
Uso del método de Monte Carlo y estadística inferencial en la estimación de áreas geográficas Casos de Bogotá y Burlington

Use of the Monte Carlo method and inferential statistics in the estimation of geographic areas Cases of Bogotá and Burlington

Oscar Fabián Melo Cabra^a
oscarmelo1091@gmail.com

María Carolina Rojas Velásquez^b
carolinaesp2024@gmail.com

Erika Yulieth Clavijo Miranda^c
Erycla14@gmail.com

Resumen

Este artículo busca por medio de la implementación estadística del método de Monte Carlo determinar áreas de superficies geográficas, tomando a las ciudades de Bogotá y Burlington como casos de estudio usando shapefiles de las mismas. Para cumplir con este propósito se hace uso del paquete estadístico de R Studio, el cual como lenguaje y entorno de programación cuenta con varias ventajas importantes al ejecutar los métodos paramétricos de inferencia estadística tales como estimación y contraste de hipótesis, el primero se aplica mediante la estimación por intervalo de confianza puesto que el resultado obtenido es una variable aleatoria con probabilidad de ocurrencia determinada y error estándar respectivo, el segundo método se emplea al comprobar que los resultados sean válidos respecto a la evidencia empírica obtenida, es decir, comparando las áreas estimadas con las áreas de las ciudades obtenidas de fuentes oficiales. Es de considerar que al generar pocos números aleatorios el error en la estimación del área será bajo, pero la varianza será mayor y la precisión del cálculo al realizar las réplicas será baja, lo cual puede llevar a resultados inexactos.

Palabras clave: lenguaje de programación R, método de Montecarlo, Shapefile, simulación estadística, Superficie geográfica..

^aEgresado Universidad Distrital

^bEgresada Universidad Distrital

^cEgresada Universidad de Manizales

Abstract

This article seeks, through the statistical implementation of the Monte Carlo method, to determine geographic surface areas, taking the cities of Bogotá and Burlington as case studies using their shapefiles.

To fulfill this purpose, the R Studio statistical package is used, which as a programming language and environment has several important advantages when executing parametric methods of statistical inference such as estimation and hypothesis testing, the first is applied through estimation by confidence interval since the result obtained is a random variable with a determined probability of occurrence and respective standard error, the second method is used to verify that the results are valid with respect to the empirical evidence obtained, that is, by comparing the estimated areas with city areas obtained from sources officers. It should be considered that by generating few random numbers the error in the area estimate will be low, but the variance will be greater and the precision of the calculation when carrying out the replications will be low, which can lead to inaccurate results.

Keywords: R programming language, Montecarlo method, Shapefile statistical simulation, geographical surface..

1. Introducción

La simulación estadística busca resolver problemas probabilísticos particularmente complejos, según López (2022) La Simulación se entiende como la reproducción de un fenómeno real mediante otro más sencillo y más adecuado para ser estudiado. En el caso de los fenómenos estadísticos la consideramos como una técnica que consiste en realizar experimentos de muestreo sobre el modelo de un sistema bajo determinadas condiciones.

La simulación entonces permite evaluar diferentes cursos de acciones de un evento dado que responden a suposiciones y ambientes preestablecidos, cuyo objetivo es dar facilidad en la toma de decisiones para eventos reales, situación derivaba de sus resultados los cuales describen el comportamiento de una variable (aleatoria uniforme en nuestro caso). Los modelos simulados son más fáciles de entender que muchos modelos analíticos (Azofeifa-Z, 2004). Esta técnica es ideal cuando se tienen experimentos en los que es aconsejable utilizar algún procedimiento de muestreo, pero tomar físicamente la muestra es imposible o demasiado costoso como en este caso puntual en el cual se pretende hallar el área de una superficie geográfica extensa, implicando el reemplazo de elementos reales por un universo teórico dando solución a un planteamiento por medio de la generación de números aleatorios con una distribución de probabilidad dada, de esta manera se concluyen exitosamente cuestiones por medios no determinísticos ideales para casos donde es imposible conjurar una ecuación precisa.

En este punto surge la necesidad de implementar el método de Monte Carlo, del cual implementamos en este artículo una nueva aplicación práctica al utilizarlo para estimar áreas geográficas teniendo en cuenta además la latitud a la que se en-

cuentre la extensión geográfica de interés, aprovechando que entre las aplicaciones de este método se encuentra obtener una buena aproximación de áreas de figuras geométricas planas y también se utiliza para evaluar integrales multidimensionales muy complejas. (Pengelly, 2002).

El aspecto más importante de esta aplicación práctica radica en la posibilidad de dar a conocer el error que tendría la estimación y el intervalo de confianza en el que se presume estará el valor real de la superficie con un error aleatorio ya establecido, siendo este un indicativo de la bondad de los resultados, contrario a lo que sucede con ciertos programas en los que el procedimiento del cálculo de áreas se basa en el principio de encapsulamiento que oculta la complejidad del código (Madrigal, 2020).

En los resultados se evidencia que entre más puntos sean generados, el área geográfica calculada para cada ciudad tiende a acercarse al valor definido por las entidades territoriales correspondientes que se usan como fuente de verificación de los resultados, mostrando así que para ambas ciudades se obtienen resultados aceptables generando un intervalo de entre 5000 y 100.000 puntos más allá de este último valor el resultado se estabiliza y su variación disminuye. Para el caso de Bogotá el error aumenta entre más puntos se generen hasta aproximarse a 0.8 el cual es su límite incluso con más de 100.000 puntos. En el caso de Burlington el error siempre oscila entre 0,95 y 0,97 generando desde 100 puntos hasta más de 100.000.

2. Preliminares

2.1. Simulación de Monte Carlo

Este método es propicio para ser aplicado en este caso debido a que emplea números aleatorios uniformemente distribuidos en el intervalo $[0,1]$ que es utilizado para resolver problemas donde la evolución de una variable en el tiempo no es de importancia (Tarifa, 2005). También es posible utilizar números pseudoaleatorios, los cuales son aquellos conjuntos de números contruidos de manera que todos los dígitos tienen la misma probabilidad de aparición (Peña, 2014). Este método se usa para calcular numéricamente expresiones matemáticamente complejas y difíciles de evaluar con exactitud, o que no pueden resolverse analíticamente (Campoy, 2015).

Bajo el contexto estadístico Diharce (2008) explica que una simulación consta de conocer primero las distribuciones de las variables de entrada y generar un valor al azar de cada una de ellas obteniendo el resultado deseado evaluando el modelo de medición en estas entradas aleatorias. Este proceso se repite un número grande de veces, y así con las realizaciones de la variable explicada se obtiene una función de distribución empírica, a partir de la cual, podemos obtener cualquier estadística del mensurando con su respectivo intervalo de confianza. Este tipo de simulación es la de Monte Carlo.

La simulación de Monte Carlo no depende del estado inmediatamente anterior del

sistema contrario a los métodos MCMC (Markov Chain Monte Carlo) los cuales son métodos Monte Carlo usando cadenas de Markov como los muestreos de Gibbs y de Metrópolis Hastings, por ejemplo el segundo es un muestreo por rechazo generalizado, donde los valores aleatorios se toman de distribuciones escogidas razonablemente y corregidas de tal manera que se comporten asintóticamente como valores de la distribución objetivo (Jimenez , 2015). Sin embargo, ambos constan de un algoritmo que simula una muestra de una cadena de Markov con una distribución estacionaria, en esencia una cadena constituye un proceso en tiempo discreto en el cual una variable aleatoria va cambiando conforme este avanza y por lo tanto es un requisito para cumplir la propiedad de Markov la cual hace referencia a que todo el proceso histórico de una variable se resume en la posición actual de dicho proceso con el fin de estimar la probabilidad de que la variable cambie a otro estado, lo que diferencia a este proceso del método de Monte Carlo.

2.2. Números Aleatorios y Pseudoaleatorios

Los números pseudoaleatorios se determinan por una semilla por lo cual no son exactamente aleatorios, pero si cumplen la propiedad de estos últimos al verificar la equiprobabilidad de aparición de cada dígito (Peña, 2014). Por tal motivo los números pseudoaleatorios son generados por medio de una función (ya que se generan mediante un algoritmo, todo generador podría repetirse luego de largas sucesiones de números), pero no cumplen la propiedad de que la elección de un número no dependa de la selección de otro, por lo tanto, se puede decir en términos prácticos que un número de naturaleza pseudoaleatoria es el valor de una variable aleatoria que tiene una distribución de probabilidad uniforme definida en el intervalo $[0,1]$. Este tipo de números se utilizan con mayor frecuencia en la práctica debido a que generar números aleatorios es más engorroso y demorado.

2.3. Simulación de Montecarlo para hallar áreas de figuras utilizando integrales

Se debe recordar que sea una función $f : [a, b] \rightarrow \mathbb{R}$ si $y = f(x)$ es no negativa e integrable en un intervalo cerrado $[a, b]$ entonces el área bajo la función $y = f(x)$ en $[a, b]$ es (Thomas, Weir, Hass y Giordano, 2006).

$$A = \int_a^b f(x) dx$$

Representando efectivamente el área encerrada en la gráfica f y el eje de las abscisas entre las rectas $x = a, x = b$, lo cual representa una base matemática de la que se desprende conceptualmente el método de Monte Carlo, en este caso utilizado para la estimación de áreas de funciones continuas (lo que a su vez configurará una base para el cálculo de superficies que tengan formas irregulares) puesto que el cálculo de integrales definidas supone una de las aplicaciones, debido a que si consideramos esta vez a X una variable aleatoria con densidad f y $g : \mathbb{R} \rightarrow \mathbb{R}$ es

una función, entonces el valor esperado de la variable aleatoria $g(x)$ es (Kisbye, 2008):

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Para demostrar que la ecuación es válida, se parte del hecho de que el principio del método de Montecarlo para aproximar esa expresión es generar una muestra aleatoria (X_1, X_2, \dots, X_n) con $n \rightarrow \infty$ de la densidad f y proponer como una aproximación el promedio empírico (Robert y Castella, 1999), teniendo en cuenta el teorema del límite central en el cual la suma de esas n variables aleatorias independientes y de varianza finita posee una función de distribución que se aproxima a una distribución normal dado $E[x_i] = \mu$ y $var[x_i] = \vartheta^2$ se establece $\bar{X} \simeq N\left(\mu, \frac{\vartheta^2}{n}\right)$ que al reemplazado en la ecuación del valor esperado se obtiene:

$$E[g(X)] \simeq \overline{g(x)_n}$$

La teoría se sustenta en la ley de los grandes números puesto que de esta se sabe que, un buen estimador del valor esperado de una variable aleatoria continua X con distribución F es el valor promedio de una muestra finita de variables aleatorias, independientes con distribución F (Barrera y Mercado, s. f.) como se expuso anteriormente y generalizando este concepto para estimar el valor esperado de la función continua g con una variable aleatoria de densidad f como argumento, se puede tomar una muestra de estas variables independientes e idénticamente distribuidas con densidad f y llegar a:

$$\overline{g(x)_n} \simeq \frac{1}{n} \sum_{i=1}^n g(x_i)$$

En donde cada x_i es un número aleatorio, que en conjunto obedecen a la densidad $f(x)$, permitiendo así afirmar que la esperanza de una variable aleatoria continua es una integral y la media muestral se puede usar para hallar el valor de una integral siendo la noción que sustenta la idea del método de Monte Carlo. Partiendo de lo anterior para el caso del presente artículo, esencialmente este método consiste en generar puntos aleatorios dentro de una región determinada cuya área sea conocida y que encierre una superficie de magnitud desconocida la cual es objeto de análisis determinando la cantidad de estos puntos que se generen dentro de esta superficie incógnita (cantidad de éxitos n') y la razón que los relaciona con el número total de puntos aleatorios generados (n) que viene dada de la siguiente manera:

$$\frac{n'}{n} \simeq E[I_A(X)] = \int_{-\infty}^{\infty} I_A(x)f(x)dx \quad (1)$$

De esta forma se introducen matemáticamente todas las variables necesarias para determinar el área de una superficie, debido al hecho de que la probabilidad de generar un punto dentro de la figura con área desconocida es igual al cociente de dicha área dividida el área de la superficie conocida, las variables quedan

relacionadas de la siguiente manera:

$$\frac{Af}{AF} = \frac{Pf}{PT} \quad (2)$$

En dónde “Af” es el área de la figura que se desea hallar, “AF” es el área de una figura conocida, “Pf” es el número de puntos dentro de la función y “PT” son los puntos totales. Nótese que el término del lado derecho en la igualdad de la ecuación (2) es $\frac{Pf}{n}$. De esta manera se puede calcular el área de cualquier figura simple (por ejemplo, cuadrado o círculo) utilizando la aplicación de las integrales para calcular áreas, sin embargo, para estimar áreas de figuras geométricas y superficies reales es necesario ir más allá y utilizar un algoritmo empleado en código de lenguaje R que se presentará y explicará más adelante.

2.4. Determinación del área de figuras regulares e irregulares utilizando el método de Montecarlo

Calcular el área de alguna figura geométrica como por ejemplo un círculo es trivial utilizando el método analítico, pues es bien conocido que dicha área se encuentra utilizando la ecuación $A = \pi r^2$, sin embargo, también es posible calcular áreas utilizando métodos numéricos como la simulación tomando la figura a la cual queremos estimar su área y encerrándola dentro de una figura cuya área ya es conocida.

Este método también funciona para determinar el área de una forma irregular como por ejemplo una superficie de agua o el área de alguna ciudad, el mencionado método implica realizar los siguientes pasos:

Inscribir la figura a la cual se le desea estimar el área dentro de un polígono con dimensiones conocidas, preferiblemente un cuadrado o rectángulo debido a la facilidad que proveen.

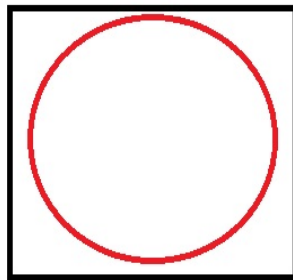


Figura 1: Círculo inscrito en cuadrado

Generar n puntos aleatorios dentro del cuadrado, por lo cual a su vez varios quedarán dentro de la figura cuya área pretende ser estimada.

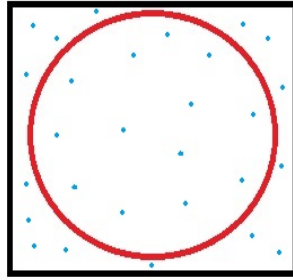


Figura 2: Generación de puntos aleatorios dentro de las figuras

Contar el número de puntos que quedaron dentro de la figura con área desconocida.

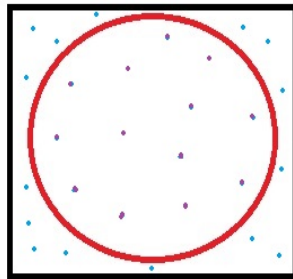


Figura 3: Conteo de puntos dentro de la figura de interés

Utilizar la siguiente ecuación que se origina a partir de la ecuación (2) despejando la variable incógnita:

$$A = \frac{N_i}{N} A_{\text{cuadrado}} \quad (3)$$

En donde “A” es el área estimada de la figura, “ n_i ” es el número de puntos que quedaron dentro de la figura con área desconocida, “ n ” es el número de puntos aleatorios dentro del cuadrado y “ A_{cuadrado} ” es el área de una figura cuya superficie es conocida, por facilidad se escogió un cuadrado.

2.5. Determinación de los intervalos de confianza

2.5.1. Intervalo de confianza para la cantidad de aciertos

Como se puede observar en la ecuación (3), debido a que “A” es proporcional a la probabilidad de que un punto aleatorio caiga en la superficie tal y como se constató en la ecuación (2), se estima esa probabilidad como:

$$\hat{p} = \frac{N'}{N} \quad (4)$$

Siendo N' el número de éxitos sobre n intentos cuyo resultado aproximado se observó en la ecuación (1) y es bien sabido que esta expresión viene dada por una distribución discreta binomial, esta variable discreta generada que cuenta el número de éxitos en N pruebas es muy importante además de la estimación de la superficie como tal, también en el establecimiento del intervalo de confianza, cada una de ellas con la misma probabilidad de éxito como se aclaró anteriormente, ya que el modelo se aplica a una población finita de puntos generados de las que se toman elementos al azar con reemplazo. la probabilidad de N' éxitos en N intentos viene dada por la distribución binomial (Rodríguez, 2011):

$$P(N' \text{ aciertos en } N) = \binom{N'}{N} p^{N'} q^{N-N'} \quad (5)$$

Recordando que p es la probabilidad de acierto y q la probabilidad de fracaso siendo igual a:

$$q = 1 - p \quad (6)$$

Tal como se mencionó en la introducción y en el apartado anterior de este marco teórico, el intervalo de confianza tiene una importancia latente en todo este procedimiento, pues al determinar un nivel de significancia se le puede otorgar a un usuario de la metodología el saber la probabilidad de fallar en la estimación por medio de ese intervalo o en otras palabras conocer el intervalo de aciertos que debe existir para que se pueda estimar de manera precisa la superficie en cuestión dependiendo de la cantidad total de puntos generados. Resulta que la distribución binomial recién expuesta puede aproximarse mediante una normal cuando se tiene un número de pruebas grande (Rodríguez, 2011):

$$np > 5$$

y

$$nq > 5 \quad (7)$$

La mencionada distribución tendrá media $\mu = np$ y varianza $\sigma^2 = npq$ (Rodríguez, 2011) y $\sigma^2 = np(1 - p)$, si a esto le incluimos que se sabe que para una distribución normal $N(\mu, \sigma^2)$ el 95 % de las observaciones se encuentran en el intervalo

$(\mu - 2\sigma, \mu + 2\sigma)$ y suponiendo la expresión (7), se tendrá que el intervalo de confianza al 95 % de la cantidad de aciertos n en el área de la figura cuya superficie se desea estimar “A” se encuentra en (Rodríguez, 2011):

$$\left(Np - 2\sqrt{Npq}, Np + 2\sqrt{Npq}\right) \quad (8)$$

Siendo este el intervalo de confianza para la cantidad de aciertos. Ahora debe hallarse el intervalo de confianza para la superficie estimada.

2.5.2. Intervalo de confianza para la estimación

Para determinar un intervalo de confianza es necesario tener como requisito primordial el cálculo del error una vez propuesto un nivel de confianza, el error se calcula por medio de:

$$E = t_{1-\frac{\alpha}{2}; n-1} \frac{S_g}{\sqrt{n}} \quad (9)$$

Dónde se toma el valor de la tabla T-Student una vez se haya escogido el nivel de significancia α , tomando en cuenta que el error viene dado por la raíz cuadrada de la cantidad de puntos generados; dato que se utiliza una vez se haya calculado la varianza por medio de:

$$S_g = \sqrt{\frac{(g(x_i) - g(\bar{x}))^2}{n-1}} \quad (10)$$

En la cual se tiene como $n - 1$ porque la varianza es insesgada.

Para facilitar los cálculos se estimará un intervalo de confianza al 95 %, de esta manera la ecuación (9) del error quedaría de la siguiente manera:

$$E = \frac{2(S_g)}{\sqrt{n}} \quad (11)$$

Recordando que de la expresión (8) se tomó la varianza $\sigma^2 = npq$ y que $(S_g) = \sqrt{\sigma^2}$ se puede reescribir la ecuación (11) de la siguiente manera:

$$E = 2\sqrt{\frac{npq}{n}} \quad (12)$$

Una vez hecho esto, la estimación del intervalo de confianza constará en tomar la estimación que se obtiene previamente y restarle el error para el límite inferior y para el límite superior habrá que sumarle el error.

$$IC = estimacion \pm error \quad (13)$$

Todo este mismo proceso puede ser aplicado a áreas de figuras cuya naturaleza irregular las vuelve más complejas como por ejemplo el área de la ciudad de Bogotá, esto es posible conociendo la escala del mapa de la región y fijando un cuadrado que encierre dicho mapa como se mostrará a continuación.

2.6. Cálculo de la escala del rectángulo

2.6.1. Valor de un grado de latitud en kilómetros

Según los parámetros del sistema geodésico mundial WGS84 el radio ecuatorial es igual a 6.378.137 m por lo tanto el diámetro es 12.756.274m, a partir de este dato se puede calcular la circunferencia del ecuador mediante $12.756.274 * \pi = 40.075.016,69\text{m}$ o $40.075,01669\text{km}$, dividiendo dicho valor entre 360° se obtiene la distancia que representa en el terreno cada grado de latitud siendo esta equivalente a $111,3195\text{km}$. La variación de este valor es despreciable en proximidad a los polos.

2.6.2. Valor de un grado de longitud en kilómetros

En este caso la circunferencia de los paralelos se reduce conforme se aproximan a alguno de los polos debido a que se van estrechando ya que allí convergen, por lo que el valor de un grado de longitud depende de la latitud a la que se encuentre la ciudad o región geográfica en específico según la relación $111,3195\text{km} * \cos(\text{latitud})$. Basándose en estos conceptos se puede calcular el área real del rectángulo generado que encerrará a la figura y al cual llamaremos (H), esto se hará hallando la distancia entre latitudes y entre longitudes correspondientes a cada uno de sus vértices estableciendo su diferencia y multiplicándolos por su respectiva equivalencia en el terreno en kilómetros, de esta manera solo queda hallar su área multiplicando entre el valor de sus lados según los parámetros expuestos anteriormente, de la siguiente manera:

$$H = (||long_1| - |long_2|| * (111,3195 * \cos(\text{lat})) * (||lat_1| - |lat_2|| * 111,3195)) \quad (14)$$

Se debe tener en cuenta que la ecuación (14) es válida para aquellas regiones que se encuentren por completo dentro de un mismo cuadrante, siendo así, dicha ecuación es susceptible a cambios en cuanto a los signos tanto en la diferencia de las latitudes como de las longitudes o incluso ambas al tiempo, esto se deberá a la posición geográfica de la región a la cual se le calculará el área, puesto que si dicha región se encuentra atravesada por el paralelo 0° (Ecuador) como por ejemplo la ciudad de Quito en Ecuador, las latitudes tendrán que sumarse debido a que la parte superior se encuentra en el hemisferio norte y la parte inferior se encuentra en el hemisferio sur. Caso equivalente ocurriría con las longitudes para alguna región que sea atravesada por el meridiano de referencia (el cual para este artículo será el que pasa por el Real Observatorio de Greenwich siendo este el más aceptado).

3. Estimación del área de las ciudades de Bogotá y Burlington como casos de estudio utilizando el método de Monte Carlo

Dentro del paquete estadístico, una vez establecida la conexión entre el archivo (con extensión `.shp`) y la interfaz, se visualizará la imagen escalada de Bogotá. Es importante recordar que toda digitalización de una superficie geográfica debe realizarse en un sistema de referencia establecido; por consenso mundial, se recomienda el **WGS84**. Este aspecto es clave para definir la escala del mapa a utilizar, teniendo en cuenta la equivalencia entre un grado de latitud y uno de longitud en kilómetros sobre la superficie terrestre, como se explicó anteriormente.

A partir de estas imágenes digitalizadas en formato **shapefile**, se determinará un rectángulo que las encierre. Esto se logra permitiendo al usuario visualizar la dimensión necesaria para delimitar la imagen y especificar las coordenadas de sus vértices. Dichas coordenadas se definen en función de ejes coordenados previamente establecidos mediante instrucciones específicas en el código.

El propósito de determinar áreas de esta manera —si bien en la actualidad es común recurrir a herramientas de Sistemas de Información Geográfica (SIG)— surge de la necesidad de evaluar la confiabilidad de los resultados obtenidos. En muchos casos, no es posible determinar con certeza qué tan precisos son dichos resultados. Por ello, se plantea esta metodología, que permite al usuario no solo conocer el área estimada, sino también evaluar su nivel de confianza. Para ello, se define un **intervalo de confianza**, tanto para la estimación como para la probabilidad de éxito, lo que facilita la cuantificación del error. Todo el proceso se lleva a cabo dentro de un mismo software, integrando distintas tareas de análisis.

Se emplearán las siguientes salidas gráficas: **escala 1:500.000 para Bogotá y 1:40.000 para Burlington**.

La imagen del terreno para cuestiones de los cálculos se tomará como un polígono de nombre “mapa”, es decir, todo lo contenido en este será el área objeto de estudio. Para el caso práctico de Bogotá, se toma el área de la ciudad la cual en la actualidad está estimada en 1636,35 kilómetros cuadrados (Peñalosa, Ortiz, Avendaño y Burbano, 2018). Para el caso de las coordenadas de Bogotá la latitud según el sistema de referencia geocéntrico MAGNA-SIRGAS es de $4^{\circ}35'46,3215'' N$ o 4,596200417 su equivalente decimal.

Respecto a la ciudad de Burlington, esta se ubica en la Provincia de Ontario en el país de Canadá, su área estipulada es de $185,66 km^2$ (Statistics Canada [Government of Canada], 2023), su latitud corresponde a $43^{\circ}23'16,22'' N$, esta ciudad se ha seleccionado para la aplicación del método de Monte Carlo debido a la variación en coordenadas ya que se encuentra al norte del planeta en una latitud relativamente elevada comparada con Bogotá y así ratificar la eficiencia del cálculo del área para cualquier territorio del planeta.

La selección del rectángulo como medio de contención de la imagen del terreno se realizó debido a que es una figura geométrica simple que ofrece: primero comodidad

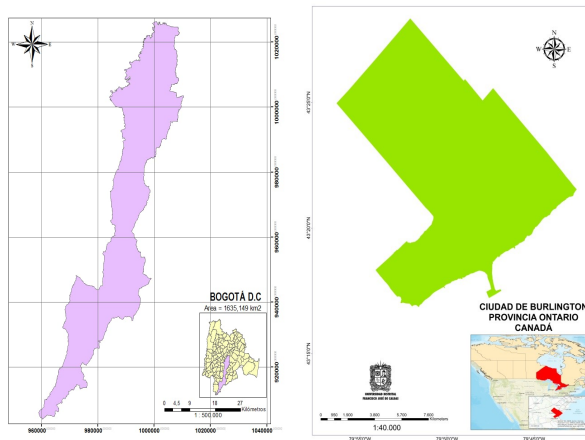


Figura 4: Salidas gráficas de las ciudades caso de estudio

en cuanto a su visualización puesto que es más fácil observar la dimensión necesaria para contener la imagen en su interior comparada con otras figuras geométricas, segundo, también presenta facilidad para ser construida en R por medio de una corta y clara instrucción, tercero y último, la sencilla manera de calcular su área, ya que es un paralelogramo y se define por excelencia en el plano.

4. Resultados

Se realizó una tabla para cada ciudad con el fin de mostrar los resultados de los cálculos.

Mediante el método de Monte Carlo, el mejor resultado para Bogotá se obtuvo al generar 50.000 puntos, ya que se estimó un área de $1637,269 \text{ Km}^2$ y su respectivo intervalo al 95 % de confianza se encuentra entre $[1636,478 \text{ Km}^2 - 1638,059 \text{ Km}^2]$, con un error de 0,7905. Esta estimación supera en apenas $0,919 \text{ Km}^2$ el área calculada por la Secretaría Distrital de Planeación, lo que representa una diferencia del 0,059 % respecto al valor total de la superficie.

Para la ciudad de Burlington, el mejor resultado se obtuvo generando 50.000 puntos, estimando un área de $185,601 \text{ Km}^2$ y su respectivo intervalo al 95 % de confianza entre $[184,605 \text{ Km}^2 - 186,597 \text{ Km}^2]$, con un error de 0,966 (valor considerablemente mayor al error calculado para Bogotá). Esta estimación es inferior en apenas $0,059 \text{ Km}^2$ al área reportada por Statistics Canada, lo que representa una diferencia del 0,032 % respecto al valor total de la superficie.

La metodología muestra resultados satisfactorios para calcular áreas de superficies geográficas equivalentes en tamaño hasta ciudades como Bogotá. Sin embargo, es importante destacar que, a medida que aumenta la superficie, se deben generar más puntos. Por este motivo, en las pruebas fue necesario generar menos puntos para obtener un resultado cercano al área real de Burlington en comparación con Bogotá, cuya mayor extensión requería un número más alto de puntos.

En el caso del cálculo de áreas de mayor tamaño, como ciudades de gran extensión (Los Ángeles, Tokio, Delhi, etc.) o incluso países, es necesario considerar que estas superficies abarcan múltiples latitudes. Esto podría representar una limitación para la metodología, ya que la latitud influye en los cálculos y, por lo tanto, sería necesario realizar pruebas adicionales en cada superficie desconocida con el fin de determinar cuál es la mejor latitud a utilizar.

Por otro lado, es esencial realizar varios ensayos en la ejecución del código para determinar la cantidad óptima de puntos que permita aproximar el valor obtenido al área real. Como se observó en Burlington, con solo 500 puntos se obtuvo un área calculada con un aceptable grado de exactitud, mientras que en el caso de Bogotá fue necesario aumentar el número de puntos a 5000 para alcanzar un nivel de precisión similar.

A pesar de estas limitaciones, la metodología demuestra ser confiable para el cálculo de superficies geográficas de gran extensión, proporcionando tanto el error como el intervalo de confianza de los cálculos. Además, permite realizar múltiples pruebas hasta obtener una estimación satisfactoria. Esta propuesta elimina el principio de encapsulamiento, ya que el presente artículo expone todo el proceso metodológico, lo que permite a cualquier usuario modificarlo y adaptarlo a necesidades específicas.

5. Conclusiones

- El resultado de la simulación, que incluye la estimación de la superficie geográfica junto con los intervalos de confianza y el error, está compuesto por números aleatorios, ya que depende del número de puntos generados dentro de las figuras, tanto la geométrica (rectángulo) como el mapa. Por lo tanto, cada vez que se ejecute la simulación, el resultado será diferente, oscilando alrededor del valor real.
- El área estimada de una superficie real se acercará más a su valor verdadero a medida que se generen más puntos aleatorios dentro del rectángulo, lo que disminuye la varianza del resultado y aproxima su media a la solución

Bogotá				
Tamaño de muestra (n)	Número de réplica	Área estimada (km ²)	Intervalo de confianza	Error
100	1	852,3	851,7-852,9	0,6
100	2	2045,52	2044,66-2046,37	0,85
100	3	1704,06	1703,8-1705,4	0,8
200	1	1576,76	1575,99-1577,53	0,78
200	2	1747,22	1746,41-1748,02	0,81
200	3	1789,83	1789,01-1790,65	0,81
500	1	1670,51	1669,72-1671,3	0,79
500	2	1500,05	1499,29-1500,81	0,76
500	3	1551,19	1550,42-1551,96	0,77
1000	1	1542,66	1541,89-1543,43	0,77
1000	2	1602,32	1601,54-1603,11	0,78
1000	3	1704,6	1703,8-1705,4	0,8
5000	1	1682,44	1681,64-1683,23	0,79
5000	2	1638,12	1637,33-1638,91	0,78
5000	3	1637,27	1732,77-1734,38	0,81
50.000	1	1643,23	1642,44-1644,02	0,78
50.000	2	1639,99	1639,21-1640,79	0,79
50.000	3	1637,27	1636,48-1638,06	0,79
100.000	1	1639,91	1639,12-1640,7	0,79
100.000	2	1649,63	1648,84-1650,41	0,79
100.000	3	1644,51	1643,73-1645,3	0,79

Tabla 1: Resultados para Bogotá

analítica. Sin embargo, en el caso de Bogotá, se evidenció que el error se incrementa hasta estabilizarse alrededor de 0,8. Además, tras múltiples repeticiones de la simulación con diferentes cantidades de puntos generados, se observó que el área estimada comienza a estabilizarse en torno al valor real cuando se alcanzan **5000 puntos**. En los casos en que este número fue considerablemente menor (por ejemplo, desde 100), el área estimada presentó desviaciones respecto al valor definido por la Secretaría Distrital de Planeación en un intervalo aproximado de $[400 \text{ Km}^2 - 1000 \text{ Km}^2]$, considerando todas las ejecuciones de la simulación.

- Para la ciudad de Burlington, no fue necesario generar tantos puntos como en Bogotá para estabilizar la estimación. De hecho, con **500 puntos**, los resultados en promedio ya se acercan al área real, con intervalos de $[1 \text{ Km}^2 - 20 \text{ Km}^2]$, considerando todas las ejecuciones de la simulación. Además, el error en Burlington varía siempre entre **0,95 y 0,97**, siendo significativamente mayor que en Bogotá. Estos resultados pueden atribuirse al hecho de que Bogotá es casi nueve veces más grande que Burlington, lo que obliga a utilizar una escala gráfica más pequeña en el caso de Bogotá y, en consecuencia, a generar un mayor número de puntos para cubrir

Burlington				
Tamaño de muestra (n)	Número de réplica	Área estimada (km ²)	Intervalo de confianza	Error
100	1	209,92	208,95-210,88	0,96
100	2	192,89	191,99-193,85	0,95
100	3	170,2	169,29-171,12	0,92
200	1	218,43	217,46-219,4	0,97
200	2	173,04	172,12-173,97	0,92
200	3	212,76	211,79-213,72	0,97
500	1	197,44	196,49-198,4	0,95
500	2	184,96	184,02-185,89	0,94
500	3	205,38	204,42-206,34	0,96
1000	1	203,68	202,72-204,64	0,96
1000	2	196,87	195,92-197,82	0,95
1000	3	204,25	203,29-205,2	0,96
5000	1	204,7	203,74-205,66	0,96
5000	2	199,02	198,07-199,98	0,95
5000	3	205,04	204,08-206	0,96
50.000	1	199,41	198,46-200,36	0,95
50.000	2	199,78	198,83-200,74	0,95
50.000	3	185,6	184,6-186,6	0,97
100.000	1	199,74	198,79-200,7	0,95
100.000	2	200,78	199,82-201,74	0,95
100.000	3	200,93	199,97-201,88	0,96

Tabla 2: resultados para Burlington

adecuadamente la superficie y realizar los cálculos.

- Debido a que Bogotá se encuentra muy cerca del ****Ecuador****, el valor de un grado de latitud es casi idéntico al de un grado de longitud, siendo esta diferencia prácticamente despreciable. Sin embargo, para latitudes mayores, esta diferencia se vuelve más significativa a medida que aumenta la proximidad a los polos, como se observa en la ciudad de Burlington.
- Al estimar el área de cualquier figura basada en información espacial, todos los datos generados a partir del software deben estar referenciados en un ****sistema de coordenadas común****. En caso contrario, no será posible realizar análisis adecuados sobre los datos.

Recibido: Julio 15 de 2024
Aceptado: Febrero 10 de 2025

Referencias

- [Azofeifa(2004)] C. E. Azofeifa. Aplicación de la simulación monte carlo en el cálculo del riesgo usando excel. *Revista Tecnología en Marcha*, 17(1):ág-97, 2004.
- [Barrera and Mercado()] P. S. Barrera and V. H. I. Mercado. El método monte-carlo y su aplicación a finanzas. *Departamento de Matemáticas, Universidad Autónoma Metropolitana Iztapalapa*.
- [Campoy et al.(2015)Campoy, Alcarria, and González-Marcos] J. Campoy, R. Alcarria, and A. González-Marcos. Estudio del comportamiento óptico de pieles utilizando técnicas de monte carlo. *Revista Internacional de Métodos Numéricos Para Cálculo y Diseño En Ingeniería*, 31(3):161-170, 2015.
- [Diharce(2008)] E. V. Diharce. Técnicas de simulación para el análisis estadístico de datos de medición. *Centro de investigación en Matemáticas*, 2008.
- [Jiménez Luna(2015)] J. Jiménez Luna. Métodos monte carlo basados en cadenas de markov. 2015.
- [Kisbye(2008)] P. Kisbye. Generación de eventos en procesos de poisson. *Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba*, 2008.
- [López et al.(2022)] M. López et al. Simulación estadística. *Estudio Static*, 2022.
- [Madrigal Chaves(2020)] W. Madrigal Chaves. Polimorfismo. 2020.
- [Peña(2014)] D. Peña. *Fundamentos de estadística*. Alianza editorial, 2014.
- [Peñalosa et al.(2018)Peñalosa, Ortiz, Avendaño, and Burbano] L. E. Peñalosa, G. A. Ortiz, A. A. Avendaño, and G. L. Burbano. Monografía 2017, diagnóstico de los principales aspectos territoriales de infraestructura demográficos y socioeconómicos. *Alcaldía Mayor de Bogotá, secretaría Distrital de Planeación*, 2018.
- [Pengelly(2002)] J. Pengelly. Monte carlo methods. *University of Otago*, 2002.
- [Robert et al.(1999)Robert, Casella, and Casella] C. P. Robert, G. Casella, and G. Casella. *Monte Carlo statistical methods*, volume 2. Springer, 1999.
- [Rodríguez-Aragón(2011)] L. J. Rodríguez-Aragón. Simulación, método de montecarlo. *Recuperado de: https://previa.uclm.es/profesorado/licesio/docencia/mcoi/tema4_guion.pdf*, 2011.
- [Statistics Canada(15 de Junio de 2023)] Statistics Canada. www12.statcan.gc.ca, 15 de Junio de 2023.
- [Tarifa(2005)] E. Tarifa. Teoría de modelos y simulación. introducción a la simulación. *Facultad de Ingeniería-Universidad Nacional de Jujuy*, 2005.
- [Thomas et al.(2005)Thomas, Weir, Hass, and Giordano] G. B. Thomas, M. D. Weir, J. Hass, and F. R. Giordano. *Thomas' calculus*. Addison-Wesley, 2005.

A. Códigos

```

Librerías utilizadas:\\
library(RColorBrewer) ‘‘Librería para la discriminación por colores’’.\\
library(classInt) ‘‘Librería para la toma de atributos de archivos tipo shapefile’’.\\
library(rgdal) ‘‘Librería para la lectura del archivo anterior y la generación del mapa’’.\\
library(sp) ‘‘Librería para la asignación de sistemas de referencia coordenados’’.\\

Ploteo del mapa:\\
archivo <- choose.files() ‘‘Ruta para buscar el archivo tipo shape’’.\\
mapa <- readOGR(archivo)\\
m <- geometry(mapa) ‘‘Reconoce el archivo a mostrar’’.\\
plot(mapa, border="blue", axes=F) ‘‘Genera el polígono objeto de estudio’’.\\
title(main = "Mapa de Bogotá", cex=3)\\

Generación del rectángulo:\\
axis(2) ‘‘Genera eje coordenado y’’.\\
axis(1) ‘‘Genera eje coordenado x’’.\\
D <- -74.5 ‘‘Solicita el valor para asignarlo en x al punto D (longitud costado izquierdo del rectángulo)’’.\\
E <- 3.7 ‘‘Solicita el valor para asignarlo en y al punto E (latitud parte inferior del rectángulo)’’.\\
F <- -73.9 ‘‘Solicita el valor para asignarlo en x al punto F (longitud costado derecho del rectángulo)’’.\\
G <- 4.85 ‘‘Solicita el valor para asignarlo en y al punto G (latitud parte superior del rectángulo)’’.\\
b <- rect(D, E, F, G, border="red", lwd=2) ‘‘Crea el polígono entre los puntos anteriores’’.\\

Cálculo del área real del rectángulo:\\
LatCiudad <- 4.596200417 ‘‘Solicita el valor de latitud en decimales del lugar objeto de estudio’’.\\
H <- ((abs(abs(D)-abs(F))*111.3195*cos(LatCiudad*(pi/180)))*((abs(abs(E)-abs(G))*111.3195)))\\
H ‘‘Se aplica ecuación 18, para llevar el polígono a las dimensiones ‘‘reales’’ ’’.\\

Generación de puntos aleatorios:\\
N <- 50000 ‘‘Asigna el número de puntos a generar dentro del rectángulo’’.\\
X <- runif(N, -74.5, -73.9) ‘‘Asigna valores en x a recorrer horizontalmente por los puntos generados’’.\\
Y <- runif(N, 3.7, 4.85) ‘‘Asigna valores en y a recorrer verticalmente por los puntos generados’’.\\

Asignación de atributos espaciales:\\
z <- data.frame(X, Y) ‘‘Guarda los valores de los puntos generados en el paso anterior’’.\\
coordinates(z) <- ~X + Y ‘‘Asigna estos valores como coordenadas’’.\\
proj4string(z) <- proj4string(m) ‘‘Asigna la misma proyección del polígono a los puntos guardados’’.\\
points(z) ‘‘Permite visualización de los puntos’’.\\

Diferenciación de los puntos dentro del mapa de Bogotá:\\
pt.in.poly <- rgeos::gIntersection(z, m) ‘‘Diferencia los puntos generados dentro del polígono y fuera’’.\\
plot(z, pch = 19, cex = 1, add = TRUE) ‘‘Validación de los puntos dentro del polígono’’.\\
plot(pt.in.poly, pch = 19, cex = 0.5, col = "green", add = TRUE) ‘‘Asigna un color para diferenciar los

Conteo de puntos dentro del polígono:\\

```

```
a <- data.frame(pt.in.poly, Id=1) ‘‘Convierte los datos en una lista’’.\nfor (Id in a) ‘‘Bucle for para realizar el conteo de datos contenidos en la lista generada’’.\nconteo.pts <- sum(Id) ‘‘Cuenta puntos totales y los guarda como dato numérico’’.\n\nCálculo del intervalo de confianza para la cantidad de aciertos:\n p <- conteo.pts/N ‘‘Calcula variable p (probabilidad de acierto) según la ecuación 8’’.\n q <- 1 - p ‘‘Calcula variable q (probabilidad de fracaso) mediante la ecuación 10’’.\n intervinf <- ((N*p) - 2*sqrt(N*p*q)) ‘‘Determinación del intervalo inferior’’.\n intervsup <- ((N*p) + 2*sqrt(N*p*q)) ‘‘Determinación del intervalo superior’’.\n print(c(intervinf, intervsup)) ‘‘Intervalo de confianza con significancia del 5% usando la ecuación 12\n\nEstimación de la superficie:\n estimacion <- p*H ‘‘Cálculo de la estimación mediante ecuación 7’’.\n estimacion ‘‘Muestra el área estimada del polígono objeto de estudio’’.\n error <- 2*sqrt((N*p*q)/N) ‘‘Se usa la ecuación 16 para determinar el error’’.\n error\n liminf <- estimacion - error\n limsup <- estimacion + error\n print(c(liminf, limsup)) ‘‘Muestra resultado con error al 95% de confianza mediante la ecuación 17’’.\n
```