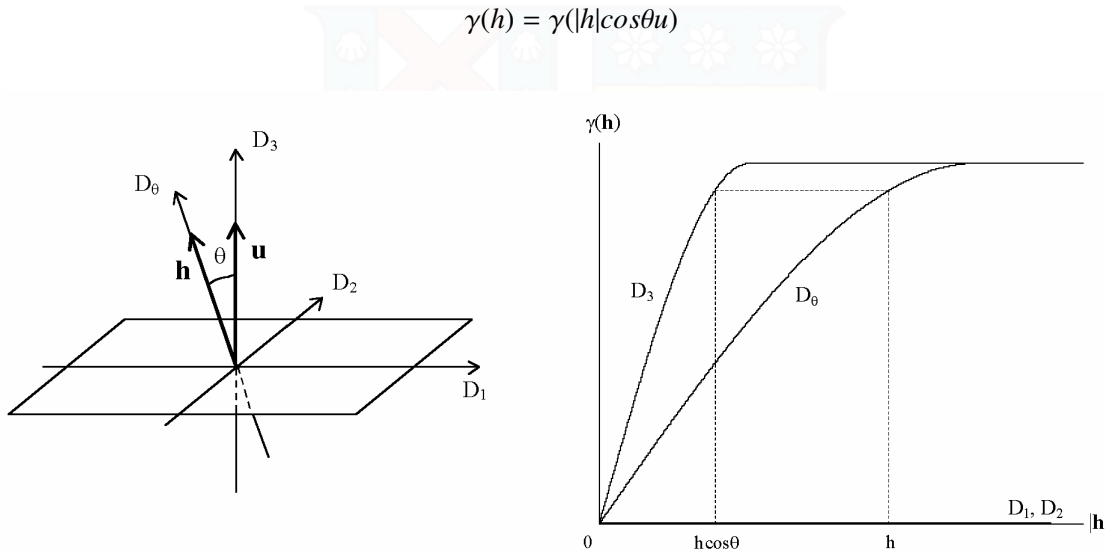


Figura 2.6: Ejemplo (3D) de anisotropía zonal en dirección vertical. Derecha: variogramas en las direcciones horizontales (D_1, D_2), vertical (D_3) y una dirección oblicua (D_θ) (Fuente: Emery, 2007)

Modelos elementales de variograma

Una vez que se ha obtenido el variograma experimental y se ha estudiado su comportamiento, el paso siguiente es encontrar algún modelo paramétrico que ajuste adecuadamente los datos muestrales (Hernández y Corvalán, 2000). El ajuste debe realizarse debido que el variograma experimental se encuentra incompleto, esto debido a que sólo está definido para ciertas distancias y direcciones y está sujeto a aproximaciones (debido al número limitado de datos y a los parámetros de tolerancia utilizado en el cálculo) (Emery, 2007). A continuación se presentan los modelos de variogramas más utilizados:

- Efecto pepita: El variograma pepítico de meseta C se define como:

$$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C & \text{en caso contrario} \end{cases}$$

Este modelo de variograma alcanza inmediatamente su meseta: hay ausencia total de corre-

lación espacial, de modo que la geoestadística encuentra todos los resultados de la estadística clásica.

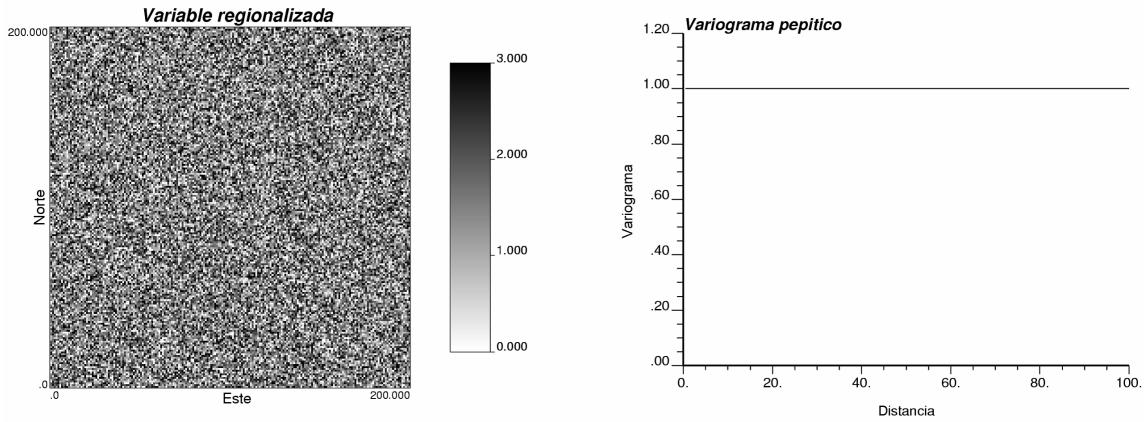


Figura 2.7: Variograma pepítico (derecha) y ejemplo de variable regionalizada asociada (izquierda)
(Fuente: [Emery, 2007](#))

- Modelo esférico: El variograma esférico de alcance a y meseta C se define como:

$$\gamma(h) = \begin{cases} C \left\{ \frac{2}{3} \frac{|h|}{a} - \frac{1}{2} \left(\frac{|h|}{a} \right)^3 \right\} \\ C \end{cases} \quad \text{en caso contrario}$$

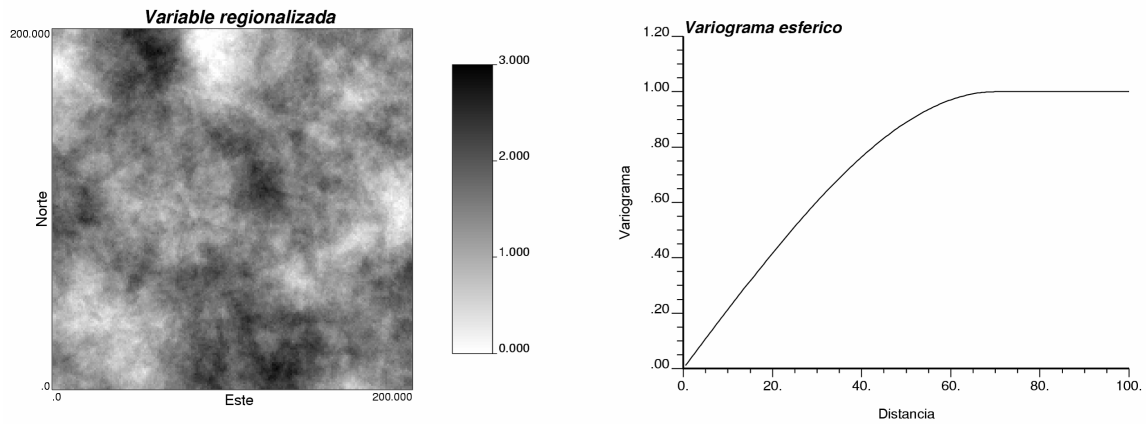


Figura 2.8: Variograma esférico (derecha) y ejemplo de variable regionalizada asociada (izquierda)
(Fuente: [Emery, 2007](#))

- Modelo exponencial: . El variograma exponencial de alcance a y meseta C se define como:

$$\gamma(h) = C \left\{ 1 - \exp \left(-\frac{|h|}{a} \right) \right\}$$

Contrariamente al modelo esférico que llega a la meseta exacta para $|h| = a$, el modelo es asintótico. En todo caso, se puede considerar un alcance práctico igual a $3a$ para el cual el variograma llega al 95 % del valor de su meseta

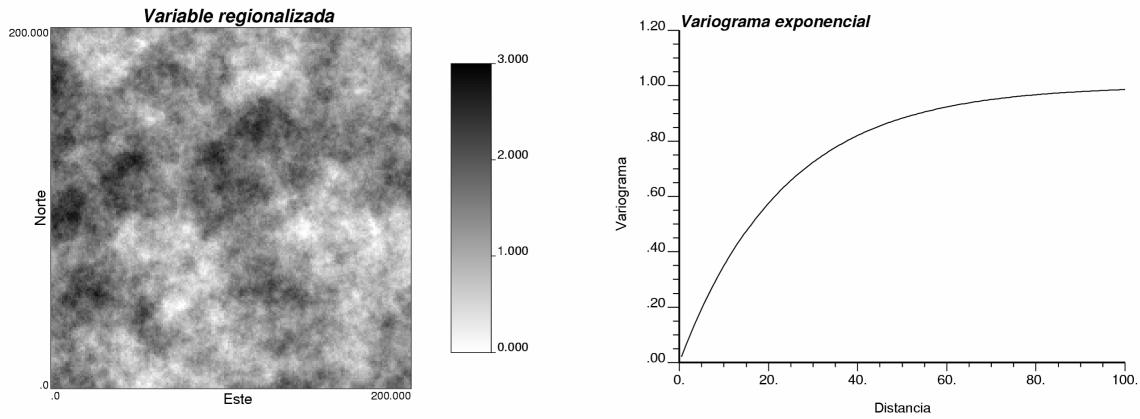


Figura 2.9: Variograma exponencial (derecha) y ejemplo de variable regionalizada asociada (izquierda)
(Fuente: [Emery, 2007](#))

- Modelo Gaussiano: El variograma Gaussiano de alcance a y meseta C se define como:

$$\gamma(h) = C \left\{ 1 - \exp\left(-\frac{|h|^2}{a^2}\right) \right\}$$

La meseta se alcanza asintóticamente y el alcance práctico puede considerarse igual a $\sqrt{3}$.

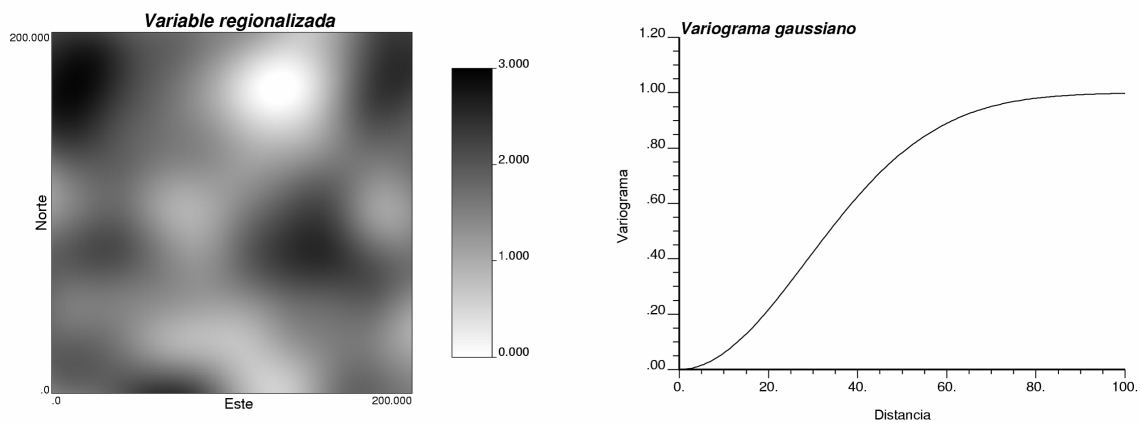


Figura 2.10: Variograma Gaussiano (derecha) y ejemplo de variable regionalizada asociada (izquierda)
(Fuente: [Emery, 2007](#))

- Modelo seno cardinal: El variograma seno cardinal de alcance a y meseta C se define como:

$$\gamma(h) = C \left\{ 1 - \frac{a}{|h|} \operatorname{sen} \left(\frac{|h|}{a} \right) \right\}$$

Este modelo presenta oscilaciones, las cuales tienen generalmente una interpretación física (fenómeno periódico amortiguado, por ejemplo). El alcance práctico vale $20.37 D$ y la mitad del pseudo-período es igual a $4.49D$, distancia para la cual el variograma vale $1.21 C$ (Emery, 2007).

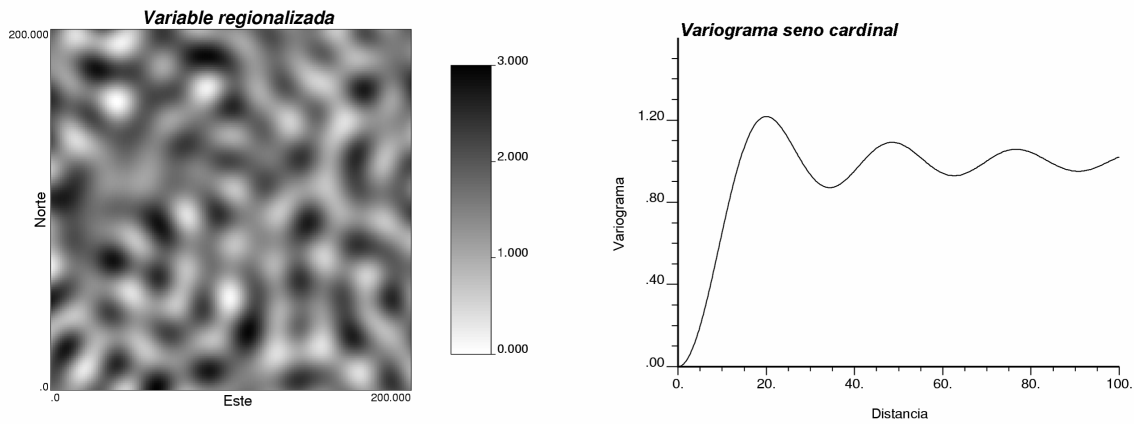


Figura 2.11: Variograma seno cardinal (derecha) y ejemplo de variable regionalizada asociada (izquierda) (Fuente:Emery, 2007)

- Modelo potencia El variograma potencia de pendiente ω y exponente θ se define como:

$$\gamma(h) = \omega |h|^\theta$$

Este variograma no posee ni meseta ni alcance, sino que crece en forma indefinida. El exponente θ puede variar entre 0 (variograma péptico) y 2 (variograma parabólico). El modelo se llama lineal cuando el exponente vale 1.

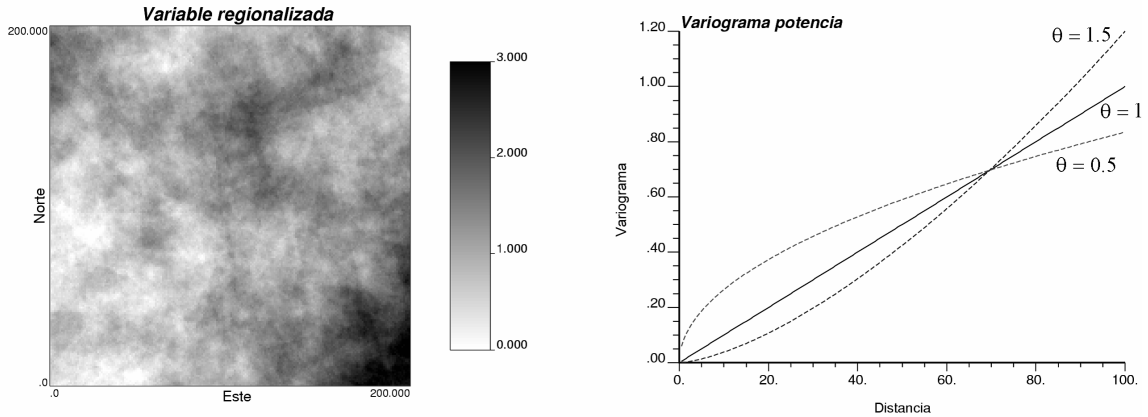


Figura 2.12: Variograma potencia (derecha) y ejemplo de variable regionalizada asociada a un variograma lineal (izquierda)
(Fuente: [Emery, 2007](#))

Kriging: predicción e interpolación

El muestreo y posterior análisis de los datos se realiza con el fin de poder predecir los valores del predio. Mediante la estimación local se busca predecir el valor de la variable regionalizada en un sector del campo estudiado, a partir de los datos disponibles (problema de interpolación espacial). El kriging es el método óptimo en el sentido estadístico de interpolación y extrapolación ([Gutierrez-Lopez et al., 2011](#)).

La resolución de un problema de estimación por kriging se articula siempre en torno a las mismas etapas. Las diferentes variantes sólo radican en las hipótesis realizadas sobre la función $Z = \{Z(x), x \in D\}$ que representa la variable regionalizada.

La primera etapa corresponde a la restricción de linealidad. Esta se refiere a que el estimador debe ser una combinación lineal ponderada de los datos:

$$Z^*(x_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha})$$

Donde x_0 es el sitio donde se busca tener una estimación, $\{x_{\alpha}, \alpha = 1, n\}$ son los sitios con datos, mientras que los ponderadores $\{\lambda_{\alpha}, \alpha = 1, n\}$ y el coeficiente a son las incógnitas del problema de kriging.

La segunda etapa corresponde a la restricción de insesgo. Esta consiste en expresar que el error de estimación tiene esperanza nula, es decir:

$$E[Z^*(x_0) - Z(x_0)] = 0$$

Luego de estas etapas el estimador se encuentra sometido a una o varias restricciones, pero no está totalmente especificado. La última etapa corresponde a buscar los ponderadores que minimicen la varianza del error de estimación:

$$Var[Z^*(x_0) - Z(x_0)] \text{ es mínima}$$

Es importante mencionar que los problemas de estimación local generalmente no involucran la totalidad del campo ni tampoco utilizan todos los datos disponibles ya que esto puede significar aumentar considerablemente los tiempos de cálculo. Con el fin de reducir este tiempo se define la vecindad del kriging, esta corresponde al dominio del espacio que contiene el sitio a estimar y los datos utilizados en esta estimación (Emery, 2007). Se pueden considerar dos tipos de vecindad:

- Vecindad única: se utilizan todos los datos e incluso datos muy alejados intervienen en la estimación.
- Vecindad móvil: sólo se utilizan los datos cercanos al sitio a estimar. El tamaño de la vecindad debe permitir un equilibrio entre varios factores (precisión, tiempos de cálculo, confiabilidad del modelo de variograma para grandes distancias, cambios en la continuidad espacial de la variable regionalizada) y su forma debe tomar en cuenta la anisotropía de la variable regionalizada.

A continuación se describirán los modelos de kriging más utilizados:

Kriging de media conocida (Kriging Simple) (SK)

Se supone que la variable regionalizada z es la realización de una función aleatoria Z estacionaria tal que:

$$\begin{cases} \forall x \in V, & E[Z(x)] \text{ conocida} \\ \forall x, x+h \in V, & cov[Z(x+h), Z(x)] = C(h) \end{cases}$$

Donde V representa a la vecindad del kriging.

Siguiendo las etapas antes mencionadas:

- Linealidad: se asegura esta restricción al tomar como estimador en x_0

$$Z^*(x_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha})$$

- Insésigo: el valor esperado del error de la estimación es:

$$E[Z^*(x_0) - Z(x_0)] = a + \sum_{\alpha=1}^n \lambda_{\alpha} E[Z(x_{\alpha})] - E[Z(x_0)] = a + \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right) m$$

Este valor esperado es nulo si:

$$a = \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right) m$$

- Optimalidad: se debe calcular la varianza del error de estimación:

$$Var[Z^*(x_0) - Z(x_0)] = Var \left\{ \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}) - Z(x_0) \right\}$$

Como la constante a no influye en la varianza, se omite de la expresión. Utilizando la siguiente regla de cálculo:

$$Var \left\{ \sum_{k=1}^K \lambda_k Z_k \right\} = \sum_{k=1}^K \sum_{k'=1}^K \lambda_k \lambda_{k'} Cov \{Z_k, Z_{k'}\}$$

Se obtiene:

$$Var[Z^*(x_0) - Z(x_0)] = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} Cov(x_{\alpha} - x_{\beta})$$

El mínimo de esta expresión se obtiene igualando sus derivadas parciales a cero con respecto a las incógnitas $\{\lambda_{\alpha}, \alpha = 1, n\}$. De donde se obtiene el sistema de ecuaciones:

$$\sum_{\beta=1}^n \lambda_{\beta} Cov(x_{\alpha} - x_{\beta}) = Cov(x_{\alpha} - x_0) \quad \forall \alpha = 1, n$$

Permitiendo determinar los ponderadores de kriging $\{\lambda_{\alpha}, \alpha = 1, n\}$. Debido a la condición de insésigo, el estimador se pone de la forma:

$$Z^*(x_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}) + a = \left(1 - \sum_{\alpha=1}^n \lambda_{\alpha} \right) m$$

De modo que el valor de la media aparece como si fuera un dato adicional, al cual se le asigna una

ponderación igual al complemento de la ponderación acumulada de otros datos. El rol de la media es de compensar la falta de información cuando los datos son escasos o alejados (Emery, 2007). La varianza de la predicción será:

$$\sigma_{SK}^2(x_0) = \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} \text{Cov}(x_{\alpha} - x_0)$$

Kriging de media desconocida (Kriging Ordinario) (OK)

Se supone que la variable regionalizada z es la realización de una función aleatoria Z estacionaria tal que:

$$\begin{cases} \forall x \in V, & E[Z(x)] \text{ desconocida} \\ \forall x, x+h \in V & \text{cov}[Z(x+h), Z(x)] = C(h) \end{cases}$$

Donde V representa a la vecindad del kriging. Siguiendo las etapas antes mencionadas:

- Linealidad: se asegura esta restricción al tomar como estimador en x_0

$$Z^*(x_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha})$$

- Insesgo: el valor esperado del error de la estimación es:

$$E[Z^*(x_0) - Z(x_0)] = a + \sum_{\alpha=1}^n \lambda_{\alpha} E[Z(x_{\alpha})] - E[Z(x_0)] = a + \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right) m$$

Este valor esperado es nulo si:

$$a = 0 \text{ y } \sum_{\alpha=1}^n \lambda_{\alpha} = 1$$

- Optimalidad: como en el caso del kriging simple, la varianza del error de estimación es:

$$\text{Var}[Z^*(x_0) - Z(x_0)] = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \text{Cov}(x_{\alpha} - x_{\beta}) + \text{Cov}(0) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \text{Cov}(x_{\alpha} - x_0)$$

Para minimizar esta expresión bajo la condición de insesgo se introduce una incógnita adi-

cional llamada multiplicador de Lagrange (μ):

$$\begin{aligned} Var[Z^*(x_0) - Z(x_0)] = & \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} Cov(x_{\alpha} - x_{\beta}) + Cov(0) \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right) \\ & - 2 \sum_{\alpha=1}^n \lambda_{\alpha} Cov(x_{\alpha} - x_0) + 2\mu \end{aligned}$$

Y se minimiza la función de las $n + 1$ variables $\lambda_1, \dots, \lambda_n, \mu$. Calculando las $n + 1$ derivadas parciales e igualando a cero, se obtiene el sistema:

$$\begin{cases} \frac{\delta}{\delta \lambda_{\alpha}} = 0 : \sum_{\beta=1}^n \lambda_{\beta} Cov(x_{\alpha} - x_{\beta}) + \mu = Cov(x_{\alpha} - x_0) \quad \forall \alpha = 1, n \\ \frac{\delta}{\delta \mu} = 0 : \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \end{cases}$$

Siendo el variograma una herramienta equivalente a la covarianza las ecuaciones pasan a ser:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} Cov(x_{\alpha} - x_{\beta}) + \mu = \gamma(x_{\alpha} - x_0) \quad \forall \alpha = 1, n \\ \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \end{cases}$$

La varianza de la predicción será:

$$\begin{aligned} \sigma_{OK}^2(x_0) &= \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} Cov(x_{\alpha} - x_0) - \mu \\ &= \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(x_{\alpha} - x_0) - \mu \end{aligned}$$

Otros

Existen otros tipos de kriging utilizados en diversas situaciones, estos corresponden a:

- Kriging con derivas: se sustenta en la hipótesis que la función aleatoria ya no es estacionaria y su valor esperado varía en el espacio, reflejando una tendencia sistemática en la distribución espacial de los valores.
- Kriging de bloques: consiste en estimar directamente el valor promedio de la variable sobre un soporte mayor que el soporte de los datos (bloque)
- Co-Kriging: corresponde a la versión multivariable del kriging donde se busca estimar el

valor de una variable tomando en cuenta los datos de esta variable y de otras variables correlacionadas.

En un problema de optimización de muestreo el uso del kriging como método de estimación resulta muy útil ya que permite realizar mapas de curvas de nivel a través de la estimación de valores de sitios de los cuales no se posee información, mediante puntos, ya sea regular o irregularmente distribuidos, sin sesgo y con una varianza mínima conocida.

Medidas de autocorrelación

Como fue mencionado anteriormente, el modelamiento de una variable regionalizada no es considerado como un proceso unívoco, por lo que el uso de ciertas medidas de correlación puede ser útil al momento comparar distintos modelos de optimización.

Para representar los contrastes de dependencia o autocorrelación se puede utilizar una noción de vecindad binaria entre dos unidades espaciales. Es decir, si dos unidades espaciales tienen una frontera común de longitud no nula, se considera que son contiguas y se les asigna el valor 1 (Yrigoyen, 2003).

Cuando los datos están regularmente determinados la vecindad no es única, es decir se pueden considerar diversas formas para obtener una frontera común. Por ejemplo, se puede utilizar un borde en común para la obtención de fronteras, como también se puede utilizar los vértices para clasificar qué datos corresponden a contiguos, así como también generar una red que contempla tanto los vértices como los bordes en común. Para estos tres casos se hace una analogía con el juego de ajedrez, siendo el criterio de la torre, el criterio del alfil y el criterio de la reina, respetivamente (Ledo, 2013).

De forma similar, pueden considerarse varios órdenes de contigüidad de forma recursiva, definiendo el orden de contigüidad k -ésimo cuando las unidades espaciales son de primer orden contiguas a una unidad espacial de orden k -ésimo, y no contiguas de orden inferior. Por ejemplo, en la figura 2.13, las celdas c y d son contiguas de segundo orden con a siguiendo el criterio de la torre, siendo también contiguas de primer orden con b (Yrigoyen, 2003).

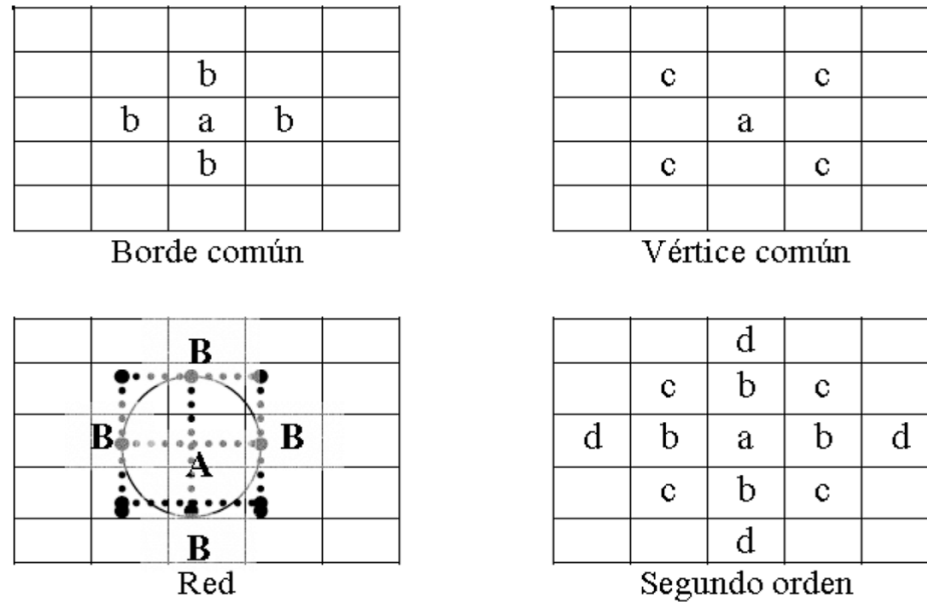


Figura 2.13: Contigüidad en un entramado regular
(Fuente: [Anselin, 1988](#))

La estructura espacial se puede representar mediante una matriz de interacciones espaciales. Para esta matriz cada unidad espacial es representada mediante una fila y columna a la vez. Al utilizar el criterio de la torre, en cada fila se representará con el número 1 las observaciones que cumplan con el criterio de contigüidad y con el valor cero en caso contrario. Por ejemplo, para el dato A, según el criterio de la torre, los datos que tienen contigüidad son B y D, entonces para la fila A solo tendrá valores 1 para B y D el resto estará con valores 0, como se muestra en la figura 2.14 continuación ([Ledo, 2013](#)).

	1	2	3						
	4	5	6						
	7	8	9						

	1	2	3	4	5	6	7	8	9
1	0	1	0	1	0	0	0	0	0
2	1	0	1	0	1	0	0	0	0
3	0	1	0	0	0	1	0	0	0
4	1	0	0	0	1	0	1	0	0
5	0	1	0	1	0	1	0	1	0
6	0	0	1	0	1	0	0	0	1
7	0	0	0	1	0	0	0	1	0
8	0	0	0	0	1	0	1	0	1
9	0	0	0	0	0	1	0	1	0

Figura 2.14: Matriz de interacciones para nueve datos situados en una cuadrícula
(Fuente: [Yrigoyen, 2003](#))

Un problema de la matriz de interacciones es que no permite incorporar las influencias no recíprocas (no siempre la influencia que ejerce a sobre b es la misma que recibe a por parte de b), además registra sólo contactos físicos, no puede considerar influencias entre dos observaciones muy lejanas geográficamente (Martori et al., 2008).

Para superar estas simplificaciones se utiliza la matriz de pesos especiales, que permite al investigador elegir el conjunto de ponderadores que encuentre más apropiado para cada fenómeno, logrando una mayor flexibilidad en la definición de la estructura de independencias (Yrigoyen, 2003).

En la matriz de pesos (W), en vez de utilizar ponderaciones binarias, se recoge el efecto del dato i sobre j por medio de un peso o ponderación w_{ij} , de forma que $W = [w_{ij}]$, tiene la siguiente interpretación:

$w_{ij} = 0$ indicaría ausencia de autocorrelación espacial entre las observaciones i e j

$w_{ij} \neq 0$ indicara presencia de una interacción espacial entre las observaciones

I de Moran

Uno de los índices más utilizados para medir la correlación de manera global corresponde al I de Moran:

$$I = \frac{N}{S_0} \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}$$

Donde:

w_{ij} : Elemento de la matriz de pesos espaciales correspondientes al par (i, j)

$S_0 = \sum_i \sum_j w_{ij}$, es decir, la suma de los pesos especiales

$\bar{y}$: Valor medio o esperado de la variable

N : número de observaciones o tamaño muestral

El valor esperado de I es negativo y función únicamente del tamaño de la muestra (N), no es equivalente al coeficiente clásico de correlación, porque no está centrado en el valor cero, sino que su media teórica se encuentra en $-1/(N - 1)$. Un coeficiente I de Moran mayor que su valor esperado sería indicativo de autocorrelación espacial positiva, mientras que un valor de I inferior a la media pondría de manifiesto la existencia de autocorrelación espacial negativa. Como puede apreciarse, este estadístico estará muy afectado por aquellos puntos vecinos sensiblemente distintos de la media de la variable en estudio (Yrigoyen, 2003).

C de Geary

El test c de geary es similar al test I de moran, donde:

$$I = \frac{(N-1)}{2S_0} \frac{\sum_i \sum_j w_{ij}(y_i - y_j)}{\sum_i (y_i - \bar{y})^2}$$

La diferencia que tiene con respecto al I de Moran es que Geary considera la diferencia de los valores Y en dos puntos, por lo que el valor esperado teórico para el test es 1 (Ledo, 2013).

El uso de índices de correlación es una herramienta útil no sólo para determinar si existe correlación estadísticamente significativa en la muestra, si no, facilita la comparación de modelos en el caso de realizar una optimización del muestreo.

Distancia Inversa Ponderada (IDW): predicción e interpolación

Existe además un método de interpolación geométrica conocido como la Distancia Inversa ponderada o IDW por sus siglas en inglés (Inverse Distance Weighting). Este método a diferencia del Kriging Ordinario, que es método geoestadístico y analítico y que tiene en cuenta la autocorrelación espacial de la variable a interpolar, es determinista, exacto y local. Al igual que el OK el valor de la variable Z en la localización sin muestreos x_0 , $Z^*(x_0)$ es estimado en base a los datos de las localizaciones aledañas, $Z(x_\alpha)$, como (Sun et al., 2013; Torrecilla et al., 2010):

$$Z^*(x_0) = a + \sum_{\alpha=1}^n \lambda_\alpha Z(x_\alpha)$$

Donde λ_α es el peso asignado a cada valor $Z(x_\alpha)$ y n es el número de vecinos cercanos muestreados utilizados para la estimación. Lo que diferencia a ambos métodos es el procedimiento de cálculo de los pesos asignados a los valores muestrales. En el caso del IDW los pesos se determinan mediante la fórmula:

$$\lambda_\alpha = \frac{d_{\alpha 0}^{-p}}{\sum_{\alpha=0}^n d_{\alpha 0}^{-p}}$$

donde $d_{\alpha 0}$ es la distancia entre el lugar de predicción x_0 y cada lugar muestral x_α , a medida que la distancia se hace más grande, el peso es reducido por un factor p (Torrecilla et al., 2010). Los pesos asignados en IDW son usualmente proporcionales al cuadrado de la distancia inversa entre el punto de predicción y cada lugar muestral, con lo que p se iguala a 2 (Sun et al., 2013).

Validación del modelo de predicción

Los modelos de predicción pueden ser evaluados según el error en sus predicciones al comparar la predicción del modelo con el valor observado de la muestra. Uno de los métodos para realizar esta evaluación corresponde a la raíz del error cuadrático medio (RMSE por sus siglas en inglés, root square mean standar error). Como los errores son elevados al cuadrado antes de ser ponderados, el RMSE le da un mayor peso a los errores más grandes.

$$RMSE = \left[\frac{1}{n} \sum_{\alpha=1}^n (p_{\alpha} - o_{\alpha})^2 \right]^{\frac{1}{2}}$$

Donde p_{α} corresponde a la predicción en la localización α y o_{α} al valor observado en la misma localización (Sun et al., 2013).

Materiales y métodos

Área de estudio

El terreno estudiado se encuentra en la comuna y provincia de San Felipe, en la quinta región de Valparaíso. Corresponde a una plantación de uva de mesa (*Vitis vinifera* L, variedad Red Globe).



Figura 3.1: Área de estudio
(Fuente: Elaboración propia)

La variable estudiada corresponde al Índice de vegetación de diferencia normalizada, también conocido como Ndvi por sus siglas en inglés. Este es un índice usado para estimar la cantidad, calidad y desarrollo de la vegetación con base a la medición, por medio de sensores remotos instalados

comúnmente desde una plataforma espacial, de la intensidad de la radiación de ciertas bandas del espectro electromagnético que la vegetación emite o refleja.

Recopilación y gestión de datos

Las muestras fueron tomadas mediante el uso de dos sensores activos OptRx (Ag leader Technology, Inc., Iowa, USA), que funcionaban independientemente, en una posición ascendente, a un metro del dosel, de un ATV que pasaba por debajo de la vid. Cada sensor fue conectado a un robusto equipo de campo, con sistema de posicionamiento global diferenciado (DGPS), y fue operado con el software SMS Mobile. Las muestras fueron tomadas durante la temporada agrícola 2014-2015, específicamente el 10 de noviembre de 2015.

Las muestras fueron exportadas al formato .DAT de MapInfo, estas corresponden a 3341 valores y se pueden encontrar en el siguiente link:

[Valores muestreo](#)

Para la variable Ndvi se presenta el siguiente resumen estadístico:

Mínimo	1er Cuadrante	Mediana	Media	3er Cuadrante	Máximo	Desviación Estándar
0,4066	0,4257	0,4451	0,4418	0,4549	0,5257	0,01542347

Tabla 3.1: Resumen estadístico variable Ndvi

Los programas utilizados para la gestión de los datos corresponden a:

- MapInfo Pro 15.2
- SMS Basic
- R 3.3.2
- RStudio (entorno de desarrollo integrado (IDE) para R)

Para realizar la optimización además se utilizaron los siguientes paquetes de R 3.3.2:

- Gstat
- raster
- rgeos

- maptools
- ggplot2
- sp

La aproximación al problema se dividió en cuatro etapas sucesivas, la primera corresponde a la definición del tamaño de la muestra, la segunda corresponde a la definición de los cuadrantes sobre los cuales se realiza la optimización. Luego se definió el método a utilizar para finalmente concluir con la optimización. Las etapas de la aproximación al problema serán especificadas con mayor detalle a continuación.

Definición del tamaño de muestra

Para definir el tamaño de la muestra primero se prepararon los datos para poder ser utilizados el programa RStudio. Para esto se utilizó la herramienta Coordinate Extractor de MapInfo para obtener las coordenadas de las muestras en kilómetros a través de la proyección de las coordenadas en UTM (WGS 84) para lo cual se utiliza la zona 19 del hemisferio sur. Luego se procedió a exportar los datos en la extensión .txt para que puedan ser cargados y leídos por el programa RStudio.

A continuación, en el programa RStudio, se utilizó el código descrito en el cuadro C.1 para definir el tamaño de la muestra. En este código se utilizó la prueba t de Student en donde se iteró considerando como valores iniciales (asignados arbitrariamente):

- Poder estadístico : 0,9 (1- probabilidad de error tipo II)
- Significancia: 0,05 (probabilidad error tipo I)
- Tamaño del efecto: 0,005
- Desviación estándar: 0,01542347 (desviación estándar de toda la población)

Se iteró variando el valor del poder estadístico hasta obtener un tamaño de muestra con un valor entero.

Definición de los cuadrantes

Con el fin de restringir el problema de optimización se utilizó el sistema de muestreo sistemático en grilla. Esto con el propósito de acercar el problema de optimización a una aplicación de

muestreo real.

Los cuadrantes fueron definidos de manera que existe un muestreo inicial, con el tamaño de muestra definido en la sección anterior. El valor que tome cada muestra puede ir variando en dirección vertical, horizontal y en diagonal hacia los cuadrantes vecinos como se muestra en la Figura 3.2. En donde los cuadrantes verdes representan el muestreo inicial y se representan con flechas las posibilidades de movimiento para el cuadrante (3,4).

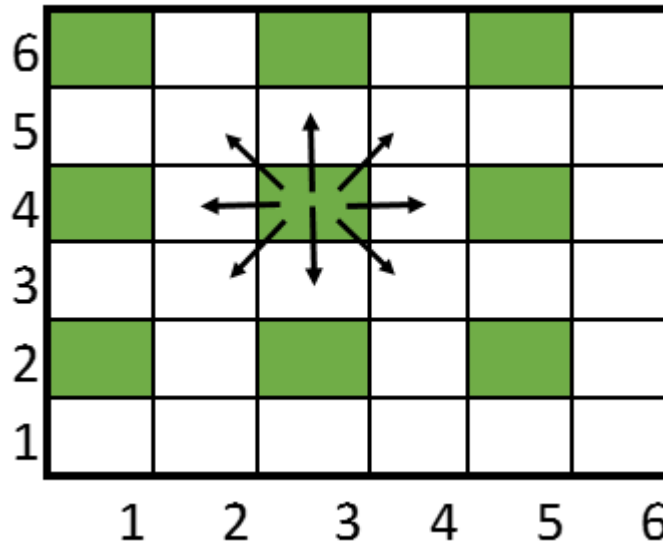


Figura 3.2: Definición de los cuadrantes para muestreo sistemático, las flechas indican posibilidad de movimiento del cuadrante
(Fuente: Elaboración propia)

Los cuadrantes fueron creados a través de la herramienta muestreo de suelos en el programa SMS Basic. Para definir los cuadrantes primero se prepararon los datos en la extensión .DAT de MapInfo para poder ser exportados. Esto se hizo transformando la capa cosmética del mapa y a través del menú Spatial se insertó un polígono a manera de limitar el perímetro del área de muestreo el cual se guardó para luego ser exportado en formato .mif.

Mediante la herramienta leer archivos de fuente genérica se exportaron los datos al programa SMS Basic en donde se utilizó la herramienta muestreo de suelos. Luego se exportó el muestreo a formato KMZ, el cual es leído por el programa MapInfo. Finalmente, en MapInfo se utilizó la herramienta SQL Select en el menú de Table para extraer sólo los valores correspondientes a los cuadrantes antes definidos para ser exportados a formato .txt para poder ser leídos por el programa RStudio.

Definición de método a utilizar

Para definir el método de predicción e interpolación a utilizar en la optimización se eligió como indicador la RMSE de los métodos Kriging Ordinario e IDW utilizando todos los valores de la población utilizando el programa RStudio con el código descrito en el cuadro C.2.

Para esto primero se definió una grilla sobre la cual se realizó la interpolación, esta corresponde a una grilla con cuadrantes de 3,2 [m] x 3,2 [m] como se puede ver representada en la Figura 3.3, en donde los círculos rojos representan la ubicación de los valores de toda la población.

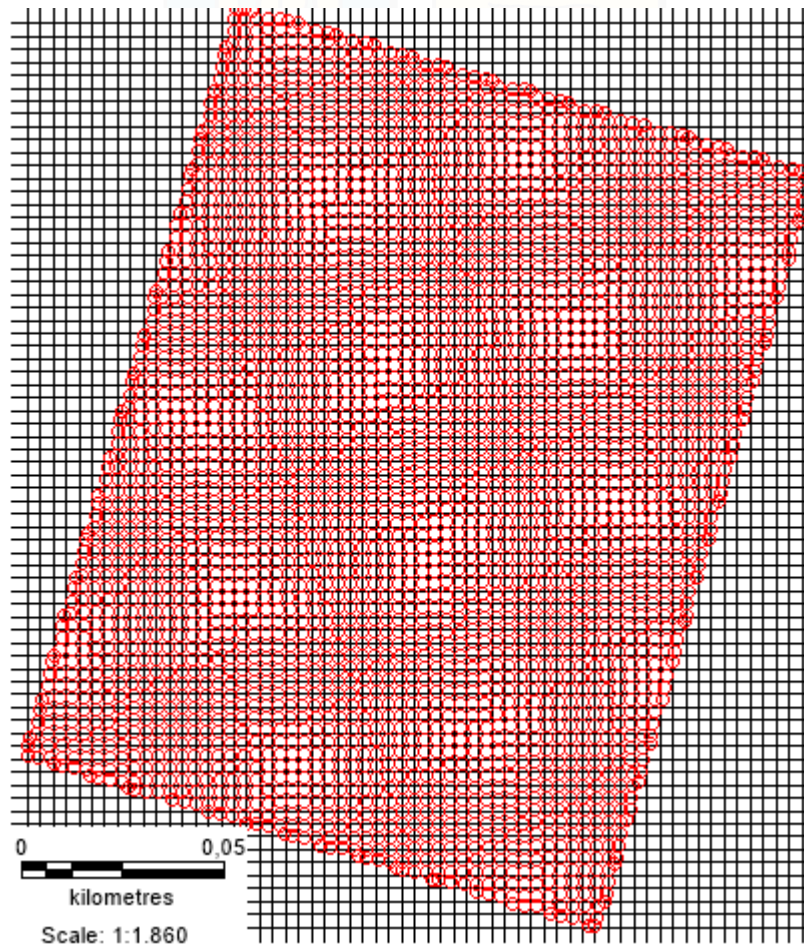


Figura 3.3: Grilla de interpolación con valores de toda la población, círculos rojos representan a los valores muestrales
(Fuente: Elaboración propia)

Luego se realizaron las predicciones con ambas opciones y se calculan sus RMSE, se eligió el método con el menor valor.

Algoritmo de optimización

Para la optimización se tomó como objetivo de minimización la RMSE entre los valores de la muestra y los valores inter y extrapolados del método de predicción elegido en el paso anterior.

Para esto se utilizó el código descrito en el cuadro C.3. En este, se parte con la muestra inicial definida en por el tamaño de muestra y cuadrantes en las secciones anteriores, como se puede ver ejemplificado en la Figura 3.4, en donde las ubicaciones en negro corresponden a la configuración de muestreo inicial para los cuadrantes definidos.

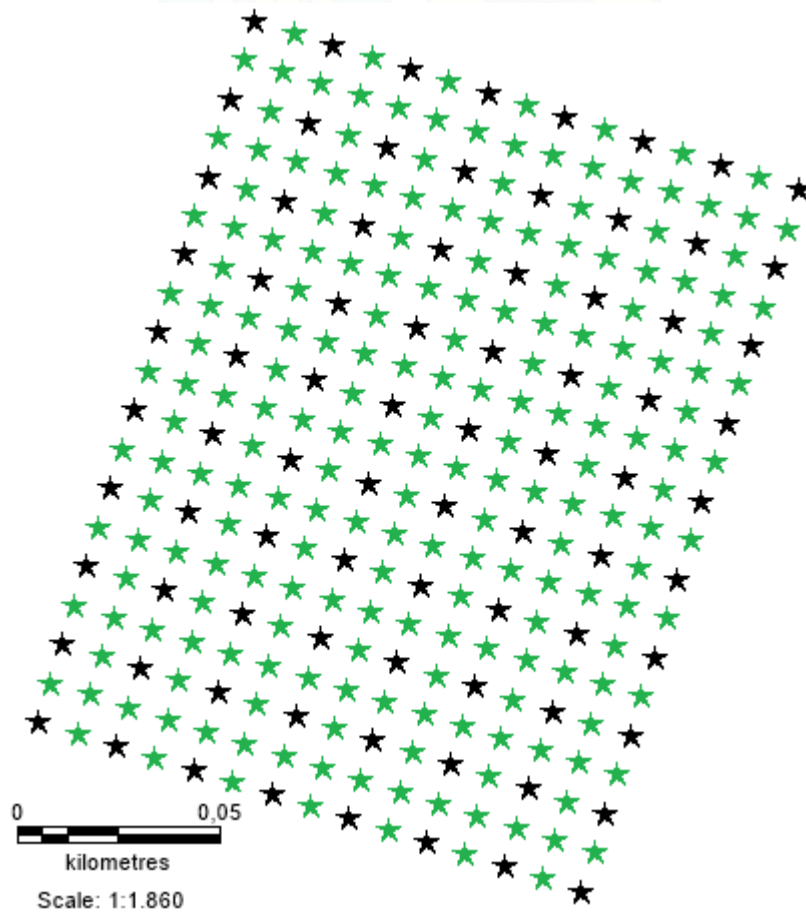


Figura 3.4: Configuración inicial de muestreo (en negro)
(Fuente: Elaboración propia)

Posteriormente se realizó la interpolación y cálculo de la RMSE. Luego se realizaron 1.000.000 iteraciones en donde las muestras se mueven de manera aleatoria por sobre los cuadrantes de muestreo, se realiza la interpolación y el cálculo de la RMSE. Si la nueva configuración de muestreo presenta una menor RMSE que la configuración inicial, las siguientes iteraciones son realizadas comparando el nuevo valor de RMSE con las nuevas configuraciones de muestreo y así sucesiva-

mente.

Finalmente se realizaron los mapas de las predicciones, para ambos métodos y configuraciones (muestra inicial y optimizada) junto con un mapa de sus MSE.



Resultados y discusión

Tamaño de la muestra y diseño de cuadrantes

Para los siguientes parámetros se obtuvo un tamaño de muestra de 82,85961:

- Poder estadístico: 0,9
- Significancia: 0,05
- Tamaño del efecto: 0,005
- Desviación estándar: 0,01542347

Se decidió, para facilitar el diseño de los cuadrantes, trabajar con un tamaño de muestra de 80. Para lo cual se tienen los siguientes parámetros:

- Poder estadístico : 0,8906023
- Significancia: 0,05
- Tamaño del efecto: 0,005
- Desviación estándar: 0,01542347

El diseño de los cuadrantes de muestreo quedó en 15 columnas y 19 filas, con un total de 285 cuadrantes de 10 [m] x 10 [m], como se aprecia en la Figura [4.1](#).

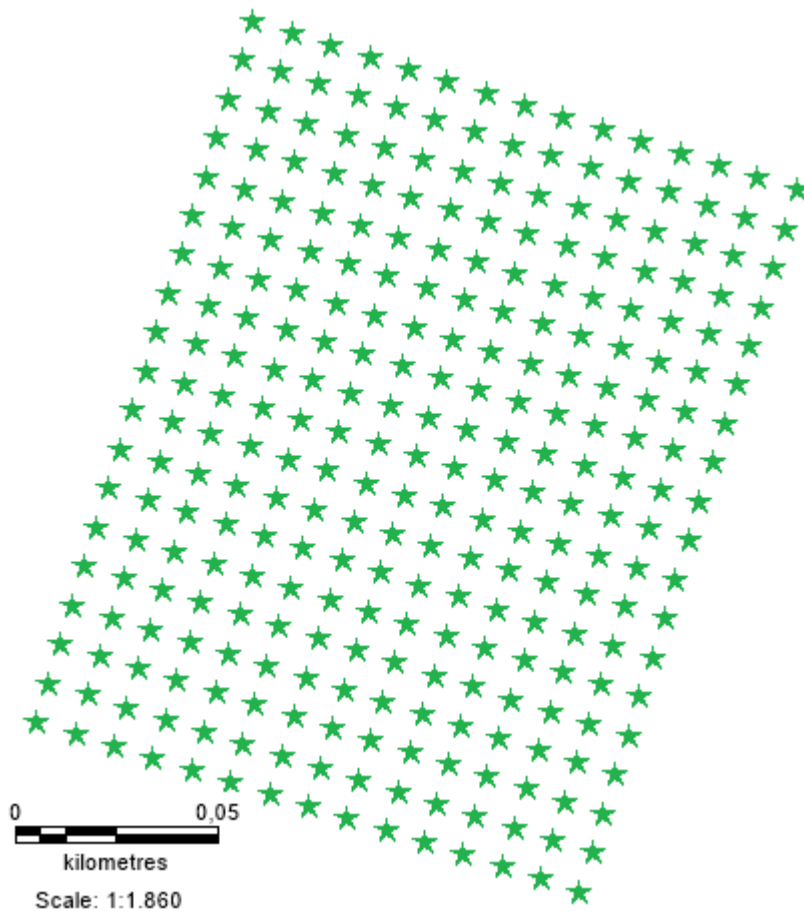


Figura 4.1: Cuadrantes optimización finales, 15 columnas y 19 filas, con un total de 285 cuadrantes (Fuente: Elaboración propia)

Cabe mencionar que al ser configurados los cuadrantes de manera manual, la metodología utilizada puede ser replicada en diversos terrenos, no es necesario que este tenga la forma de un cuadrilátero.

Método de predicción

Kriging Ordinario

Para este método primero se modeló el variograma experimental de los datos y luego se probaron diversos modelos de variogramas teóricos hasta encontrar el que mejor se ajustó al variograma experimental.

El modelo de variograma utilizado corresponde a un modelo gaussiano con la siguiente ecuación:

$$\gamma(h) = 0,00025 \left\{ 1 - \exp\left(-\frac{|h|^2}{100^2}\right) \right\}$$

Ambos variogramas se pueden observar en la Figura 4.2:

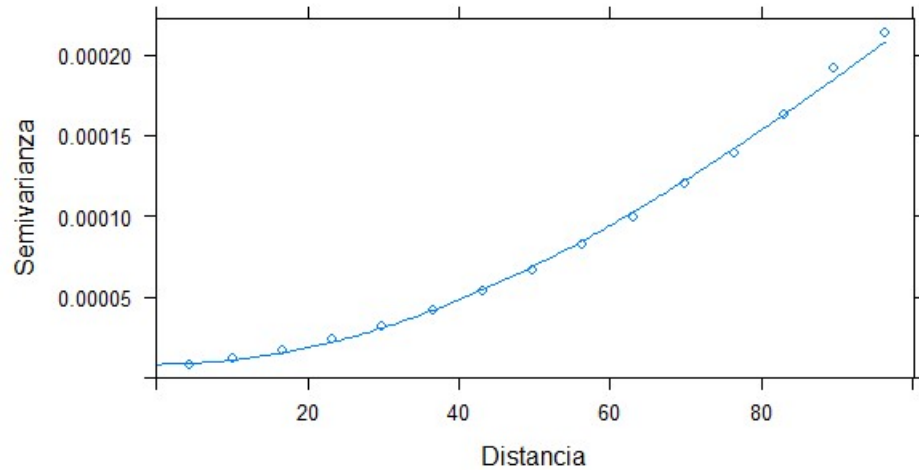


Figura 4.2: Variograma experimental y teórico para los valores de toda la población
(Fuente: Elaboración propia)

En donde la meseta se fijó en 0,00025, el rango en 100 y el nugget en 0.

Es importante mencionar que al utilizar un modelo elemental de variograma se asume que el terreno se comporta de manera isotrópica, es decir, que la correlación la correlación espacial entre puntos del espacio es igual en todas las direcciones.

Para la predicción se obtiene el siguiente mapa (imagen 4.3):

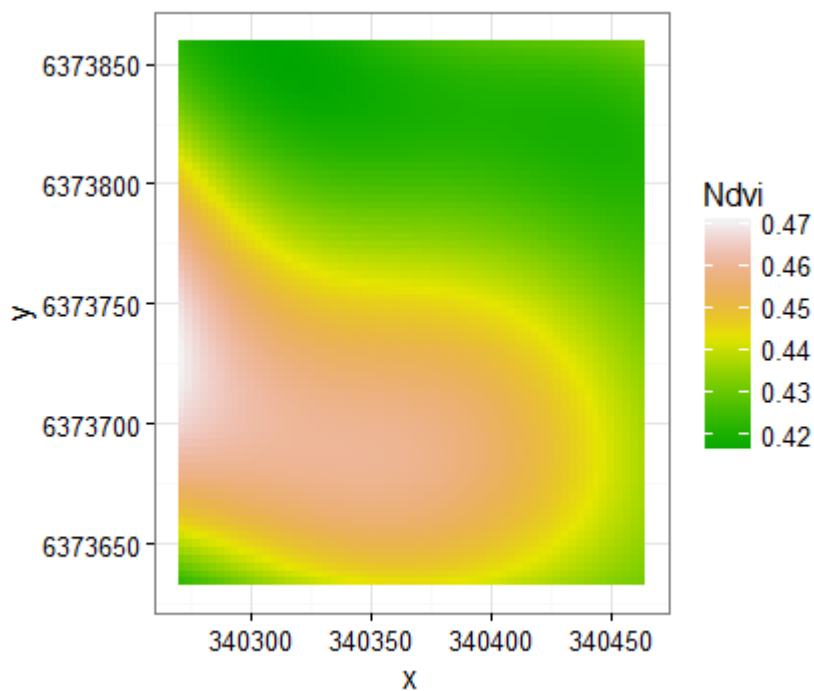


Figura 4.3: Mapa mediante K.O realizado con valores de toda la población
(Fuente: Elaboración propia)

Para la cual se obtuvo una RMSE de $8,630 \times 10^{-6}$

IDW

Para este método se obtiene el siguiente mapa de predicción:

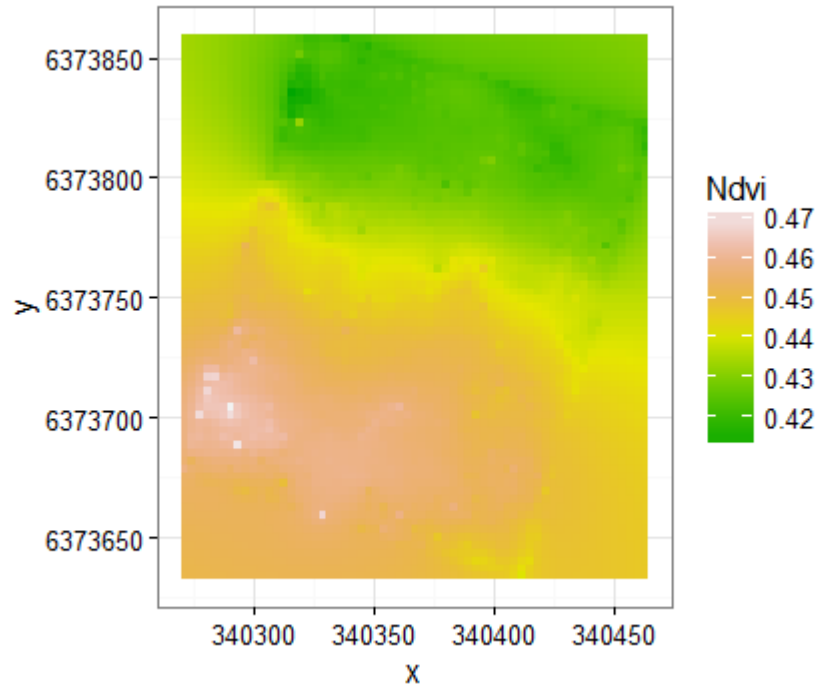


Figura 4.4: Mapa mediante IDW realizado con valores de toda la población
(Fuente: Elaboración propia)

La RMSE obtenida fue menor a la resolución del programa (RStudio) la cual corresponde a 10^{-11} . Es por este motivo que se decidió utilizar este método para el desarrollo de la optimización. Además, al tomar un menor número de datos, es posible que el variograma teórico antes definido para todos los valores, no describa el comportamiento de la muestra.

Algoritmo de optimización

Para poder realizar la optimización primero se obtuvieron los mapas y valor de la RMSE de la muestra inicial (tabla B.1) mediante el método de elección elegido (IDW). A continuación se presentan los mapas de predicción (Figura 4.5), predicción más la ubicación de los puntos de muestra (Figura 4.6) y por último un mapa de los MSE al comparar la predicción con los valores de la predicción mediante IDW con toda la población (imagen 4.7).

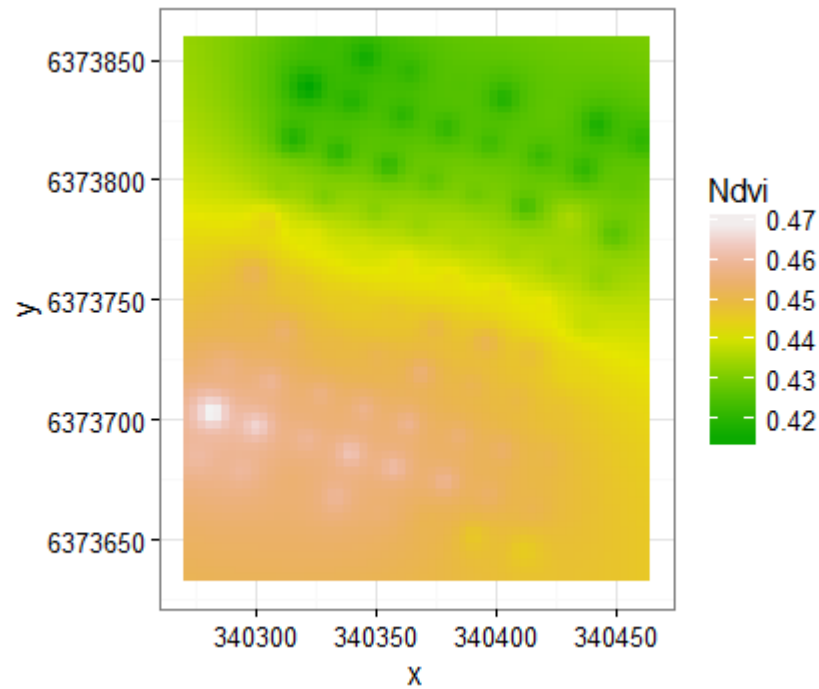


Figura 4.5: Mapa mediante IDW realizado con la muestra inicial
(Fuente: Elaboración propia)

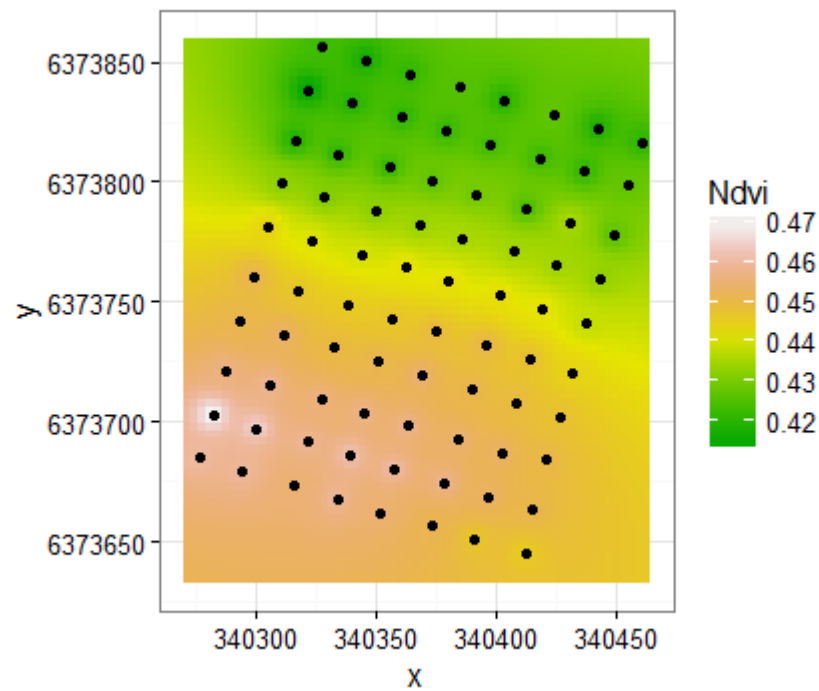


Figura 4.6: Mapa mediante IDW realizado con la muestra inicial con puntos de muestreo en negro
(Fuente: Elaboración propia)

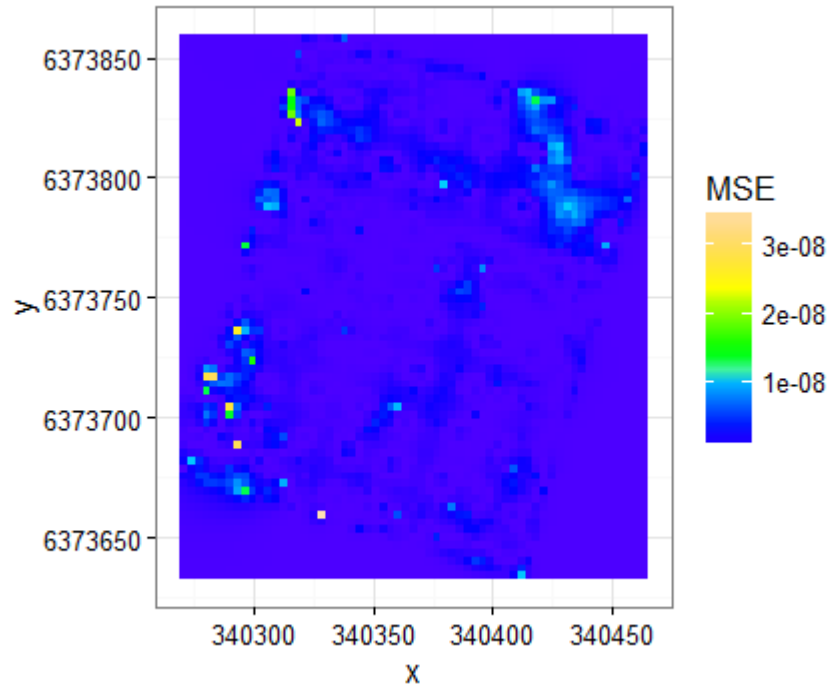


Figura 4.7: Mapa MSE mediante IDW realizado con la muestra inicial, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población
(Fuente: Elaboración propia)

La RMSE obtenido corresponde a 0,001935787. Se decidió utilizar los valores de la predicción con toda la población para poder así comprar los valores extrapolados también.

A continuación el algoritmo produjo nuevas configuraciones de muestreo y comparó sus RMSE con el valor de la muestra inicial, en el apéndice D se pueden observar los mapas obtenidos para una configuración aleatoria de muestreo (tabla B.2) con una RMSE de 0,002196255.

Como resultado de la optimización se obtuvo una configuración de muestreo (tabla B.3) con una RMSE de 0,001788088. A continuación se presentan los mapas de predicción (Figura 4.8), predicción más la ubicación de los puntos de muestra (imagen 4.9) y por último un mapa de los MSE al comparar la predicción con los valores de la predicción mediante IDW con toda la población (imagen 4.10).

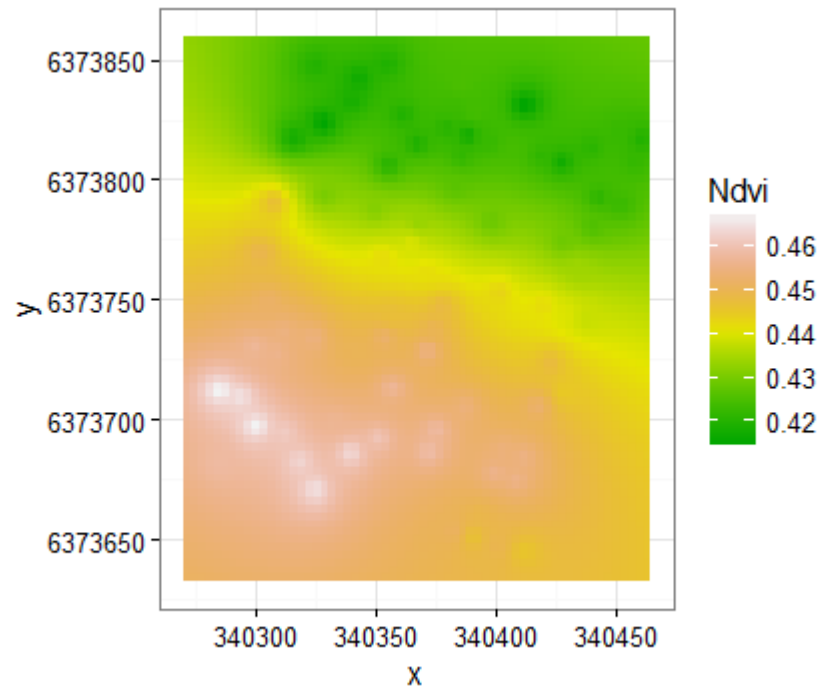


Figura 4.8: Mapa mediante IDW realizado con la muestra optimizada
(Fuente: Elaboración propia)

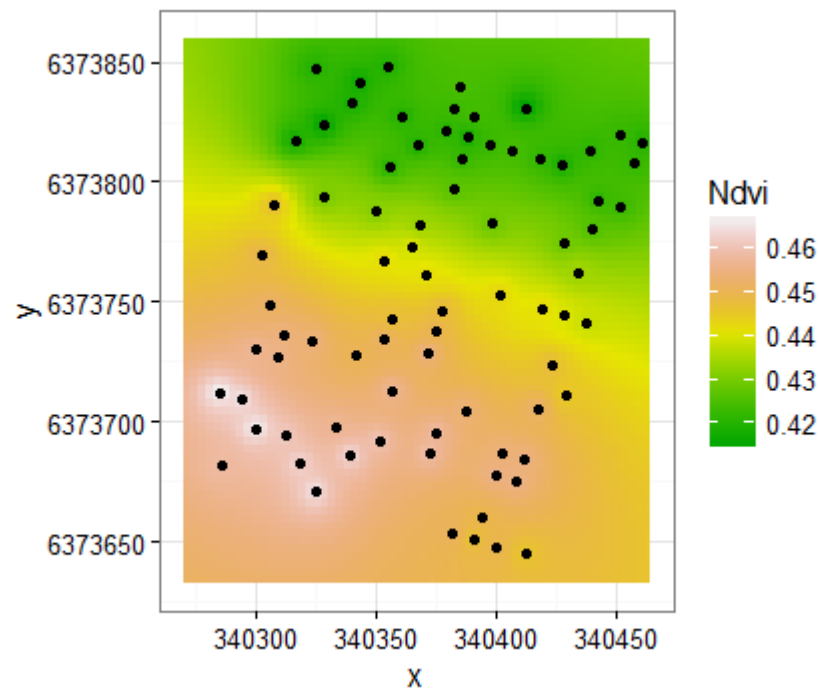


Figura 4.9: Mapa mediante IDW realizado con la muestra optimizada con puntos de muestreo en negro
(Fuente: Elaboración propia)

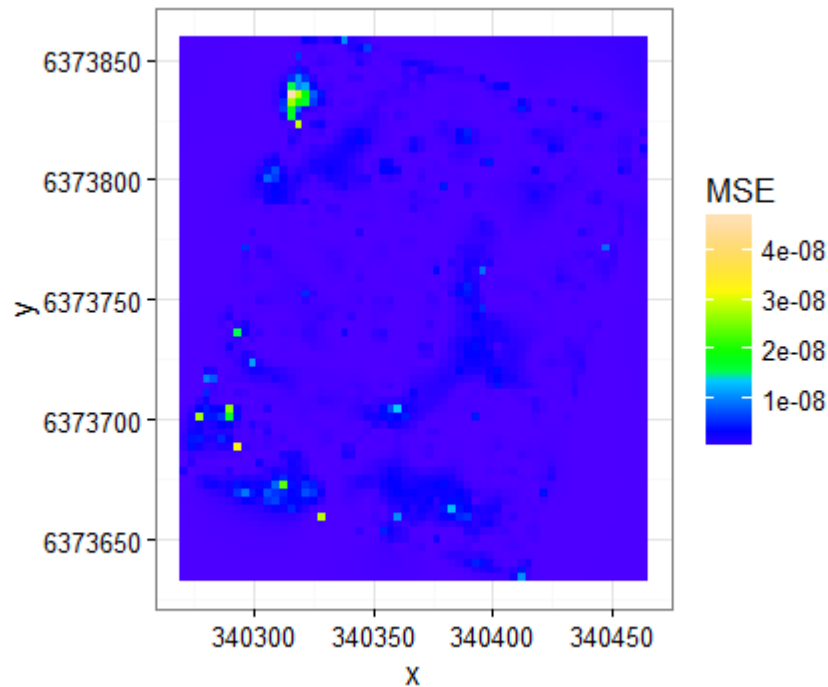


Figura 4.10: Mapa MSE mediante IDW realizado con la muestra optimizada, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población
(Fuente: Elaboración propia)

Comparación Kriging y muestreo aleatorio

A pesar de ser IDW el método elegido para realizar la optimización, se decidió utilizar el método de predicción de K.O. para comparar los resultados obtenidos tanto como para la muestra inicial, como para la muestra optimizada. El código utilizado se encuentra en el cuadro C.7.

Tanto para la muestra inicial como para la muestra optimizada se ajustó el mismo variograma teórico que el utilizado para el caso de toda la población, en donde el variograma, teórico y experimental, de la muestra inicial corresponde a la Figura 4.11 y el variograma de la muestra optimizada, teórico y experimental, a la Figura 4.12.

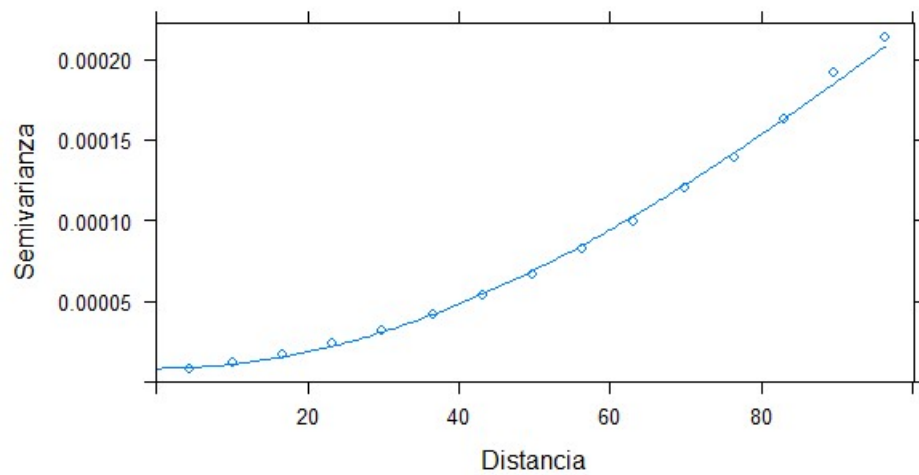


Figura 4.11: Variograma experimental y teórico para los valores de la muestra inicial
(Fuente: Elaboración propia)

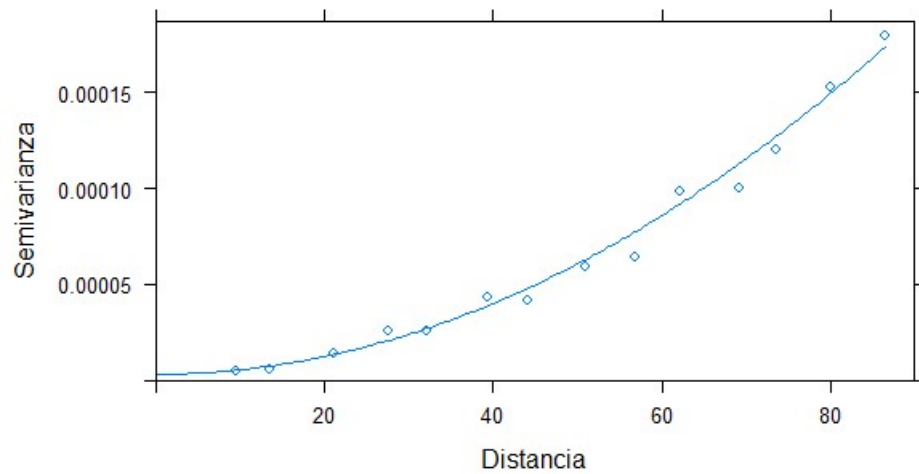


Figura 4.12: Variograma experimental y teórico para los valores de la muestra optimizada
(Fuente: Elaboración propia)

Para ambas configuraciones se obtuvieron los siguientes mapas de predicción, siendo el de la muestra inicial la Figura 4.13 y 4.14, y para la muestra optimizada la Figura 4.15 y 4.16.

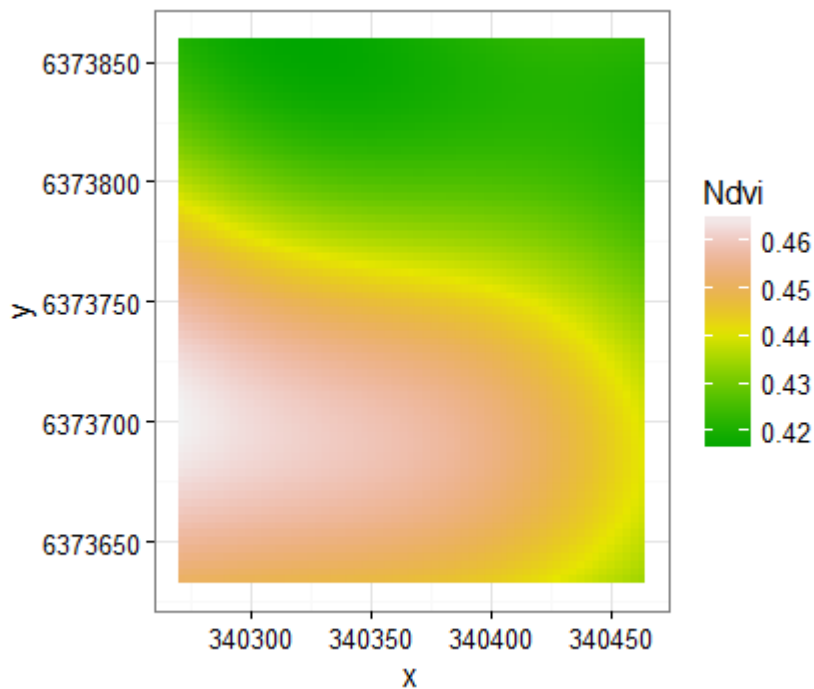


Figura 4.13: Mapa mediante K.O. realizado con la muestra inicial
(Fuente: Elaboración propia)

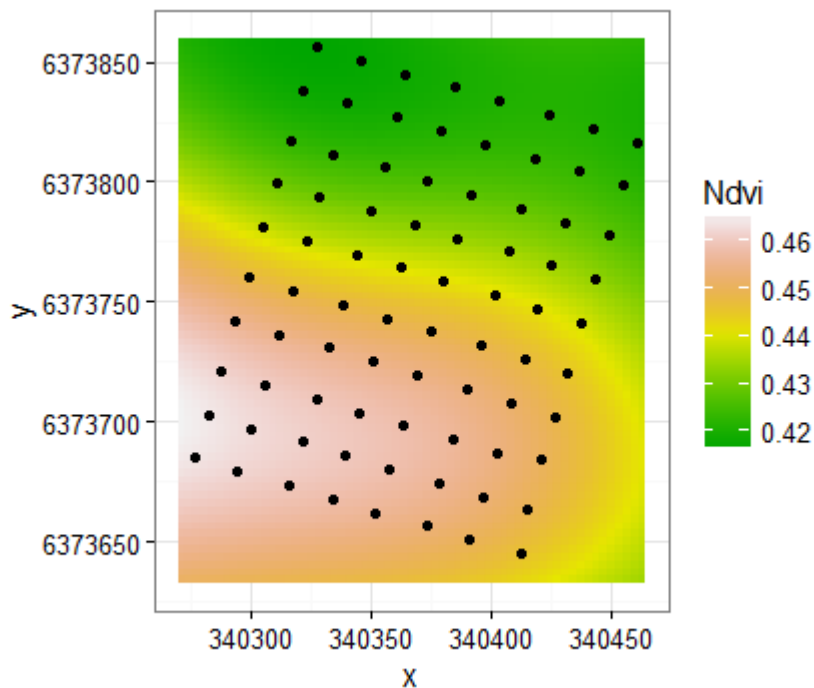


Figura 4.14: Mapa mediante K.O. realizado con la muestra inicial con puntos de muestreo en negro
(Fuente: Elaboración propia)

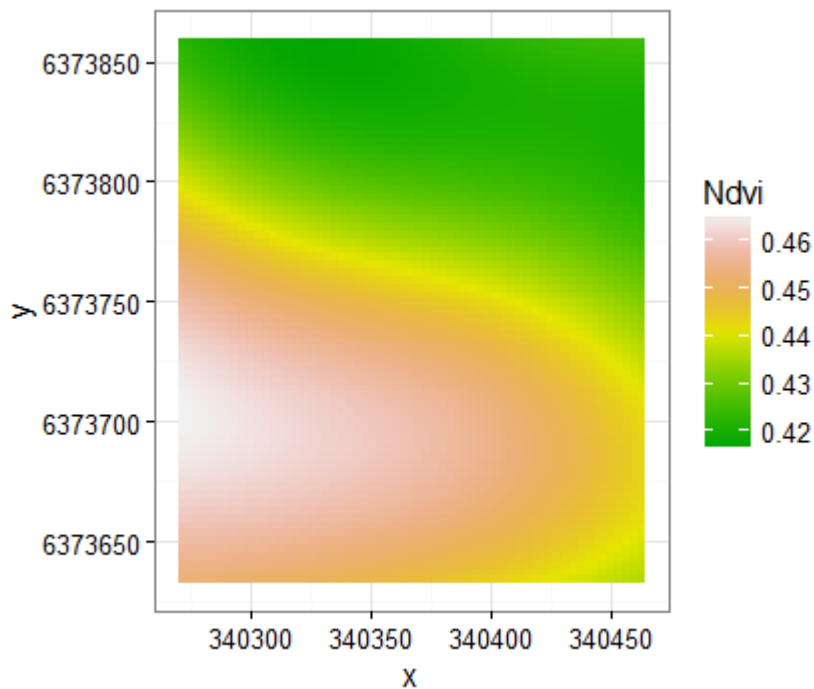


Figura 4.15: Mapa mediante K.O. realizado con la muestra optimizada
(Fuente: Elaboración propia)

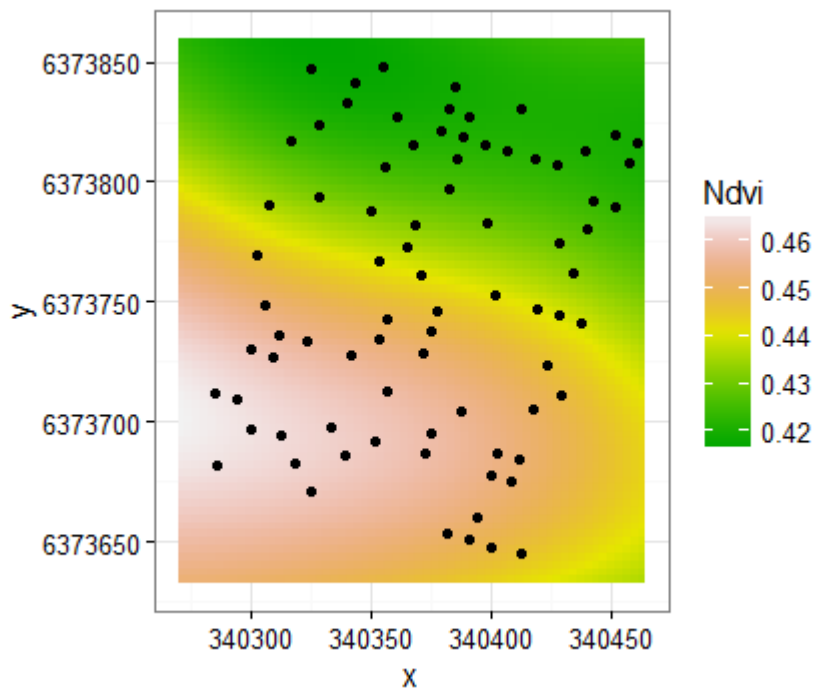


Figura 4.16: Mapa mediante K.O. realizado con la muestra optimizada con puntos de muestreo en negro
(Fuente: Elaboración propia)

Para el cálculo de la RMSE se obtuvo, al comparar con las predicciones de IDW para toda la población, una RMSE de:

- Muestra inicial: 0,003586413
- Muestra optimizada: 0,003479626

En donde se obtuvieron los siguientes mapas del MSE, en donde la Figura 4.17 corresponde a la muestra inicial y la Figura 4.18 a la muestra optimizada.

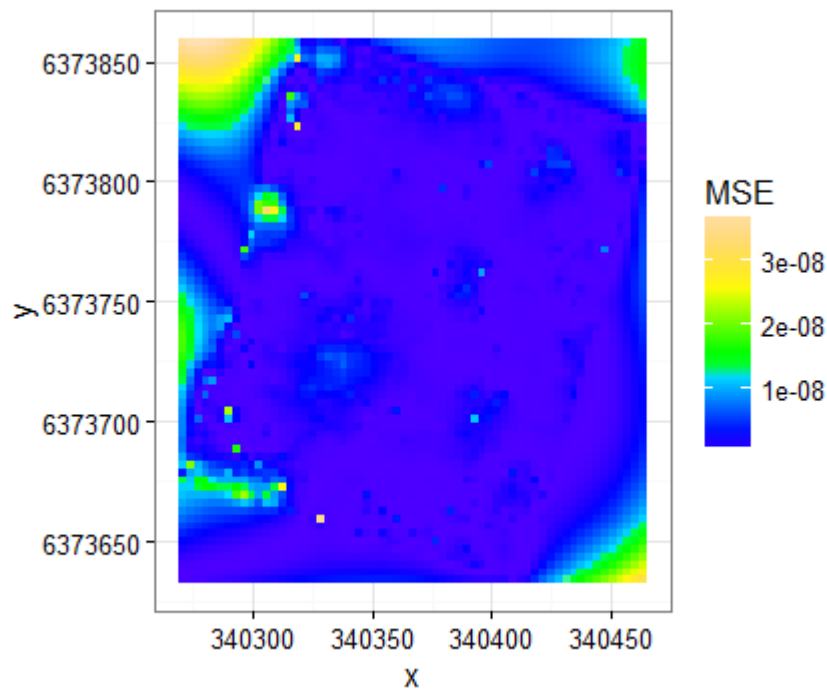


Figura 4.17: Mapa MSE mediante K.O realizado con la muestra inicial, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población
(Fuente: Elaboración propia)

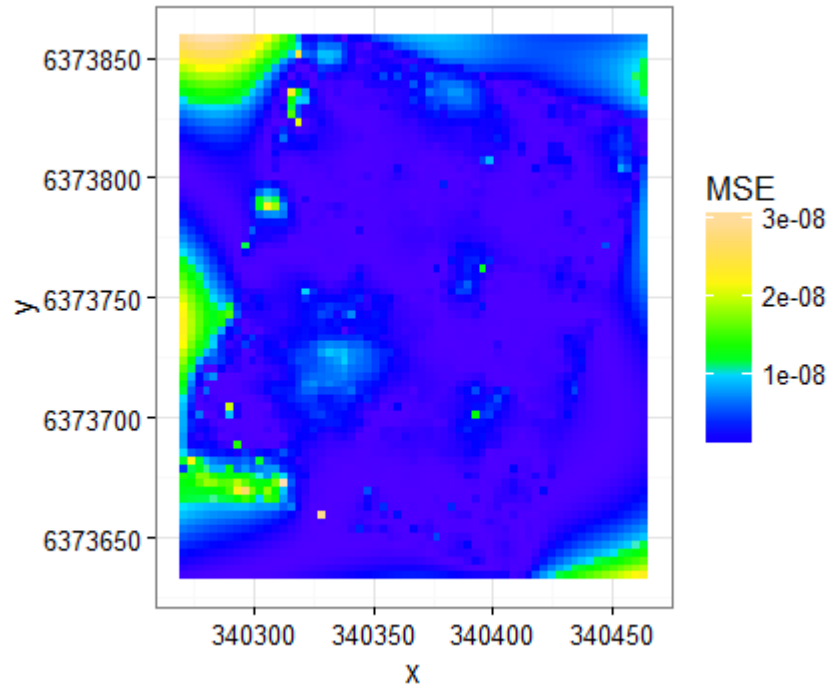


Figura 4.18: Mapa MSE mediante K.O. realizado con la muestra optimizada, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población
(Fuente: Elaboración propia)

Además se decidió comparar ambas configuraciones con los valores de la predicción hecha mediante K.O. con toda la población, esto con el fin de vislumbrar si existe una disminución del error de la predicción al comparar los datos obtenidos mediante el mismo método de predicción. Con lo que se obtuvieron los valores de:

- Muestra inicial: 0,003565634
- Muestra optimizada: 0,003682524

Luego se realizaron los siguientes mapas de la MSE, en donde la Figura 4.19 corresponde a la muestra inicial y la Figura 4.20 a la muestra optimizada.

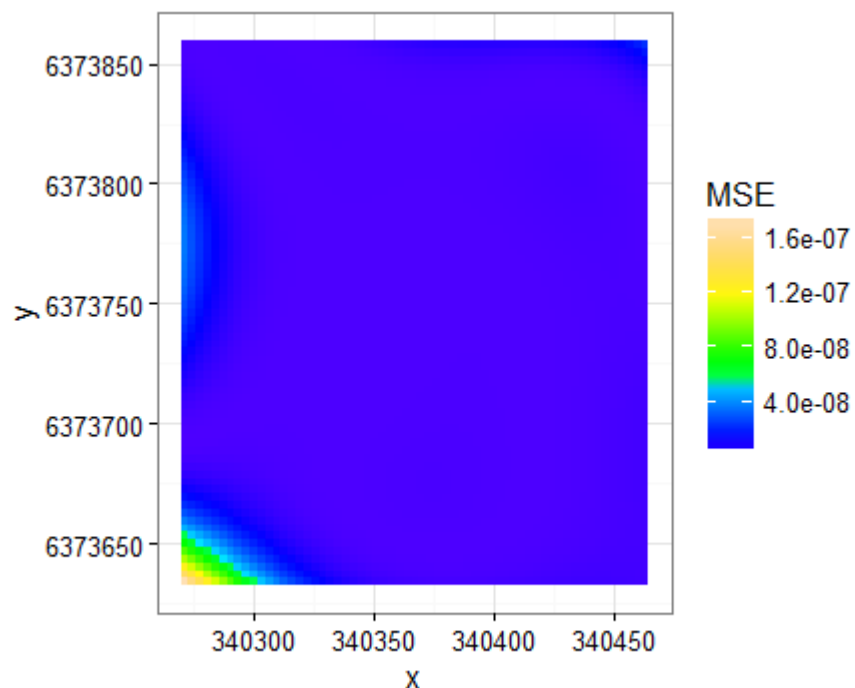


Figura 4.19: Mapa MSE mediante K.O. realizado con la muestra inicial, donde el MSE se calcula con los valores obtenidos mediante K.O para toda la población
(Fuente: Elaboración propia)

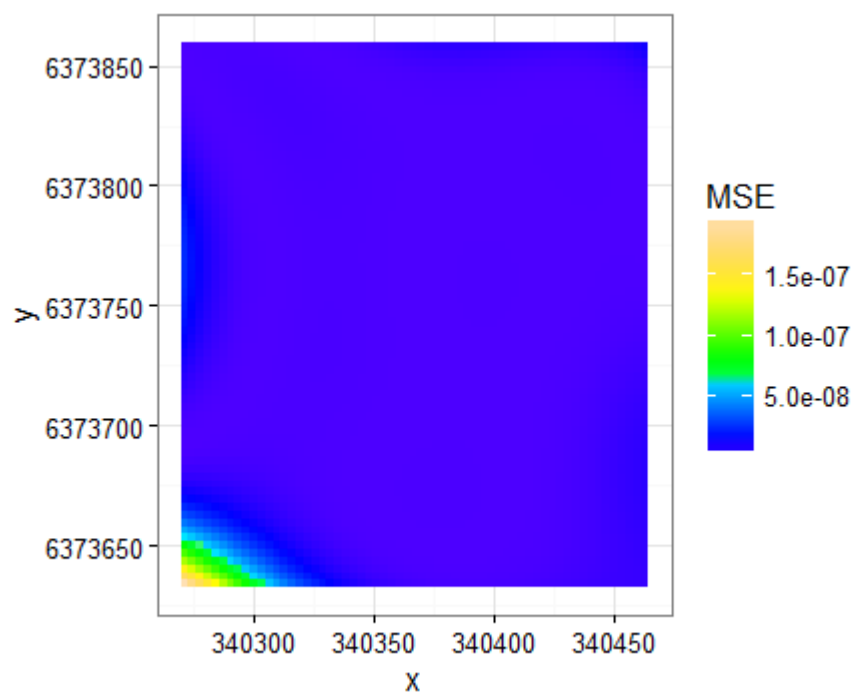


Figura 4.20: Mapa MSE mediante K.O. realizado con la muestra optimización, donde el MSE se calcula con los valores obtenidos mediante K.O para toda la población
(Fuente: Elaboración propia)

Además se decidió comparar los resultados obtenidos versus un muestreo aleatorio, cuyos puntos se encuentran en la tabla B.4. Para estos valores se obtuvo una RMSE de 0,002471089. Los mapas para esta configuración de muestreo se encuentran en el apéndice D.

A continuación se presentan las RMSE ordenadas de menor a mayor:

$$RMSE_{optIDW-IDW} < RMSE_{MI_{IDW}-IDW} < RMSE_{Al_{IDW}-IDW} < RMSE_{optKO-IDW} <$$

$$RMSE_{MI_{KO}-KO} < RMSE_{MI_{KO}-IDW} < RMSE_{optKO-KO}$$

De donde se observa que efectivamente la menor RMSE se obtiene con los valores de la optimización al ser comparados con los valores obtenidos mediante IDW con toda la población. Además se observa que la RMSE siempre es mayor al utilizar el método de predicción de K.O. a pesar de ser comparada versus la predicción mediante K.O con toda la población.

A continuación se presenta un cuadro comparativo del error de interpolación y extrapolación según MSE para la muestra optimizada utilizando ambos métodos de predicción y comparando el método de K.O. con las predicciones de ambos métodos para toda la población:

	Ndvi vs Ndvi	Ndvi vs K.O.	K.O. vs K.O.
Interpolación	$2,8557 \times 10^{-6}$	$4,355714 \times 10^{-6}$	$1,02359 \times 10^{-6}$
Extrapolación	$3,37026 \times 10^{-7}$	$7,75209 \times 10^{-6}$	$1,25374 \times 10^{-5}$
Total	$3,197258 \times 10^{-6}$	$1,21078 \times 10^{-5}$	$1,356098 \times 10^{-5}$

Tabla 4.1: Comparación MSE de intra y extrapolación

De donde se observa que a pesar que al realizar la predicción y comparación mediante K.O, este método presenta el mayor error de extrapolación.

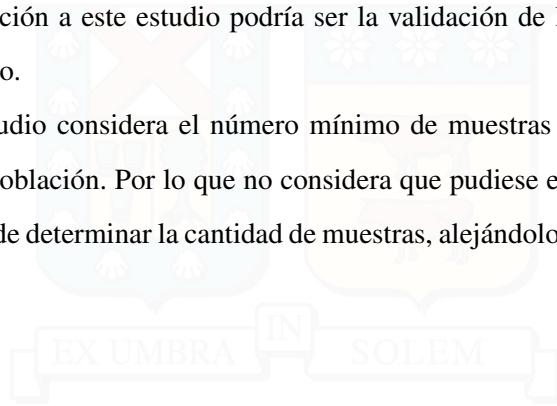
A pesar que la literatura sugiere que el método de predicción de Kriging es el método óptimo en el sentido estadístico de interpolación y extrapolación, los resultados obtenidos podrían ser explicados por diversos motivos. Primero, en el modelamiento de la dependencia espacial se asume que la variable a analizar se comporta de manera isotrópica con el fin de facilitar el estudio. Segundo, al comparar ambas metodologías de predicción, los valores extrapolados también han sido predichos, por lo que no se cuenta con los valores reales para determinar la RMSE.

Al analizar las fuentes errores que presentan los muestreos, la optimización llevada a cabo permite disminuir el error de selección de las muestras, causado por la tendencia a seleccionar algunas unidades muestrales dentro de la población con mayor o menor frecuencia. Esto debido

a que el número de muestras ha sido seleccionado con el fin de asegurar la representatividad del terreno y la locación de estas con el fin de disminuir el error entre las predicciones y el valor real.

Es importante mencionar que la aproximación utilizada, a pesar de considerar la variabilidad de la variable estudiada, no considera que esta variable puede no ser el elemento nutritivo más móvil. Por lo que una continuación a este estudio podría ser la validación de la optimización mediante otras variables del terreno.

Finalmente, este estudio considera el número mínimo de muestras con el fin de asegurar la representatividad de la población. Por lo que no considera que pudiese existir una restricción presupuestaria al momento de determinar la cantidad de muestras, alejándolo de un problema de muestreo real.



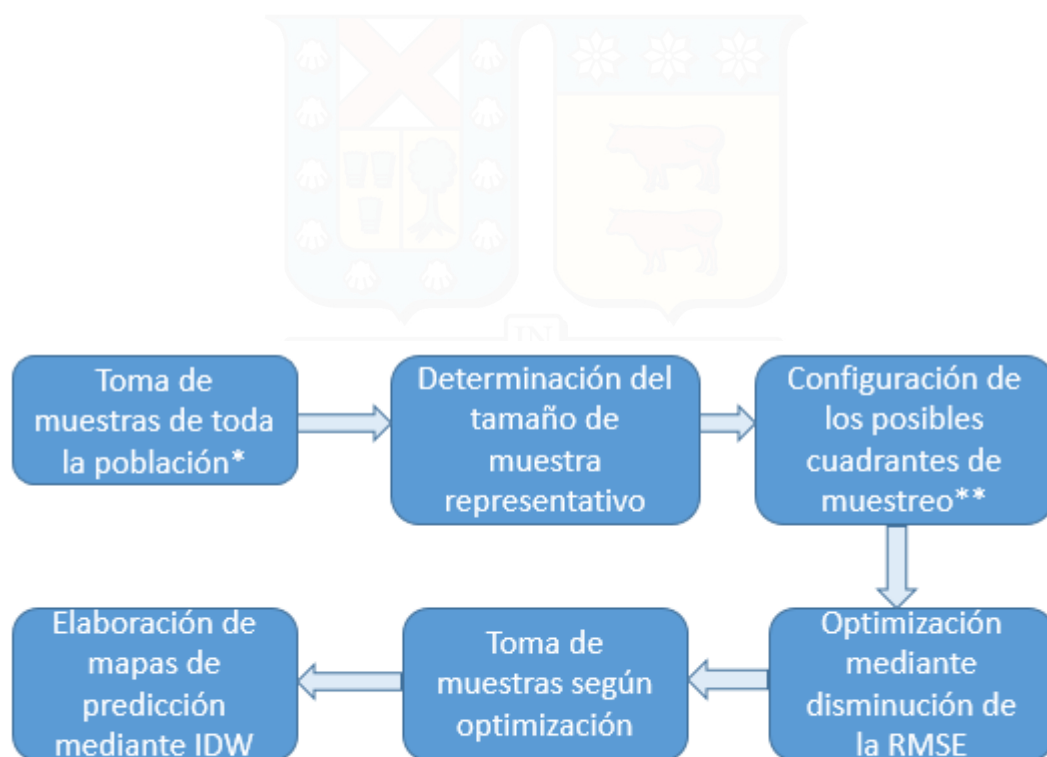
Conclusiones

Un plan óptimo de muestreo debe representar la dependencia espacial con el fin de poder interpolar los lugares no muestreados utilizando la menor cantidad de muestras posibles. Por lo que en este trabajo se presenta una aproximación a la optimización para la ubicación espacial de muestras a partir de un diseño sistemático.

El diseño de muestreo mediante un muestreo sistemático permite no sólo facilitar la aproximación del muestreo a un problema de muestreo real, esto debido a la facilidad de ubicar los puntos de muestreo mediante cuadrantes. Además disminuye el error de selección de las muestras, causado por la tendencia a seleccionar algunas unidades muestrales dentro de la población con mayor o menor frecuencia. Y por último, demuestra ser una mejor alternativa al muestreo aleatorio (esto mediante la comparación de sus RMSE (0,002196255 vs 0,001788088)).

Respecto a las técnicas de predicción utilizadas, la IDW presenta ciertas ventajas por sobre el K.O. La IDW, no requiere de la estimación de un modelo para expresar la dependencia espacial a diferentes pares de distancias (cuantificada a través del uso de variogramas). Lo que primero, simplifica la aproximación del problema y segundo, disminuye la probabilidad de errores al realizar un mal modelamiento de la dependencia espacial; como por ejemplo, considerar erradamente que la población es isotrópica.

Finalmente, al ser IDW la técnica de predicción escogida, se recomienda el siguiente diagrama para la aproximación a un problema de muestreo con fines de mapeo:



*Mediante muestreo no destructivo (ej.: muestreo mediante sensor activo)

**Según criterio detallado en la imagen 3.2

Figura 5.1: Recomendación aproximación problema de muestreo con fines de mapeo

Bibliografía

Anselin, L (1988). *Spatial econometrics: methods and models*. [2.13](#)

Bernier, Rene (1999). Curso de Capacitación para operadores del programa de Recuperación de Suelos Degradados INDAP, Décima Región. *Curso de Capacitación para operadores del Programa de Recuperación de Suelos Degradados INDAP, Décima Región*, (pp. 117). [2.3](#)

Emery, Xavier (2007). Apunte de Geoestadística. [2.8](#), [2.8.2](#), [2.9](#), [2.9.1](#), [2.5](#), [2.9.1](#), [2.6](#), [2.9.1](#), [2.7](#), [2.8](#), [2.9](#), [2.10](#), [2.9.1](#), [2.11](#), [2.12](#), [2.10](#), [2.10.1](#)

Ginzo Villamayor, María José (2009). Análisis Geostadístico de datos funcionales. (pp. 108). [2.9.1](#)

Gutierrez-Lopez, Alfonso; Ramirez, Aldo I; Lebel, Thierry; Santillan, Oscar; y Fuentes, Carlos (2011). Variogram and correlogram, two proposed estimators of the hydrological variability measurement. *Revista Facultad De Ingenieria-Universidad De Antioquia*, (59), 193–202. [2.8.1](#), [2.10](#)

Hernández, J y Corvalán, P (2000). Uso de Variogramas para la determinación del tamaño medio de las copas de pino silvestre en imagenes aereas digitales. *Ciencias Forestales*, 14, 3–18. [2.9.1](#), [2.9.1](#)

INN (1999). Norma Chilena NCh 2060-1999. Suelos-Muestreo para análisis de fertilidad. [2.4.3](#), [2.6](#)

Kitchen, N.R.; Sudduth, K.A.; y Drummond, S.T. (1999). Kitchen, N.R., K.A. Sudduth and S.T. Drummond. 1999. Electrical conductivity as a crop productivity measure for claypan soils. *Journal of Production Agriculture*, (pp. 12(4): 607–617). [2](#)

Ledo, Diego Francisco (Universidad técnica Federico Santa María) (2013). *Análisis del impacto económico y ambiental al aplicar fertilizantes a tasas variables con tecnologías de agricultura de precisión*. PhD thesis, Universidad Técnica Federico Santa María. [2.11](#), [2.11](#), [2.11.2](#)

Martori, Joan Carles; Hoberg, Karen; y Madariaga, Rafa (2008). La incorporacion del espacio en los metodos estadísticos: autocorrelación espacial y segregación. In *X Coloquio Internacional de Geocrítica* (pp. 1–8). [2.11](#)

Mason, B.J (1992). *Preparation of soil sampling protocols: Sampling techniques and strategies*. Nevada Univ., Las Vegas, NV (United States). Environmental Research Center. [2.4.4](#), [2.2](#)

Ortega, Rodrigo (2013). Muestreo de suelos para recomendación de fertilizantes.Taller-Seminario Agricultura de Precisión y Variabilidad Espacial, Guatemala. [2.1](#), [2.4.1](#), [2.4.2](#), [2.4.3](#), [2.6](#), [2.9](#), [2.4](#)

- Ortega, Rodrigo (2015). Recolección espacial de datos. [2.1](#), [2.4.3](#), [2.4.4](#), [2.5](#), [2.3](#), [2.7](#)
- Ortega B., Rodrigo y Flores M., Luis (2000). Agricultura de precisión: Introducción al manejo sitio-específico. *Informaciones Agronómicas del Cono Sur*, 7, 1–5. [2](#)
- Pantoja, Felipe Andrés (Universidad Técnica Federico Santa María) (2013). *Optimización de la predicción del nivel de producción de trigo mediante la utilización de econometría espacial*. PhD thesis, Universidad Técnica Federico Santa María. [2.7](#)
- Roberts, TL y Henry, JL (2000). El muestreo de suelos: Los beneficios de un buen trabajo. [1.1.1](#), [2.4.4](#), [2.5](#)
- Sahrawat, K L; Rego, T J; Wani, S P; y Pardhasaradhi, G (2008). Stretching Soil Sampling to Watershed: Evaluation of Soil-Test Parameters in a Semi-arid Tropical Watershed. *Communications in Soil Science and Plant Analysis*, 39(19-20), 2950–2960. [2.1](#)
- Santacruz Delgado, Alí Marcel (2010). Diseño de Redes de Muestreo Óptimas para el Monitoreo del Carbono Orgánico del Suelo en el C.I. La Libertad mediante la Aplicación de Algoritmos Genéticos Alí Marcel Santacruz Delgado. [2.9](#)
- Sun, Feixiang; Wang, Shuai; Liu, Min; Yao, Xueling; Fu, Bojie; y Lu, Yihe (2013). Comparison of Four Spatial Interpolation Methods for Estimating Soil Moisture in a Complex Terrain Catchment. 8(1). [2.12](#), [2.13](#)
- Torrecilla, Rosa Cañada; Jesús, Maria; Domínguez, Vidal; y Jiménez, Antonio Moreno (2010). Interpolación espacial y visualización cartográfica para el análisis de la justicia ambiental: Ensayo metodológico sobre la contaminación por partículas atmosféricas en Madrid. *Tecnologías de la información geográfica: La información geográfica al servicio de los ciudadanos*, (pp. 691–715). [2.12](#)
- Villalobos, Pablo; Manriquez, Richard; Acevedo, Cesar; y Ortega, Samuel (2009). Alcance de la agricultura de precisión en Chile : estado del arte, ámbito de aplicación y perspectivas. *Www.Odepa.Gob.Cl*, (pp. 112). [2](#), [2](#)
- Volke, Tania; Velasco, Juan A.; y de la Rosa, David A. (2015). *Suelos contaminados por metales y metaloides: muestreo y alternativas para su remediación*, volume 1. Instituto Nacional de Ecología (INE-Semarnat) Periférico sur 5000. Col. Insurgentes Cuicuilco, C.P. 04530. México, D.F. [www.ine.gob.mx](#). [2.4.1](#), [2.4.2](#), [2.4.4](#)
- Yrigoyen, Coro Chasco (2003). *Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales*. Madrid: Consejería de Economía e Innovación Tecnológica. [2.11](#), [2.14](#), [2.11](#), [2.11.1](#)

Marco Teórico

n-1	nivel de probabilidad			
	0.2	0.1	0.5	0.01
9	1.383	1.833	2.262	3.250
10	1.372	1.812	2.228	3.169
11	1.363	1.796	2.201	3.106
12	1.356	1.782	2.179	3.055
13	1.350	1.771	2.160	3.012
14	1.345	1.761	2.145	2.977
15	1.341	1.753	2.131	2.947
16	1.337	1.746	2.120	2.921
17	1.333	1.740	2.110	2.898
18	1.330	1.734	2.101	2.878
19	1.328	1.729	2.093	2.861
20	1.325	1.725	2.086	2.845
21	1.323	1.721	2.080	2.831
22	1.321	1.717	2.074	2.819
23	1.319	1.714	2.069	2.807
24	1.318	1.711	2.064	2.797
25	1.316	1.708	2.060	2.787

Tabla A.1: Prueba T de dos colas

Valores Muestreo



n	x	y	Ndvi	n	x	y	Ndvi
1	34.027.620.663.885	6.373.684.865.915	0.4615	41	34.036.787.568.989	63.737.821.661.099	0.4336
2	3.402.819.127.631	63.737.029.322.377	0.4717	42	34.037.289.740.906	63.736.564.816.926	0.4516
3	34.028.761.888.735	63.737.209.885.624	0.4579	43	34.037.459.605.299	6.373.737.365.229	0.4504
4	34.029.428.977.539	6.373.679.177.073	0.4607	44	34.037.358.181.414	63.738.002.324.325	0.4268
5	34.029.327.553.654	63.737.420.442.766	0.4517	45	34.037.860.353.332	63.736.745.380.173	0.4594
6	34.030.008.660.393	63.736.972.333.977	0.4668	46	34.038.034.340.646	63.737.584.109.452	0.4429
7	34.029.907.236.508	63.737.601.006.012	0.4522	47	34.037.932.916.761	63.738.212.781.487	0.421
8	34.030.579.272.818	63.737.152.897.224	0.4599	48	34.038.440.036.186	63.736.926.043.399	0.456
9	34.030.472.901.427	63.737.811.563.154	0.4448	49	34.038.604.953.072	63.737.764.772.678	0.4355
10	34.031.144.937.737	63.737.363.454.365	0.455	50	34.038.503.529.187	63.738.393.444.714	0.4254
11	34.031.043.513.852	63.737.992.126.401	0.4324	51	34.039.107.124.989	63.736.507.828.526	0.4465
12	34.031.546.510.354	63.736.735.282.228	0.4557	52	34.039.006.525.689	63.737.136.500.561	0.4519
13	34.031.715.550.162	63.737.544.017.612	0.4472	53	34.039.175.565.497	63.737.945.335.925	0.429
14	34.031.614.126.277	63.738.172.689.648	0.4186	54	34.039.677.737.414	63.736.688.491.752	0.4544
15	34.032.117.122.779	63.736.915.845.475	0.4601	55	34.039.577.138.114	63.737.317.163.788	0.4519
16	34.032.291.110.094	63.737.754.574.754	0.4415	56	34.039.741.230.416	63.738.155.793.087	0.4229
17	34.032.189.686.209	6.373.838.324.679	0.4131	57	3.404.024.834.984	63.736.869.054.999	0.4544
18	34.032.696.805.633	63.737.096.408.722	0.4578	58	34.040.142.803.033	6.373.752.762.095	0.4433
19	34.032.861.722.519	63.737.935.138.001	0.4306	59	34.040.312.667.425	63.738.336.456.314	0.4186
20	34.032.760.298.635	63.738.563.810.036	0.4225	60	34.040.814.839.343	63.737.079.612.141	0.4502
21	34.033.363.894.437	63.736.678.293.828	0.4601	61	34.040.713.415.458	63.737.708.184.197	0.4347
22	34.033.262.470.552	63.737.306.965.863	0.4502	62	34.041.224.657.804	63.736.451.440.003	0.4457
23	34.033.432.334.945	63.738.115.701.248	0.4201	63	34.041.495.946.082	63.736.631.503.352	0.4516
24	34.033.934.506.862	63.736.858.857.075	0.4636	64	34.041.394.522.197	63.737.260.175.387	0.4485
25	34.033.833.082.977	6.373.748.752.911	0.446	65	34.041.293.098.312	63.737.888.847.423	0.4239
26	34.033.997.999.863	63.738.326.258.389	0.4188	66	34.042.061.611	63.736.842.060.494	0.451
27	34.034.505.119.288	63.737.039.420.322	0.4588	67	34.041.960.187.116	63.737.470.732.529	0.4423
28	34.034.398.747.896	63.737.698.086.252	0.4401	68	34.041.858.763.231	63.738.099.304.585	0.422
29	34.034.568.612.289	63.738.506.821.636	0.4161	69	34.042.632.223.426	6.373.702.262.374	0.4481
30	34.035.172.208.091	63.736.621.305.428	0.4559	70	34.042.530.799.541	63.737.651.295.776	0.4343
31	34.035.070.784.206	63.737.249.977.463	0.4519	71	3.404.243.020.024	63.738.279.967.812	0.4271
32	34.034.969.360.321	63.737.878.649.499	0.4318	72	34.043.202.835.851	63.737.203.186.987	0.4444
33	34.035.751.890.945	63.736.801.868.674	0.4618	73	34.043.101.411.966	63.737.831.859.023	0.4358
34	34.035.651.291.645	6.373.743.054.071	0.4472	74	34.043.778.395.783	63.737.413.744.129	0.4383
35	3.403.554.986.776	63.738.059.212.746	0.4212	75	34.043.676.971.898	63.738.042.416.165	0.4207
36	3.403.632.250.337	63.736.982.531.901	0.4592	76	34.044.349.008.208	63.737.594.307.376	0.4328
37	34.036.216.956.563	63.737.641.097.852	0.4426	77	34.044.247.584.323	63.738.222.979.412	0.4175
38	34.036.115.532.679	63.738.269.769.887	0.42	78	34.044.919.620.633	63.737.774.870.623	0.4264
39	34.036.386.820.956	63.738.449.833.236	0.4207	79	34.045.485.285.552	63.737.985.427.764	0.4272
40	34.036.888.992.874	63.737.192.989.063	0.4562	80	34.046.055.897.977	63.738.165.991.011	0.4203

Tabla B.1: Puntos Originales

n	Ndvi	x	y	n	x	y	Ndvi
73	0.4186	34.033.094.255.329	63.738.354.752.589	276	34.044.247.584.323	63.738.222.979.412	0.4175
147	0.4551	34.037.174.299.086	63.737.283.270.686	134	34.036.670.478	63.736.683.592.749	0.4585
119	0.436	34.035.597.693.657	63.737.759.873.675	80	34.033.934.506.862	63.736.858.857.075	0.4636
263	0.42	34.043.048.638.563	63.738.161.191.988	115	34.035.196.945.624	63.738.388.045.813	0.4198
29	0.447	3.403.075.820.764	63.737.901.844.778	51	34.032.358.726.017	63.737.335.460.063	0.4548
149	0.4288	34.037.072.875.201	63.737.911.942.722	124	34.035.936.597.857	63.737.520.922.313	0.4453
225	0.4509	34.041.109.215.984	63.737.169.893.764	135	34.036.603.686.661	6.373.710.270.744	0.4634
47	0.4306	34.031.948.082.971	63.737.963.632.201	256	34.042.917.529.638	63.737.112.905.364	0.445
207	0.4544	3.404.024.834.984	63.736.869.054.999	208	34.040.195.576.436	63.737.198.387.964	0.4532
283	0.4272	34.045.485.285.552	63.737.985.427.764	45	34.032.063.524.792	6.373.724.517.844	0.4547
270	0.4207	34.043.676.971.898	63.738.042.416.165	274	34.044.349.008.208	63.737.594.307.376	0.4328
182	0.4552	34.039.059.299.092	63.736.807.267.576	277	34.044.634.314.421	63.737.684.588.999	0.4304
92	0.4665	34.034.552.945.185	63.736.740.081.251	262	34.043.101.411.966	63.737.831.859.023	0.4358
61	0.4204	34.032.474.992.422	63.738.473.528.413	12	34.029.327.553.654	63.737.420.442.766	0.4517
120	0.4212	3.403.554.986.776	63.738.059.212.746	57	34.032.696.805.633	63.737.096.408.722	0.4578
44	0.4601	34.032.117.122.779	63.736.915.845.475	103	34.034.568.612.289	63.738.506.821.636	0.4161
99	0.4551	34.034.785.477.994	6.373.715.969.584	5	34.028.476.582.523	63.737.119.604	0.4671
278	0.4239	34.044.580.716.433	63.738.013.921.965	210	34.040.080.134.616	63.737.916.841.725	0.4272
26	0.4589	34.030.927.247.448	63.736.854.058.051	114	34.035.264.561.547	63.737.968.931.122	0.4265
150	0.4215	34.037.020.101.798	63.738.241.275.687	228	34.040.940.176.176	6.373.821.768.049	0.4221
68	0.4601	34.033.363.894.437	63.736.678.293.828	76	34.033.547.776.765	63.737.397.247.487	0.4493
105	0.4605	3.403.512.355.761	63.736.920.644.498	169	34.038.488.686.667	63.736.626.604.349	0.4567
4	0.4584	3.402.852.440.842	6.373.682.026.493	243	34.041.960.187.116	63.737.470.732.529	0.4423
65	0.4306	34.032.861.722.519	63.737.935.138.001	195	34.039.624.964.011	63.737.017.724.738	0.4508
23	0.4599	34.030.579.272.818	63.737.152.897.224	145	34.037.289.740.906	63.736.564.816.926	0.4516
74	0.4645	3.403.364.920.065	63.736.768.575.451	22	34.030.641.941.235	63.736.763.776.428	0.4572
190	0.4535	34.039.291.831.901	63.737.226.882.164	183	34.039.006.525.689	63.737.136.500.561	0.4519
240	0.4214	34.041.573.457.018	63.738.009.022.962	66	34.032.813.072.038	63.738.234.477.071	0.4153
136	0.4492	3.403.655.503.618	6.373.740.204.651	238	3.404.167.982.841	63.737.350.457.011	0.4467
21	0.4495	34.030.192.542.721	63.737.691.287.636	49	34.032.459.325.318	63.736.706.788.028	0.4646
245	0.422	34.041.858.763.231	63.738.099.304.585	161	3.403.763.854.012	63.738.122.499.864	0.4201
7	0.4579	34.028.761.888.735	63.737.209.885.624	157	34.037.860.353.332	63.736.745.380.173	0.4594
156	0.4231	34.037.305.408.011	63.738.331.557.311	188	34.039.392.431.202	63.736.598.210.129	0.4492
174	0.4237	34.038.218.222.974	63.738.303.063.111	67	34.032.760.298.635	63.738.563.810.036	0.4225
109	0.4168	34.034.902.568.983	63.738.297.764.189	285	34.046.055.897.977	63.738.165.991.011	0.4203
176	0.4544	34.038.721.219.476	63.737.046.218.938	108	34.034.969.360.321	63.737.878.649.499	0.4318
15	0.4519	34.029.621.930.296	63.737.510.724.389	213	34.029.974.852.431	63.737.181.891.322	0.4640
282	0.423	34.045.151.328.858	63.738.194.485.211	214	34.029.969.904.925	63.737.211.785.237	0.4628
269	0.4295	34.043.729.745.301	63.737.713.083.199	215	34.029.964.957.418	63.737.241.779.132	0.4685
117	0.4573	34.035.699.117.542	6.373.713.120.164	216	34.029.960.834.496	63.737.271.673.047	0.4600

Tabla B.2: Puntos aleatorios

n	Ndvi	x	y	n	x	y	Ndvi
29	0.447	3.403.075.820.764	63.737.901.844.778	179	34.038.552.179.668	63.738.094.005.664	0.4222
174	0.4237	34.038.218.222.974	63.738.303.063.111	245	34.041.858.763.231	63.738.099.304.585	0.422
187	0.4178	34.038.837.485.881	63.738.184.287.287	31	34.031.144.937.737	63.737.363.454.365	0.455
281	0.4232	34.045.199.979.339	63.737.895.146.141	39	34.031.831.816.567	63.736.825.563.851	0.4632
147	0.4551	34.037.174.299.086	63.737.283.270.686	259	3.404.276.333.235	63.738.070.910.365	0.4182
237	0.4528	34.041.727.654.306	63.737.051.117.941	235	34.041.225.482.389	63.738.307.962.114	0.4149
168	0.421	34.037.932.916.761	63.738.212.781.487	258	34.042.816.105.753	63.737.741.577.399	0.4294
201	0.4561	34.039.963.043.627	63.736.778.773.375	193	34.039.122.792.093	63.738.274.568.911	0.4247
43	0.4186	34.031.614.126.277	63.738.172.689.648	27	34.030.859.631.524	63.737.273.172.742	0.4562
51	0.4548	34.032.358.726.017	63.737.335.460.063	121	34.035.482.251.837	63.738.478.327.436	0.4199
181	0.4465	34.039.107.124.989	63.736.507.828.526	112	34.035.365.985.432	63.737.340.259.087	0.4527
223	0.4457	34.041.224.657.804	63.736.451.440.003	143	34.036.787.568.989	63.737.821.661.099	0.4336
209	0.4433	34.040.142.803.033	6.373.752.762.095	30	3.403.121.255.366	63.736.944.339.675	0.4618
132	0.42	34.036.115.532.679	63.738.269.769.887	4	3.402.852.440.842	6.373.682.026.493	0.4584
173	0.4271	34.038.271.820.962	63.737.973.830.125	180	34.038.503.529.187	63.738.393.444.714	0.4254
265	0.4345	3.404.343.536.866	63.737.622.801.576	272	34.044.015.051.514	63.737.803.364.823	0.426
207	0.4544	3.404.024.834.984	63.736.869.054.999	120	3.403.554.986.776	63.738.059.212.746	0.4212
80	0.4636	34.033.934.506.862	63.736.858.857.075	61	34.032.474.992.422	63.738.473.528.413	0.4204
69	0.457	34.033.315.243.956	63.736.977.632.898	284	34.045.770.591.765	63.738.075.709.388	0.4226
17	0.4573	3.402.995.588.699	63.737.301.666.942	88	34.034.167.039.671	63.737.278.471.663	0.4505
282	0.423	34.045.151.328.858	63.738.194.485.211	117	34.035.699.117.542	6.373.713.120.164	0.4573
205	0.4229	34.039.741.230.416	63.738.155.793.087	159	34.037.744.911.511	63.737.463.933.913	0.4484
21	0.4495	34.030.192.542.721	63.737.691.287.636	65	34.032.861.722.519	63.737.935.138.001	0.4306
113	0.4422	34.035.312.387.444	63.737.669.592.052	224	34.041.161.989.388	63.736.840.560.799	0.4552
11	0.4649	34.029.380.327.058	63.737.911.098	108	34.034.969.360.321	63.737.878.649.499	0.4318
275	0.4216	3.404.429.541.022	63.737.923.640.341	5	34.028.476.582.523	63.737.119.604	0.4671
218	0.4558	34.040.867.612.746	63.736.750.279.175	152	34.037.526.396.638	63.736.954.537.599	0.4584
273	0.4214	34.043.962.278.111	63.738.132.697.788	91	34.033.997.999.863	63.738.326.258.389	0.4188
49	0.4646	34.032.459.325.318	63.736.706.788.028	243	34.041.960.187.116	63.737.470.732.529	0.4423
66	0.4153	34.032.813.072.038	63.738.234.477.071	268	34.043.778.395.783	63.737.413.744.129	0.4383
247	0.4479	34.042.298.266.732	63.737.231.681.187	163	34.038.203.380.454	63.736.536.322.726	0.4514
153	0.4504	34.037.459.605.299	6.373.737.365.229	146	34.037.241.090.425	63.736.864.155.996	0.459
256	0.445	34.042.917.529.638	63.737.112.905.364	105	3.403.512.355.761	63.736.920.644.498	0.4605
16	0.4668	34.030.008.660.393	63.736.972.333.977	97	34.034.283.306.076	63.738.416.540.013	0.4178
257	0.4402	34.042.864.756.235	63.737.442.238.329	204	34.039.794.828.403	63.737.826.560.102	0.4291
118	0.4472	34.035.651.291.645	6.373.743.054.071	176	34.038.721.219.476	63.737.046.218.938	0.4544
222	0.4249	34.040.645.799.535	63.738.127.298.887	24	34.030.526.499.415	63.737.482.230.189	0.4523
144	0.4197	34.036.734.795.585	63.738.150.994.064	188	34.039.392.431.202	63.736.598.210.129	0.4492
285	0.4203	34.046.055.897.977	63.738.165.991.011	137	34.036.502.262.776	63.737.731.379.475	0.4401
148	0.4419	34.037.120.701.098	63.737.612.603.652	200	34.040.011.694.108	63.736.479.334.325	0.4498

Tabla B.3: Puntos optimización

n	x	y	Ndvi	n	x	y	Ndvi
1931	34.038.208.327.961	6.373.650.642.881	0.4485	1558	34.036.188.096.108	63.737.820.761.282	0.4314
2070	34.038.802.853.335	63.736.537.322.522	0.4514	2193	34.038.847.380.894	63.738.124.399.477	0.4221
2474	34.040.036.431.641	6.373.818.628.688	0.4231	468	34.031.377.470.546	63.737.783.068.954	0.4452
3068	34.043.087.394.031	63.737.921.640.748	0.4233	1028	34.034.282.481.492	63.736.560.017.902	0.4558
1671	34.036.912.905.822	63.737.043.319.528	0.4584	245	34.030.308.809.125	63.736.972.833.875	0.4654
2167	34.038.972.717.727	63.737.346.057.906	0.4514	2892	34.042.327.951.771	63.737.052.017.757	0.449
2529	34.040.404.196.297	63.737.767.672.088	0.4319	2177	34.038.924.067.246	63.737.645.496.956	0.4433
2100	34.038.658.551.059	63.737.435.439.713	0.4502	3325	34.045.746.678.816	63.738.223.179.371	0.4203
3297	34.045.376.440.407	63.737.788.967.753	0.4229	437	34.031.526.720.328	63.736.854.957.868	0.4588
2833	34.042.042.645.559	63.736.961.736.134	0.4497	1670	34.036.917.853.329	63.737.013.425.613	0.4607
527	34.031.672.671.772	63.737.813.462.767	0.4422	1659	34.036.970.626.732	63.736.684.092.647	0.4571
1427	34.035.544.920.253	63.738.089.206.641	0.4212	1007	34.033.746.501.612	63.738.026.419.421	0.4244
2948	34.042.613.257.984	63.737.142.299.381	0.4463	234	34.029.986.396.613	63.737.807.364.009	0.444
2452	34.040.142.803.033	6.373.752.762.095	0.4433	403	34.031.125.972.295	63.737.483.229.985	0.4515
2310	34.039.552.400.581	63.737.466.833.323	0.4524	2935	34.042.120.981.079	63.738.333.756.863	0.4183
2902	3.404.227.930.129	63.737.351.456.807	0.4437	407	34.031.106.182.269	63.737.602.905.626	0.4463
425	34.031.029.495.917	63.738.141.795.936	0.4203	1335	34.035.356.090.419	63.737.400.146.897	0.448
2448	34.040.161.768.475	63.737.407.945.309	0.4501	3207	34.044.329.218.182	63.737.714.082.996	0.4289
892	34.033.668.166.092	63.736.648.799.831	0.4603	482	34.031.309.854.623	63.738.202.183.644	0.4181
510	3.403.175.430.563	63.737.304.566.352	0.4556	224	34.029.922.079.028	63.737.511.224.287	0.4516
1898	34.037.720.173.979	63.737.613.603.448	0.4294	3323	34.045.755.749.245	63.738.165.491.113	0.4231
915	34.033.557.671.778	63.737.337.359.677	0.4519	2914	3.404.222.158.038	63.737.710.683.688	0.4308
1553	34.036.212.009.057	63.737.670.991.767	0.4414	2835	34.042.032.750.545	63.737.021.623.944	0.4505
1466	34.035.994.318.767	63.737.161.595.453	0.4562	1110	34.034.524.084.729	63.736.919.744.681	0.4637
2702	34.041.476.156.055	63.736.751.278.972	0.454	3026	34.042.773.227.363	63.738.011.022.555	0.4203
1152	34.034.322.061.544	63.738.177.088.752	0.4162	2715	34.041.413.487.639	63.737.140.399.767	0.4536
672	34.032.194.633.716	63.738.353.252.895	0.4123	2583	34.040.781.031.381	63.737.289.169.486	0.4487
3256	34.044.910.550.205	63.737.655.095.003	0.4278	2518	34.040.569.697	63.737.438.339.123	0.4441
302	34.030.545.464.857	63.737.362.454.569	0.4566	138	34.029.395.169.577	63.737.001.328.075	0.4703
617	34.032.459.325.318	63.736.706.788.028	0.4646	695	34.032.715.771.075	63.736.976.733.081	0.4596
2671	34.041.002.844.593	63.737.828.459.715	0.4266	912	34.033.571.689.713	63.737.247.577.952	0.451
1956	3.403.808.711.405	63.737.254.876.466	0.4544	2322	34.039.494.679.671	63.737.826.060.203	0.4315
663	34.032.237.512.106	63.738.083.907.719	0.4221	3320	34.045.770.591.765	63.738.075.709.388	0.4226
1429	3.403.553.502.524	63.738.149.094.451	0.4189	2598	34.040.708.467.951	63.737.738.178.091	0.4343
3181	34.044.005.156.501	63.737.863.252.633	0.4236	2766	34.041.776.304.788	6.373.675.167.889	0.4566
1709	34.036.729.848.079	63.738.180.887.979	0.42	2094	34.038.687.411.515	63.737.255.776.283	0.4537
1133	34.034.413.590.416	63.737.608.304.527	0.436	1289	34.034.931.429.438	63.738.118.100.759	0.4219
829	34.033.339.981.489	63.736.827.963.363	0.4629	127	34.029.586.473.165	63.736.686.592.139	0.4481
1508	34.035.791.470.998	63.738.418.939.524	0.4211	2507	34.040.509.743.104	63.737.109.006.158	0.4478
503	34.031.788.113.592	63.737.095.009.007	0.4578	2502	34.040.533.656.052	63.736.959.336.622	0.4545

Tabla B.4: Puntos aleatorios de toda la población

Códigos de Programación

```
data <- read.table("C:\\Users\\Catalina Reynolds\\Desktop\\R\\cuartel1.txt", header = TRUE)
attach(data)
summary(Ndvi)
sd(Ndvi)
power.t.test( n=80, delta = 1, sd = 0.01542347, sig.level = 0.05, type = "one.sample", alternative
= "one.sided")
#n =number of observations (per group)
#delta =true difference in means
#sd =standard deviation
#sig.level =significance level (Type I error probability)
#power =power of test (1 minus Type II error probability)
#type =string specifying the type of t test. Can be abbreviated
```

Tabla C.1: Estimación tamaño de muestra

```

data <- read.table("C:\\Users\\Catalina Reynolds\\Desktop\\R\\cuartel1.txt", header = TRUE)
setwd("C:\\Users\\Catalina Reynolds\\Desktop\\R")
coordinates(data) <- ~ x + y
#grilla
x.range <- as.integer(range(data@coords[,1]))
y.range <- as.integer(range(data@coords[,2]))
grd <- expand.grid(x=seq(from=x.range[1], to=x.range[2], by=3.2), y=seq(from=y.range[1],
to=y.range[2], by=3.2)) coordinates(grd) <- ~ x+y
plot(grd, cex=1.5)
points(data, pch=1, col='red', cex=1)
#variogramas
variogcloud<-variogram(Ndvi~1, data=data, cloud=TRUE)
plot(variogcloud, xlab="Distancia",ylab="Semivarianza",main="Nube variográfica")
ndvi.vg <- variogram(Ndvi~1, data=data) ndvi.fit <- fit.variogram(ndvi.vg , model=vgm(0.00025,
"Gau", 100, 0))
plot(ndvi.vg,ndvi.fit,xlab="Distancia",ylab="Semivarianza",main="Variograma")
#idw
idw <- idw(Ndvi ~ 1, data, grd)
RMSE.idw <- sqrt(sum(idw$var1.var^2)/length(idw$var1.var))
RMSE.idw idw.output = as.data.frame(idw)
MSE.idw<- (sum(idw$var1.var^2)/length(idw$var1.var))
MSE.idw
names(idw.output)[1:3] <- c ("x","y","Ndvi")
#idw.outputmapa
plot <- ggplot(data=idw.output,aes(x= mpas,y=y))
layer1 <- c(geom_tile(data=idw.output,aes(fill=Ndvi)))
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() +theme_bw()
#kriging
krig <- krige(Ndvi ~ 1, data, grd, model=ndvi.fit)
krig.output=as.data.frame(krig)
RMSE.krig <- sqrt(sum(krig$var1.var^2)/length(krig$var1.var))
RMSE.krig
MSE.krig<- (sum(krig$var1.var^2)/length(idw$var1.var))
MSE.krig
names(krig.output)[1:3]<-c("x","y","Ndvi")
#kirg.outputmapa
plot <- ggplot(data=krig.output,aes(x=x,y=y))
layer1 <- c(geom_tile(data=krig.output,aes(fill=Ndvi)))
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() + theme_bw()

```

Tabla C.2: Prueba modelo de predicción e interpolación

```

datos <- read.table("C:\\Users\\Catalina Reynolds\\Desktop\\R\\Puntos_originales.txt", header =
TRUE)
data <- read.table("C:\\Users\\Catalina Reynolds\\Desktop\\R\\cuartel1.txt", header = TRUE)
#grilla
coordinates(data) <- ~ x + y
coordinates(datos) <- ~ x + y
x.range <- as.integer(range(data@coords[,1]))
y.range <- as.integer(range(data@coords[,2]))
grd <- expand.grid(x=seq(from=x.range[1], to=x.range[2], by=3.2), y=seq(from=y.range[1],
to=y.range[2], by=3.2))
coordinates(grd) <- ~ x+y
#distancia inversa
idw <- idw(Ndvi ~ 1, data, grd)
idw1 <- idw(Ndvi ~ 1, datos, grd)
#Error
RMSE.idw1 <- sqrt(sum((idw$var1.pred-idw1$var1.pred)^2)/length(idw$var1.pred))
RMSE.idw1
MSE.idw1 <- (sum((idw$var1.pred-idw1$var1.pred)^2)/length(idw$var1.pred))
MSE.idw1
#Mapas para puntos originales
idw1.output = as.data.frame(idw1)
names(idw1.output)[1:3] <- c ("x","y","Ndvi")
puntos1 = as.data.frame(datos)
plot <- ggplot(data=idw1.output,aes(x=x,y=y))
layer1 <- c(geom_tile(data=idw1.output,aes(fill=Ndvi)))
layer2 <- c(geom_point(data=puntos1, aes(x=x, y=y), color="black"))
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() +theme_bw()
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() +theme_bw()+ la-
yer2
col1 <- ((idw$var1.pred-idw1$var1.pred)^2)/length(idw$var1.pred)
nuevo1 <- cbind(MSE = col1, idw1.output)
plot <- ggplot(data=nuevo1,aes(x=x,y=y))
layer1 <- c(geom_tile(data=nuevo1,aes(fill=MSE)))
plot+layer1+scale_fill_gradientn(colours = topo.colors(400))+coord_equal() +theme_bw()
#variograma
ndvi.vg <- variogram(Ndvi~1, data=puntos_)
ndvi.fit <- fit.variogram(ndvi.vg , model=vgm(0.00025, "Gau", 100, 0))
plot(ndvi.vg,ndvi.fit,xlab="Distancia",ylab="Semivarianza",main="Variograma")
#Kriging
krig6 <- krig(Ndvi ~ 1, datos, grd, model=ndvi.fit)
krig6.output=as.data.frame(krig6)
RMSE.krig6 <- sqrt(sum((idw$var1.pred-krig6$var1.pred)^2)/length(krig$var1.pred))
RMSE.krig6
MSE.krig6 <- (sum((idw$var1.pred-krig6$var1.pred)^2)/length(krig$var1.pred))
MSE.krig6

```

Tabla C.3: Algoritmo de optimización

```

names(krig6.output)[1:3]<-c("x","y","Ndvi")
plot <- ggplot(data=krig6.output,aes(x=x,y=y))
layer1 <- c(geom_tile(data=krig6.output,aes(fill=Ndvi)))
layer2 <- c(geom_point(data=puntos1, aes(x=x, y=y), color="black"))
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() + theme_bw()
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() +theme_bw()+ la-
yer2
col6<- ((idw$var1.pred-krig6$var1.pred)^2)/length(idw$var1.pred)
nuevo6 <-cbind(MSE = col6, krig6.output)
plot <- ggplot(data=nuevo6,aes(x=x,y=y))
layer1 <- c(geom_tile(data=nuevo6,aes(fill=RMSE)))
plot+layer1+scale_fill_gradientn(colours = topo.colors(400))+coord_equal() +theme_bw()
RMSE.krig7 <- sqrt(sum((krig$var1.pred-krig6$var1.pred)^2)/length(krig$var1.pred))
RMSE.krig7
MSE.krig7 <- (sum((krig$var1.pred-krig6$var1.pred)^2)/length(krig$var1.pred)) MSE.krig7
col7<- ((krig$var1.pred-krig6$var1.pred)^2)/length(krig$var1.pred)
nuevo7 <-cbind(MSE = col7, krig6.output)
plot <- ggplot(data=nuevo7,aes(x=x,y=y))
layer1 <- c(geom_tile(data=nuevo7,aes(fill=MSE)))
plot+layer1+scale_fill_gradientn(colours = topo.colors(400))+coord_equal() +theme_bw()
#Ejemplo iteración intermedia
mysample <- valores_10x10[sample(1:nrow(valores_10x10), 80, replace=FALSE),]
coordinates(mysample) <- ~ x + y
idw2 <- idw(Ndvi ~ 1, mysample, grd)
RMSE.idw2 <- sqrt(sum((idw$var1.pred-idw2$var1.pred)^2)/length(idw$var1.pred))
RMSE.idw2
write.table(mysample, "C:\\Users\\Catalina Reynolds\\Desktop\\R\\aleatorio.txt", sep="t")
idw2.output = as.data.frame(idw2) puntos = data.frame(mysample)
names(idw2.output)[1:3] <- c ("x","y","Ndvi")
plot <- ggplot(data=idw2.output,aes(x=x,y=y))
layer1 <- c(geom_tile(data=idw2.output,aes(fill=Ndvi)))
layer2 <- c(geom_point(data=puntos, aes(x=x, y=y), color="black"))
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() +theme_bw()
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() +the-
me_bw()+layer2
col2<- ((idw$var1.pred-idw2$var1.pred)^2)/length(idw$var1.pred)
sum (col2)
nuevo2 <-cbind(MSE = col2, idw2.output)
write.table(idw2.output, "C:\\Users\\Catalina Reynolds\\Desktop\\R\\idw2outp.txt", sep="t")
plot <- ggplot(data=nuevo2,aes(x=x,y=y))
layer1 <- c(geom_tile(data=nuevo2,aes(fill=MSE)))
plot+layer1+scale_fill_gradientn(colours = topo.colors(400))+coord_equal() +theme_bw()

```

Tabla C.4: Algoritmo de optimización (continuación)

```

#iteración
a <- datos
b <- idw(Ndvi ~ 1, a, grd)
c <- sqrt(sum((idw$var1.pred-a$var1.pred)^2)/length(idw$var1.pred))
for (i in 1:1000000) {
  muestra <- valores_10x10[sample(1:nrow(valores_10x10), 80, replace=FALSE),]      coord-
dinates(muestra) <- ~ x + y
  idw3 <- idw(Ndvi ~ 1, muestra, grd)
  RMSE.idw3 <- sqrt(sum((idw$var1.pred-idw3$var1.pred)^2)/length(idw$var1.pred))
  if (RMSE.idw3 <= c)
  {
    a <- muestra
    b <- idw(Ndvi ~ 1, a, grd)
    c <- sqrt(sum((idw$var1.pred-a$var1.pred)^2)/length(idw$var1.pred))
    write.table(a, "C:\\Users\\Catalina Reynolds\\Desktop\\R\\mejor.txt", sep="\t")
  }
  i= 1+i
}
b <- idw(Ndvi ~ 1, a, grd)
c <- sqrt(sum((idw$var1.pred-b$var1.pred)^2)/length(idw$var1.pred))
c
RMSE.idw1
#mapas opt
b.output = as.data.frame(b)
puntos3 = data.frame(a)
names(b.output)[1:3] <- c ("x","y","Ndvi")
plot <- ggplot(data=b.output,aes(x=x,y=y))
layer1 <- c(geom_tile(data=b.output,aes(fill=Ndvi)))
layer2 <- c(geom_point(data=puntos3, aes(x=x, y=y), color="black"))
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() +theme_bw()
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() +the-
me_bw()+layer2
col3<- ((idw$var1.pred-b$var1.pred)^2)/length(idw$var1.pred)
sum(col3)
nuevo3 <-cbind(MSE = col3, b.output)
plot <- ggplot(data=nuevo3,aes(x=x,y=y))
layer1 <- c(geom_tile(data=nuevo3,aes(fill=MSE)))
plot+layer1+scale_fill_gradientn(colours = topo.colors(400))+coord_equal() +theme_bw()
ndvi.vg <- variogram(Ndvi~1, data=a)
ndvi.fit <- fit.variogram(ndvi.vg , model=vgm(0.00025, "Gau", 100, 0))
plot(ndvi.vg,ndvi.fit,xlab="Distancia",ylab="Semivarianza",main="Variograma")
krig4 <- krige(Ndvi ~ 1, a, grd, model=ndvi.fit) krig4.output=as.data.frame(krig4)
RMSE.krig4 <- sqrt(sum((idw$var1.pred-krig4$var1.pred)^2)/length(krig4$var1.pred))
RMSE.krig4

```

Tabla C.5: Algoritmo de optimización (continuación)

```

names(krig4.output)[1:3]<-c("x","y","Ndvi")
plot <- ggplot(data=krig4.output,aes(x=x,y=y))
layer1 <- c(geom_tile(data=krig4.output,aes(fill=Ndvi))) plot+layer1+scale_fill_gradientn(colours
= terrain.colors(400))+coord_equal() + theme_bw()
col4<- ((idw$var1.pred-krig4$var1.pred)^2)/length(idw$var1.pred)
sum(col4)
nuevo4 <-cbind(MSE = col4, krig4.output)
plot <- ggplot(data=nuevo4,aes(x=x,y=y))
layer1 <- c(geom_tile(data=nuevo4,aes(fill=MSE)))
plot+layer1+scale_fill_gradientn(colours = topo.colors(400))+coord_equal() +theme_bw()

```

Tabla C.6: Algoritmo de optimización (continuación)

```

data <- read.table("C:\\Users\\Catalina Reynolds\\Desktop\\R\\cuartel1.txt", header = TRUE)
setwd("C:\\Users\\Catalina Reynolds\\Desktop\\R")
muestra_aleatoria <- data[sample(1:nrow(data), 80, replace=FALSE),]
coordinates(data) <- ~ x + y
x.range <- as.integer(range(data@coords[,1]))
y.range <- as.integer(range(data@coords[,2]))
grd <- expand.grid(x=seq(from=x.range[1], to=x.range[2], by=3.2), y=seq(from=y.range[1],
to=y.range[2], by=3.2))
coordinates(grd) <- ~ x+y coordinates(muestra_aleatoria) <- ~ x + y
idw_aleatorio <- idw(Ndvi ~ 1, muestra_aleatoria, grd)
RMSE.idw_aleatorio <- sqrt(sum((idw$var1.pred-idw_aleatorio$var1.pred)^2)/length(idw$var1.pred))
RMSE.idw_aleatorio
write.table(muestra_aleatoria, "C:\\Users\\Catalina Reynolds\\Desktop\\R\\muestreo_aleatorio.csv",
sep="t")
idw_aleatorio.output = as.data.frame(idw_aleatorio)
puntos_aleatorios = data.frame(muestra_aleatoria)
names(idw_aleatorio.output)[1:3] <- c ("x","y","Ndvi")
plot <- ggplot(data=idw_aleatorio.output,aes(x=x,y=y))
layer1 <- c(geom_tile(data=idw_aleatorio.output,aes(fill=Ndvi)))
layer2 <- c(geom_point(data=puntos_aleatorios, aes(x=x, y=y), color="black"))
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() +theme_bw()
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() +the-
me_bw()+layer2
col_aleatoria<- sqrt((idw$var1.pred-idw_aleatorio$var1.pred)^2)/length(idw$var1.pred)
nuevo_aleatorio <-cbind(RMSE = col_aleatoria, idw_aleatorio.output)
plot <- ggplot(data=nuevo_aleatorio,aes(x=x,y=y))
layer1 <- c(geom_tile(data=nuevo_aleatorio,aes(fill=RMSE)))
plot+layer1+scale_fill_gradientn(colours = topo.colors(400))+coord_equal() +theme_bw()

```

Tabla C.8: Estimación tamaño de muestra

```

data <- read.table("C:\\Users\\Catalina Reynolds\\Desktop\\R\\cuartel1.txt", header = TRUE)
setwd("C:\\Users\\Catalina Reynolds\\Desktop\\R")
coordinates(data) <- ~ x + y
coordinates(mejor)<- ~ x + y
#grilla
x.range <- as.integer(range(data@coords[,1]))
y.range <- as.integer(range(data@coords[,2]))
grd <- expand.grid(x=seq(from=x.range[1], to=x.range[2], by=3.2), y=seq(from=y.range[1],
to=y.range[2], by=3.2))
coordinates(grd) <- ~ x+y
ndvi.vg <- variogram(Ndvi~1, data=data)
ndvi.fit <- fit.variogram(ndvi.vg , model=vgm(0.00025, "Gau", 100, 0))
krig <- krige(Ndvi ~ 1, data, grd, model=ndvi.fit) krig.output=as.data.frame(krig)
RMSE.krig <- sqrt(sum(krig$var1.var^2)/length(krig$var1.var))
RMSE.krig
names(krig.output)[1:3]<-c("x","y","Ndvi")
plot <- ggplot(data=krig.output,aes(x=x,y=y))
layer1 <- c(geom_tile(data=krig.output,aes(fill=Ndvi)))
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() + theme_bw()
ndvi1.vg <- variogram(Ndvi~1, data=mejor)
ndvi1.fit <- fit.variogram(ndvi1.vg , model=vgm(0.00025, "Gau", 100, 0))
krig1 <- krige(Ndvi ~ 1, mejor, grd, model=ndvi1.fit)
krig1.output=as.data.frame(krig1)
RMSE.krig1 <- sqrt(sum((krig$var1.pred-krig1$var1.pred)^2)/length(krig1$var1.pred))
RMSE.krig1
names(krig1.output)[1:3]<-c("x","y","Ndvi")
plot <- ggplot(data=krig1.output,aes(x=x,y=y))
layer1 <- c(geom_tile(data=krig1.output,aes(fill=Ndvi)))
layer2 <- c(geom_point(data=mejor, aes(x=x, y=y), color="black"))
plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal() + the-
me_bw() plot+layer1+scale_fill_gradientn(colours = terrain.colors(400))+coord_equal()
+theme_bw()+layer2
col5<- ((krig$var1.pred-krig1$var1.pred)^2)/length(krig$var1.pred)
sum(col5)
nuevo5 <-cbind(MSE = col5, krig1.output)
plot <- ggplot(data=nuevo5,aes(x=x,y=y))
layer1 <- c(geom_tile(data=nuevo5,aes(fill=RMSE)))
plot+layer1+scale_fill_gradientn(colours = topo.colors(400))+coord_equal() +theme_bw()

```

Tabla C.7: Comparación Kriging datos originales vs. datos optimizados

Mapas muestra aleatoria sistemática

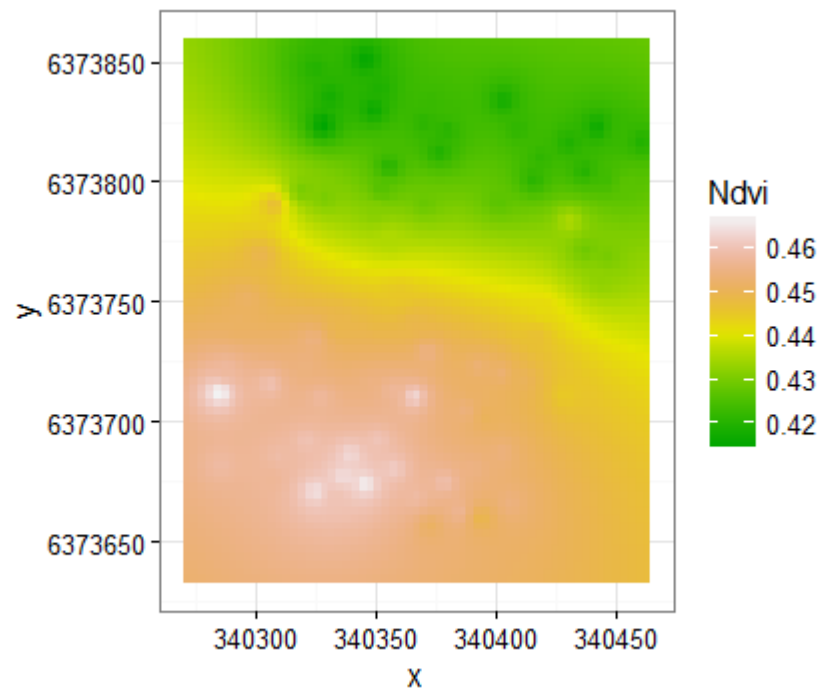


Figura D.1: Mapa mediante IDW realizado con muestra aleatoria dentro de los cuadrantes definidos
(Fuente: Elaboración propia)

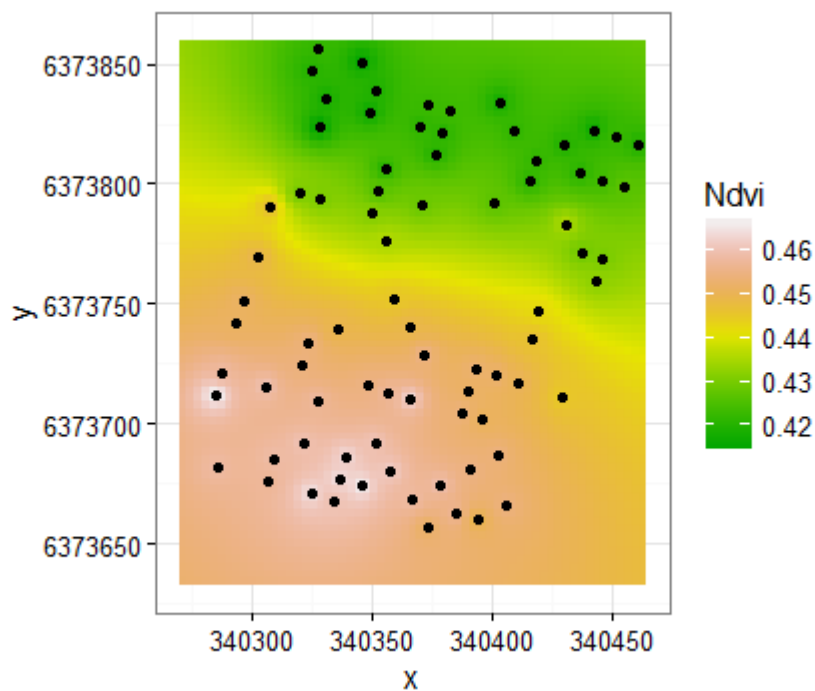


Figura D.2: Mapa mediante IDW realizado con muestra aleatoria dentro de los cuadrantes definidos con puntos de muestreo en negro
(Fuente: Elaboración propia)

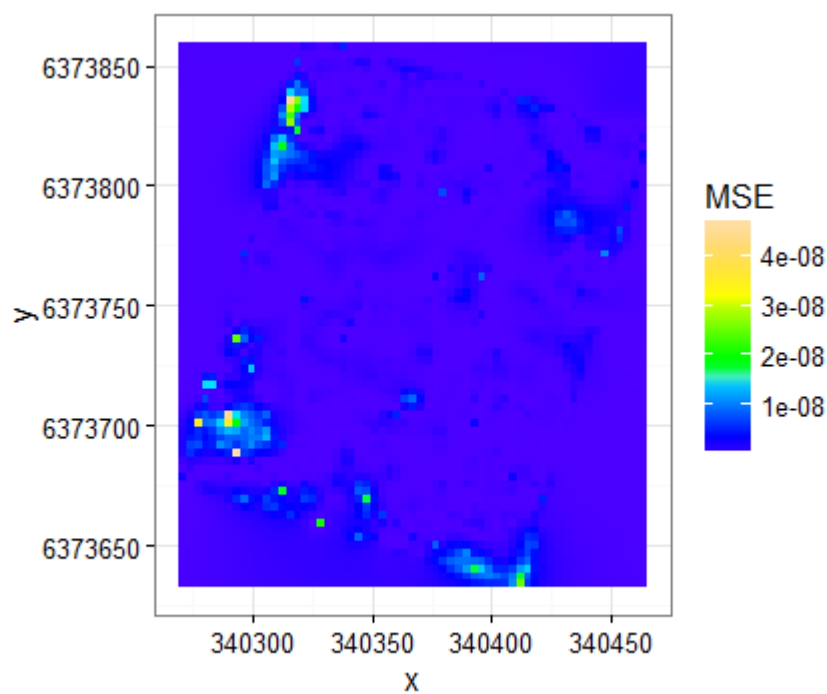


Figura D.3: Mapa MSE con muestra aleatoria dentro de los cuadrantes definidos, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población
(Fuente: Elaboración propia)

Mapas muestreo aleatorio

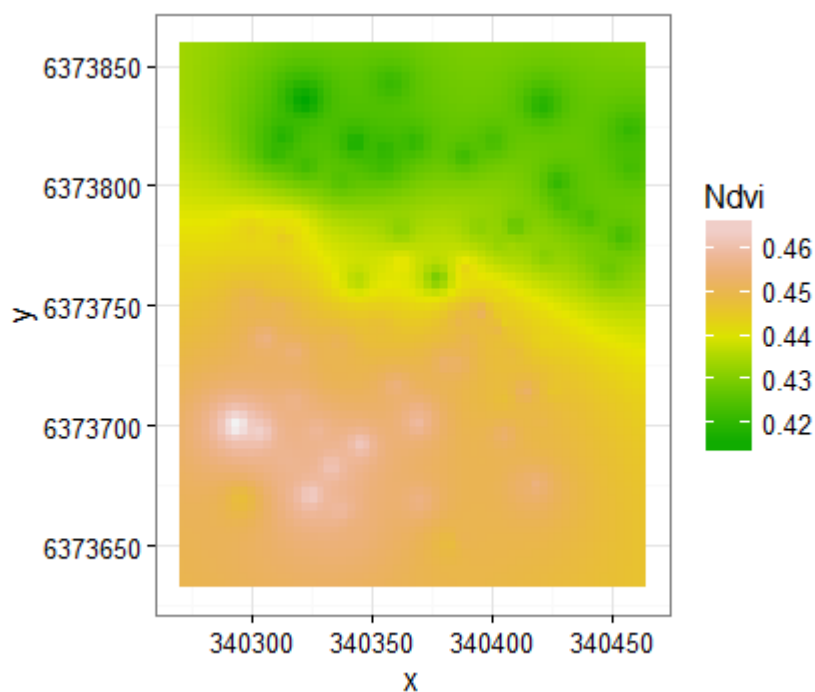


Figura E.1: Mapa mediante IDW realizado con muestreo aleatorio
(Fuente: Elaboración propia)

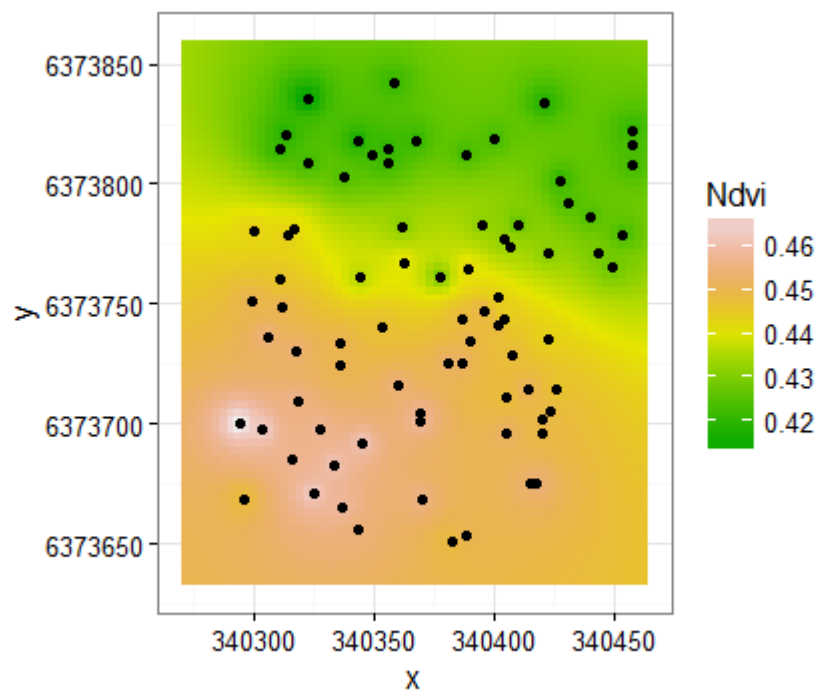


Figura E.2: Mapa mediante IDW realizado con muestreo aleatorio con puntos de muestreo en negro
(Fuente: Elaboración propia)

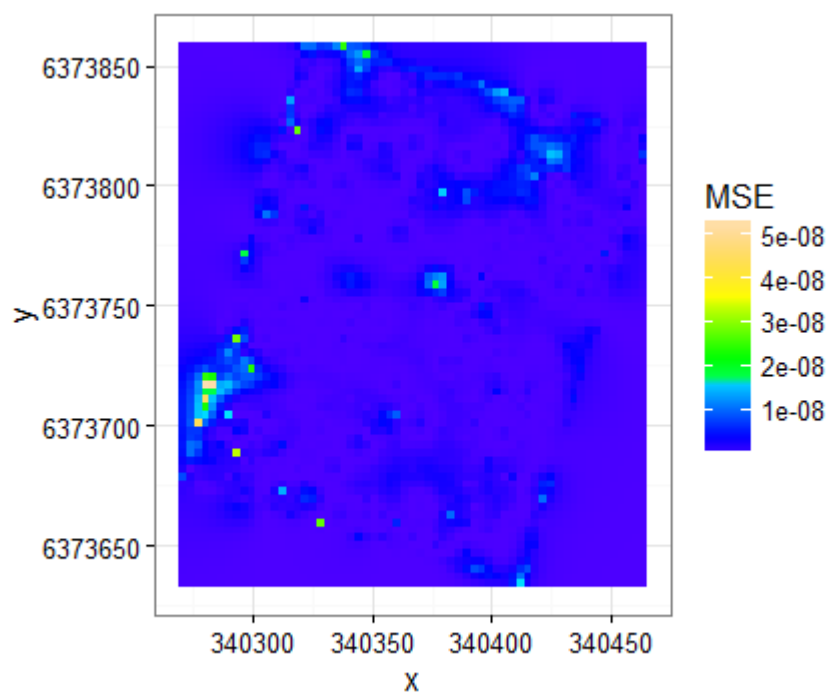


Figura E.3: Mapa MSE realizado con muestreo aleatorio, donde el MSE se calcula con los valores obtenidos mediante IDW para toda la población
(Fuente: Elaboración propia)