

INCERTIDUMBRE Y ERRORES EN HIDROGEOLOGÍA: APLICACIÓN DE LA SIMULACIÓN MONTE CARLO UTILIZANDO EXCEL

UNCERTAINTY AND ERRORS IN HYDROGEOLOGY: APPLICATION OF THE MONTE CARLO SIMULATION USING EXCEL

Rudy Machorro Sagastume
Asociación Guatemalteca de Geociencias Ambientales – ASGA
Email: rudy.machorro@yahoo.com

RESUMEN

Los monitoreos hidrogeológicos involucran mediciones ambientales que no son perfectas y difíciles de cuantificar con exactitud y precisión en un contexto con incertidumbres aleatorias y epistémicas. En la presente investigación se utiliza la Simulación Monte Carlo (SMC) basada en Excel como herramienta predictiva para evaluar la incertidumbre mensual y anual de la lluvia en el Corredor Seco de Zacapa. El análisis exploratorio de datos revela una fuerte fluctuación aleatoria de lluvia mensual y anual, con los meses de verano manifestando un fuerte contraste en la desviación estándar especificada y presencia de extremos pluviales. Se describe la configuración de la SMC en Excel para implementar 1,000 iteraciones, generar las variables de valor aleatorio, verificar los valores simulados de lluvia, y visualizar los datos aleatorios generados. La comparación de los datos de campo y los datos simulados de lluvia generados con la SMC revelan un muy buen ajuste mensual, con ciertas diferencias en junio y octubre, y un comportamiento estacional de lluvia que puede aproximarse con una distribución tipo Gaussiana con asimetría positiva.

La SMC implementada con Excel aporta una muy buena primera aproximación para la predicción de la lluvia en el Corredor Seco. Sin embargo, para generar escenarios hidrogeológicos de menor incertidumbre puede requerirse software mucho más avanzado tal como SPSS y SAS. La mayoría de los procesos hidrogeológicos pueden considerarse variables aleatorias, pero para producir un modelo integrado más eficiente es necesario combinar los métodos estadísticos con métodos determinísticos de modo que se logre una mejor planificación, desarrollo y manejo de los recursos hídricos.

Palabras clave: Excel, Simulación Monte Carlo, Corredor Seco, lluvia.

ABSTRACT

Hydrogeological monitoring involves environmental measurements that are not perfect and difficult to quantify accurately and precisely in a context with random and epistemic uncertainties. In this research, Excel-based Monte Carlo Simulation (MCS) is used as a predictive tool to evaluate the monthly and annual uncertainty of rainfall in the Zacapa Dry Corridor. Exploratory data analysis reveals a strong random fluctuation of monthly and annual rainfall, with the summer months manifesting a strong contrast in the specified standard deviation and presence of rainfall extremes. The configuration of the MCS in Excel is described to implement 1,000 iterations, generate the random value variables, verify the simulated rainfall values, and visualize the generated random data. The comparison of the field data and the simulated rainfall data generated with the SMC reveals a very good monthly fit, with certain differences in June and October, and a seasonal rainfall behavior that can be approximated with a Gaussian-type distribution with positive asymmetry.

The MCS implemented with Excel provides a very good first approximation for predicting rainfall in the Dry Corridor. However, much more advanced software such as SPSS and SAS may be required to generate lower uncertainty hydrogeological scenarios. Most hydrogeological processes can be considered random variables, but to produce a more efficient integrated model it is necessary to combine statistical methods with deterministic methods so that better planning, development and management of water resources is achieved.

Key words: Excel, Monte Carlo Simulation, Dry Corridor, rainfall.

Introducción

“Quien parte de certezas, acaba con dudas, pero quien se contenta con partir de dudas, acaba con certezas.” Francis Bacon (1605), *The Advancement of Learning*.

El contexto de incertidumbre y variabilidad inherente que predomina en contextos hidrogeológicos promueve el uso de métodos estadísticos. La mayoría de procesos hidrogeológicos tal como lluvia, niveles de agua subterránea, evaporación, y calidad del agua son algunos ejemplos de variables aleatorias (Yevjevich, 1972). Estas variaciones aleatorias que impulsan el sistema hidrogeológico resultan de procesos físicos y ambientales que resulta muy difícil medir (McCuen, 2003). Las variables hidroclimatológicas están asociadas con aleatoriedad/incertidumbre por lo que deben tratarse como variables aleatorias (Maity, 2018). Los métodos estadísticos evalúan y cuantifican la incertidumbre realizando inferencias con base a datos disponibles y considerando que variables como la lluvia son aleatorias lo cual está relacionado con la complejidad de los procesos atmosféricos que la ocasionan.

¿Pero qué es una variable aleatoria?

Una variable aleatoria es aquella cuyo valor es incierto o difícil de predecir y cuya variabilidad no se explica adecuadamente mediante mediciones analíticas o procesos físicos (Viessman y Lewis, 2003). Una definición más amplia de variable aleatoria relacionada con la teoría de la probabilidad es la siguiente (Yevjevich, 1972): cualquier cantidad X de un número real dado x para la cual existe una probabilidad P de que X es igual o menor a x . Este planteamiento se denota como $P(X \leq x)$. La variable aleatoria frecuentemente se indica con la letra mayúscula X y la letra minúscula x representa el valor específico de la variable aleatoria. El valor que

toma la variable aleatoria depende de un evento particular el cual está sujeto a observación o medición. Un ejemplo para visualizar esta definición probabilística de variable aleatoria es el siguiente. En términos de calidad de agua puede considerarse que la concentración de oxígeno disuelto es la cantidad X y que para cualquier valor positivo x existirá una probabilidad P que $X \leq x$. En este caso particular el oxígeno disuelto se considera una variable aleatoria con valor positivo ya que puede asumir únicamente valores positivos, incluyendo cero. Existen diferentes tipos de aleatoriedad. La Simulación Monte Carlo (SMC) permite calcular numerosos escenarios de un modelo mediante la selección repetitiva de valores a partir de una distribución de probabilidad predefinida por el científico. Para efectos de la presente investigación se utilizan datos de lluvia de la estación La Fragua que el INSIVUMEH administra en Zacapa con fines de evaluar la incertidumbre mensual y anual de la lluvia generando la SMC con Excel.

Marco teórico del problema

Teoría de las mediciones

Cualquier medición que se realice es parte de una dispersión aleatoria cuyas probabilidades se describen mediante una función matemática, la función de densidad de probabilidad (Salsburg, 2001). La ciencia de las mediciones se conoce como Metrología y puede considerarse una rama de las matemáticas. Una medición es el proceso de determinar experimentalmente el valor de una cantidad con el apoyo de una técnica especial, frecuentemente un instrumento de medición. El resultado de una medición tiene que ser un número concreto que se expresa en unidades de medición sancionadas. La exactitud de una medición refleja la proximidad entre el resultado medido y el valor real del mensurando. La desviación del resultado medido

en relación con el valor real se denomina error de medición. El valor real de un mensurando se conoce únicamente cuando se cuenta con calibración de instrumentos de medición (Rabinovich, 2013). En este caso, el valor real es el valor del estándar de medición usado en la calibración. Algunos estándares de medición han sido establecidos arbitrariamente. De este modo, el segundo se definió con base al periodo de revolución de la Tierra alrededor del sol. Según avances recientes en física cuántica, el segundo se reproduce ahora mediante el estándar atómico de cesio el cual es tan exacto que toma casi 20 millones de años en acumular la deriva de 1 segundo. El metro se definió inicialmente con base a la longitud del Meridiano de París. Actualmente la definición del metro se basa en la velocidad de la luz en el vacío que se considera una constante física conocida con exactitud. Del mismo modo, se han creado estándares primarios para el voltio y el amperio.

Errores en la medición de la lluvia

La medición de la lluvia parece ser un proceso simple y, de hecho, se han realizado durante largo tiempo, en algunos casos durante siglos, mediciones de precipitación puntual propensas a errores sistemáticos en el orden del 10% (Shuttleworth, 2012). La medición del promedio de lluvia sobre un área determinada es mucho más difícil, particularmente si se necesita el promedio de un área grande con relieve montañoso.

La mayoría de los pluviómetros no están equipados con escudos de viento que permitan enfrentar la turbulencia que influye en la exactitud del instrumento. Por lo tanto, las mediciones pluviométricas tienden a subestimar la lluvia real entre un 5% y 40% con un sesgo promedio de 9% (Shelton, 2009).

El diccionario Webster define error como “la diferencia entre el valor observado o calculado y el valor real.” Frecuentemente no se conoce el valor real ya que de otro modo no se realizarían observaciones o cálculos. Se puede tener una aproximación del valor real, pero tiene que determinarse a partir de los datos y condiciones en el campo que tan confiable es el resultado (Bevington y Robinson, 2003). Los instrumentos de medición no son perfectos por lo que ocasionan errores de medición. Debido a que la medición es un procedimiento experimental los errores de medición son inevitables por lo que el valor real de la cantidad medible es un concepto abstracto. El valor real es el valor que se obtiene mediante una medición perfecta. En áreas científicas como química analítica no existe una medición perfecta de modo que nunca se conoce el valor real de una medición. En cuanto a los monitoreos hidrogeológicos, el nivel freático no es un concepto simple ni una variable fácil de cuantificar. A la fecha no existe un tratado detallado de lo que significa “nivel estático”. El nivel estático es una de las variables hidrogeológicas que se mide con frecuencia y posiblemente se mide de manera inadecuada (Baird y Low, 2022). El aspecto crítico resulta ser que tan mal tienen que estar las mediciones para que dejen de ser útiles. Los términos relacionados con los errores de medición en el contexto hidrogeológico son los siguientes (Rau y otros, 2019):

- Exactitud es una medida de que tan cerca se encuentra la media de la presión hidráulica medida en relación con la presión real. La desviación entre el valor real y la media de las mediciones constituye el error sistemático (o absoluto (Figura 1).
- Precisión es la propagación de las presiones medidas alrededor de su valor medio. Cuando las mediciones tienen una

distribución normal la precisión se puede expresar mediante la desviación estándar. También se conoce como error aleatorio (Figura 1).

- Resolución es la separación numérica más pequeña a la cual no puede distinguirse el cambio del valor real.
- Rango es la diferencia entre el valor máximo y mínimo que un instrumento puede medir.

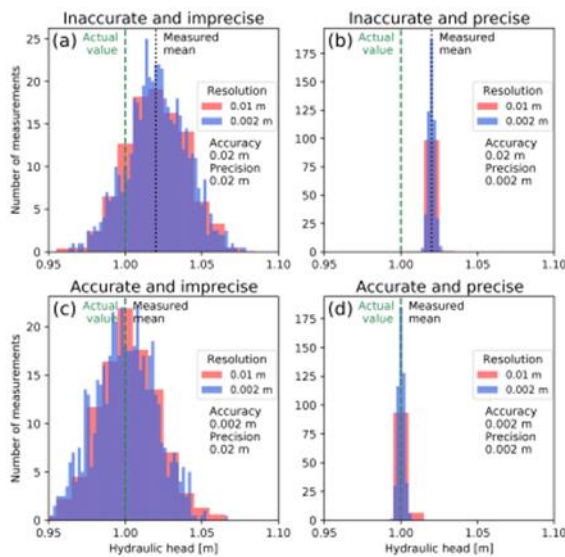


Figura 1. Combinaciones posibles de exactitud, precisión y resolución para un total de 1,000 mediciones del mismo nivel estático (traducido de Rau y otros, 2019). Las mediciones son (a) inexactas e imprecisas, (b) inexactas y precisas, (c) exactas e imprecisas, y (d) exactas y precisas. Actual value = valor real; measured mean = media medida.

Cualquier medición se puede considerar como la suma del valor real más las contribuciones de cada uno de los dos componentes del error:

Valor medido = valor real + error sistemático + error aleatorio.

Los errores sistemáticos se originan a partir de un equipo de medición sin calibración adecuada, por sesgos del observador, y/o por errores durante la medición. La magnitud de los errores sistemáticos tiende a ser mucho más grande que la de los errores

aleatorios por lo que el no reconocer la existencia de errores sistemáticos puede afectar significativamente los resultados de la investigación. Los errores sistemáticos tienen que evitarse o identificarse y corregirse. En gran parte, los errores sistemáticos se derivan de las condiciones de medición en el campo las cuales no son fáciles de controlar y afectan negativamente el desempeño del equipo de monitoreo. Otros factores que también afectan las mediciones incluyen el uso inadecuado del equipo y la falta de protocolos de medición. No todos los errores sistemáticos pueden eliminarse, pero para minimizar el error humano las mediciones deben realizarse con personal capacitado. Para la validación de los datos es crítico documentar el procedimiento de medición. Debe diseñarse un protocolo detallado de medición y evaluarse periódicamente, contando con procedimientos de control de calidad para asegurar la confiabilidad de los datos.

El error aleatorio varía entre mediciones, está asociado al error durante el muestreo, y tiende a ser igual a cero en el largo plazo. Pueden registrarse fluctuaciones aleatorias en las observaciones que producen resultados diferentes cada vez que se realizan las mediciones por lo que frecuentemente se requiere repetir las mediciones para generar resultados precisos. Para reducir los errores aleatorios debe mejorarse el método de medición y repetir o realizar más mediciones lo cual permitirá reducir la fluctuación estadística. Debido a que parte del error es aleatorio puede utilizarse métodos estadísticos para evaluar los errores de medición. En términos estadísticos se destacan dos aspectos en el proceso de medición: a) la exactitud determinada por el error sistemático y que constituye la diferencia entre la media estadística μ de la medición y el valor real de la medición. Si la media μ es igual al valor real el error sistemático es igual a 0;

b) la precisión que constituye el grado en que tienden a coincidir las mediciones repetidas de una misma cantidad. Si las mediciones realizadas resultan cercanas entre sí la precisión es alta y si las mediciones resultan dispersas la precisión es baja. Por lo tanto, la precisión se establece mediante la desviación estándar σ del proceso de medición. Mientras más pequeño sea el valor de σ más precisa es la medición. A σ también se le conoce como incertidumbre aleatoria o incertidumbre estadística del proceso de medición. Con σ también denota un intervalo de confianza el cual es un intervalo calculado que posiblemente contiene el valor real. Una de las mejores maneras de reducir la incertidumbre es tomar varias mediciones independientes y determinar el promedio de ellas. En este caso, las mediciones son una muestra aleatoria simple de una población y su promedio es la media de la muestra. Las mediciones erróneas producen un error en el valor calculado por lo que el error tiende a propagarse de las mediciones erróneas hacia el valor calculado (Navidi, 2006). Si se determina el tamaño de los errores en las mediciones se puede utilizar métodos específicos para conocer la magnitud del error en la cantidad calculada.

Definición de incertidumbre

“La incertidumbre no es únicamente una señal de ignorancia humana; la incertidumbre es de lo que está hecho el mundo”. (Stewart, 2019).

La incertidumbre puede definirse simplemente como la ocurrencia de eventos que están fuera de nuestro control lo cual se atribuye a la falta de información relacionada con datos y procesos involucrados en la definición y resolución de un problema. Las incertidumbres son inevitables en los proyectos hidrogeológicos y pueden ser de tipo geológico, hidráulico, hidrológico, climatológico, estructural, ambiental, y socioeconómica.

Tipos y fuentes de incertidumbre

Se distinguen dos tipos de incertidumbres (Tung y Yen, 2005): a) incertidumbres objetivas asociadas a procesos aleatorios o deducibles de muestras estadísticas, y b) incertidumbres subjetivas para las cuales no existe información cuantitativa objetiva. La Figura 2 ilustra que en el contexto de ingeniería de hidrosistemas existen dos fuentes de incertidumbres: a) aleatoria, relacionada con la variabilidad natural y b) epistémica vinculada con la deficiencia de conocimiento. Ambos tipos de incertidumbre se expresan en términos de probabilidades (Dowd, 2018). En muchos casos el énfasis se aplica en la incertidumbre aleatoria considerando que la incertidumbre epistémica es despreciable. En general, la incertidumbre derivada de la aleatoriedad inherente a los procesos geofísicos tal como la ocurrencia de deslizamientos, inundaciones, y terremotos no puede eliminarse y debe de abordarse mediante un proceso de adaptación. Esta variabilidad natural puede manifestar contrastes significativos en el tiempo y el espacio y su magnitud puede ser muy difícil de pronosticar. Las incertidumbres asociadas con la deficiencia de conocimiento pueden reducirse mediante la investigación científica y toma de datos.

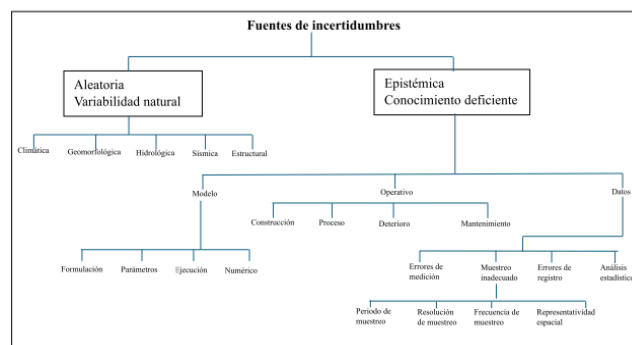


Figura 2. Diagrama que ilustra las fuentes de incertidumbres en la ingeniería de hidrosistemas (ligeramente modificado de Tung y Yen, 2005).

Medidas de la incertidumbre

Una manera práctica de cuantificar el nivel de incertidumbre para un parámetro determinado consiste en utilizar los momentos estadísticos de una variable aleatoria asociados con una cantidad sujeta a la incertidumbre. En particular, el momento de segundo orden denominado varianza es una medida de la dispersión de una variable aleatoria. Puede utilizarse ya sea la varianza o la desviación estándar (la raíz cuadrada positiva de la varianza). Otra medida de incertidumbre sobre el rango de una variable consiste en expresar la incertidumbre en términos de un dominio de confiabilidad, tal como el intervalo de confianza. El intervalo de confianza es un intervalo numérico que incluiría el valor real de una variable sujeta a incertidumbre, con una confianza probabilística especificada. La descripción más completa de las características de incertidumbre de una cantidad lo aporta la función de densidad de probabilidad (FDP). La FDP es una relación matemática que describe como puede variar la densidad de probabilidad de una variable aleatoria continua sobre un rango permisible de valores. La FDP permite establecer restricciones en los diferentes tipos de aleatoriedad y predecir eventos futuros aleatorios los cuales tienen una estructura que puede describirse matemáticamente.

Método Monte Carlo

El nombre Monte Carlo proviene de la ciudad en la costa mediterránea de Francia, en el Principado de Mónaco (208 Ha) con su famoso casino. El método Monte Carlo pudo haberse originado cuando un matemático trató de establecer cuantos pasos tendría que caminar un borracho para alejarse una distancia específica de una cantina, asumiendo que cada uno de sus pasos tenía la misma probabilidad de ser lanzado en cualquiera de las cuatro direcciones principales (Stoian, 1965). Este escenario conduce al concepto clásico de

“caminata aleatoria” el cual consiste en simular el movimiento de un objeto que no sigue un patrón o tendencia discernible sino una ruta que consiste en una secuencia de pasos aleatorios. El método Monte Carlo fue inventado por el matemático de Polonia Stanislaw Ulam a finales de la década de 1940 en el proyecto relacionado con la construcción de la bomba de hidrógeno. El método lo denominó Monte Carlo en honor a un tío suyo que frecuentaba el casino en esa localidad.

La Simulación Monte Carlo (SMC) permite calcular numerosos escenarios de un modelo mediante la selección repetitiva de valores a partir de una distribución de probabilidad predefinida por el científico. La SMC se basa en la generación de ensayos múltiples para determinar el valor esperado de una variable aleatoria. Los datos artificiales se producen mediante un generador de números aleatorios y la distribución de probabilidad acumulada seleccionada. El generador de números aleatorios tiende a producir números reales con distribución uniforme entre 0 y 1 obtenidos a partir de distribuciones de probabilidad específicas. Se utiliza el término números pseudoaleatorios para los números generados por el algoritmo de una computadora el cual ha sido diseñado para producir una secuencia de números sin correlación aparente los cuales se distribuyen uniformemente sobre un rango predefinido (Bevington y Robinson, 2003).

Simulación Monte Carlo (SMC) usando Excel para el cálculo de incertidumbre

Excel puede ser una herramienta predictiva poderosa para implementar la SMC (Farrance y Frenkel, 2014). Excel permite generar múltiples ensayos mediante la propagación de una fórmula básica según el número de iteraciones que requiera el modelo. El cálculo de incertidumbre de las mediciones usando Excel utiliza un enfoque general

denominado “propagación de distribuciones” el cual se implementa numéricamente mediante la SMC. El resultado de la medición se calcula mediante un procedimiento numérico iterativo con base a un modelo de medición que incluya todos los factores individuales que afectan el resultado y que pueda expresarse mediante una ecuación. Adicionalmente, se requiere para las cantidades de entrada una función de densidad de probabilidad (FDP) tal como la distribución normal, gamma, etc. Los valores de entrada se seleccionan aleatoriamente a partir de la FDP predefinida. La SMC implementa un muestreo aleatorio a partir de la FDP de las cantidades de entrada. Para cada entrada, la SMC genera un valor numérico extraído al azar de la respectiva FDP. Este proceso se repite un gran número de veces (ensayos Monte Carlo) para producir un conjunto de resultados simulados de salida.

Los estadísticos asocian aleatoriedad con probabilidad. La emisión de rayos gamma desde una riolita enriquecida en uranio puede considerarse aleatoria en el sentido de que sería imposible predecir el valor de la intensidad de la radiación a partir de emisiones anteriores, y no presenta ningún patrón ni correlación con emisiones vecinas. Del mismo modo, la intensidad de la lluvia puede considerarse una variable aleatoria con un valor mínimo y máximo y con distribución normal. En muchos casos la FDP Gaussiana se considera apropiada ya que representa el comportamiento de muchos fenómenos naturales. Para una FDP normal, la propagación de números aleatorios está predeterminada por la media específica μ y la distribución estándar específica σ . La media y desviación estándar de los resultados de salida constituyen los estimados respectivos del mensurando y su incertidumbre estándar. La SMC siempre producirá una FDP para la cantidad de salida que será consistente con la FDP de las

distintas entradas. La SMC calcula el resultado correspondiente a un valor de cada cantidad de entrada seleccionado de manera aleatoria a partir de la FDP, y repite este cálculo un gran número de veces, con la cantidad de ensayos normalmente de 10^5 a 10^6 . Este proceso produce un grupo de resultados simulados los cuales, bajo ciertos supuestos, forman una aproximación a la FDP para el valor del mensurando. Mientras mayor sea el número de ensayos más “estable” será la desviación estándar de salida. La variación de la desviación estándar disminuirá a medida que el número de ensayos incrementa.

Metodología

Simulación Monte Carlo (SMC)

Para ilustrar el cálculo de la incertidumbre con la SMC usando Excel se utilizan los datos de lluvia mensual en el Corredor Seco de Guatemala, estación La Fragua, que administra el INSIVUMEH en Zacapa. Se considera que las variaciones aleatorias se superponen en las fluctuaciones estacionales o periódicas (Yevjevich, 1972). La formulación de la SMC en Excel para este contexto hidrogeológico particular sigue los siguientes pasos:

Paso 1. Configuración en Excel de la SMC

Los requerimientos para implementar la SMC son los siguientes:

- Un modelo de medición que describa el proceso de medición en términos de las entradas al mensurando.
- Una ecuación que describa el mensurando (salida), en este caso lluvia anual, en función de las entradas relevantes, para este caso particular los valores de lluvia mensual.
- Una evaluación de las FDPs que aplican a las varias incertidumbres de entrada. Las

incertidumbres en las entradas deben de expresar incertidumbres estándar. La incertidumbre de entrada se obtiene a partir del análisis estadístico de mediciones repetidas (lluvia).

La SMC genera una serie de observaciones aleatorias basadas en un tipo específico de FDP las cuales pueden representarse como curvas de probabilidad que determinan la posibilidad de un resultado particular. La FDP de una variable aleatoria normal con media μ y varianza σ^2 está dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Cuando se utiliza la FDP normal se espera que un 68% de los datos caigan dentro de una desviación estándar (σ) de la media, el 95% dentro de dos desviaciones estándar, y aproximadamente el 99.7% dentro de tres desviaciones estándar (Figura 3).

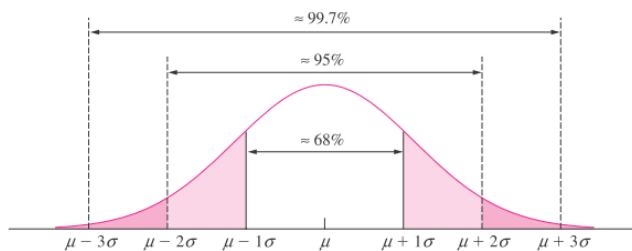


Figura 3. Diagrama que ilustra la función de densidad de probabilidad de una variable aleatoria normal con media μ y desviación estándar σ (Navidi, 2006).

Se requieren los parámetros que determina la FDP para cada una de las cantidades de entrada; para una distribución normal estos parámetros son la media y la desviación estándar. Para generar los valores aleatorios de lluvia basados en una distribución normal se utilizan el promedio mensual de lluvia (media especificada) y la desviación estándar respectiva (desviación estándar especificada).

Paso 2. Crear filas para las iteraciones (ensayos) de la SMC donde se generarán los números aleatorios. La configuración de la SMC en Excel incluye la creación de dos columnas etiquetadas “Ensayo Monte Carlo #” y “Variable Aleatoria Normal”. Posteriormente se crean dos campos con etiquetas “Media Simulada” y “Desviación Estándar Simulada”.

Paso 3. Generar las variables de valor aleatorio

La hoja electrónica se construye con columnas destinadas para cada una de las cantidades de entrada (lluvia mensual), la ecuación que en este caso es la sumatoria de los promedios mensuales de lluvia, y la cantidad de salida (mensurando, lluvia anual). En cada una de las columnas asignadas para las distintas cantidades de entrada debe de ingresarse la fórmula Excel adecuada para generar números aleatorios según la FDP especificada y arrastrarse hacia abajo según el número de iteraciones requeridas.

La SMC en Excel utiliza dos funciones específicas: ALEATORIO() y INV.NORMAL. La función ALEATORIO() introduce variabilidad para simular aleatoriedad mediante una fórmula incorporada que genera valores aleatorios numéricos decimales entre 0 y 1. La función INV.NORMAL aporta el inverso de la FDP normal para la media y desviación estándar especificada. La fórmula que utiliza la media especificada y la desviación estándar especificada en celdas fijas mediante el símbolo \$ para generar los valores aleatorios es la siguiente:

$$\text{INV.NORM(ALEATORIO();\$G\$13;\$G\$14)}$$

La fórmula se arrastra hacia abajo en la columna hasta generar el número de ensayos (iteraciones MC) requerido. La iteración consiste en repetir varias veces el proceso con fines de alcanzar la meta deseada o el resultado solicitado. Los resultados de

una iteración se usan como punto de partida para la siguiente iteración. Cada vez que se recalcula la formula se generan nuevos valores de lluvia que siguen la FDP. Frecuentemente se requieren 105 a 106 iteraciones ya que la exactitud de la simulación aumenta con el tamaño de la muestra. En este ejemplo para el Corredor Seco se han implementado 1,000 iteraciones.

Paso 4. Verificar los valores

Para verificar si los resultados están de acuerdo con la FDP seleccionada se calcula la media y desviación estándar para el conjunto de datos generados con la SMC obteniendo la media simulada y la desviación estándar simulada. La SMC genera un conjunto de datos hipotéticos que se pueden comparar con los datos reales de campo. Al comparar los datos específicos con los simulados se debe estar en línea con los esperados.

Paso 5. Visualizar los resultados de la SMC

Los datos en la columna de salida (mensurando, lluvia anual) puede ser sujetos a una evaluación que incluya construcción de un histograma y cálculo de parámetros estadísticos tal como media, moda, mediana y desviación estándar. Debido a que se utilizó la función INV.NORM para generar los valores de la variable aleatoria la gráfica debería de tener una curva en forma de campana. Esto puede verificarse graficando los datos aleatorios generados en un histograma.

Resultados y Discusión

Análisis Exploratorio de Datos

El Cuadro 1 muestra el resumen estadístico para la lluvia mensual y anual (1990-2017) para los datos que administra el INSIVUMEH para la estación de La Fragua. Los parámetros estadísticos calculados indican una fuerte fluctuación de la lluvia ya sea a

escala temporal mensual (Cuadro 1) o anual (Figura 4).

Cuadro 1. Resumen estadístico (1990-2017) para datos de lluvia de La Fragua.

	En	Fe	Ma	Ab	Ma	Ju	Jul	Ag	Se	Oct	No	Dic	Anual
μ media especificada	1.9	1.2	2.6	28.7	92.7	175.8	104.4	122.0	172.2	89.5	17.4	4.0	811.7
σ desviación estándar especificada	2.8	2.0	3.0	16.1	66.9	66.6	80.8	67.4	83.6	72.9	27.2	6.6	239.6
Mínimo	0.0	0.0	0.0	0.0	15.0	17.9	21.8	24.3	79.5	11.9	0.0	0.0	43.0
Máximo	9.7	8.6	13.3	142.5	250.3	327.4	398.1	333.3	430.1	331.5	102.0	31.2	1494.0

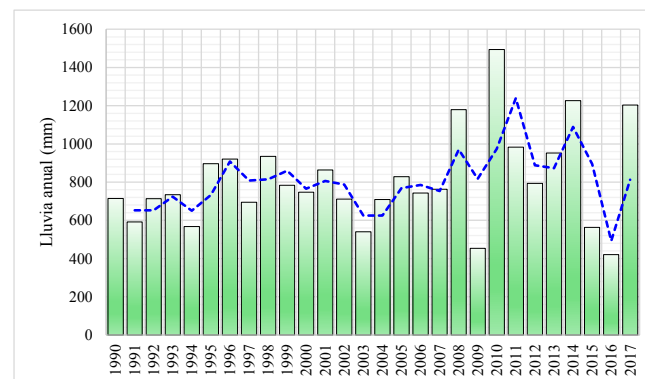


Figura 4. Lluvia anual con fuertes contrastes temporales. La línea de tendencia media móvil resume el comportamiento a largo plazo de la lluvia anual.

En términos de lluvia anual puede notarse una fuerte variación aleatoria asociada con mínimos en los años 2009 y 2016 y máximos registrados en los años 2008, 2010, 2014 y 2017. La variabilidad anual de lluvia registró menos contrastes para el periodo 1990 a 2007.

Los meses de verano manifiestan un fuerte contraste en la desviación estándar especificada σ lo cual indica una notable dispersión en los valores de lluvia con presencia de extremos pluviales (Figura 5). Para enero y febrero se manifiesta una fluctuación aleatoria en las intensidades de lluvia mientras que para marzo la línea de tendencia media móvil para la intensidad de lluvia registra un comportamiento cíclico.

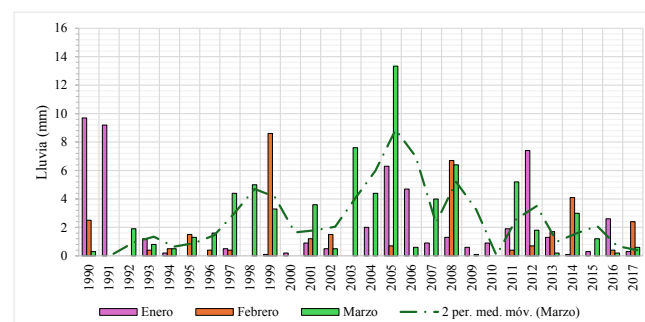


Figura 5. Diagrama de barras que muestra la lluvia para los meses del verano con notable manifestación de valores atípicos y fuerte dispersión.

Los resultados crudos de las SMC realizadas para 1,000 iteraciones siguiendo la metodología descrita se incluyen en el Anexo I. El Cuadro 2 muestra la comparación entre los datos de lluvia de campo (especificados) y los datos hipotéticos de lluvia (simulados) generados con la SMC.

Cuadro 2. Verificación entre datos de campo y valores generados con la SMC.

Datos reales de campo. Precipitación mensual o anual de lluvia (1998-2017) para la estación del INSVUMER en La Fragua, Zangaza											
	En	Fe	Ma	Ab	Ma	Ju	Ju	Ag	Se	Oct	Nov
Especificado	1.0	1.2	4	20.7	62.7	175.0	63.5	122	172.2	88.5	17.4
Desviación (incertidumbre)	2.9	7	8.3	28.1	63.9	68.6	38.9	67.4	83.5	72.9	17.2
El valor esperado (simulado) de la variable aleatoria generada de la simulación de las 1,000 iteraciones realizadas.											
Simulado	2.0	1.2	3.6	30.3	69.3	178.3	61.7	121.7	169.4	86.7	17.4
Desviación (incertidumbre)	2.9	7.0	8.4	26.5	67.0	68.1	38.4	67.6	84.7	73.3	17.8

Cada vez que el archivo Excel se abre, se guarda, o se resetea los valores de la media y desviación estándar simulados son recalculados y cambian ligeramente. En la Figura 6 se ilustra la comparación entre los datos de campo y los datos simulados con la SMC. La línea de tendencia media móvil suaviza la fluctuación de los datos y muestra que tanto para los datos de campo como para los datos simulados el comportamiento estacional de la lluvia se aproxima a una distribución tipo Gaussiana. El descenso mensual de lluvia para julio y agosto suele estar asociado con la canícula. Los datos especificados se cargan a la izquierda de los datos simulados para meses de verano con un valor medio ligeramente mayor en los valores especificados de lluvia. Para la salida del invierno el valor medio especificado mensual de lluvia tiende a ser menor que el simulado, con la línea de tendencia media móvil no siendo representativa de la recesión hidrometeorológica de fin de año, tanto para los valores de campo especificados como para los valores simulados. Sin embargo, puede notarse en la Figura 6 que los valores simulados de lluvia se encuentran bastante cercanos a los valores

especificados de campo, con las mayores diferencias en junio y octubre.

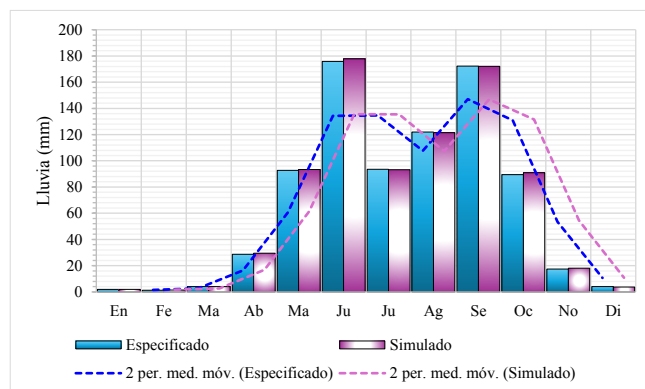


Figura 6. Diagrama de barras lluvia mensual comparando datos de campo especificados con datos teóricos simulados con SMC.

Para la visualización de los resultados de lluvia anual simulados con la SMC se construyó un histograma el cual tiene forma de campana (Figura 7) ya que fue generado con la función INV.NORM. Se registra una ligera asimetría positiva hacia la derecha lo cual sugiere que la distribución de los valores simulados no es completamente normal. Sin embargo, esta desviación ligera de la normalidad no altera significativamente los valores de incertidumbre obtenidos para la lluvia anual. Por lo tanto, puede esperarse que el 68% de los datos de lluvia anual generados sintéticamente caigan dentro de 1 desviación estándar de la media simulada, el 95% dentro de 2 desviaciones estándar, y aproximadamente el 99.7% de los valores de lluvia anual caigan dentro de 3 desviaciones estándar de la media simulada (Figura 3 y Figura 7).

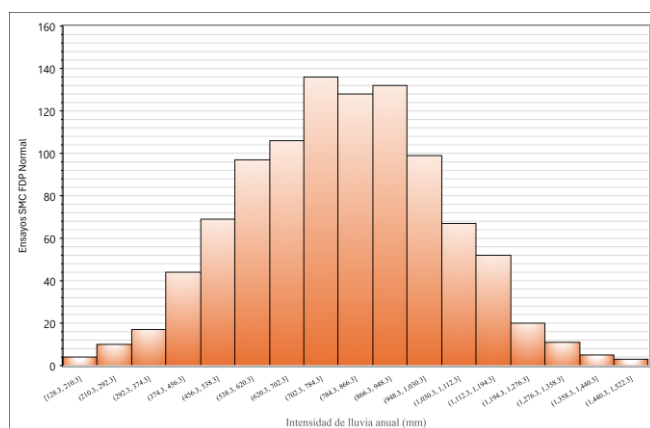


Figura 7. Histograma de lluvia anual sintético generado con 1,000 iteraciones de la SMC para estación La Fragua.

Conclusiones

La evaluación de la lluvia constituye un proceso esencial y desafiante para la planificación, desarrollo y manejo de recursos hídricos. Con la presente investigación se realiza un esfuerzo para enfrentar el problema de la incertidumbre hidroclimatológica en el Corredor Seco. La SMC realizada con Excel aporta un conjunto de datos hipotéticos derivados científicamente los cuales pueden compararse con los datos de campo y utilizarse para predicción de eventos mensuales y anuales de lluvia en el Corredor Seco de Zacapa. Los datos hipotéticos de lluvia generados por la SMC simulan bastante cerca los datos reales de campo que reporta la estación del INSIVUMEH en La Fragua.

La SMC implementada con Excel puede ser inadecuada para modelos más avanzados que involucran interdependencias complejas, relaciones no lineales, y sistemas altamente dinámicos para los cuales se requiere lenguajes de programación o software mucho más especializados tal como SPSS y SAS que demandan una curva pronunciada de aprendizaje. El primer modelo de SMC que incluyó 1,000 iteraciones se considera bastante representativo pero el tamaño de la muestra para la SMC puede ser aún mayor para disminuir los niveles de incertidumbre lo cual depende de varios factores,

incluyendo el nivel esperado de precisión y las características del escenario específico de simulación.

Varios procesos hidrogeológicos como la escorrentía se consideran un proceso aleatorio, pero cuando se calcula la escorrentía en función de la lluvia se puede adoptar un enfoque determinístico. Normalmente los procesos hidrogeológicos son una combinación de varios procesos determinísticos y estocásticos por lo que para producir un modelo de menor incertidumbre es recomendable combinar los métodos determinísticos y estadísticos para la exploración, desarrollo, y manejo de los recursos hídricos.

Referencias

- Baird, A., and Low, R., 2022, The water table: Its conceptual basis, its measurement and its usefulness as a hydrological variable: *Hydrological Processes*, v. 36, 18 p. <https://doi.org/10.1002/hyp.14622>.
- Bevington, P., & Robinson, D.K., 2003, *Data Reduction and Error Analysis for the Physical Sciences*. Third Edition. McGraw-Hill. 338 p.
- Dowd, P., 2018, Quantifying the Impacts of Uncertainty, in Daya Sagar, B., Cheng, Q., & Agterberg (eds.), *Handbook of Mathematical Geosciences*, p. 349-373.
- Farrance, I., & Frenkel, R., 2014, Uncertainty in Measurement: A Review of Monte Carlo Simulation Using Microsoft Excel for the Calculation of Uncertainties Through Functional Relationships, Including Uncertainties in Empirically Derived Constants: *Clinical Biochemist Reviews* v. 35, p. 37-61.
- Maity, R., 2018, *Statistical Methods in Hydrology and Hydroclimatology*: Springer. 444 p.
- McCuen, R., 2003, *Modeling Hydrologic Change: Statistical Methods*. CRC Press. 433 p.
- Navidi, W., 2006, *Estadística para ingenieros y científicos*: McGraw-Hill. 868 p.

- Rabinovich, S.G., 2013, Evaluating Measurement Accuracy: Springer. 324 p.
- Rau, G., Post, V., Shanafield, M., Krekeler, T., Banks, E., and Blum, P., 2019, Error in hydraulic head and gradient time-series measurements: a quantitative appraisal: Hydrology and Earth System Sciences, v. 23, p. 3603-3629.
- Salsburg, D., 2001, The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century. Holt Paperbacks. 340 p.
- Shelton, M.L., 2009, Hydroclimatology: Perspectives and Applications. Cambridge University Press. 426 p.
- Shuttleworth, W.J., 2012, Terrestrial Hydrometeorology. Wiley-Blackwell. 448 p.
- Stewart, I., 2019, Do Dice Play God? The Mathematics of Uncertainty. Profile Books Ltd. 292 p.
- Stoian, E., 1965, Fundamentals and Applications of the Monte Carlo Method: The Journal of Canadian Petroleum, v. 4, p. 120-129.
- Tung, Y.K., and Yen, B.C., 2005, Hydrosystems Engineering Uncertainty Analysis: McGraw-Hill. 274 p.
- Viessman, W., and Lewis, G., 2003, Introduction to Hydrology: Prentice Hall. 612 p.
- Yevjevich, V., 1972, Probability and Statistics in Hydrology: Water Resources Publications. 302 p.

