

UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA DE MINAS

**INFERENCIA DE LA DISTRIBUCIÓN REPRESENTATIVA DE COBRE SOLUBLE
EN PRESENCIA DE MUESTREO PREFERENCIAL**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL DE MINAS

JORGE LUIS SAA REYES

**PROFESOR GUÍA:
JULIÁN ORTIZ CABRERA**

**MIEMBROS DE LA COMISIÓN
EDUARDO MAGRI VARELA
ANTONIO COUBLE CERVIÑO**

**SANTIAGO DE CHILE
MARZO 2011**

RESUMEN.

Es bastante común, en el proceso de evaluación de yacimientos, encontrar varias especies de interés asociadas en un mismo cuerpo mineralizado, generalmente una ellas es más importante que las demás. Usualmente la variable más importante, que en este caso de estudio es el cobre total (CuT), es informada en todas las muestras obtenidas, mientras que la variable secundaria, en este estudio el cobre soluble (CuS), solo se analiza en ciertas zonas, a menudo las de más altas leyes de la variable primaria.

Esta decisión genera un sesgo en la distribución de la variable secundaria y en la distribución bivariable conjunta (CuT-CuS). Esto produce ciertas dificultades en el proceso de simulación, ya que estos métodos usan como información de entrada, las distribuciones (histogramas) de ambas variables, por ende, si esta no son representativas, es decir, si presentan sesgo debido al muestreo preferencial, los resultados obtenidos de la simulación reflejaran estos sesgos.

En esta memoria de título se presentan la metodología de una corrección de la distribución sesgada de cobre soluble, a partir de la distribución bivariable y las distribuciones de cobre total, enfatizando principalmente en dos aspectos: cómo cambian los resultados en la medida que el sesgo aumenta, y por otro lado, cómo cambian los resultados al aumentar o disminuir el grado de refinación con el cual se realiza la corrección.

Los resultados muestran que la corrección propuesta de la distribución sesgada, corrige el efecto del muestreo preferencial, en cuanto a estadísticas básicas, variabilidad espacial, distribución de valores tanto en CuT como en CuS, etc. Por esta razón, la corrección se considera exitosa, ya que estimaciones hechas a partir de distribuciones no corregidas sobrestiman los recursos del yacimiento hasta en un 100% para algunas leyes de corte.

Sin embargo, analizando los tonelajes reportados en las realizaciones se puede ver que, para una ley de corte que pertenezca a una de las clases que se vieron sesgadas, los tonelajes informados son levemente mayores que en el caso insesgado. Por el contrario, para leyes superiores a las clases sesgadas, el tonelaje medio informado es menor al caso sin sesgo.

El suavizamiento de la distribución sesgada parece ser necesario solo en casos extremos en que el sesgo sea excesivo o se tengan muy pocos datos para aplicar la corrección. El suavizamiento depende fuertemente de la decisión del usuario en cuanto al ajuste del histograma experimental y por esto los resultados pueden ser bastante distintos dependiendo de esta elección.

ABSTRACT.

It is quite common in ore deposit evaluation, finding several interesting species associated in the same ore body, usually one of these species is more important than others. Commonly the most important variable, which in this case study is the total copper grade (CuT) is fully informed, while the secondary variable, the soluble copper grade (CuS), is only sampled in certain areas, often of higher grades. This decision creates a bias in the distribution of the secondary variable and the bivariate distribution (CuT-CuS). This causes some difficulties in the simulation process.

The multivariate Gaussian model is widely used to perform simulation, requiring the grade data to be transformed into normal scores, i.e. into data with a standard Gaussian distribution. To this end, a representative distribution of the original grade data is needed, in order to avoid biases in the simulation outputs. Such a representative distribution is difficult to infer when the grade variable is not regularly sampled, in particular when low grade sectors are under sampled (preferential sampling design).

This situation is frequently found in oxide copper deposits, for which soluble copper grades are assayed preferentially at samples with high total copper grades. To infer a representative distribution of soluble copper grade, we propose a declustering technique that relies on the distribution of total copper grade, assumed representative, and on the bivariate distribution between total and soluble copper grades at the samples where both variables are assayed. The effects of using the original (biased) and corrected (representative) distributions to simulate soluble copper grades are illustrated through a case study.

The results show that due to the correction of the biased distribution, the ore body appears to have the same characteristics as the case in which no preferential sampling had been done in terms of basic statistics, spatial variability, distribution of values of CuT o CuS, etc. So the correction was considered successful, as estimates made from biased distributions overestimate the resources of the deposit up to 100% for some cut-off grades.

The smoothing of the distribution seems to be necessary only in extreme cases where the bias is excessive or a few data are available. The smoothing is heavily dependent on the user to decide how fit the curve with the experimental histogram and therefore the results can be quite different depending on that choice.

AGRADECIMIENTOS.

Agradezco principalmente a mi familia que me ha apoyado durante toda mi carrera, gracias a mis padres especialmente, a mi hermano y mi polola, son los que me han soportado todo este tiempo y me han facilitado las cosas para que haya llegado hasta este punto, sin duda, sin ellos hubiera sido todo mas difícil, soy realmente afortunado.

Agradezco también a los profesores de la comisión, por todos los comentarios y correcciones que recibí, especialmente a Julián que siempre estuvo disponible para las innumerables preguntas que tuve, siempre muy amable y paciente.

Mención especial tienen también los académicos y funcionarios del departamento de ingeniería de minas de la escuela, que me enseñaron de alguna u otra forma gran parte de lo que se.

Mención honrosa tienen los amigos que hice en el transcurso de la carrera, no solo los de nuestra portentosa generación, sino de los mineros del departamento que conocí en el camino, sin duda hicieron más llevadera las partes más difíciles de la carrera e hicieron que me sintiera a gusto.

Agradezco finalmente a todos los que han participado de este logro directa o indirectamente, ya que no solo me pertenece a mí.

Índice

1. INTRODUCCIÓN.....	1
1.1 Motivación del trabajo.....	1
1.2 Objetivos.....	2
1.3 Alcances.....	3
2. ANTECEDENTES.....	4
2.1 Evaluación de recursos: kriging y simulación.....	4
2.2 Simulación multivariable.....	5
2.3 Análisis de las variables en estudio, cobre total (CuT) y cobre soluble (CuS).	7
2.4 Corrección del sesgo debido a muestreo preferencial.....	11
3. METODOLOGIA.....	15
3.1 Detalles de la corrección del sesgo espacial.....	15
3.2 Implementación en el lenguaje python.	20
3.3 Implementación de simulación con transformación condicional por pasos.....	21
3.4 Potenciales problemas.	23
3.5 Suavizamiento de la distribución bivariable – Polinomios de Bernstein.....	24
4. CASOS DE ESTUDIO.....	28
4.1 Base de datos.....	28
4.2 Construcción de las bases de datos sesgadas.	29
4.3 Estudio exploratorio de datos.....	30
4.3.1 Despliegue de datos.	30
4.3.2 Estadísticas básicas.....	33
4.3.3 Otros análisis.....	35
4.4 Estudio variográfico.	36
5. IMPLEMENTACION DE LA CORRECCIÓN.....	38
5.1 Dependencia del factor de corrección con el número de clases.	43
6. SIMULACIÓN CONDICIONAL.....	46
6.1 Transformación de las variables CuT y CuS en variables gaussianas.....	46
6.2 Estudio variográfico de las variables transformadas.	48
6.2.1 Estudio variográfico variable gaussiana CuT.....	48
6.2.2 Estudio variográfico variable gaussiana CuS.....	49
6.2.3 Variograma cruzado CuT – CuS	50

6.3	Validación cruzada.	51
6.4	Detalles de la simulación condicional.	52
6.5	Validación de la simulación.	53
6.5.1	Resultados de la simulación –Soporte puntual.	53
6.6	Transformación de vuelta a los datos originales.	57
6.7	Validación de los resultados en las variables originales.	59
6.7.2	Resultados de la simulación – Soporte puntual.	59
6.7.2	Resultados de la simulación – Soporte de bloques.	61
7.	ANÁLISIS DE LOS RESULTADOS.	63
7.1	Evaluación de las diferencias debido al grado de sesgo.	63
7.1.1	Evaluación gráfica.	66
7.1.2	Análisis de incertidumbre.	68
7.1.3	Análisis de varianzas.	70
7.2	Comparación utilizando la distribución sesgada de CuS en la estimación.	71
7.3	Aplicación del suavizamiento mediante polinomios de Bernstein.	74
7.4	Resultados de la simulación con suavizamiento.	77
8.	DISCUSIÓN DE LOS RESULTADOS.	78
9.	CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO.	80
10.	REFERENCIAS.	81
11.	ANEXOS.	82
	ANEXO A. Estudio exploratorio- Despliegue de datos.	82
	ANEXO B. Dependencia del factor de corrección con el número de clases.	83
	ANEXO C. Transformación gaussiana – Bases de datos sesgadas.	84
	ANEXO D. Estudio variográfico – Variables gaussianas.	87
	ANEXO E. Realizaciones de los casos con sesgo corregidos.	89
	ANEXO F. Resultados de la simulación.	90
	ANEXO G. Comparación entre los casos de estudio.	93
	ANEXO H. Análisis de incertidumbre - Ley media.	95

INDICE DE TABLAS.

Tabla 1 - Diferencia entre el muestreo homotópico y heterotópico.....	11
Tabla 2 - Metodología para asignar un orden a valores originalmente iguales.	22
Tabla 3 - Compositos sin ninguna información en la base de datos.	28
Tabla 4 - Estadísticas básicas de las bases de datos.....	29
Tabla 5 - Distancias máximas en cada dirección.....	31
Tabla 6 - Parámetros del variograma experimental.	36
Tabla 7 - Ejemplo de archivo de datos.....	38
Tabla 8 - Ejemplo de archivo de parámetros.....	39
Tabla 9 - Comparación de las estadísticas básicas entre el caso base y los casos sesgados corregidos.....	45
Tabla 10 - Parámetros para generar el variograma experimental.	48
Tabla 11 - Comparación de las estadísticas básicas reportadas de los diferentes casos de estudio.	63
Tabla 12 - Comparación de los tonelajes reportados por cada uno de los casos de estudio para una ley de corte de 0.1% CuS.	67
Tabla 13 - Comparación de los tonelajes y leyes medias reportados por cada uno de los casos de estudio para una ley de corte de 0.35% CuS.	68
Tabla 14 - Estadísticas básicas de los casos sin corrección y el caso base.....	73
Tabla 15 - Estadísticas básicas, comparación entre el caso base y los casos suavizados.....	77
Tabla 16 - Comparación entre el caso base y los casos sesgados, corrección normal.	93
Tabla 17 - Comparación entre el caso base y los casos sesgados, casos sin corrección.	93
Tabla 18 - Comparación entre el caso base y los casos sesgados, usando suavizamiento.	94

INDICE DE FIGURAS

Figura 1 – Construcción del histograma acumulado de la variable original.	7
Figura 2 - Proceso de transformación a valores gaussianos.	8
Figura 3 - Histogramas de las variables gaussianas.	9
Figura 4 - Nube de correlación entre las variables gaussianas transformadas independientemente.	10
Figura 5 - Histograma de frecuencias absolutas, comparación entre el caso sin sesgo y el con sesgo.	12
Figura 6 - Histogramas de frecuencia relativa, comparación entre el caso sin sesgo y el con sesgo.	13
Figura 7 - Grafico cuantil contra cuantil entre la distribución real y la distribución sesgada.	13
Figura 8 - Primer paso de la corrección: Generación de las distribuciones.	17
Figura 9 - Segundo paso de la corrección: Calcular la corrección para cada una de las clases.	18
Figura 10 - Tercer paso de la corrección: Actualización la distribución bivariable y obtención de la distribución corregida de CuS.	19
Figura 11 - Construcción del histograma acumulado para la transformación condicional a valores gaussianos.	21
Figura 12 - Esquema del suavizamiento de los histogramas.	23
Figura 13 - Interpolación exacta de los puntos extremos usando polinomios de Berstein.	26
Figura 14 - Vista isométrica del yacimiento en estudio. Ley de CuT.	30
Figura 15 – Proyección de los datos en planta. Ley de CuT (izquierda) y CuS (derecha).	31
Figura 16- Plantas para determinar el espaciamiento entre datos.	32
Figura 17 - Histogramas y estadísticas básicas, base de datos completa (Caso base).	33
Figura 18 - Histogramas y estadísticas básicas, base de datos sesgada 50% (izq) y 80% (der).	34
Figura 19 - Comparación de las distribuciones de CuT y CuS con una distribución lognormal.	35
Figura 20 - Mapas variográficos cobre total (CuT).	36
Figura 21 - Mapas variográficos cobre soluble (CuS).	37
Figura 22 - Variogramas experimentales de CuT y CuS.	37
Figura 23 - Histogramas de las distribuciones necesaria para hacer la corrección.	40
Figura 24 - Histogramas de las distribuciones de CuS, caso base (izq) y caso sesgado 50% (der).	40
Figura 25 - Comparación entre la distribución corregida y la distribución sesgada de cobre soluble (Caso sesgado 50%).	41
Figura 26 - Comparación entre la distribución corregida y la distribución sin corregir.	41
Figura 27 - Comparación entre la distribución corregida y la distribución sesgada de cobre soluble (Caso sesgado 50%).	42
Figura 28 - Comparación entre la distribución corregida y la distribución sin corregir.	42
Figura 29 - Dependencia del factor de corrección del número de clases.	43

Figura 30 - Comparación de las distribuciones de CuS, caso base y casos sesgadas.	44
Figura 31 - Comparación en detalle de las distribuciones.....	45
Figura 32 - Distribución bivariable, variables gaussiana CuT y gaussiana CuS.	46
Figura 33 - Histogramas y estadísticas básicas de las variables gaussiana CuT (izq) y gaussiana CuS (der).	47
Figura 34 - Comparación de las variables gaussianas con una distribución normal de media 0 y varianza 1.	47
Figura 35 - Mapas variográficos variable gaussiana CuT.....	48
Figura 36 - Variogramas experimental y modelado de la variable gaussiana CuT.....	49
Figura 37 - Mapas variográficos variable gaussiana CuS corregida y desagrupada.....	49
Figura 38 - Variogramas experimental y modelado, variable gaussiana CuS.....	50
Figura 39 - Variograma cruzado, variables gaussiana CuT y gaussiana CuS.	50
Figura 40 - Validación cruzada.....	51
Figura 41 - Distribución bivariable de los valores simulados en soporte puntual.	53
Figura 42 - Histogramas y estadísticas básicas de los valores simulados en cada uno de los casos de estudio, a soporte puntual y como promedio de 100 realizaciones.....	54
Figura 43 – Histogramas y estadísticas básicas de 3 realizaciones de la variable gaussiana CuT.....	55
Figura 44 - Histogramas y estadísticas básicas de 3 realizaciones de la variable gaussiana CuS en las distintas bases de datos.	55
Figura 45 - Histograma y estadísticas básicas de las 100 realizaciones, variable gaussiana de CuT.	56
Figura 46 - Histograma y estadísticas básicas de las 100 realizaciones, variable gaussiana de CuS.....	56
Figura 47 - Transformación de vuelta, interpolación lineal.	57
Figura 48 - Transformación de vuelta, extrapolación hiperbólica.	58
Figura 49 - Histogramas y estadísticas básicas de 3 realizaciones de la variable CuT a soporte puntual..	59
Figura 50 - Histogramas y estadísticas básicas de 3 realizaciones de la variable gaussiana CuS a soporte puntual.....	59
Figura 51 - Visualización grafica de 3 realizaciones, ley de CuS.....	60
Figura 52 - Comparación entre las distribuciones de valores y simulados y los datos originales. Arriba comparando con los resultados del caso sesgado 50% y abajo comparando con el caso sesgado 80%.....	60
Figura 53 - Histogramas y estadísticas básicas de 3 realizaciones de la variable CuT a soporte de bloques.	61
Figura 54 - Histogramas y estadísticas básicas de 3 realizaciones de la variable CuS a soporte de bloques.	61
Figura 55 - Nubes de correlación CuT - CuS para los distintos casos de estudio.....	62
Figura 56 - Comparación de los histogramas de las estimaciones para cada uno de los casos de estudio..	64

Figura 57 - Grafico cuantil contra cuantil. Caso base y caso sesgado 50% .	65
Figura 58 - Grafico cuantil contra cuantil. Caso base y caso sesgado 80% .	65
Figura 59 - Comparación grafica, leyes de CuS. Caso base (izq), caso sesgado 50% (centro) y caso sesgado 80% (der).	66
Figura 60 - Comparación gráfica, probabilidad de superar la ley de corte 0,2% CuS. Caso base (izq), caso sesgado 50% (centro) y caso sesgado 80% (der).	66
Figura 61 - Comparación de las curvas tonelaje ley para cada uno de los casos de estudio.	67
Figura 62 – Histogramas de tonelajes para una ley de corte 0.1 % (de derecha a izquierda, caso base, sesgo del 50%, sesgo del 80%).	69
Figura 63 - Histogramas de tonelajes para una ley de corte 0.1 % (de derecha a izquierda, caso base, sesgo del 50%, sesgo del 80%).	69
Figura 64 - Comparaciones de los histogramas de las varianzas de cada realización en cada uno de los casos de estudio.	70
Figura 65 - Comparación de las curvas tonelaje ley, usando la corrección (caso sesgado 50%) y sin usar la corrección.	71
Figura 66 - Comparación de las curvas tonelaje ley, usando la corrección (caso sesgado 80%) y sin usar la corrección.	72
Figura 67 - Grafico cuantil contra cuantil. Caso base y caso mal hecho sesgado 50%.	72
Figura 68 - Gráfico cuantil contra cuantil. Caso base y caso mal hecho sesgado 80%.	73
Figura 69 - Comparación entre el histograma experimental y el histograma suavizado.	74
Figura 70 - Histograma suavizado, 100 puntos de control.	75
Figura 71 - Histograma suavizado, 400 puntos de control.	75
Figura 72 - Histograma suavizado, 2500 puntos de control.	76
Figura 73 - Curva tonelaje - ley, comparación entre el caso base y el caso suavizado.	77
Figura 74 - Vista Este - Oeste. Ley de CuT (arriba) y CuS (abajo).	82
Figura 75 - Vista Norte - Sur. Ley de CuT (arriba) y CuS (abajo).	82
Figura 76 - Dependencia del factor de corrección del grado de discretización, caso sesgado 80%.	83
Figura 77 - Distribución bivariable, variables gaussianas, base de datos sesgada 50%.	84
Figura 78 - Estadísticas básicas e histogramas de las distribuciones marginales, variables gaussianas, caso sesgado 50%.	84
Figura 79 - Comparación de las distribuciones marginales con una normal de media 0 y varianza 1.	85
Figura 80 - Distribución bivariable, variables gaussianas, base de datos sesgada 80%.	85
Figura 81 - Estadísticas básicas e histogramas de las distribuciones marginales, variables gaussianas, caso sesgado 80%.	86
Figura 82 - Comparación de las distribuciones marginales con una normal de media 0 y varianza 1.	86

Figura 83 - Mapas variográficos, variable gaussiana CuS, caso sesgado 50%.....	87
Figura 84 - Variogramas experimental y modelado, variable gaussiana CuS, caso sesgado 50%.	87
Figura 85 - Mapas variográficos, variable gaussiana CuS, caso sesgado 80%.....	88
Figura 86 - Variogramas experimental y modelado, variable gaussiana CuS, caso sesgado 80%.	88
Figura 87 - Histogramas y estadísticas básicas de 3 realizaciones de la variable CuS a soporte puntual. Caso sesgado 50%.	89
Figura 88 - Histogramas y estadísticas básicas de 3 realizaciones de la variable CuS a soporte puntual. Caso sesgado 80%.	89

1. INTRODUCCIÓN.

1.1 Motivación del trabajo.

Un problema frecuente en la estimación de recursos es que los datos de entrada (testigos de campañas de sondajes, muestras de pozos de tronadura, etc.) a partir de los cuales se hará la estimación, presenten algún tipo de sesgo espacial. Este sesgo generalmente se debe a que se decide muestrear preferentemente ciertas porciones de la población (usualmente las zonas de altas leyes), dejando pobremente representadas las zonas que no son de interés. Esta decisión repercute finalmente en los resultados de la estimación a partir de estas muestras.

Un ejemplo típico de este problema se presenta en el muestreo de las variables ley de cobre total (CuT) y ley de cobre soluble (CuS), en este caso la primera variable se estudia en detalle, generalmente toda el área de interés es muestreada y, por lo tanto, se tiene plena confianza en el conocimiento que se tiene de esta variable. Sin embargo, la segunda variable no se muestrea como la primera. Usualmente solo se analizan las muestras que presentan mayor potencial (por ejemplo, aquellas con ley de CuT sobre cierta ley de corte) dejando sin información partes importantes de la población (generalmente las zonas de bajas leyes). De esta manera las estimaciones que se hacen de esta última variable serán menos representativas debido al sesgo que presentan, y por ende, se tiene una distribución experimental sesgada a los valores altos.

Con el fin de solucionar este problema se ha propuesto un método que permite corregir las distribuciones de las variables menos informadas, en este caso CuS, a partir de otras variables mejor informadas (CuT), usando la correlación que existe entre ambas variables.

Para probar el método se utilizará una base de datos que contiene datos de CuT y CuS, se truncará esta base de datos, es decir, se eliminará la información de CuS en cierto número de datos, de manera de obtener la situación con sesgo, para luego comparar los resultados de la simulación (casos sesgados y caso insesgado), verificando la calidad de la solución propuesta.

Se estimarán también los efectos que trae consigo el grado de sesgo (porcentaje de la muestra sin información de CuS) en los resultados de la simulación, así como también, el efecto en el grado de refinación con la cual se realiza la corrección.

Entonces la motivación principal es corregir la distribución sesgada de manera de obtener resultados que no presenten el sesgo que tenía la variable original, ya que las estimaciones pueden diferir enormemente si no se realiza dicha corrección.

1.2 Objetivos.

Los objetivos generales del trabajo son:

- Verificar y comparar lo que actualmente se realiza para solucionar el problema planteado, por ejemplo, comparar la corrección planteada con un caso donde no se aplica la corrección.
- Corregir las distribuciones que alimentan al proceso de simulación condicional, para generar estimaciones que no estén sesgadas.
- Generar un programa de computación que realice la corrección y entregue un set de nuevos ponderadores para cada dato, que solucione el problema del sesgo espacial.

Los objetivos específicos del trabajo son:

- Proponer soluciones para determinar la distribución representativa de la variable poco informada (CuS) a partir de otra variable mejor informada (CuT) y la correlación entre ambas.
- Implementar la solución en algún lenguaje de programación y generar un programa configurable por el usuario para reproducir la corrección según sus preferencias (principalmente el grado de discretización con el que se quiera realizar la corrección).
- Aplicar los métodos propuestos a una base de datos real, bajo distintos escenarios, principalmente asociados al grado de sesgo de la muestra.
- Comparar los resultados obtenidos por el método y lo que actualmente se realiza (un ejemplo sin corrección).
- Recomendar a partir de estos análisis el mejor método a utilizar según sea el caso en estudio.

1.3 Alcances.

El alcance de este tema de memoria es generar un algoritmo que corrija distribuciones sesgadas de manera de que la base de datos que se ingrese al proceso de simulación condicional sea la más representativa posible de la distribución real, buscando comparar este método con las practicas que se aplican actualmente.

El problema planteado es relevante para tener un conocimiento más acabado de la variable en estudio, quizá no se pueda obtener la mejor solución (pueden existir otras mejores), pero se intenta mejorar lo que actualmente se realiza.

El tema es de suma importancia ya que una mala estimación de una variable podría acarrear problemas en el futuro de un proyecto, principalmente debido a la sobrestimación o subestimación de la variable en estudio. Por ejemplo, en este caso la variable sesgada es el cobre soluble, debido al submuestreo de las leyes bajas las estimaciones que se hagan de esta estarán sesgadas a valores altos. Esto producirá una sobrestimación de los recursos asociados a esta variable, tanto en tonelaje como en la ley media asociada a una cierta ley de corte.

Se restringirán los análisis a cierto porcentaje de sesgo de la distribución (25% de los datos sin información de CuS como máximo) ya que se considera suficiente para probar el método. Además la distribución solo será sesgada en la parte baja de esta ya que es lo que pasa en la práctica.

2. ANTECEDENTES.

2.1 Evaluación de recursos: kriging y simulación.

En la estimación de recursos usualmente se usan dos herramientas: el método de kriging y las simulaciones condicionales. Un problema del kriging, es el suavizamiento de la distribución de valores estimados, por lo que los resultados no son representativos de la variabilidad real. Las simulaciones, en cambio, mantienen las mejores propiedades del kriging, a saber:

- Interpolación exacta cuando es simulación condicional.
- El error de estimación tiene esperanza nula (insesgo).
- Precisión: Aunque se deben usar muchas realizaciones para lograr este objetivo.

Y además no tienen los problemas que generan algunas propiedades del kriging, entonces las simulaciones:

- No suavizan: La dispersión de los valores estimados es la misma que la de los valores originales.
- La varianza de kriging depende ahora también de los datos y no solo de sus posiciones relativas por lo que aparece el efecto proporcional.

El kriging se usa principalmente cuando el objetivo principal es estimar la variable en estudio. Las simulaciones, por otra parte, se usan principalmente para cuantificar la incertidumbre en la estimación, evaluación de riesgos, estimar funciones de la variable en estudio, etc.

La principal característica del proceso de simulación condicional es la creación de una variable que intente reproducir la continuidad y las propiedades de la variable en estudio, considerada como una variable aleatoria regionalizada. La simulación condicional a diferencia del kriging requiere conocer completamente la variable en cuestión. Conocer esta variable implica principalmente determinar el variograma y sus funciones de distribución multivariadas, por ejemplo, para el caso de 2 variables se deben caracterizar las distribuciones monovariadas como la bivariable. (En el caso de la estimación por kriging solo basta con los variogramas simples y cruzados).

Por tanto, las razones para usar simulación condicional, en contraposición al kriging, se engloban principalmente en las siguientes:

- Evitar el suavizamiento de las estimaciones, usualmente el rango de los valores estimados por kriging, es menor que el rango de los valores originales, por tanto, en el método de kriging no se puede estimar la posible ocurrencia de valores extremos.
- La simulación entrega una medida de la incertidumbre en la estimación, mucho más útil que la varianza de kriging, ya que esta última es independiente de los datos.
- Permite hacer análisis de riesgo a partir de los resultados de la estimación, se puede, por ejemplo, calcular la frecuencia y con esto la probabilidad de ocurrencia de algún escenario determinado, lo que apoya la toma de decisiones, por ejemplo, en cuanto al diseño y planificación minera.
- Las estimaciones del kriging minimizan la varianza del error (local) pero fallan en la reproducción de la distribución original de la variable. Por otro lado la simulación condicional genera realizaciones que reproducen las estadísticas globales, los valores condicionales de los datos y la representación espacial de estas (histogramas, variogramas, etc.) [5].

Por esta y otras razones la simulación es cada vez más usada por sus propiedades de representar la verdadera continuidad de la variable y de medir de mejor forma la incertidumbre en la estimación (debido al procesamiento de las realizaciones de la variable en estudio).

2.2 Simulación multivariable.

Existen métodos de simulación multivariable que permiten reproducir las relaciones espaciales de cada variable y entre variables, usando los variogramas simples y cruzados, que pueden ser descritas por el modelo lineal de correogionalización.

El principal supuesto es que el conjunto de valores presenta una distribución multivariable gaussiana $\{Y_i(x), i = 1, \dots, N\}$, entonces la densidad de probabilidad de los valores en los sitios con datos es:

$$g(y_1, y_2, \dots, y_n) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det(C)}} \exp\left(-\frac{1}{2} y \cdot C^{-1} y^t\right)$$

Donde el vector Y está compuesto de los valores gaussianos y C es la matriz de varianza-covarianza, de esta manera, solo con la matriz C (o variograma) se caracteriza completamente la función aleatoria gaussiana, esto es un punto fuerte del modelo debido a su simplicidad.

[5] Emery, X, 2008. *Simulación Estocástica y Geoestadística No Lineal*. Universidad de Chile, facultad de ciencias físicas y matemáticas, departamento de ingeniería civil de minas.

Si se tienen varias variables de esperanza 0 y cuyas covarianzas simples y cruzadas son combinaciones lineales de las mismas estructuras, entonces:

$$C_{ij}(h) = \sum_u b_{ij}^u \cdot C_u(h) \quad \text{Para } i,j = 1, \dots, n.$$

Donde la matrices $B_u = (b_{ij}^u)$ son de tipo positivo. La simulación se basa en la interpretación del modelo lineal de correogionalización, que descompone las variables de la siguiente manera:

$$Y_i(x) = \sum_u x_i^u(x) \quad \text{Para todo } i \in [1, n]$$

Donde los grupos de componentes: $X_i^u(x)$ y $X_i^v(x)$ no están correlacionados si $u \neq v$, pero los componentes de un mismo grupo (por ejemplo, $X_i^u(x)$) sí tienen correlación intrínseca, siendo su matriz de covarianzas simple y cruzadas $B_u \cdot C_u(h)$.

Para cada estructura elemental, de índice u, se simulan los componentes $X_i^u(x)$ con el método anterior. Entonces las variables simuladas serán la suma de los componentes:

$$Y_i^S(x) = \sum_u x_i^{uS}(x) \quad \text{Para todo } i \in [1, n]$$

En el modelo multigaussiano esta técnica permite reproducir no solo las covarianzas simples y cruzadas sino también las distribuciones espaciales propias y conjuntas de las funciones aleatorias.

2.3 Análisis de las variables en estudio, cobre total (CuT) y cobre soluble (CuS).

Una de las condiciones de los métodos de simulación que se basan en el modelo multigaussiano, es que las distribuciones monovariantes, bivariantes y de orden superior tengan distribución gaussiana, sin embargo, esto muchas veces no se cumple debido a la naturaleza de las variables.

Las variables analizadas generalmente no tienen una distribución normal, por ende, el primer paso para la simulación, es la transformación de los datos originales en datos que tengan una distribución normal monovariante (de media 0 y varianza 1).

Considerando los histogramas acumulados la transformación consiste en asociar a cada valor de la variable original un valor Gaussiano que tiene la misma frecuencia acumulada. A la función que relaciona estos valores se le denomina anamorfosis gaussiana. El primer paso entonces es construir la función de probabilidad acumulada, como se muestra en la siguiente figura y asociarle el valor gaussiano que tiene esa misma probabilidad acumulada.

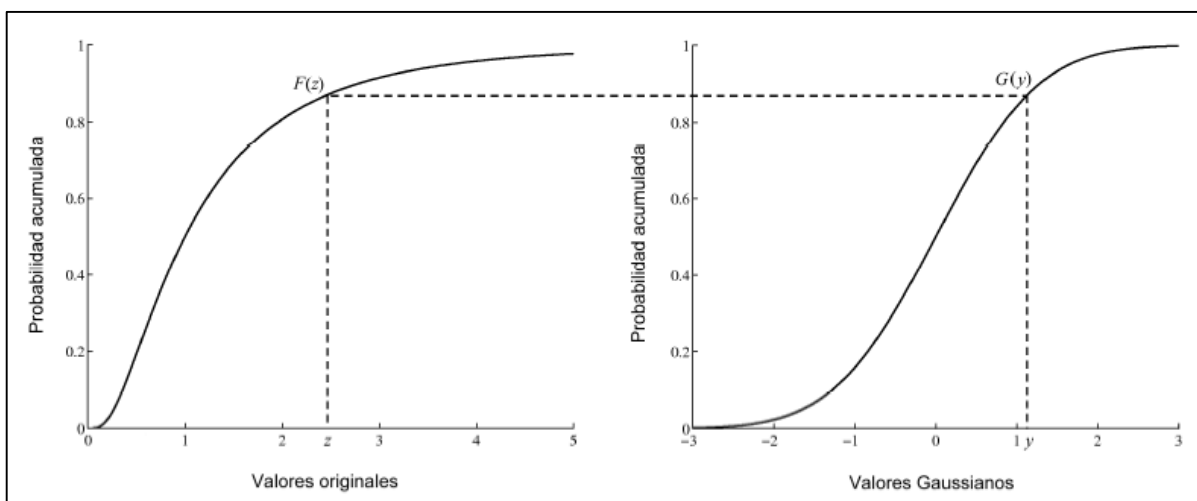


Figura 1 – Construcción del histograma acumulado de la variable original.

Entonces para cada valor de la variable original se tiene un valor gaussiano, como se muestra en la siguiente figura [5].

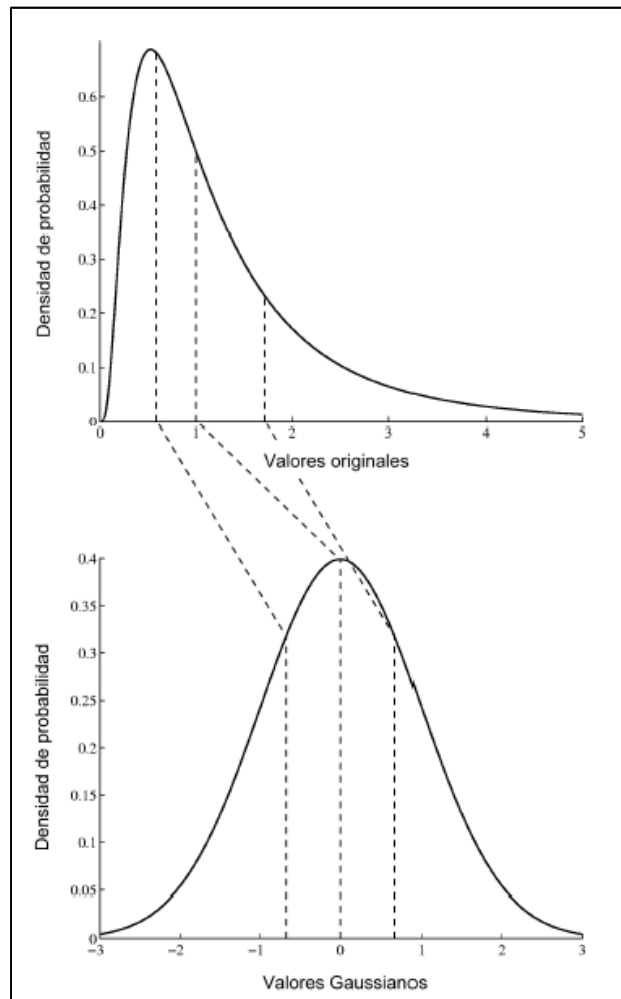


Figura 2 - Proceso de transformación a valores gaussianos.

[5] Emery, X, 2008. *Simulación Estocástica y Geoestadística No Lineal*. Universidad de Chile, facultad de ciencias físicas y matemáticas, departamento de ingeniería civil de minas.

A modo de ejemplo, en las figuras siguientes se ven las distribuciones de las variables transformadas:

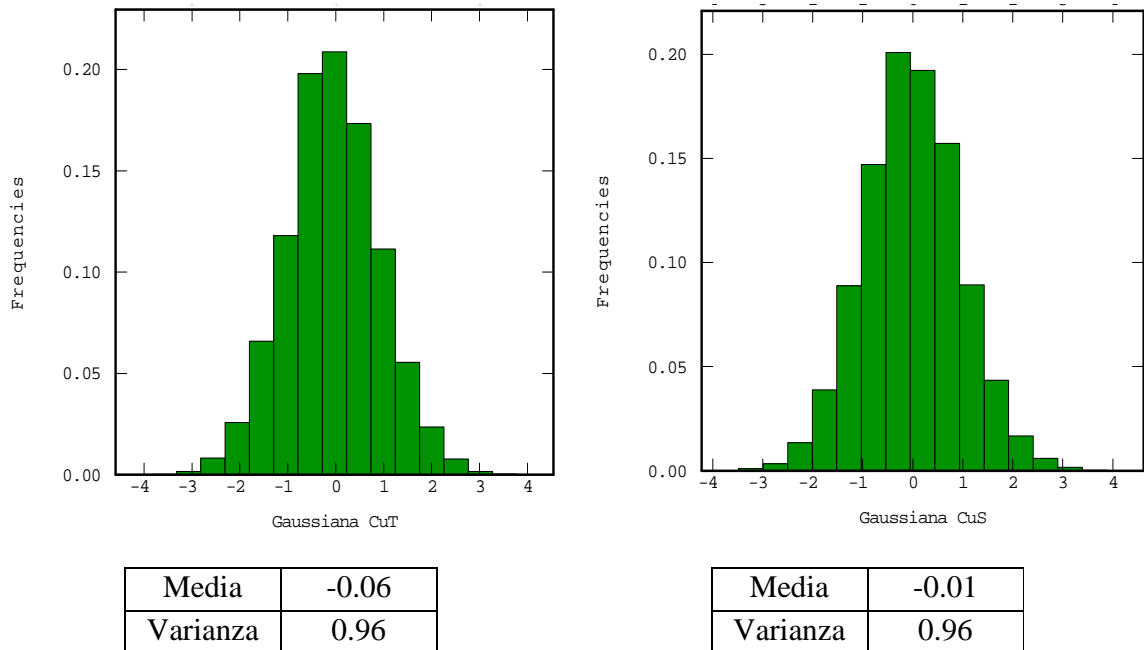


Figura 3 - Histogramas de las variables gaussianas.

Como se ve, las distribuciones monovariadas sí tienen un comportamiento gaussiano (media cercana a 0 y varianza aproximadamente 1). Sin embargo, si se transforman independientemente dos variables que tienen correlación, las variables gaussianas resultantes no necesariamente muestran un comportamiento bigaussiano. En el caso de CuT y CuS esto ocurre debido a una restricción de índole mineralógica, a saber, la ley de CuT debe ser mayor o igual que la ley de CuS.

En la figura siguiente se muestra la nube de correlación entre las variables gaussianas de CuT y CuS.

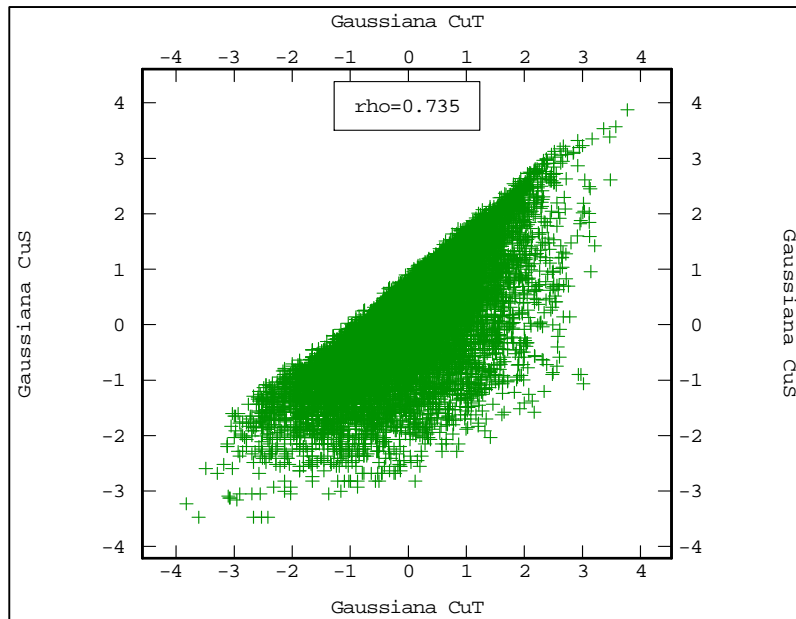


Figura 4 - Nube de correlación entre las variables gaussianas transformadas independientemente.

Por ende, existe el problema de transformar las variables en estudio de manera que respeten las hipótesis de multigaussianidad.

Una de las posibles soluciones es la transformación condicional paso a paso [6]. El método consiste en transformar primero a una variable gaussiana una de las dos variables, usualmente, si se está en presencia de muestreo heterotópico, es conveniente transformar primero la variable más informada (en este caso CuT) y la segunda variable se transforma condicionalmente a la primera. En términos prácticos se transforma la primera variable de la forma usual, luego de transformada esta se discretiza la primera variable en clases y se transforma la segunda variable independientemente en cada una de esas clases, de esta manera la segunda variable es gaussiana en cada una de estas clases y la suma de estas también será gaussiana.

El resultado final, por tanto, es que las dos variables gaussianas presentan este comportamiento tanto en las distribuciones monovariadas como en las bivariadas.

[6] Leuangthong, O., Deutsch, C. V., 2003. *Stepwise Conditional Transformation For Simulation Of Multiple Variables*.

2.4 Corrección del sesgo debido a muestreo preferencial.

Una base de datos se dice homotópica si en cada sitio con datos se tiene información de todas las variables medidas, por otro lado, una base de datos o un muestreo se dice heterotópico si en algunos sitios no existe información de alguna variable. Esta última clase puede ser subdividida dependiendo si el muestreo es preferencial o no, es decir, si la variables que no se midieron obedecen a algún criterio o decisión de la persona o equipo que realiza el muestreo. En la siguiente tabla se muestran ejemplos de estos 3 tipos de base de datos, donde las variables son la ley de cobre total (CuT) y cobre soluble (CuS).

Homotópica		Heterotópica aleatoria		Heterotópica preferencial	
CuT	CuS	CuT	CuS	CuT	CuS
[%]	[%]	[%]	[%]	[%]	[%]
0.06	0.03	0.06	-	0.06	-
0.19	0.15	0.19	0.15	0.19	-
0.32	0.13	0.32	0.13	0.32	-
0.37	0.22	0.37	-	0.37	0.22
0.55	0.27	0.55	0.27	0.55	0.27
0.78	0.72	0.78	-	0.78	0.72

Tabla 1 - Diferencia entre el muestreo homotópico y heterotópico.

En el primer caso se tiene información de CuT y CuS en todos los sitios con datos, en cambio en el muestreo heterotópico existen regiones donde no se toman muestras de cobre soluble, mientras que en la base de datos heterotópica aleatoria parece no haber ningún criterio para no muestrear la ley de cobre soluble, en el caso preferencial este no se mide si el cobre total esta bajo la ley 0,35%, diferencia fundamental que da origen al problema que se trata en este documento.

Uno de los pasos iniciales en el proceso de simulación es la transformación de los datos originales en datos gaussianos, si la distribución de los datos no es la correcta, está sesgada o tiene algún otro tipo de imperfecciones, sin duda afectará los resultados obtenidos al final del proceso. Entonces es una necesidad ajustar los histogramas y las estadísticas para que sean representativas de toda la población en estudio.

La estimación por kriging genera efectivamente estimaciones que están desagrupadas ya que, por construcción, otorga menor ponderación a los datos que son más redundantes y otorga mayor ponderación a los datos que están más aislados. Por contrapartida, los algoritmos de simulación que requieren transformar los datos no corrigen el impacto de sesgos o agrupamiento de los datos originales, lo que repercute en el histograma de los datos resultantes, por tanto, estos algoritmos requieren como parámetros de entrada una distribución que sea representativa de la población.

Para enfatizar la situación se debe tener presente que la simulación en un área con escasos datos se basa en la distribución global de los datos, la cual debe ser representativa de toda la población, para estimar de manera correcta los recursos del depósito. Al tomar muestras de zonas particularmente más ricas se produce un sesgo en la distribución de valores de esa variable, en términos de frecuencia absoluta (cuantos valores en cada clase del histograma), la distribución de valores en las bajas leyes no es representativa de la distribución real ya que no se midieron las leyes en las clases bajas, por otro lado, en términos de la frecuencia relativa (cuantos valores en cada clase dividido por el número total de datos), la distribución completa no es representativa, en particular las clases bajas están pobremente representadas y en las clases altas sucede justamente lo contrario, están sobre representadas en términos relativos.

En la figura siguiente se muestran los histogramas de frecuencias absolutas, la figura de la izquierda corresponde a una base de datos con información completa de CuT y CuS, el histograma de la derecha, corresponde a una base de datos en la que falta información de CuS en ciertas ubicaciones, principalmente en las zonas más bajas (específicamente bajo la ley 0.2%).

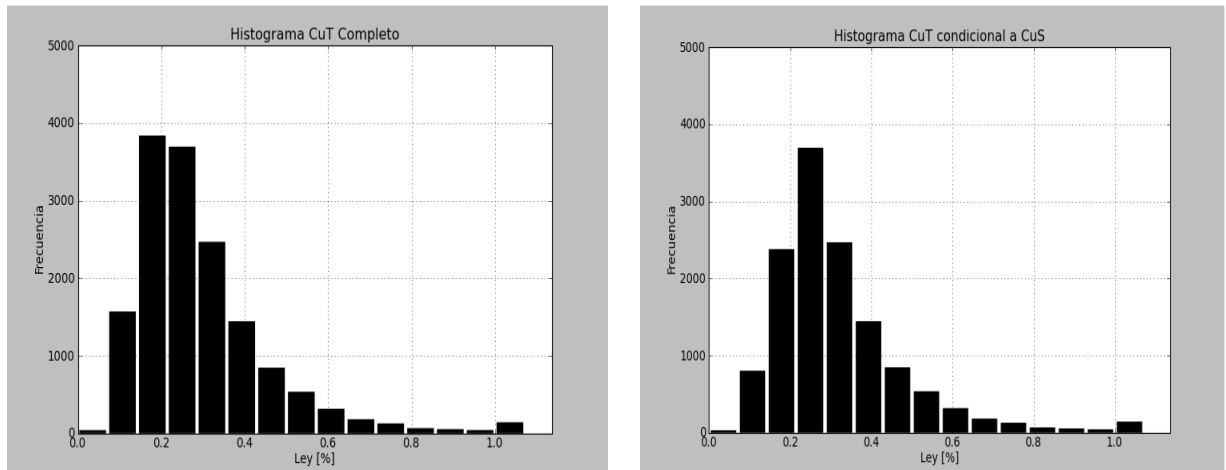


Figura 5 - Histograma de frecuencias absolutas, comparación entre el caso sin sesgo y el con sesgo.

Como se ve en las clases de leyes altas no hay ninguna diferencia, sin embargo, en las clases de leyes bajas, la frecuencia absoluta cambia significativamente. Este cambio está en directa relación a la intensidad del sesgo en esas clases, mientras mayor es el sesgo, mayor es la diferencia en términos de frecuencia absoluta.

En la figura siguiente se muestran los mismos histogramas pero ahora considerando la frecuencia relativa.

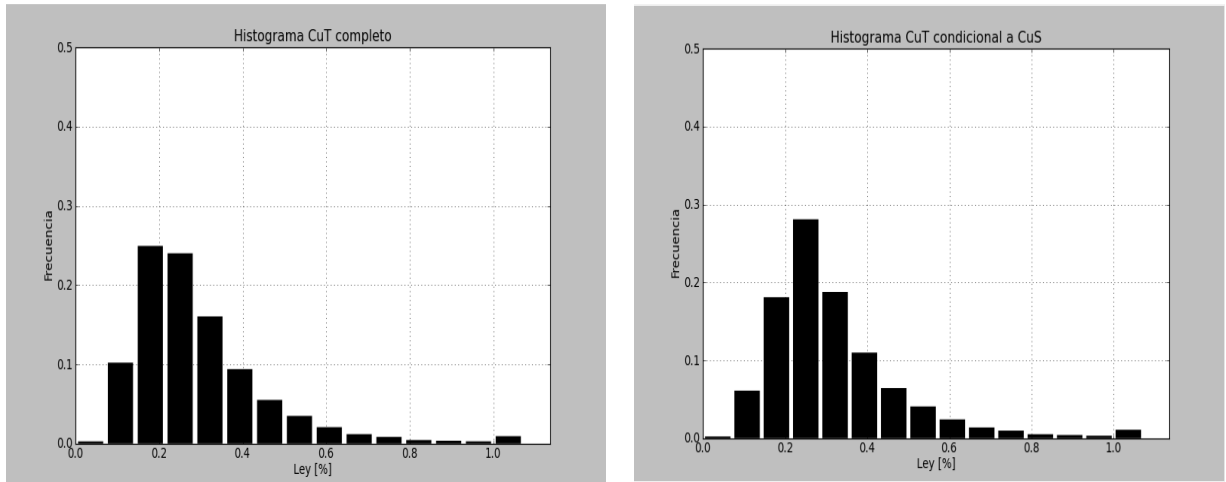


Figura 6 - Histogramas de frecuencia relativa, comparación entre el caso sin sesgo y el con sesgo.

En este caso tanto las clases de las altas leyes como las de las bajas leyes se ven afectadas por la falta de información, sin embargo, el efecto es mayor en las zonas de bajas leyes (ver por ejemplo las 3 primeras clases de los histogramas de la figura 6). Lo que se ve mejor representado en el siguiente grafico cuantil contra cuantil.

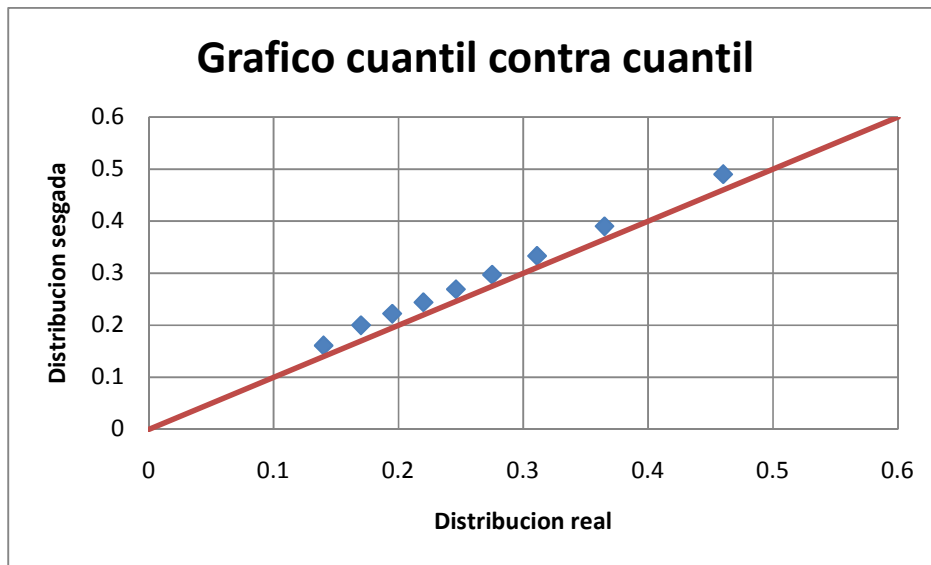


Figura 7 - Grafico cuantil contra cuantil entre la distribución real y la distribución sesgada.

Como se ve en la figura 7, la distribución sesgada presenta una leve tendencia hacia las leyes altas, sin embargo, la diferencia podría ser más significativa si el grado de sesgo de la distribución se hiciera más importante, ya que en este caso solo el 15 % de los datos no tiene información de CuS.

Entonces el problema radica en que si se usa esta distribución sesgada como la representativa de la variable en estudio, se producirá un sesgo en la distribución de valores que genera cualquier algoritmo de simulación.

Por ende, es necesario corregir la distribución sesgada antes de transformar los datos a valores gaussianos y usar los ponderadores generados de la corrección para construir las funciones de probabilidad acumuladas necesarias para la transformación.

3. METODOLOGIA.

3.1 Detalles de la corrección del sesgo espacial.

En el caso en que hay sesgo, es necesario corregir las distribuciones multivariadas (distribución de CuT-CuS) para que se tengan distribuciones monovariadas representativas (CuS). Existen varios métodos presentados en la literatura, pero muchos de ellos son demasiado complicados o no se pueden aplicar directamente en el presente problema.

El siguiente método intenta actualizar (corregir) la distribución de la variable de interés a partir de una variable relacionada con esta (CuT), mediante el siguiente formulismo [4],[7],[8].

$$f_{CuT,CuS}^2(CuT, CuS)_{ij} = f_{CuT,CuS}^1(CuT, CuS)_{ij} \cdot \frac{f_{CuT}^2}{f_{CuT}^1}$$
$$f_{CuS}^2 = \int_{-\infty}^{+\infty} f_{CuT,CuS}^2 dCuT$$

Donde:

f_{CuT}^1 : Corresponde a la frecuencia relativa de la distribución de cobre total en los sitios donde se tienen datos de cobre soluble (dist. sesgada).

$f_{CuT,CuS}^1$: Corresponde a la frecuencia relativa de la distribución bivariable usando los datos donde se tiene tanto el cobre total como el soluble (dist. sesgada).

f_{CuT}^2 : Corresponde a la frecuencia relativa de la distribución de cobre total en todos los puntos (distribución que se supone insesgada usando ponderadores de desagrupamiento).

$f_{CuT,CuS}^2$: Corresponde a la frecuencia relativa de la distribución bivariable actualizada (insesgada).

f_{CuT}^2 : Corresponde a la frecuencia relativa de la distribución de cobre soluble actualizada (insesgada).

[4] Deutsch, C. V., 2002. *Geostatistical Reservoir Modeling*, Oxford University Press.

[7] Leuangthong, O., Khan, K. D., y Deutsch, C. V. 2008. *Solved Problems In Geostatistics*, pag. 28 – 34.

[8] Leuangthong, O., Ortiz, J., y Deutsch, C. V. 2005. *On The Scaling And Use Of Multivariate Distributions In Geostatistical Simulation*.

Las relaciones anteriores deben cumplir todas las restricciones impuestas sobre las funciones de densidad de probabilidad, lo que se traduce finalmente en:

$$f_{CuS}^2(y) = \int_{-\infty}^{+\infty} f_{CuT,CuS}^2 dCuT = 1$$

A partir de las distribuciones donde se tienen las 2 variables (CuS y CuT) se construye la relación bivariable. Esta relación bivariable debe ser actualizada por la verdadera distribución de cobre total, la que se considerara representativa de la población (supuesto). Esta actualización se hace mediante el cociente $\frac{f_{CuT}^2}{f_{CuT}^1}$ el que en definitiva relaciona la frecuencia real de datos que hay en cierto rango (una clase de un histograma por ejemplo), con la frecuencia de datos que aparecen en la distribución condicionada a los sitios con información de cobre soluble.

La actualización de la distribución bivariable ($f_{CuT,CuS}^2$) permite finalmente obtener la distribución corregida de CuS “integrando” en cada una de las clases del histograma.

Gráficamente la corrección se realiza de la siguiente manera:

- 1) Generar la distribución desagrupada de CuT en todos los sitios con datos y la distribución de CuT solo donde está disponible la información de CuS. Además construir la distribución bivariable CuT-CuS, con los puntos donde se tiene información de ambos.

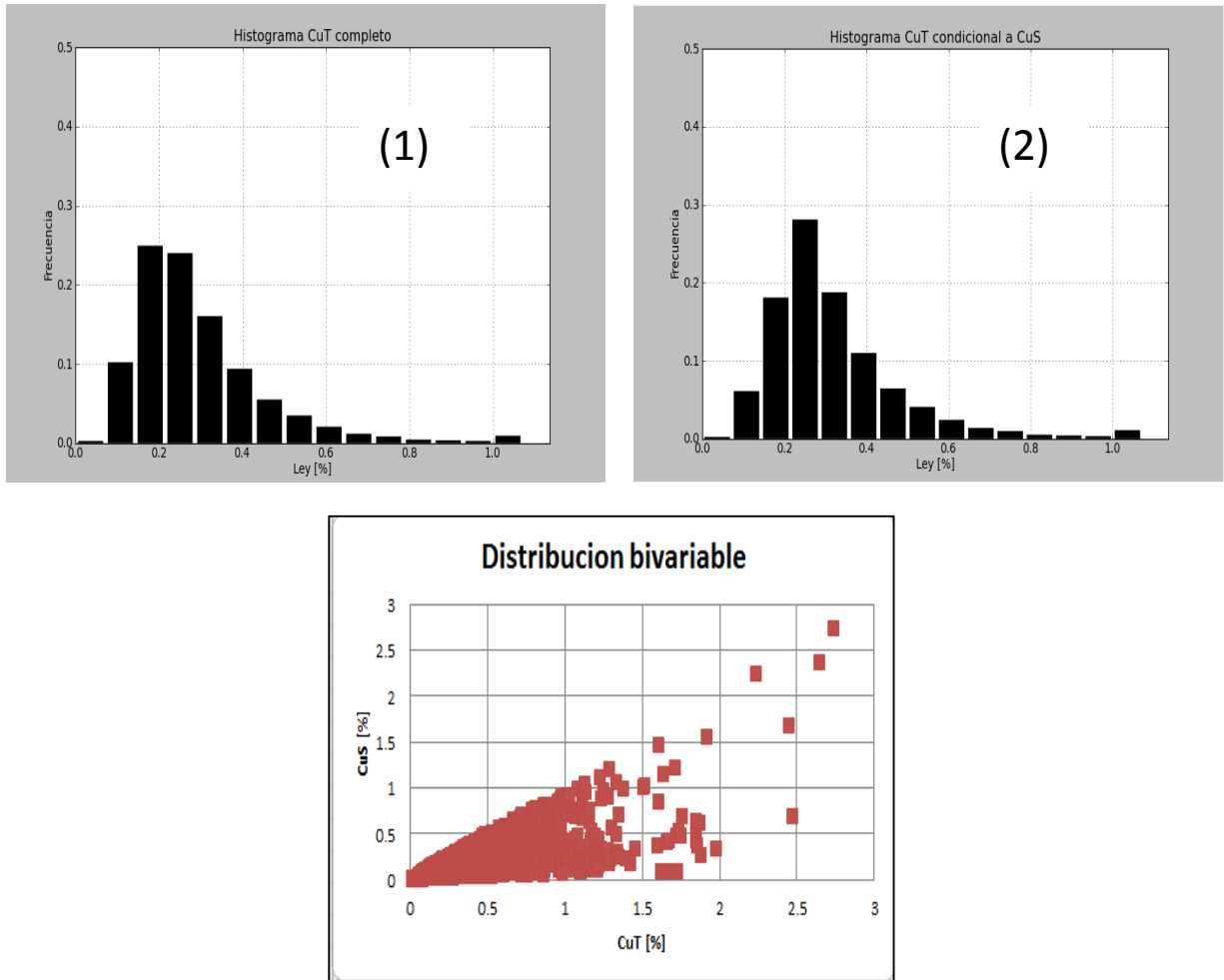


Figura 8 - Primer paso de la corrección: Generación de las distribuciones.

En este paso el usuario debe definir un ancho de clase para generar las distribuciones anteriores. Un mayor número de clases es lo recomendado debido a que se reproduce de mejor manera el histograma de la variable. Sin embargo, se debe tener en cuenta que no se debe tener una clase con frecuencia 0 ya que la corrección no se podrá calcular.

2) Generar la corrección por clase usando las distribuciones (1) y (2) mediante la fórmula:

$$\text{Corrección} = \frac{f_{CuT}(1)}{f_{CuT}(2)}$$

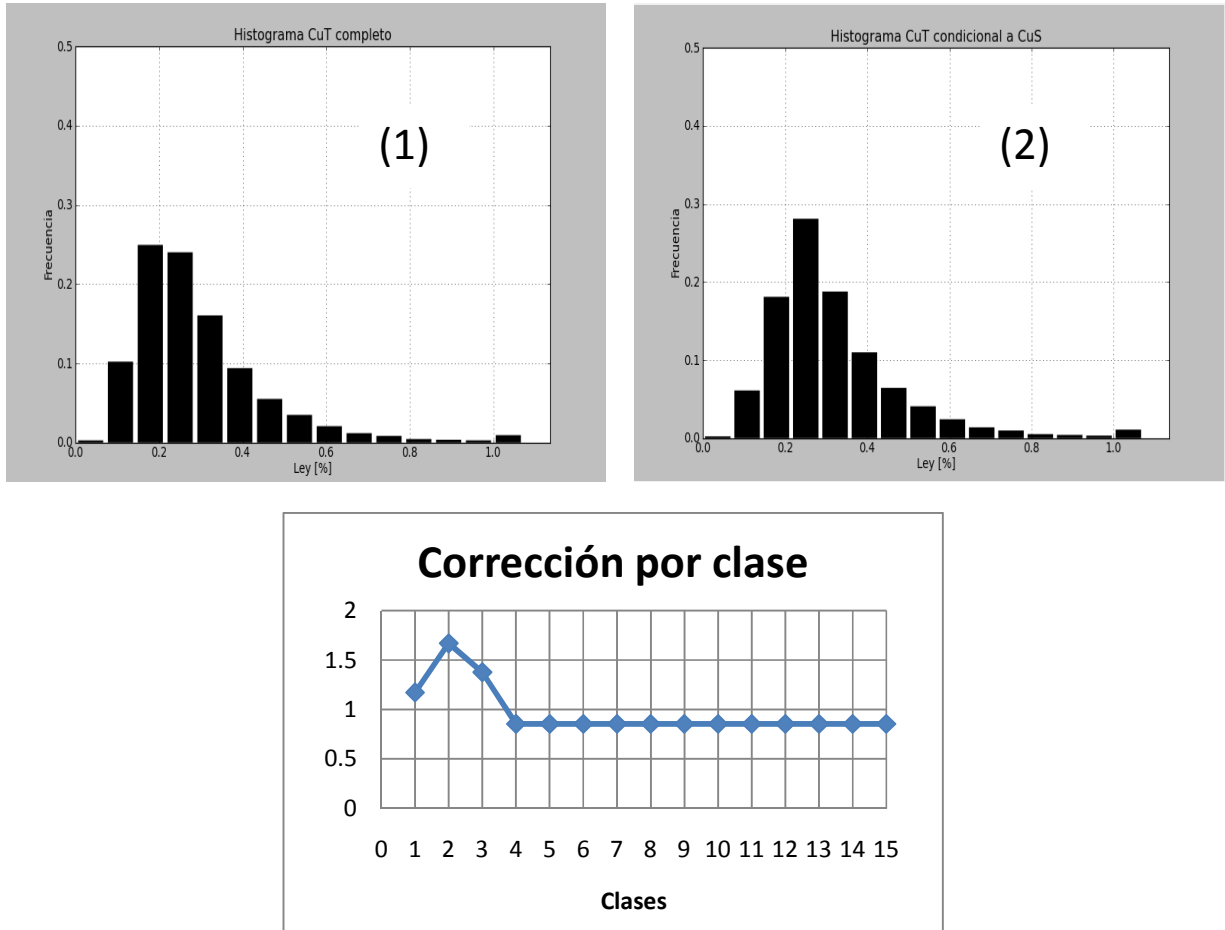


Figura 9 - Segundo paso de la corrección: Calcular la corrección para cada una de las clases.

Se puede apreciar que la corrección intenta ponderar en mayor medida las zonas donde ocurrió el sesgo (valores mayores que 1) y pondera menos a los valores que están en clases sin sesgo, intentando restituir la verdadera distribución de la variable sesgada.

- 3) Actualizar la distribución bivariable con la corrección y obtención de las distribución insesgada de CuS.

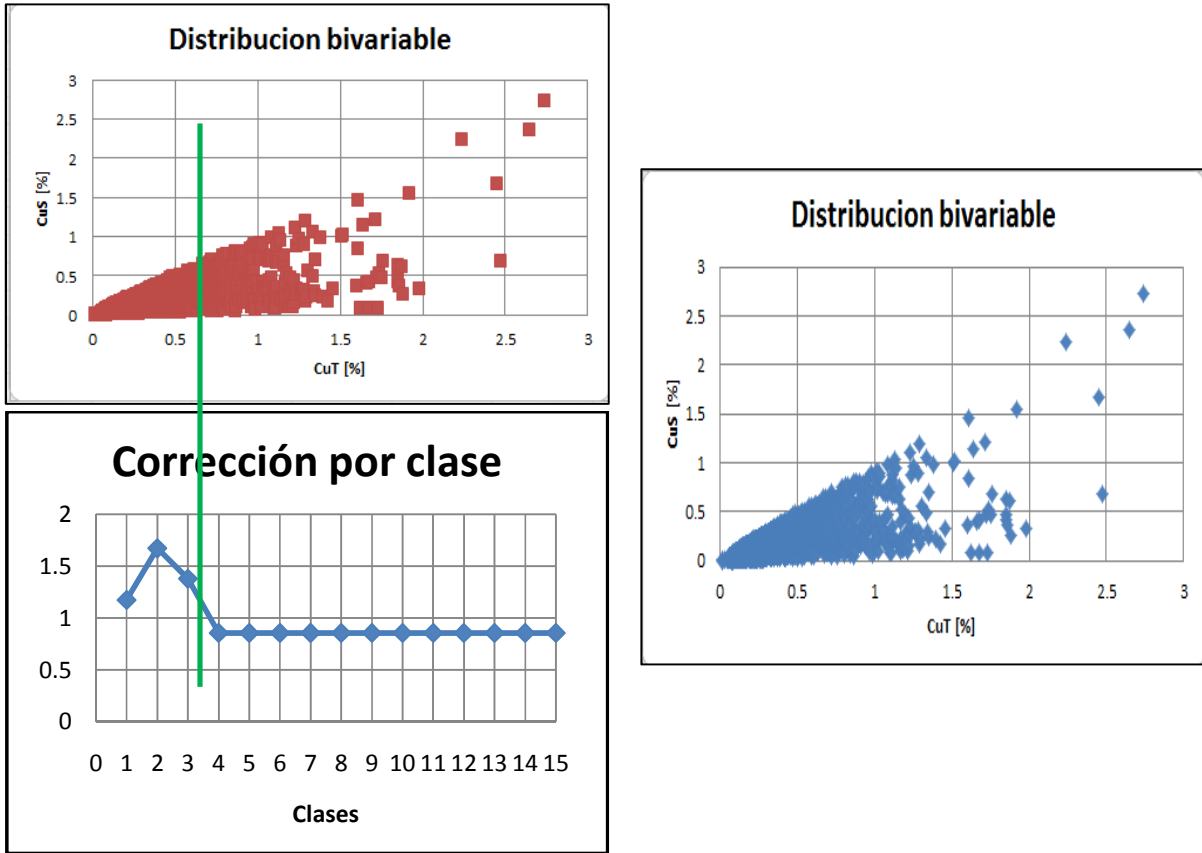


Figura 10 - Tercer paso de la corrección: Actualización la distribución bivariable y obtención de la distribución corregida de CuS.

Entonces los datos de la distribución bivariable que están por debajo de la ley en la cual se sesgó la distribución estarán ponderados por un factor mayor, mientras que los que están sobre la ley del sesgo tendrán una ponderación menor. De esta manera se construye una distribución bivariable actualizada, que considere esta nueva ponderación.

Para obtener la distribución corregida de CuS simplemente se debe integrar (en este caso sumar) la distribución bivariable actualizada. Esta distribución corregida es la que se usara como dato de entrada para el proceso de simulación condicional.

3.2 Implementación en el lenguaje python.

Para llevar a cabo la corrección se decidió implementarla en algún lenguaje de programación para generar un programa configurable por el usuario de manera que se pudiera decidir el grado de discretización.

Se escogió el lenguaje python debido a lo simple que resulta programar en él, además de que cuenta con varias librerías gráficas y matemáticas útiles al propósito de este trabajo.

En la práctica para implementar la corrección se deben discretizar todas las distribuciones en cierto número de clases de igual tamaño. El caso ideal es usar un gran número de clases ya que de esta manera se ajusta de mejor manera la distribución real. Sin embargo, si se usan clases de

igual tamaño el número de clases está restringido debido al cociente $\frac{f_{CuT}^2}{f_{CuT}}$, ya que si el número

de clases aumenta, el tamaño de esta disminuye consecuentemente, entonces para algún tamaño de estas habrán clases que no tengan ningún dato, por lo tanto, el denominador del cociente anterior es 0. Este hecho no permite usar un gran número de clases, sin embargo esto no es tan problemático cuando se tiene una gran cantidad de datos (como es el caso actual). Queda por evaluar el efecto con bases de datos más pequeñas que obliguen a una discretización pobre.

3.3 Implementación de simulación con transformación condicional por pasos.

La transformación condicional paso a paso requiere el uso de los ponderadores obtenidos de la corrección con el objetivo de generar el histograma acumulado experimental, mediante este, es posible transformar las variables originales en variables con distribución gaussiana, como se ve en la siguiente figura [5].

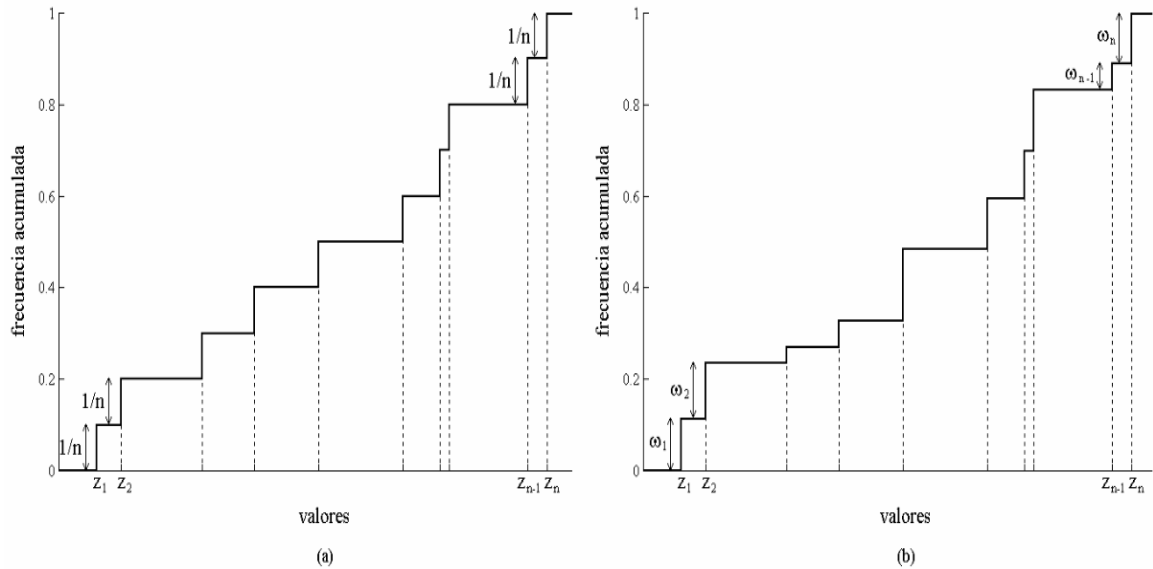


Figura 11 - Construcción del histograma acumulado para la transformación condicional a valores gaussianos.

Sin embargo, el método tiene ciertos problemas en particular con 2 situaciones:

- a) Valor Gaussiano de los valores extremos:

En la práctica al realizar el histograma acumulado se deben ordenar los datos de menor a mayor (o a la inversa), de esta manera la probabilidad acumulada del último valor (o el primero) será 1 por construcción, esto es un problema porque el valor gaussiano asociado a una probabilidad de 1 es infinito, para sortear esta dificultad se decide usar la siguiente solución:

Si se tienen datos ordenados, la probabilidad acumulada $F(z_i)$ para el dato z_i es:

$$F(z_i) = \sum_{k=1}^{i-1} p_k + \frac{p_i}{2}$$

Donde p_k es la probabilidad acumulada calculada de la forma usual de todos los valores anteriores a z_i . De esta manera se resuelve el problema de obtener valores gaussianos infinitos.

[5] Emery, X, 2008. *Simulación Estocástica y Geoestadística No Lineal*. Universidad de Chile, facultad de ciencias físicas y matemáticas, departamento de ingeniería civil de minas.

b) Datos que tienen el mismo valor:

Existen varias opciones para asignarle un valor distinto a estos datos, asignándole un orden indirectamente. En este estudio se decidió agregar a cada dato un pequeño ruido, de esta manera los datos resultantes podrían ser ordenados de forma unívoca.

En la base de datos los valores tienen 3 decimales por lo que los valores aleatorios que se agregaron marcan diferencias a partir del cuarto decimal, como se ve en el siguiente ejemplo:

#	Valor original	Valor aleatorio	Valor final
1	0,123	0,000570	0,123570
2	0,335	0,000124	0,335124
3	0,335	0,000687	0,335687
4	0,338	0,000313	0,338313
5	0,411	0,000895	0,411895

Tabla 2 - Metodología para asignar un orden a valores originalmente iguales.

De esta manera se puede asignar valores gaussianos distintos a datos originalmente iguales, como es el caso en este ejemplo de los datos 2 y 3.

3.4 Potenciales problemas.

Como ya se mencionó, cuando se tienen pocos datos se pueden presentar clases donde se tiene ausencia total de información. Si hay muestreo preferencial las clases con menos datos serán las de más baja ley, por lo tanto los histogramas generados deberían subsanar este hecho y por lo tanto someterse a un proceso de suavizamiento (Figura 12). Este problema se vuelve mucho peor en el caso de histogramas bivariados debido al número de clases que deben ser informadas, si se tienen n clases en el caso mono variable en el caso bivariado se necesitan n^2 clases informadas.

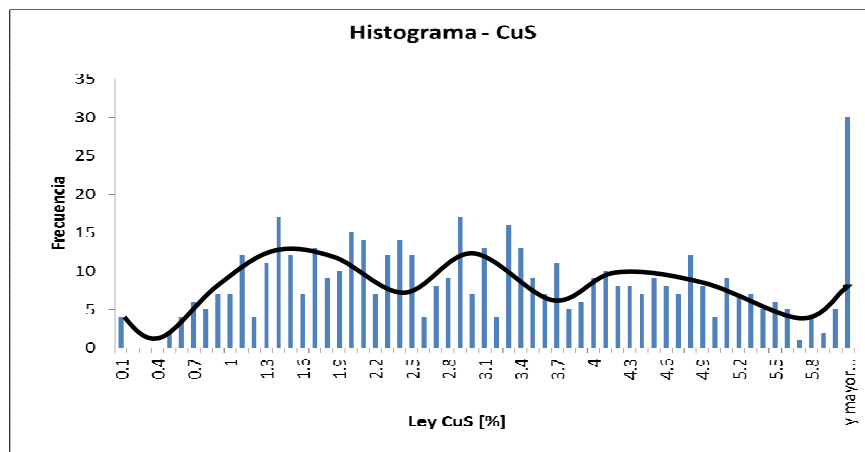


Figura 12 - Esquema del suavizamiento de los histogramas.

Como se ve en la figura anterior, algunas de las primeras clases no tienen datos por lo que la corrección no se podría aplicar en esa situación. Sin embargo si se suaviza el histograma (curva negra), si es posible implementar la corrección.

El primer enfoque para solucionar estos problemas consiste en ajustar el histograma mediante distribuciones conocidas, como la distribución normal, la distribución lognormal, etc. Estas resuelven fácilmente los problemas mencionados, sin embargo, en general los datos no se ajustan bien a estos modelos con pocos parámetros.

El segundo enfoque es remplazar cada dato por una función kernel [9], es decir, una distribución de probabilidad con media igual al valor del dato y una varianza que depende de cada problema. Los problemas de este enfoque es que generalmente no reproducen las características de la distribución original (media varianza, mínimo, máximo, etc.), además de que existe la probabilidad de que aparezcan valores anómalos (valores negativos cuando todos los datos son mayores que 0 por ejemplo).

[9] Laurini, M., Pereira, P., 2009. *Conditional Stochastic Kernel Estimation By Nonparametric Methods*.

Un tercer enfoque es resolver un problema de optimización [10], donde se busque minimizar el error entre el histograma suavizado y el histograma experimental, sujeto a que se mantengan estadísticas básicas como la media, la mediana, etc.

Un cuarto enfoque considera los polinomios de Bernstein [2],[3] para suavizar la distribución bivariable y a partir de ella generar las distribuciones marginales para implementar la corrección (distribuciones marginales también suavizadas).

3.5 Suavizamiento de la distribución bivariable – Polinomios de Bernstein.

Un polinomio de Bernstein $P(x)$ de grado n se puede expresar de la siguiente manera en el caso de una dimensión:

$$P(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \cdot B_k^n(x)$$

Donde $f\left(\frac{k}{n}\right)$ son los puntos de control con los cuales se construye la curva de suavizamiento, mientras que los términos $B_k^n(x)$ son las llamadas bases de Bernstein definidas como:

$$B_k^n(x) = \binom{n}{k} \cdot x^k \cdot (1-x)^{n-k} \quad \text{con } x \text{ en } [0,1]$$

Y en forma más general, cuando x pertenece al intervalo $[a,b]$:

$$B_k^n(x) = \binom{n}{k} \cdot \frac{(b-x)^{n-k} \cdot (x-a)^k}{(b-a)^n} \quad \text{con } x \text{ en } [a,b]$$

Por ende si se quiere saber el valor de la imagen de la curva suavizada para un x dado, se deben usar las fórmulas anteriores y calcular todas las bases de Bernstein. Esta tarea parece un poco costosa computacionalmente, pero debido a ciertas propiedades de estos polinomios, permiten un cálculo rápido.

[10] Salama, A., 2002. *A Technique For Interpolation And Smoothing Of Mass Size Density Data.*

[2] Babu, G. J., Cauty, A. J., Chaubey, Y. P., 2001. *Application Of Bernstein Polynomials For Smooth Estimation Of A Distribution And Density Function.*

[3] Chak, P. M., Madras, N., Smith, J. B., 2001. *Flexible Functional Forms: Bernstein Polynomials.*

Algunas propiedades los polinomios de Bernstein (para todo $k = 0, \dots, n$):

Positividad de las bases de Bernstein:

$$B_k^n(x) > 0 \quad \text{Para todo } k = 0, \dots, n.$$

Unicidad de las bases de Bernstein:

$$\sum_{k=0}^n B_k^n(x) = 1 \quad \text{Para todo } k = 0, \dots, n$$

Simetría:

$$B_k^n(x) = B_{n-k}^n(1-x) \quad \text{Para todo } k = 0, \dots, n$$

Fórmula de recurrencia:

$$B_k^n(x) = \begin{cases} \rightarrow B_k^n(x) = (1-x) \cdot B_k^{n-1}(x) & \text{Para } i = 0 \\ \rightarrow B_k^n(x) = (1-x) \cdot B_k^{n-1}(x) + x \cdot B_{k-1}^{n-1}(x) & \text{Para } i = 1, \dots, n-1. \\ \rightarrow B_k^n(x) = x \cdot B_{k-1}^{n-1}(x) & \text{Para } i = n. \end{cases}$$

Otras propiedades de los polinomios de Bernstein:

- El grado de la base de polinomios es uno menos que el número de puntos de control para construir la curva.
- Los puntos de control extremos son interpolados exactamente.
- El vector tangente en los extremos de la curva también coincide con el vector de los puntos en los extremos.
- Cuando se tienen muchos puntos, el grado de la base es elevado y se tiende a suavizar demasiado la curva de control
- Se torna insensible a los cambios pequeños en los puntos de control.

Los polinomios de Bernstein constituyen una herramienta para aproximar o suavizar una función dada, definida en el intervalo $[0,1]$, por ende no es un método que restituya el valor de los datos que se usan para construir la curva, solo interpolan exactamente los puntos de los límites del dominio, como se muestra en la siguiente figura [1].

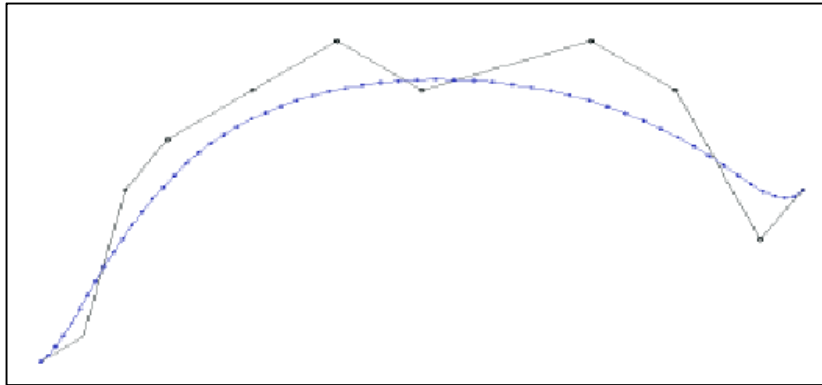


Figura 13 - Interpolación exacta de los puntos extremos usando polinomios de Bernstein.

En el caso de dos variables, los polinomios de Bernstein toman la siguiente forma.

$$P_{n_1, n_2}^f(x_1, x_2) = \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} f\left(\frac{k_1}{n_1}, \frac{k_2}{n_2}\right) \cdot B_{k_1}^{n_1}(x_1) \cdot B_{k_2}^{n_2}(x_2)$$

La fórmula es la misma que para el caso de 1 dimensión, pero esta vez se usan 2 bases de Bernstein, una para cada variable (en este caso de estudio correspondería a una base para el CuT y otra para el CuS).

El objetivo de usar esta herramienta es suavizar la distribución bivariable sesgada, lo que permitirá discretizar la corrección al nivel que el usuario quiera (a priori, una mayor discretización generaría una mejor corrección) y resolvería el problema de encontrar clases sin datos (donde la corrección no se podría aplicar).

[1] Abbate, H. *Curvas*.

El resultado del suavizamiento probablemente producirá por lo menos dos tipos de problemas:

- En ciertas clases, en especial en las que $CuS > CuT$ el suavizamiento otorgará una frecuencia relativa mayor que 0, siendo esto inconsistente por restricciones mineralógicas ($CuT > CuS$), para arreglar este hecho se redefinen todas las clases donde esto suceda como 0.
- La suma total de las frecuencias relativas de la distribución bivariable debería ser 1, probablemente esto no se cumpla por lo que deben normalizarse los resultados.

Las distribuciones marginales para construir la corrección se obtienen a partir de la distribución bivariable suavizada simplemente integrando esta última.

4. CASOS DE ESTUDIO.

4.1 Base de datos.

La base de datos original contempla 15622 compositos de 1,5 [m] de largo, con muestras de CuT y CuS, el peso de desagrupamiento, además de la razón de solubilidad (CuS/CuT) y los valores de las transformaciones gaussianas de cada uno de estos.

Sin embargo, en la base de datos se encontraron 8 compositos que no contenían absolutamente ninguna información, estos son los datos cuyas coordenadas son:

Este	Norte	Cota
[m]	[m]	[m]
518.831,5	7.411.606,3	2.488,9
518.440,1	7.412.303,7	2.497,1
517.988,8	7.412.298,8	2.600,9
518.468,2	7.411.499,5	2.548,4
518.463,0	7.411.500,7	2.548,4
518.750,2	7.412.004,4	2.528,0
518.570,7	7.412.398,0	2.574,0
518.566,4	7.412.605,3	2.573,8

Tabla 3 - Compositos sin ninguna información en la base de datos.

Se eliminaron, por tanto, estos compositos de la base de datos.

Se encontraron además 191 compositos en los que solo estaba informada la ley de cobre total CuT, en la gran mayoría de estos (187) la ley de cobre total no superaba el 0.1%. Estos compositos también fueron extraídos de la base de datos, ya que no servían para conformar la base de datos del caso base.

Por tanto, la base de datos a utilizar corresponde a 15424 compositos, con la información de coordenadas, ley de CuT, ley de CuS y el peso de desagrupamiento. Las otras variables no se consideran necesarias dados los alcances del proyecto. Esta base de datos conforma la base de datos con la que se trabajará en el caso base, los resultados obtenidos a partir de esta serán comparados con los resultados de las bases de datos sesgadas.

4.2 Construcción de las bases de datos sesgadas.

El objetivo de sesgar la base de datos es comprobar los efectos de la corrección, comparando los resultados obtenidos usando la corrección con los resultados al usar la base de datos completa, que será usada como referencia (caso base) para validar los resultados.

Con el fin de sesgar la base de datos se definió una ley de corte para definir el rango de leyes donde se extraería la información de CuS de los compositos, ya que es una práctica común muestrear solamente la variable secundaria (CuS) a partir de cierta ley en la variable primaria (CuT), la ley mencionada es 0,2% de CuT. Las proporciones de datos de CuS extraídos aleatoriamente corresponden a un 50 y 80 % de los datos bajo esta ley de corte. En la siguiente tabla se muestran las estadísticas de las distintas bases de datos.

	CuT	CuS		
		Base	LC 0.2 Sesgo (50%)	LC 0.2 Sesgo (80%)
Unidad	[%]	[%]	[%]	[%]
Media	0.29	0.17	0.19	0.20
Mediana	0.25	0.16	0.17	0.18
Moda	0.19	0.13	0.14	0.18
Desv. Estándar	0.17	0.11	0.11	0.12
Varianza	0.03	0.01	0.01	0.01
Coef. Curtosis	23.77	47.25	47.53	49.76
Coef. de asimetría	3.42	3.87	3.90	4.05
Rango	2.73	2.72	2.72	2.72
Mínimo	0.02	0.01	0.01	0.01
Máximo	2.74	2.72	2.72	2.72
N° de datos	15424	15424	13036	11552
% de datos con info. de CuS	-	100	84.5	74.9

Tabla 4 - Estadísticas básicas de las bases de datos.

Naturalmente las bases de datos sesgadas muestran una ley media mayor en el cobre soluble debido al sesgo hacia las leyes altas.

Como se ve, aunque la segunda base de datos tiene un fuerte sesgo bajo la ley de 0.2% solo deja sin información de CuS al 25 % de la población total. Como tema a discutir quedan las opciones de sesgar aún más la base de datos (a una ley mayor por ejemplo) y la opción de no restringir el sesgo a la parte de bajas leyes de la distribución. Sin embargo, el procedimiento anterior servirá sin duda para mostrar la utilidad de la corrección.

4.3 Estudio exploratorio de datos.

4.3.1 Despliegue de datos.

En la siguiente sección se muestran vistas isométricas y perfiles de la base de datos completa, donde se aprecian las variables CuT y CuS.

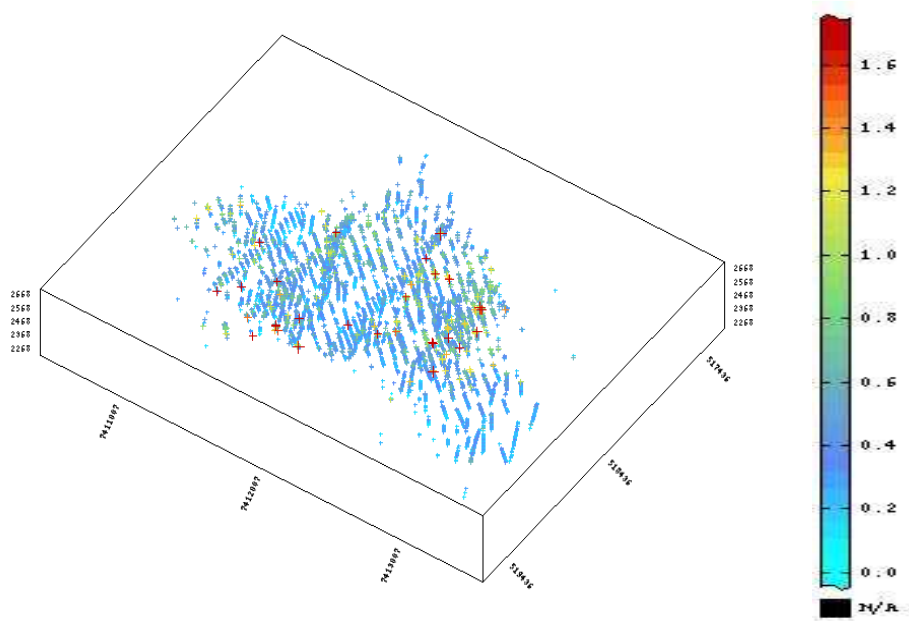


Figura 14 - Vista isométrica del yacimiento en estudio. Ley de CuT.

En la siguiente figura se ve una vista en planta del yacimiento.

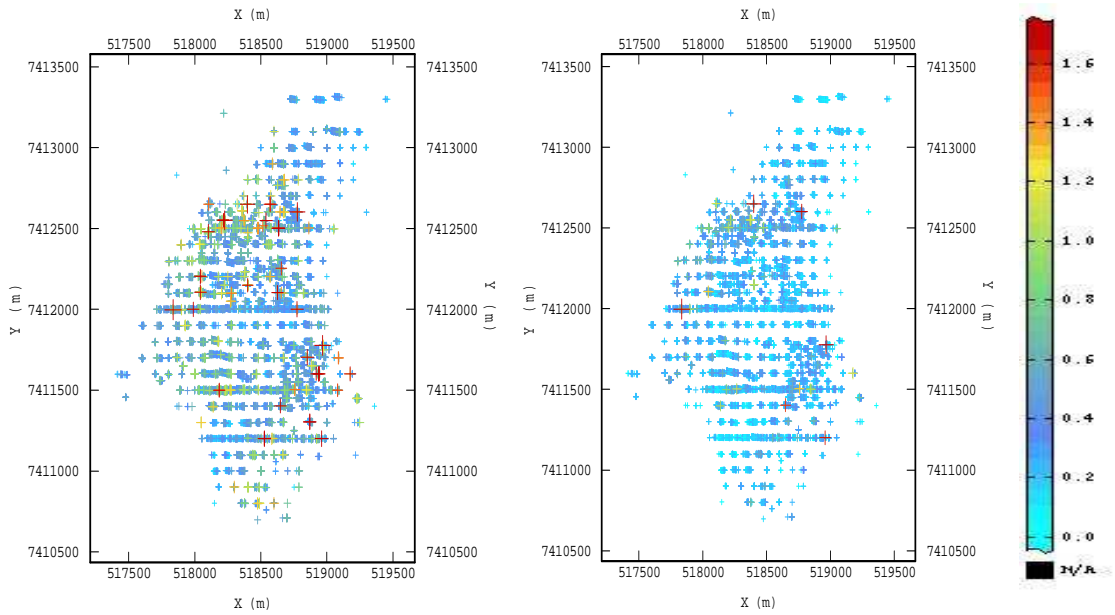


Figura 15 – Proyección de los datos en planta. Ley de CuT (izquierda) y CuS (derecha).

Del despliegue de datos se puede observar una malla de muestreo relativamente regular, el área abarcada tiene una distancia máxima de 2600 [m] en la dirección norte, aproximadamente 2000 [m] en la dirección este, en profundidad los sondajes llegan hasta los 400 [m]. En la siguiente tabla se muestran los límites del yacimiento.

	Este	Norte	Cota
Unidad	[m]	[m]	[m]
Rango	2.046,6	2.619,0	411,1
Mínimo	517.413,1	7.410.697,7	2262,4
Máximo	519.459,7	7.413.316,7	2673,5

Tabla 5 - Distancias máximas en cada dirección.

Con unas vistas en planta del yacimiento se puede apreciar en detalle la malla de muestreo.

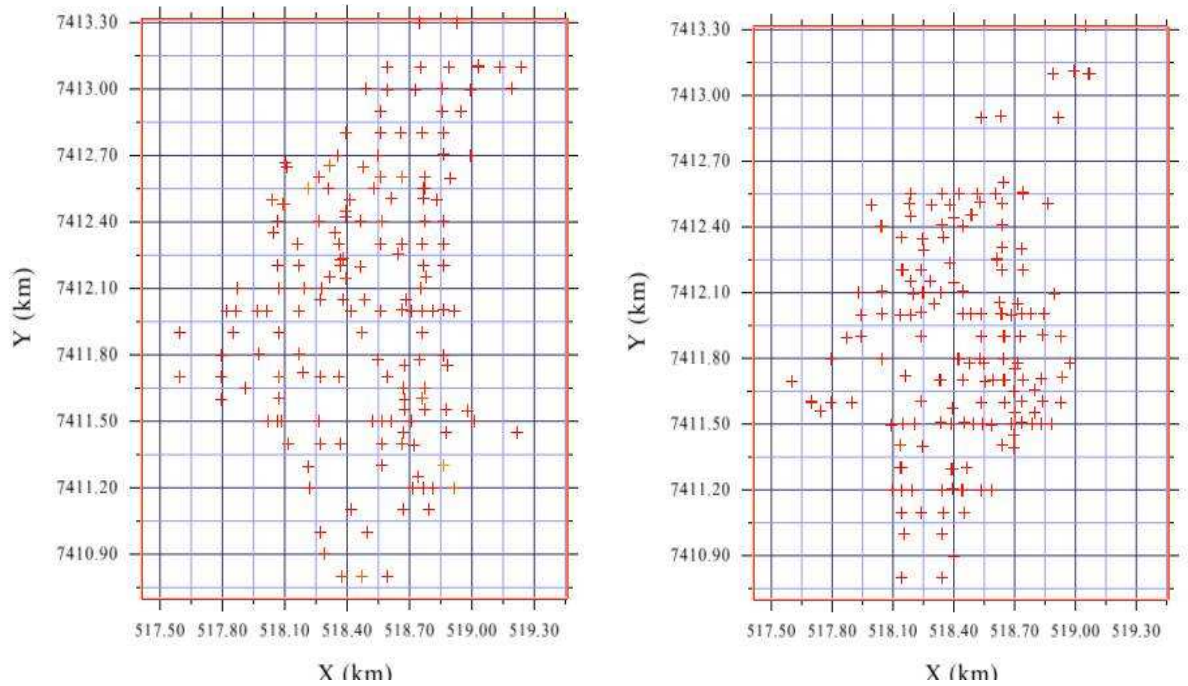


Figura 16- Plantas para determinar el espaciamiento entre datos.

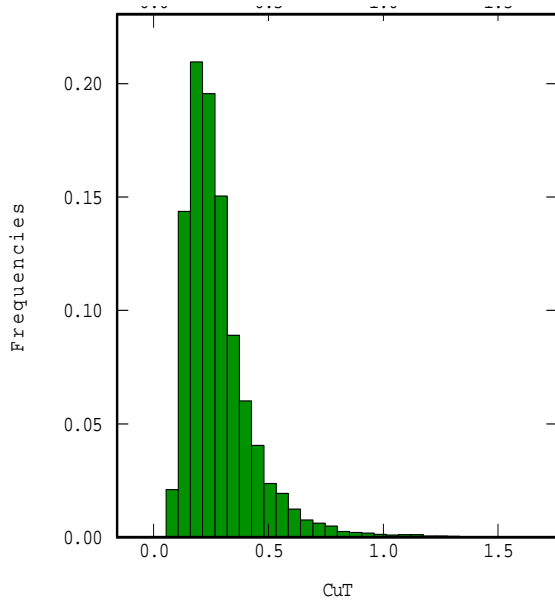
La grilla es de 150 x 150 metros (la menor subdivisión), por lo tanto, los datos están espaciados aproximadamente entre 50 – 75 [m]. El espaciamiento vertical está dado por el largo de los compositos que en este caso es de 1,5 [m]. Con esto en consideración se generaran los variogramas experimentales necesarios para el proceso de simulación.

Antes de las estadísticas se realizó el proceso de búsqueda de duplicados, se encontraron 7 de estos datos, la información siguiente es generada a partir de una base de datos sin estos últimos. Se consideraron duplicados aquellos datos que estaban a una distancia menor a 0,5 [m], ya que se tienen compositos cada 1,5 [m].

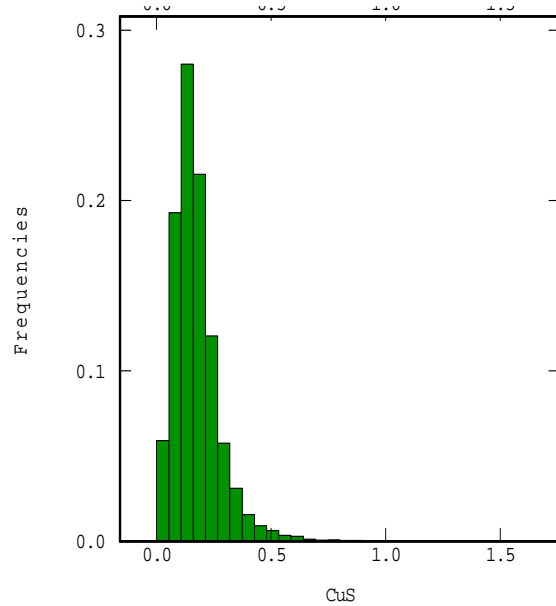
4.3.2 Estadísticas básicas.

A continuación se presentan las estadísticas básicas, así como también otras herramientas necesarias para analizar los datos.

Base de datos completa



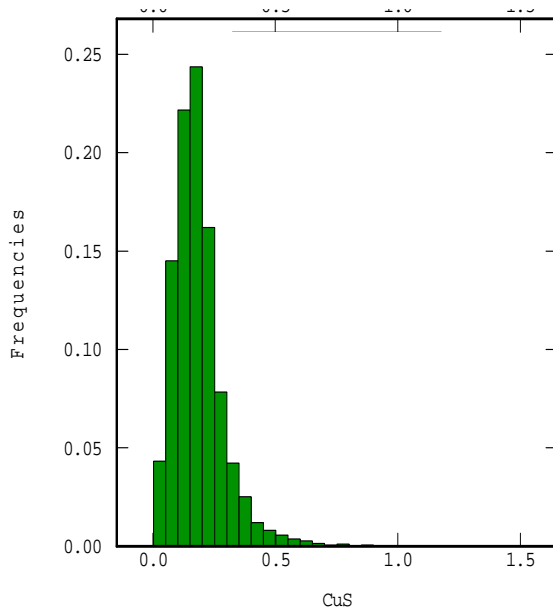
Media	0,29
Varianza	0,029
Mínimo	0,01
Máximo	2,74
n	15.424



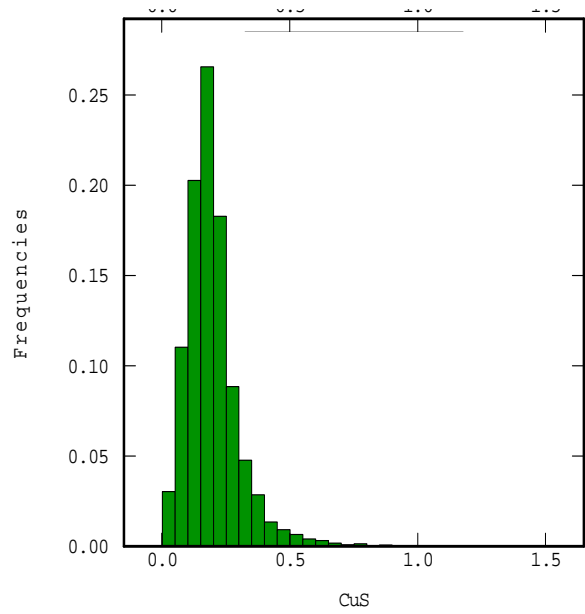
Media	0,17
Varianza	0,012
Mínimo	0,01
Máximo	2,72
n	15.424

Figura 17 - Histogramas y estadísticas básicas, base de datos completa (Caso base).

Base de datos sesgadas.



Media	0,19
Varianza	0,029
Mínimo	0,01
Máximo	2,72
n	13.036



Media	0,20
Varianza	0,014
Mínimo	0,01
Máximo	2,72
n	11.552

Figura 18 - Histogramas y estadísticas básicas, base de datos sesgada 50% (izq) y 80% (der).

Es apreciable fácilmente que las medias son más altas para la ley de CuS en los casos sesgados, debido a la falta de información en la parte baja de la distribución. Así también los histogramas están levemente desplazados a la derecha. Queda en evidencia entonces la necesidad de corregir esta distribución antes de aplicar la transformación a variables gaussianas.

4.3.3 Otros análisis.

En la siguiente figura se muestra el gráfico cuantil contra cuantil de las distribuciones de CuT y CuS comparado a una distribución lognormal.

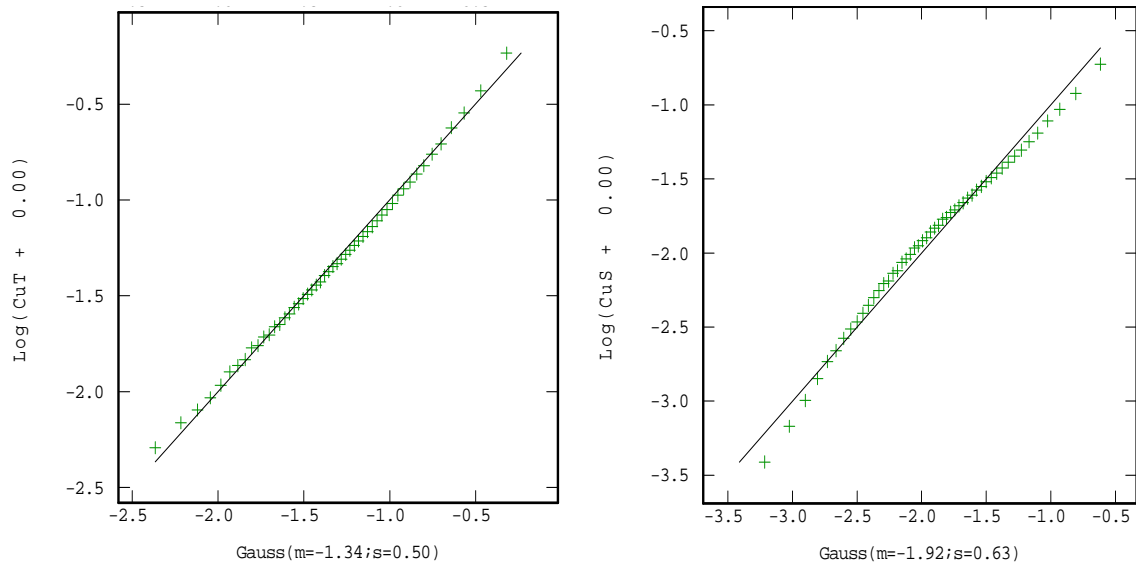


Figura 19 - Comparación de las distribuciones de CuT y CuS con una distribución lognormal.

Aunque se probaron otros tipos de distribuciones la única que se ajustaba de manera relativamente aceptable era la distribución lognormal. En general las leyes de cobre presentan un comportamiento aproximadamente lognormal lo que se comprueba en este estudio. La distribución de CuS en los casos sesgados presenta casi la misma distribución.

4.4 Estudio variográfico.

Se calculan los mapas variográficos con los siguientes parámetros para CuT y CuS.

Dirección		Horizontal	Vertical
Largo del paso	[m]	70	3
N° de pasos	[N°]	20	65
Tolerancia en el paso	[m]	35	1.5
Tolerancia en el ángulo	[°]	30	20
Ancho de banda	[m]	50	30

Tabla 6 - Parámetros del variograma experimental.

En las siguientes figuras se muestran los mapas variográficos para CuT y CuS. Se debe indicar que estos variogramas no son necesarios para el proceso de simulación ya que estos usan los variogramas modelados de las variables gaussianas, sin embargo se considera necesario realizarlos para posteriores análisis de los resultados.

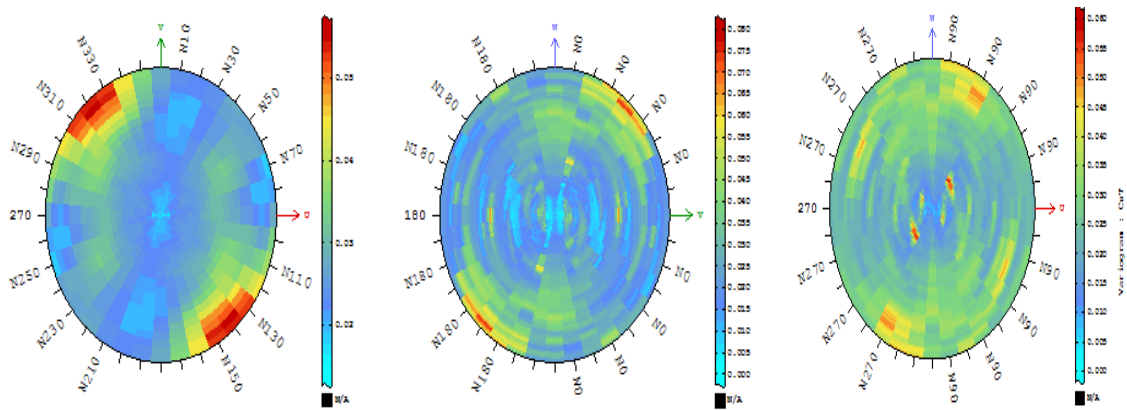


Figura 20 - Mapas variográficos cobre total (CuT).

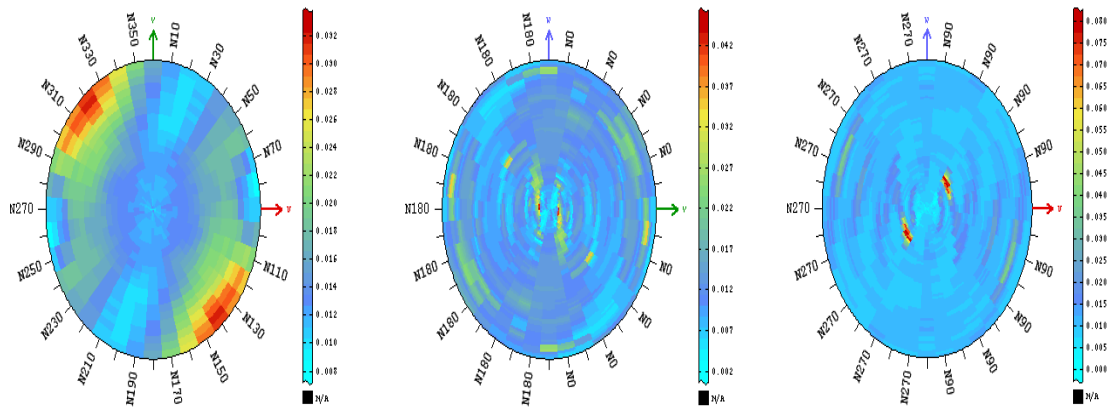


Figura 21 - Mapas variográficos cobre soluble (CuS).

De la visualización de los mapas variográficos se identificaron 3 posibles direcciones de anisotropía en la horizontal (probablemente 2 pero se prueban 3) en la vertical no se encuentran direcciones preferentes de mayor o menor continuidad por lo que se realizará un solo variograma en la dirección vertical.

Con los mismos parámetros usados para los mapas variográficos se generan los variogramas experimentales en las direcciones de interés, a saber, en la horizontal N20, N80, N140 además de la dirección vertical. El resultado se muestra en las siguientes figuras.

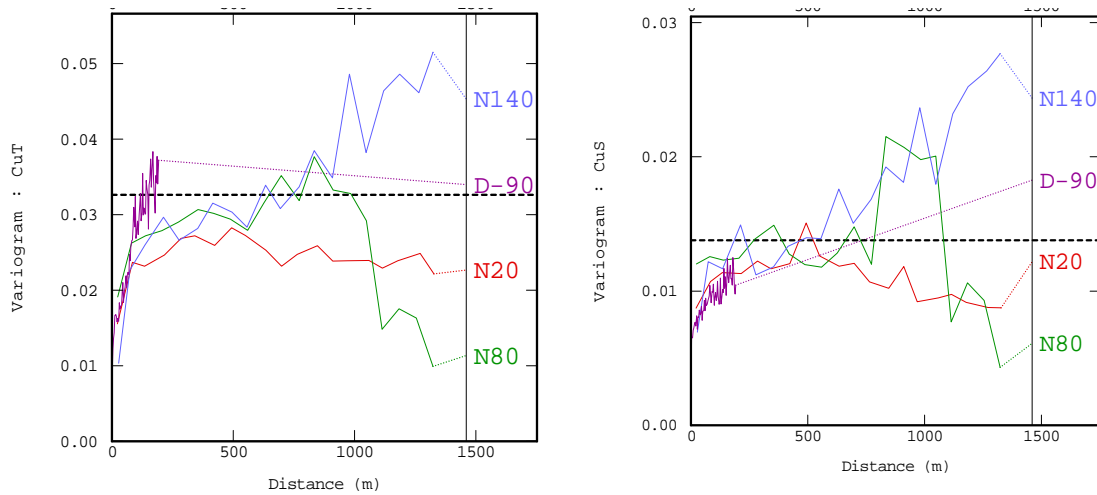


Figura 22 - Variogramas experimentales de CuT y CuS.

5. IMPLEMENTACION DE LA CORRECCIÓN.

El primer paso para llevar a cabo la corrección es la implementación de esta en algún lenguaje de programación. El lenguaje escogido fue python debido a su simplicidad y versatilidad

Como datos de entrada el programa recibe tanto las coordenadas como las variables de interés (CuT y CuS) y algún ponderador de desagrupamiento.

Este	Norte	Cota	CuT	CuS	Peso Desagrupado
[m]	[m]	[m]	[%]	[%]	
518338.7	7412650.7	2626.8	0.015	0.007	2.45287
518869.8	7411749.9	2626.4	0.021	0.005	1.16562
518641.4	7412549.7	2632	0.032	-9999	2.45287
518843.1	7411501.4	2504.8	0.045	0.017	0.73586
519040.4	7412497.6	2558.4	0.052	0.021	2.45287

Tabla 7 - Ejemplo de archivo de datos.

Además del archivo con estos datos el programa recibe otro archivo con parámetros e instrucciones a modo de opciones iniciales de ejecución. Principalmente las variables más importantes son el mínimo, el máximo y el número de clases para realizar la discretización de la distribución bivariable (y también de las marginales). El mínimo y máximo por defecto son los valores extremos de los datos, sin embargo para una mejor visualización de la distribución a menudo es conveniente reducir el máximo o aumentar el mínimo (en el presente caso conviene hacer lo primero).

Otras variables a considerar son establecer el valor que identifica la ausencia de datos (-9999 en nuestro caso) y la designación de las columnas donde están la variable primaria (CuT, la mejor informada), la variable secundaria (CuS, sin información en un porcentaje importante) y los ponderadores de desagrupamiento si los hay.

Parámetros	Valor
Columna_Este	1
Columna_Norte	2
Columna_Cota	3
Variable_primaria	4
Variable_secundaria	5
Ponderadores	6
N°de_clases	20
Minimo	0
Maximo	1
Sin_dato	-9999

Tabla 8 - Ejemplo de archivo de parámetros.

El primer paso para la corrección es la construcción de las distribuciones marginales, usando los datos obtenidos del archivo de parámetros, son necesarias por tanto 3 de ellas:

- 1) El histograma de los valores de CuT de todos los datos disponibles (distribución que suponemos insesgada).
- 2) El histograma de los valores de CuT donde existe información de CuS (distribución sesgada).
- 3) El histograma de los valores de CuS (distribución sesgada).

Con 2) y 3) se puede construir la distribución bivariable sesgada, para esta base de datos y para 15 clases entre 0 y 1, usando la base de datos de sesgada 50 % bajo la ley 0.2 % CuT se tienen los siguientes resultados.

Las distribuciones 1 (izquierda) y 2 (derecha) se presentan en la figura 23.

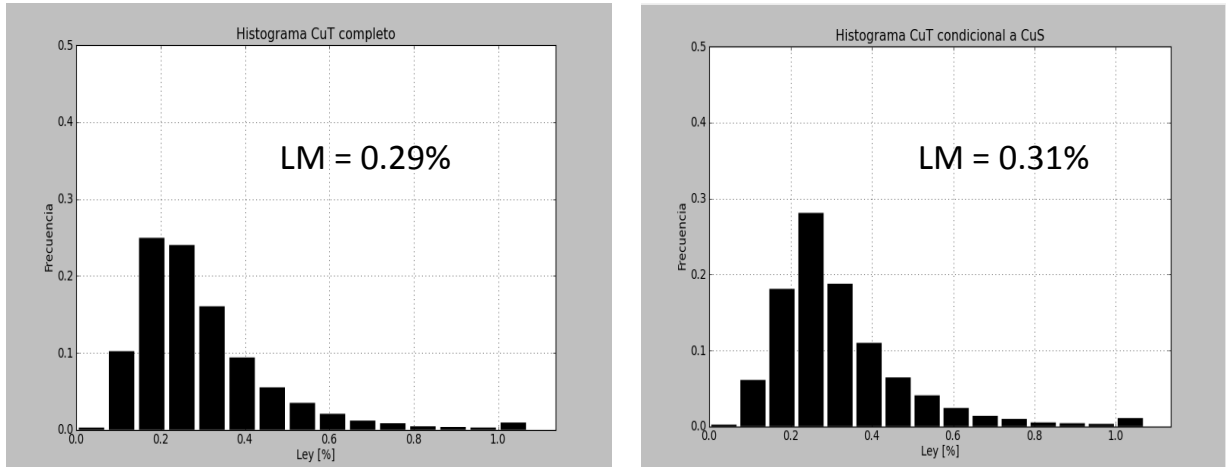


Figura 23 - Histogramas de las distribuciones necesaria para hacer la corrección.

Como era esperable toda la distribución sufrió cambios, las clases que están por debajo de la ley del sesgo (0.2%) están menos representadas y las que están más arriba de ésta, están sobre representadas. La ley media informada también sube ya que en la distribución sesgada los pares con bajos valores de CuT no aparecen en ella.

La distribución 3 se presenta en la siguiente figura (derecha), la distribución de la izquierda es la de CuS en la base de datos completa.

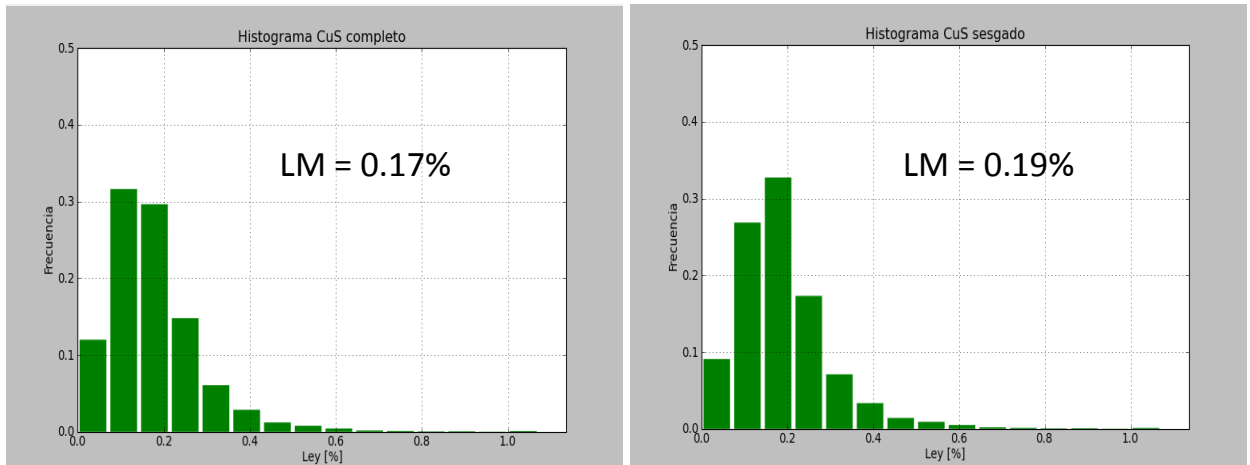


Figura 24 - Histogramas de las distribuciones de CuS, caso base (izq) y caso sesgado 50% (der).

En este caso también se nota la influencia del sesgo tanto en la forma de la distribución como en las estadísticas básicas.

Con estas distribuciones es posible realizar la corrección, el resultado de esta se muestra en las siguientes figura, siguiendo el análisis para la base de datos sesgada 50% bajo la ley de corte de 0.2% y usando 15 clases para realizar la corrección.

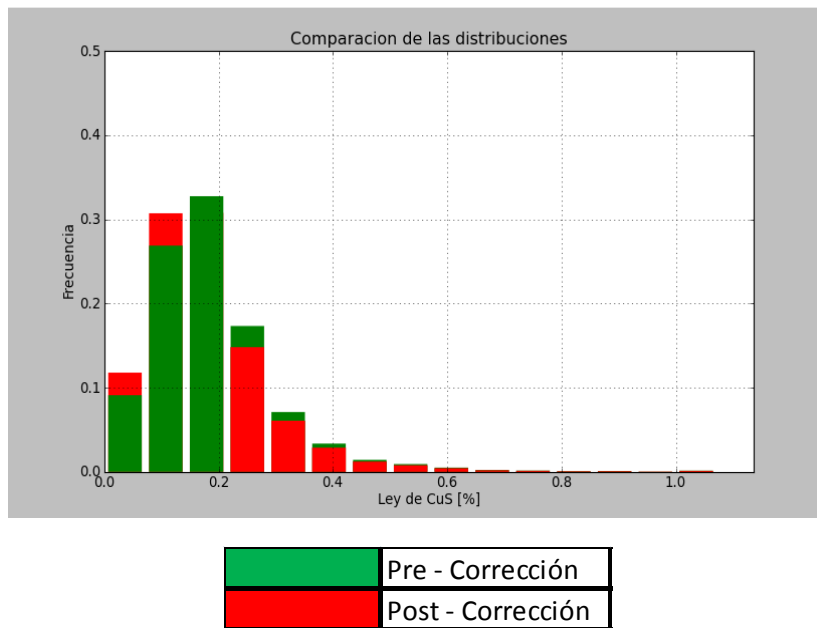


Figura 25 - Comparación entre la distribución corregida y la distribución sesgada de cobre soluble (Caso sesgado 50%).

Se puede ver que la distribución corregida se ajusta mucho mejor a la distribución real, lo que se mas claramente en la siguiente figura.

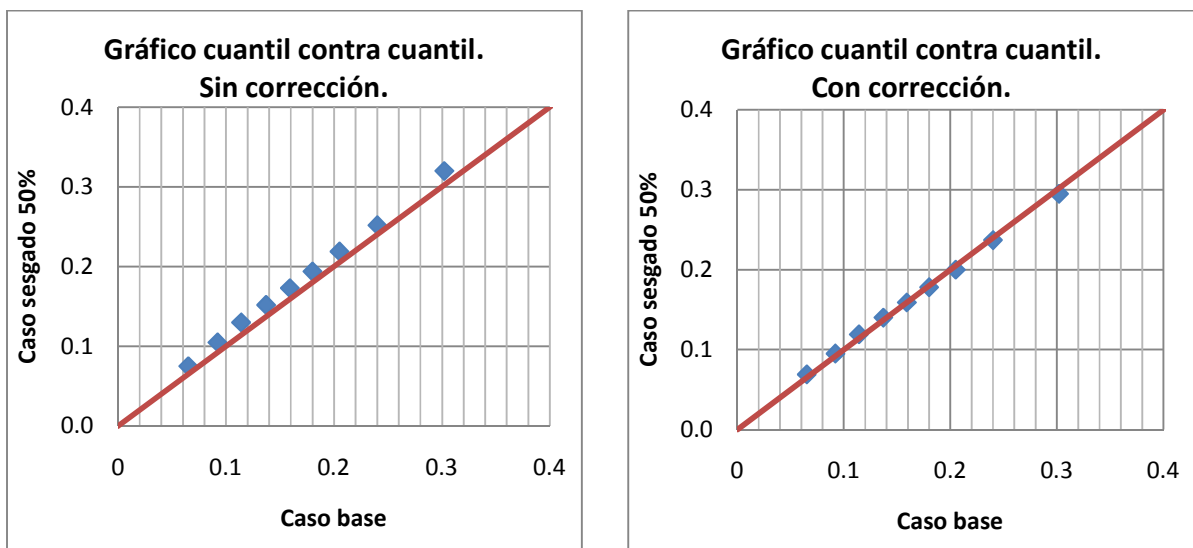


Figura 26 - Comparación entre la distribución corregida y la distribución sin corregir.

Sesgo del 50%.

Como se ve en la figura 26 la distribución corregida se ajusta mejor a la distribución del caso base, por lo tanto, la corrección de la distribución sesgada cumple el objetivo propuesto anteriormente.

Lo mismo sucede al utilizar la base de datos sesgada un 80% bajo la ley 0.2 %, como se ve en las figuras 27 y 28, también usando 15 clases para realizar la corrección.

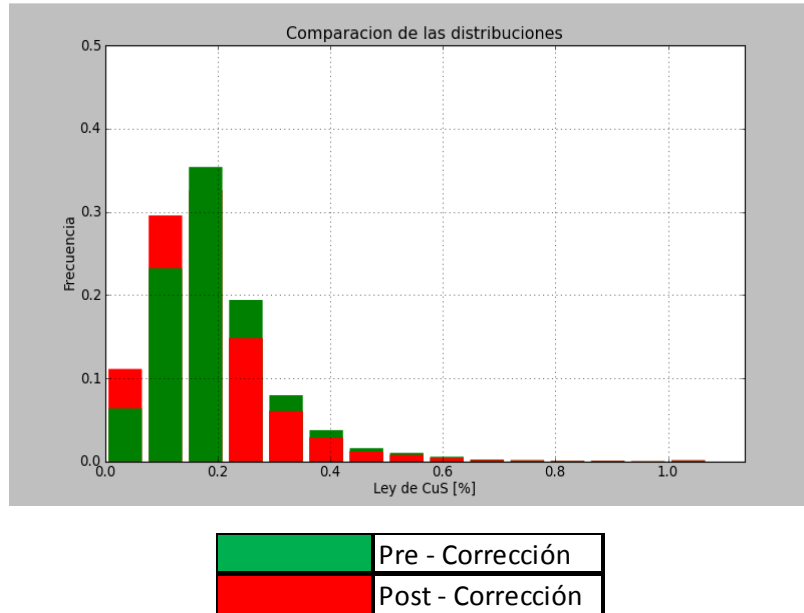


Figura 27 - Comparación entre la distribución corregida y la distribución sesgada de cobre soluble (Caso sesgado 50%).

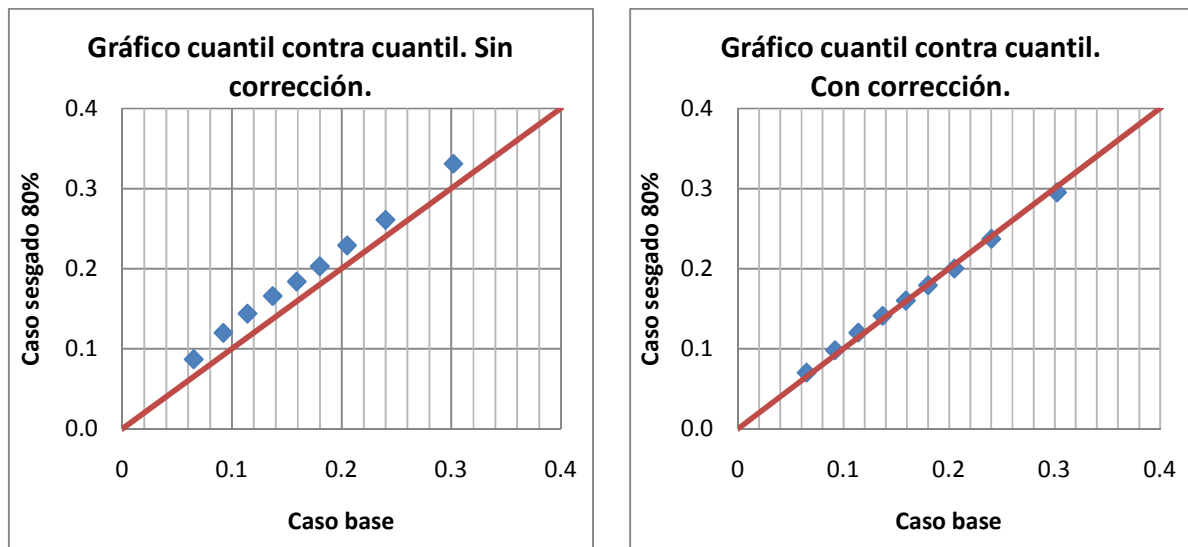


Figura 28 - Comparación entre la distribución corregida y la distribución sin corregir.

Sesgo del 80 %.

En general, la distribución generada por la corrección se ajusta mucho mejor que las distribuciones sesgadas a la distribución del caso base. Hecho que queda más en evidencia entre mayor sea el sesgo de la distribución como se ve al comparar las figuras 26 y 28.

5.1 Dependencia del factor de corrección con el número de clases.

En las siguientes figuras se muestra el factor de corrección que afecta a cada clase y la distribución de CuS corregida resultante, en la base de datos 50% de sesgo. En la primera fila se tienen 15 clases en la segunda se tienen 10 clases y en la tercera solo 5 clases.

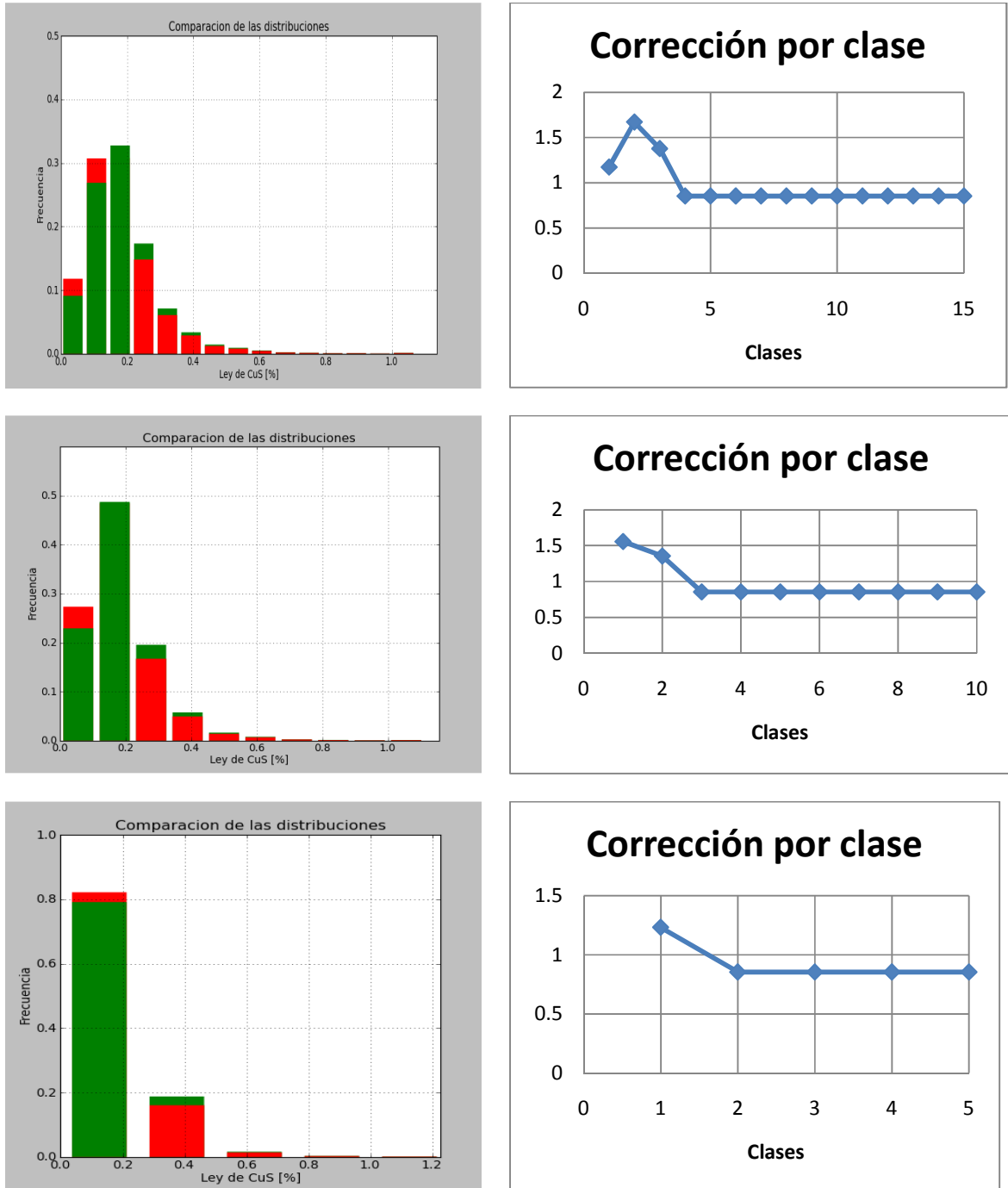


Figura 29 - Dependencia del factor de corrección del número de clases.

La corrección mostrada anteriormente corresponde a la razón sobre CuT por eso la diferencia entre la distribución post y pre corrección no corresponde exactamente a la mostrada en el factor de corrección. Por ejemplo, para la segunda clase del primer grafico la corrección casi alcanza un valor de 2, sin embargo, en la segunda clase del grafico de la izquierda no se ve que la distribución post corrección sea casi el doble que la frecuencia pre corrección, esto se debe a que el histograma presentado ahí es de la distribución de CuS y no la del CuT, en la distribución de CuT se espera ver el comportamiento presentado en los gráficos de corrección por clase.

Además se debe mencionar que la corrección sobre la ley del sesgo es constante y menor que uno ya que no existe sesgo sobre esa distribución y la corrección intenta solo restablecer su real peso en la distribución.

En la figura 30 se muestra la comparación entre las distribuciones de CuS en el caso base y las resultantes de la corrección, al usar 15 clases en la discretización.

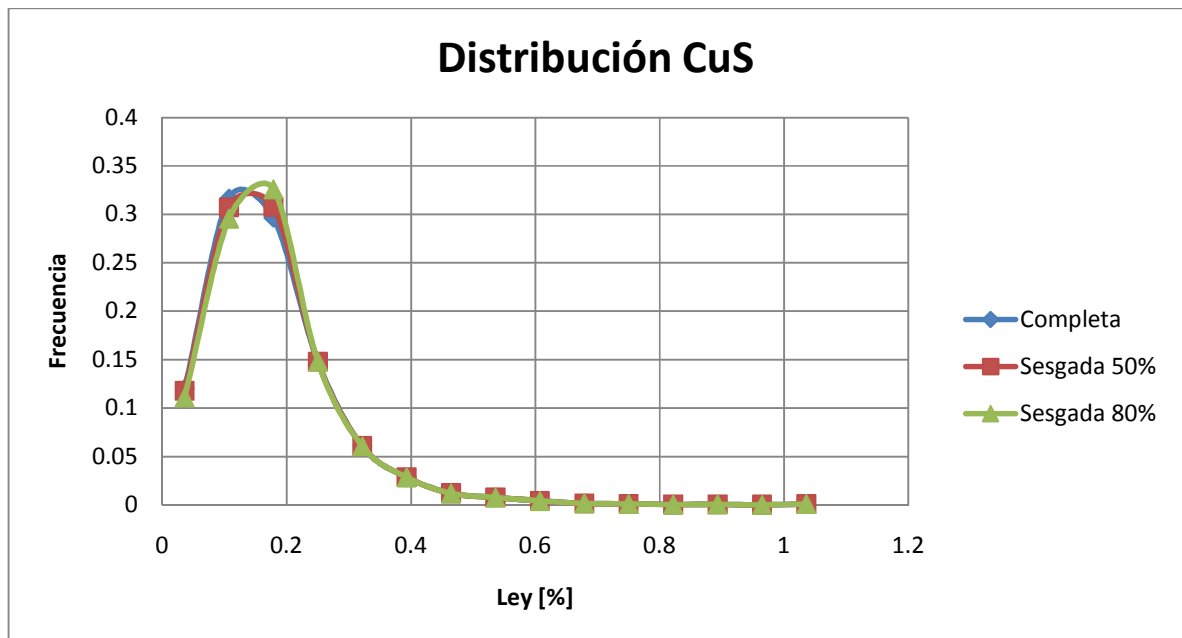


Figura 30 - Comparación de las distribuciones de CuS, caso base y casos sesgadas.

Las diferencias sobre la ley de corte del sesgo son mínimas mientras que bajo ella pueden resultar significativas, sin embargo, las diferencias observadas por clase no sobrepasan el 5%. En la siguiente figura se puede ver la parte inferior de la distribución (4 primeras clases) donde se aprecian dichas diferencias.

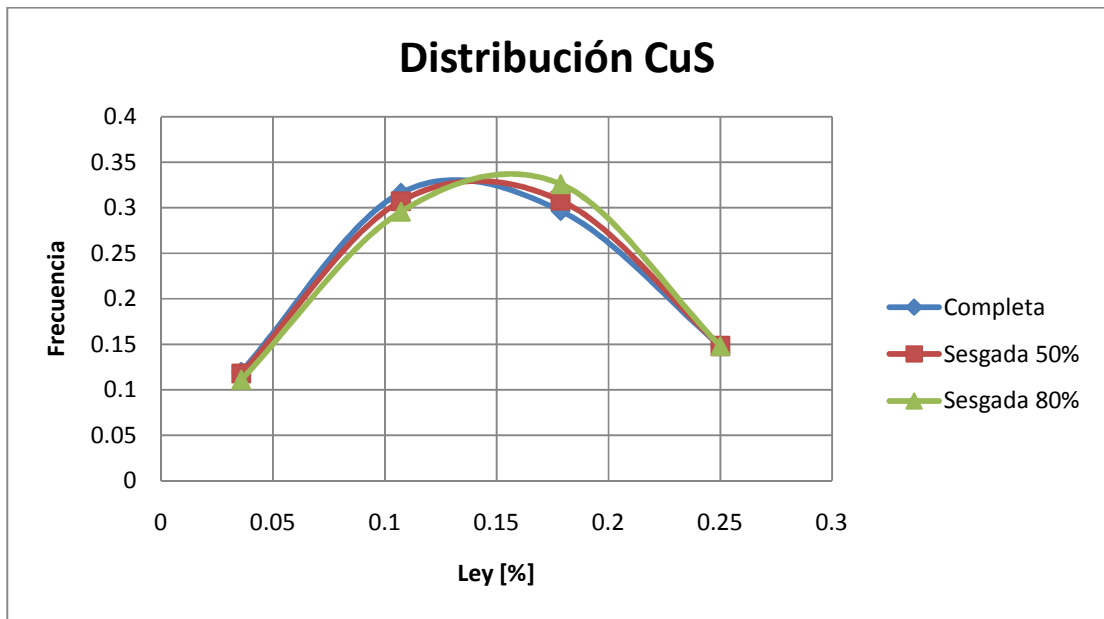


Figura 31 - Comparación en detalle de las distribuciones.

Puede ser necesario estudiar lo que pasa si se agranda el intervalo sesgado o se aumenta el sesgo aún más.

Como último punto a considerar es que las distribuciones corregidas en los puntos donde se tiene información de las 2 variables deben reproducir las estadísticas básicas usando los ponderadores de corrección, es decir, si se calcula la media en la base de datos completa usando los ponderadores de desagrupamiento esta debería ser igual a la media usando la base de datos sesgada usando los ponderadores actualizados de la corrección. En la siguiente tabla se muestran las estadísticas básicas del caso base y de los casos sesgados corregidos.

	Unidades	Caso		
		Base	Sesgo 50%	sesgo 80%
Media	[%]	0.176	0.177	0.178
Varianza	[%]	0.014	0.012	0.012
Mínimo	[%]	0.005	0.005	0.005
Máximo	[%]	2.72	2.72	2.72

Tabla 9 - Comparación de las estadísticas básicas entre el caso base y los casos sesgados corregidos.

Como se ve en la tabla 9 aunque existen pequeñas diferencias en las estadísticas básicas, debido principalmente al grado de refinación de la corrección, esta logra su objetivo, vale decir, reproducir la forma y las estadísticas básicas de la distribución de CuS.

Con los ponderadores de corrección se puede iniciar el paso siguiente, la transformación de las variables originales en variables gaussianas, usando la transformación condicional paso a paso.

6. SIMULACIÓN CONDICIONAL.

6.1 Transformación de las variables CuT y CuS en variables gaussianas.

Dada la restricción mineralógica $CuT \geq CuS$ los valores gaussianos de estas dos variables no mostrarían un comportamiento bivariable gaussiano, por esto se debe usar el método de transformación descrito anteriormente de manera de asegurar el carácter bigausiano de la distribución bivariable (Gauss CuT, Gauss CuS).

En las siguientes figuras se muestra el resultado de la transformación, aplicada a las 3 bases de datos, usando los correspondientes ponderadores de la corrección.

Base de datos completa

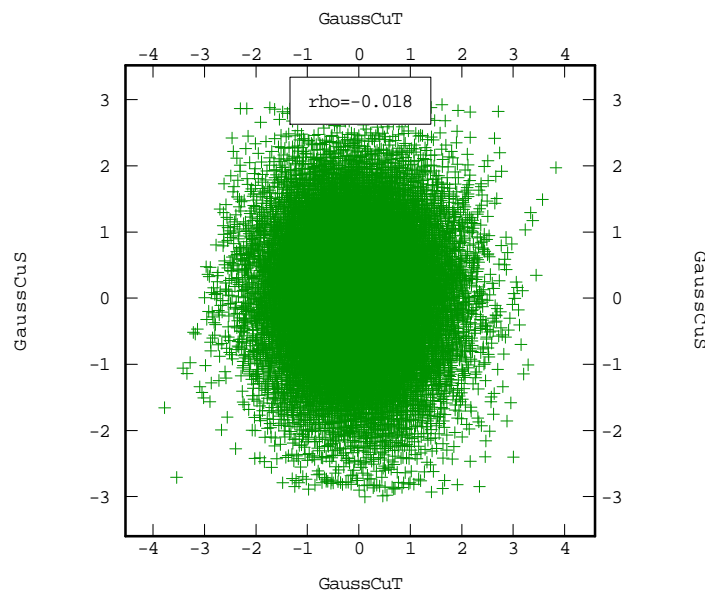
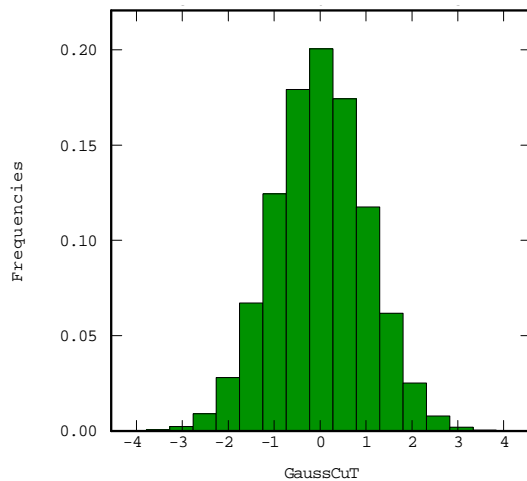
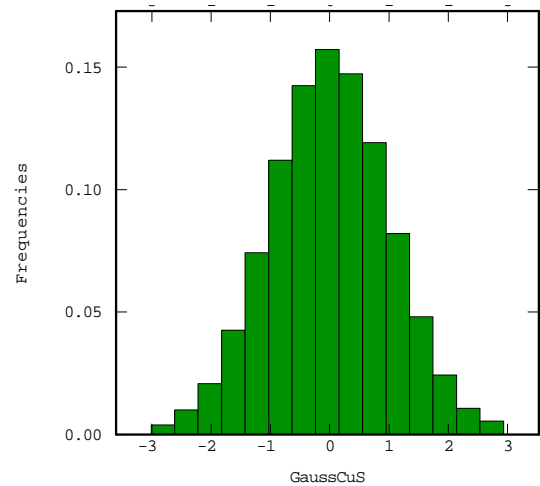


Figura 32 - Distribución bivariable, variables gaussianas CuT y gaussianas CuS.

Como se ve la distribución bivariable sí tiene un comportamiento bigausiano, y el grado de correlación entre las dos variables es bastante bajo (0.018). En las siguientes figuras se ven las distribuciones marginales de las variables gaussianas.



Media	0.00
Desv. Estandar	1.00
Mínimo	-3.78
Máximo	3.83



Media	0.00
Desv. Estandar	1.00
Mínimo	-3.00
Máximo	2.93

Figura 33 - Histogramas y estadísticas básicas de las variables gaussiana CuT (izq) y gaussiana CuS (der).

Como se puede ver la transformación resulta exitosa, ya que tanto las distribuciones marginales como la bivariable muestran un comportamiento gaussiano. Además en el gráfico bivariable se puede ver el coeficiente de correlación (0.018) lo que da una idea de la independencia de las variables gaussianas, prácticamente no tienen ninguna correlación, idea que será confirmada más adelante.

En la siguiente figura se comparan las distribuciones transformadas con una normal de media 0 y desviación estándar de 1. Se puede ver claramente que son prácticamente iguales, tanto para la variable gaussiana de CuT como la gaussiana de CuS.

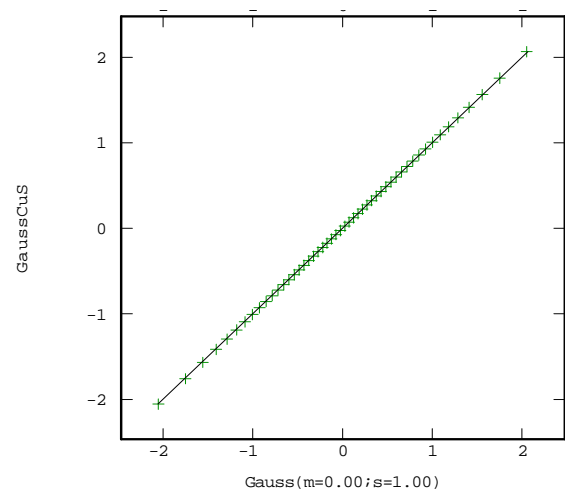
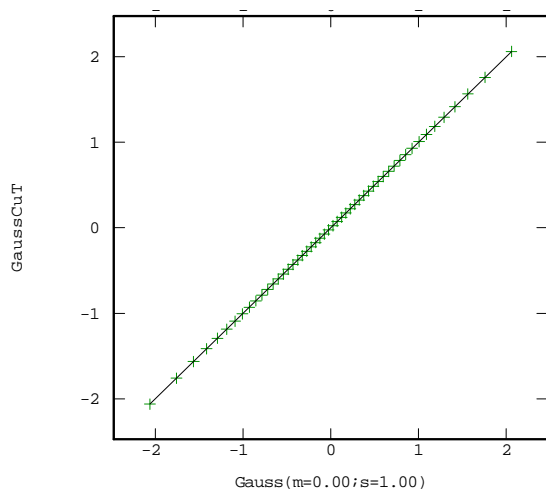


Figura 34 - Comparación de las variables gaussianas con una distribución normal de media 0 y varianza 1.

6.2 Estudio variográfico de las variables transformadas.

6.2.1 Estudio variográfico variable gaussiana CuT.

Los mapas variográficos y los variogramas experimentales se realizaron usando los mismos parámetros mencionados anteriormente, los que se muestran nuevamente en la siguiente tabla.

Dirección		Horizontal	Vertical
Largo del paso	[m]	70	3
N° de pasos	[N°]	20	65
Tolerancia en el paso	[m]	35	1.5
Tolerancia en el ángulo	[°]	30	20
Ancho de banda	[m]	50	30

Tabla 10 - Parámetros para generar el variograma experimental.

En las figuras siguientes se muestran los mapas variográficos de la variable gaussiana CuT.

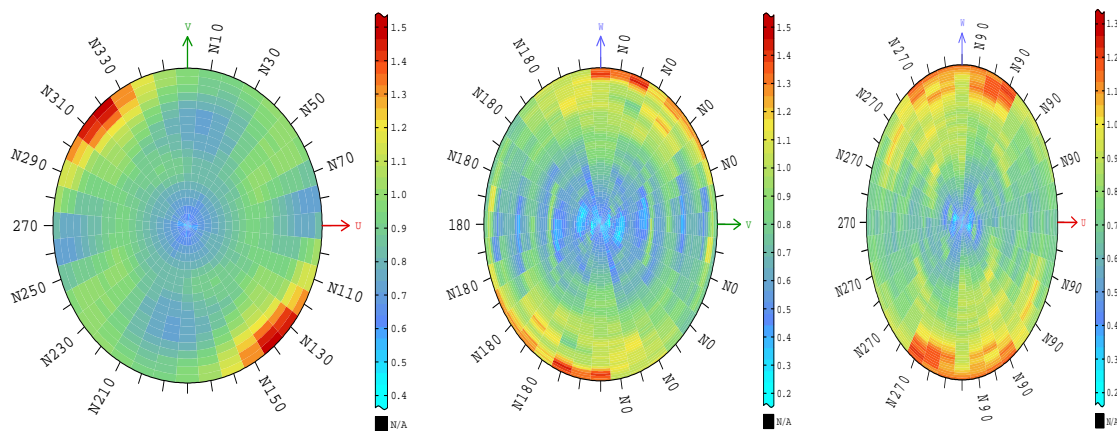


Figura 35 - Mapas variográficos variable gaussiana CuT.

Se puede apreciar que mantiene prácticamente la misma variabilidad que la variable CuT, ya que la transformación de la variable primaria es igual que la transformación común. Por tanto se escogen las mismas direcciones para calcular los variogramas. El variograma experimental y el variograma modelado se presentan en la siguiente figura.

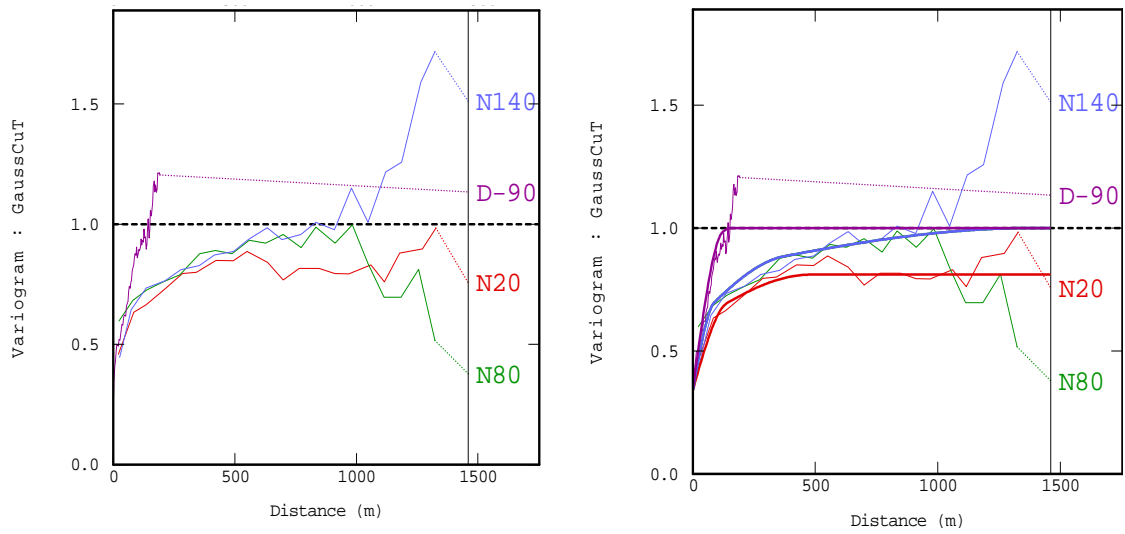


Figura 36 - Variogramas experimental y modelado de la variable gaussiana CuT.

Donde el modelo del variograma de la variable gaussiana CuT es el siguiente:

$$\gamma(h) = 0.34 + 0.25 \cdot \text{esf}(140,75,130) + 0.21 \cdot \text{esf}(470,550,140) + 0.2 \cdot \text{esf}(\infty,1100,160)$$

6.2.2 Estudio variográfico variable gaussiana CuS.

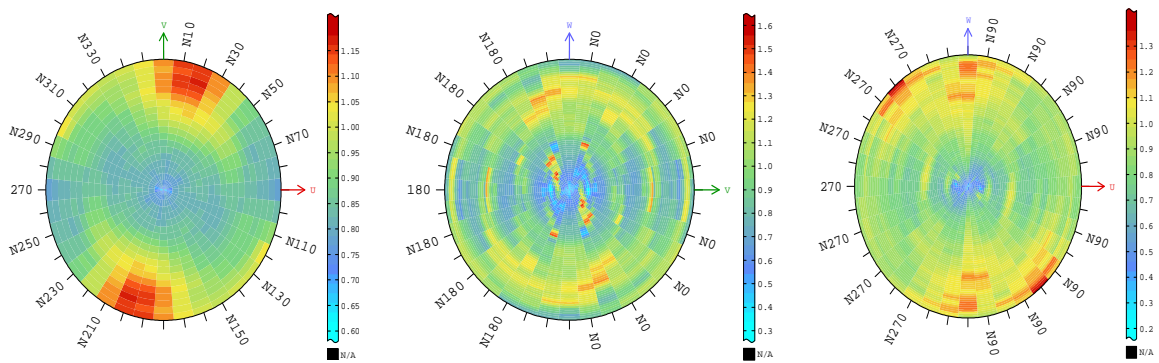


Figura 37 - Mapas variográficos variable gaussiana CuS corregida y desagrupada.

Se puede apreciar que no se mantiene la variabilidad que la variable CuS, esto se debe a que la variable secundaria se transforma condicionalmente a la primera (CuT), cabe hacer notar que la dirección de mayor variabilidad de la variable gaussiana CuS se rota justo donde está la dirección de mayor continuidad en la variable gaussiana CuT. En este caso se escoge realizar el variograma solo en 2 direcciones (N10 y N100) en la horizontal y mantener la dirección vertical.

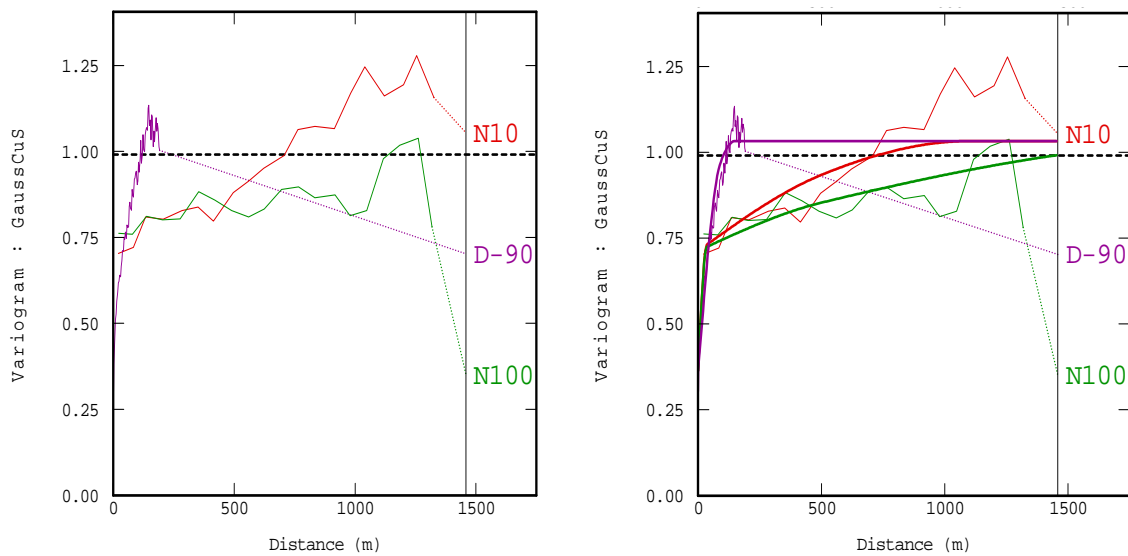


Figura 38 - Variogramas experimental y modelado, variable gaussiana CuS.

Donde el modelo del variograma de la variable gaussiana CuS es el siguiente:

$$\gamma(h) = 0.36 + 0.35 \cdot \text{esf}(35,35,90) + 0.051 \cdot \text{esf} \cdot (700,300,140) + 0.27 \cdot \text{esf}(1100,2300,150)$$

6.2.3 Variograma cruzado CuT – CuS

En la figura siguiente se muestra el variograma cruzado entre las variables gaussianas CuS y CuT.

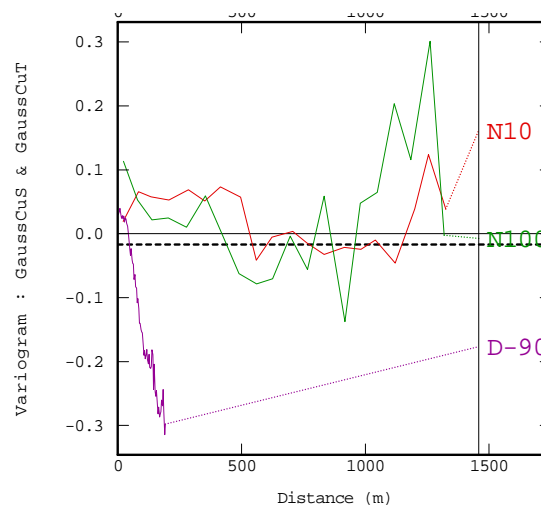


Figura 39 - Variograma cruzado, variables gaussiana CuT y gaussiana CuS.

Como se puede ver la correlación entre las dos variables es más bien baja, entonces se justifica que se pueden simular ambas variables de manera independiente.

6.3 Validación cruzada.

Antes de iniciar el proceso de simulación es necesario validar los modelos de variogramas y las vecindades que se usarán. Para este estudio se considera una vecindad de $300 \times 300 \times 30$ [m] compuesta de 4 cuadrantes con un número óptimo de datos de 10, lo que da en total 40 datos para estimar un punto.

Se hizo entonces validación cruzada considerando estos parámetros obteniéndose los siguientes resultados.

- Histograma de errores con media 0.
- Cerca del 97% de los datos son bien estimados considerando como criterio que un dato está mal estimado si supera 2.5 veces la desviación estándar (intervalo de confianza), los puntos mal estimados están marcados en rojo.
- El gráfico valor estimado contra valor real muestra una nube de puntos en torno a la diagonal, lo que da una idea de que tan bien se reproducen los valores reales a partir de los estimados.

Estos puntos se muestran en las siguientes figuras.

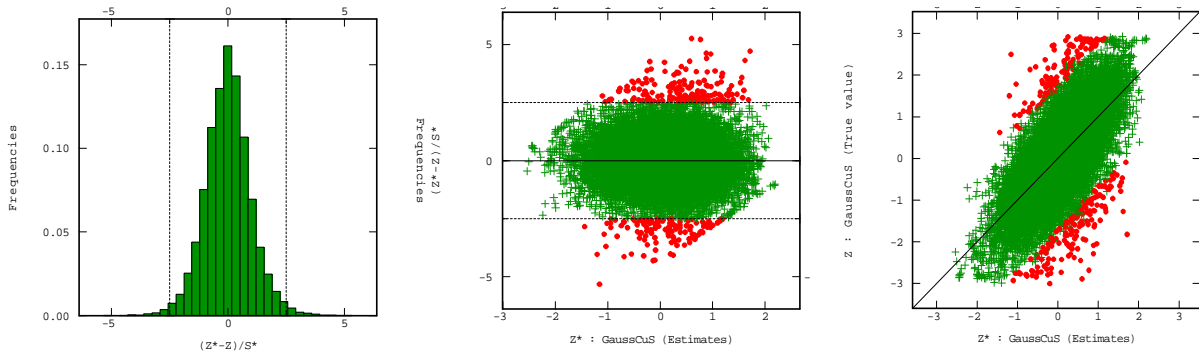


Figura 40 - Validación cruzada.

Por lo tanto, validados los modelos se procede con la simulación condicional de las variables en estudio.

6.4 Detalles de la simulación condicional.

El algoritmo usado para simular las variables en estudio (gaussiana de las variables CuT, CuS, CuS 50% sesgo y CuS 80% de sesgo) fue el de bandas rotantes, se usaron 2000 bandas rotantes y se hicieron 100 simulaciones para cada una de ellas.

Se realizó una simulación puntual en una malla regular de 6.66 x 6.66 x 5 [m] para luego rebloquear los resultados a una malla de 20 x 20 x 15 (soporte de bloques). La simulación no se hizo a lo largo y ancho de todo el yacimiento si no solo sobre una parte de este, considerando que era innecesario simular todo el depósito ya que bastaba solo una parte que fuera representativa.

En total se simularon casi 4 millones de puntos y en el soporte de bloques se tienen cerca de 140.000 datos estimados.

6.5 Validación de la simulación.

6.5.1 Resultados de la simulación –Soporte puntual.

En las figuras siguientes se puede ver la nube de correlación entre los valores simulados de la variable gaussiana CuT y los valores simulados de la variable gaussiana CuS en cada uno de los casos vistos anteriormente para soporte puntual. (Arriba Caso base, izquierda caso 50% de sesgo y derecha caso 80% de sesgo).

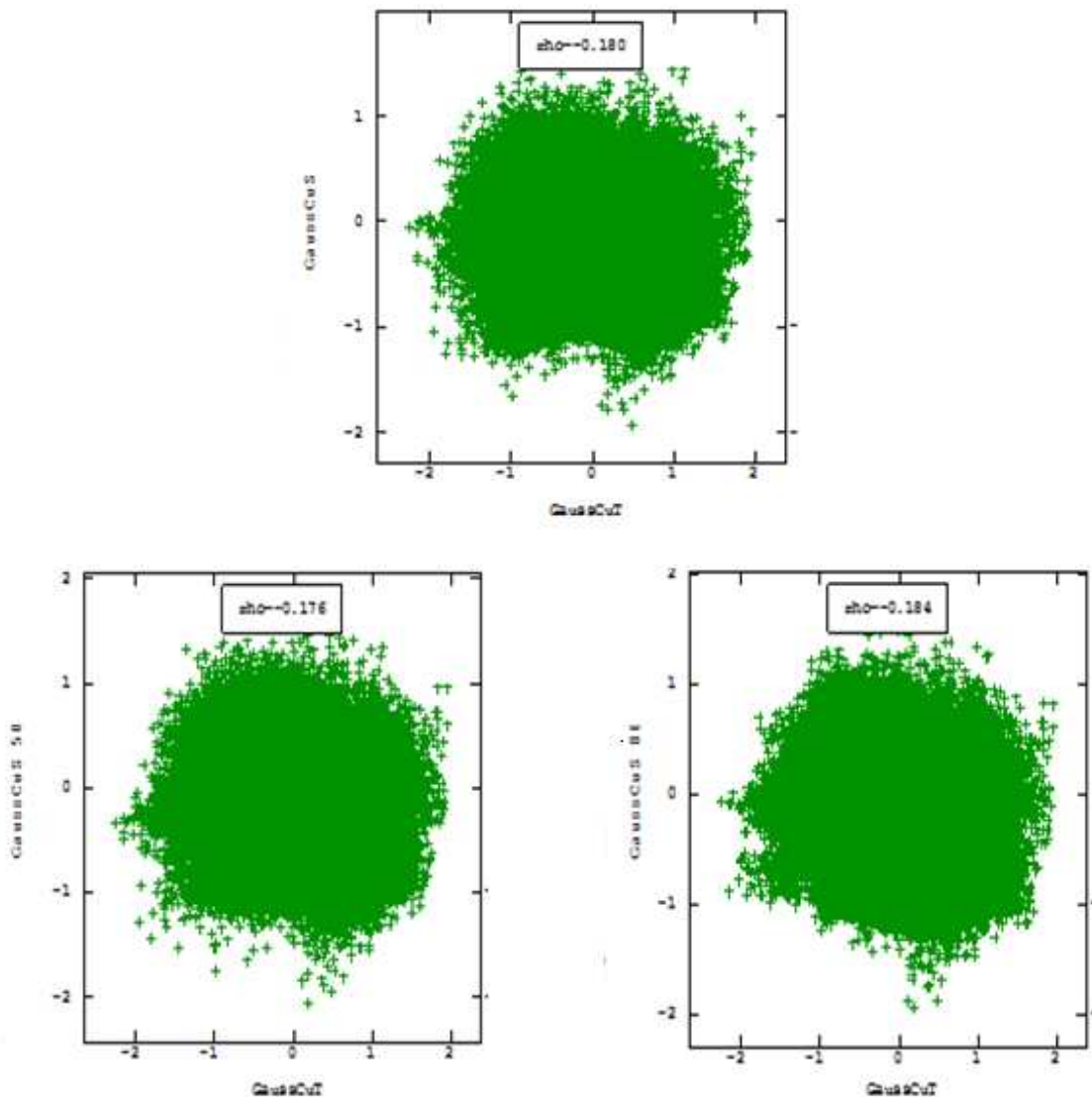
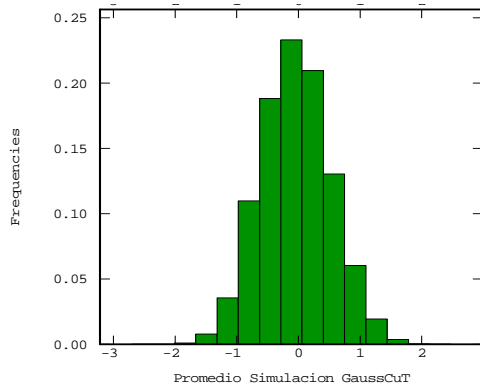
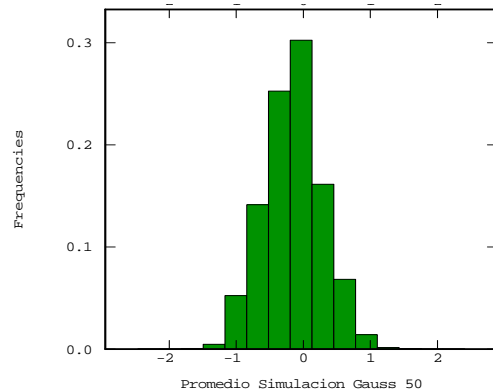


Figura 41 - Distribución bivariable de los valores simulados en soporte puntual.

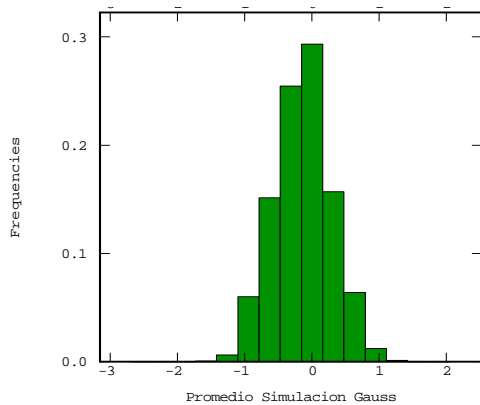
En los histogramas siguientes se ven los resultados todavía sin transformar. (arriba izquierda gaussiana CuT, abajo izquierda gaussiana CuS caso base, derecha arriba gaussiana CuS sesgada 50% y abajo derecha con un sesgo del 80%). Cabe recalcar que no se hace la transformación de vuelta a la variable original a partir de los promedios de las realizaciones y la figura siguiente pretende dar una idea de que tan buena son los resultados de la simulación no condicional.



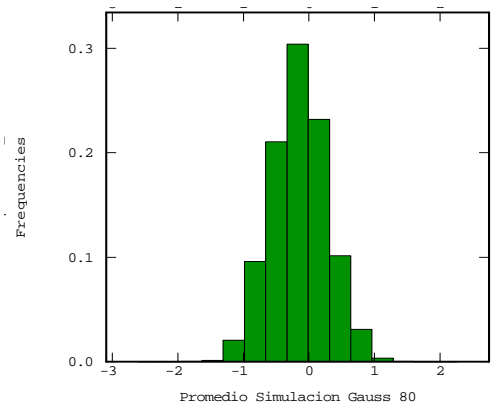
Media	-0.04
Desv. Estandar	0.57
Mínimo	-2.70
Máximo	2.47



Media	-0.15
Desv. Estandar	0.45
Mínimo	-2.48
Máximo	2.40



Media	-0.14
Desv. Estandar	0.42
Mínimo	-2.68
Máximo	2.05



Media	-0.15
Desv. Estandar	0.42
Mínimo	-2.60
Máximo	2.25

Figura 42 - Histogramas y estadísticas básicas de los valores simulados en cada uno de los casos de estudio, a soporte puntual y como promedio de 100 realizaciones.

En la siguiente figura se ven los histogramas y estadísticas básicas de 3 realizaciones de las variables gaussiana CuT y gaussiana CuS (figuras 43 y 44 respectivamente).

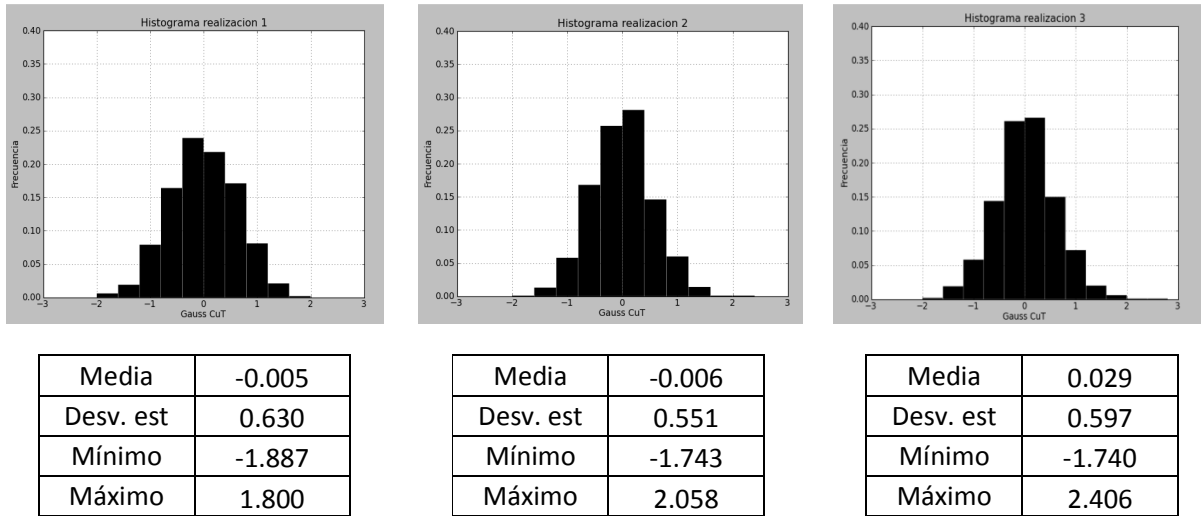


Figura 43 – Histogramas y estadísticas básicas de 3 realizaciones de la variable gaussiana CuT.

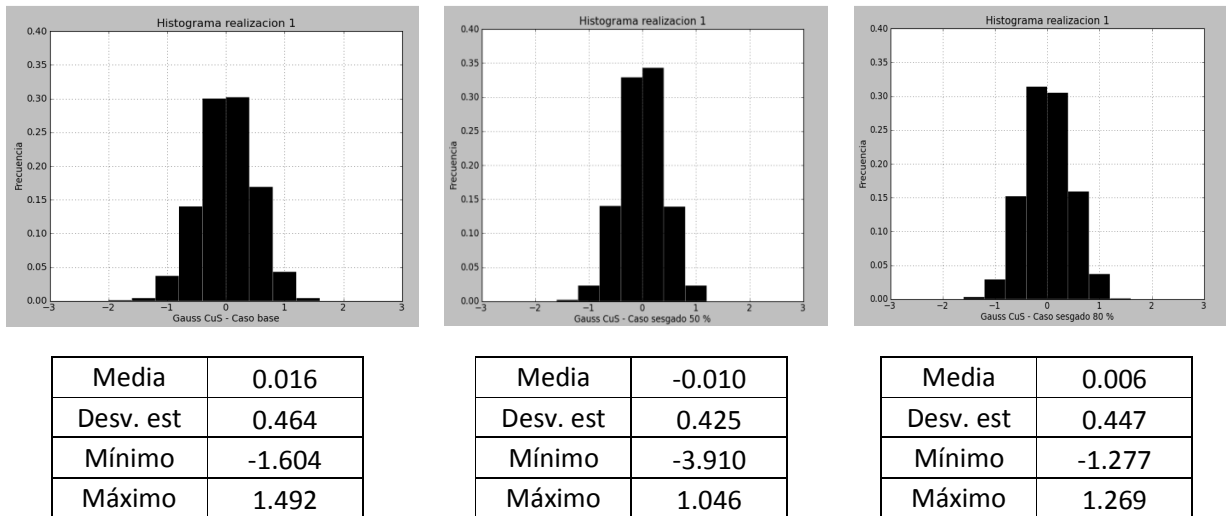


Figura 44 - Histogramas y estadísticas básicas de 3 realizaciones de la variable gaussiana CuS en las distintas bases de datos.

Como se puede ver en las figuras anteriores la media de los valores simulados es cercana a 0, sin embargo, la desviación estándar está por debajo del valor esperado (desviación estándar igual a 1) principalmente debido a que no se simuló en todo el dominio si no que se restringió la simulación a una zona acotada por la cantidad de valores involucrados que haría más difícil el análisis de resultados posterior.

En las figuras siguientes se muestran también los histogramas y las estadísticas básicas de la media y las varianzas de las 100 realizaciones de la variable gaussiana de CuT y gaussiana CuS. (Histograma de medias a la izquierda e histograma de varianzas a la derecha).

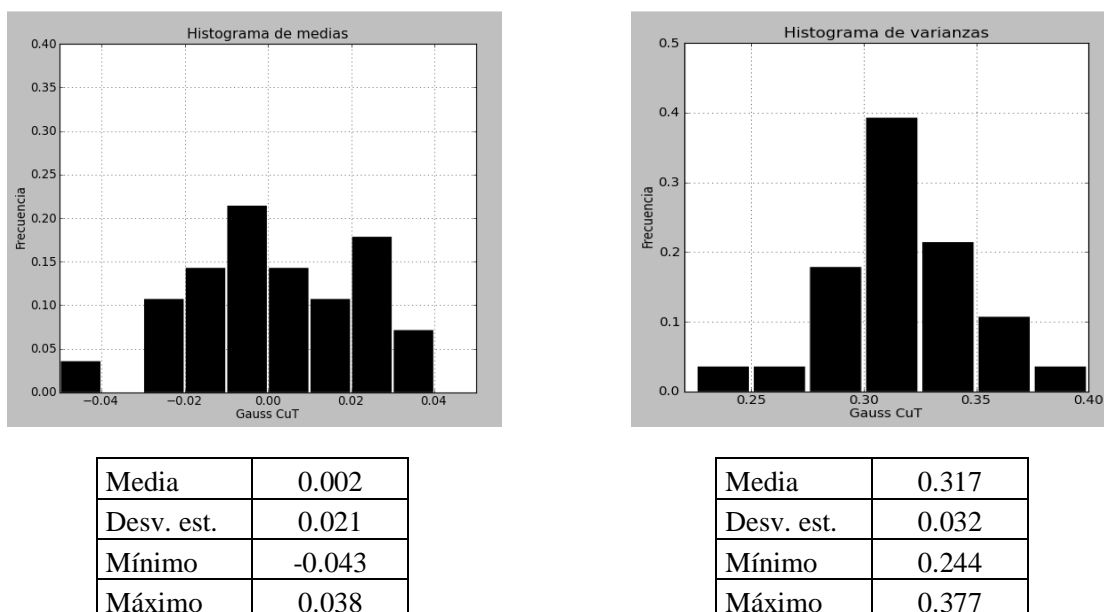


Figura 45 - Histograma y estadísticas básicas de las 100 realizaciones, variable gaussiana de CuT.

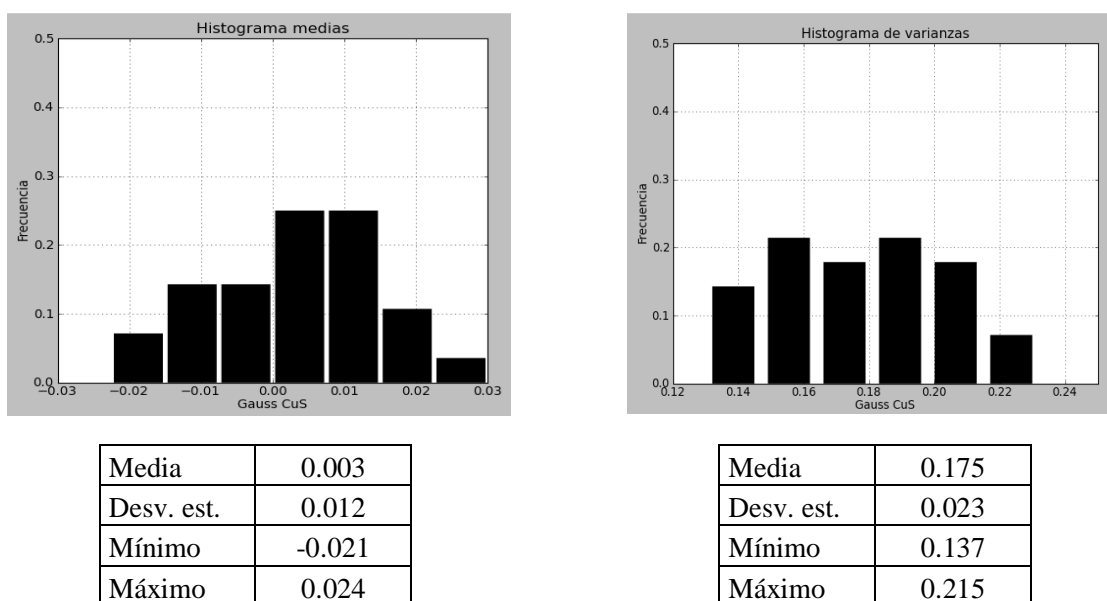


Figura 46 - Histograma y estadísticas básicas de las 100 realizaciones, variable gaussiana de CuS.

En concordancia con las figuras anteriores, se ve que la media del histograma de medias de las 100 realizaciones es bastante cercana a 0 tanto en las simulaciones de la variable CuT como en la variable CuS, sin embargo, la media de las desviaciones estándar es bastante distinta a 1, por lo explicado anteriormente.

6.6 Transformación de vuelta a los datos originales.

La transformación de vuelta a los valores originales se lleva a cabo usando las tablas de transformación, las que asocian a cada valor gaussiano un valor de la variable original, en el caso del CuS es un poco más complicado ya que la transformación a la variable gaussiana se hizo por clases por lo que la transformación inversa debe seguir el mismo procedimiento. La transformación de vuelta se debe hacer para cada realización y para cada una de las variables involucradas.

Si un valor gaussiano simulado no está en la tabla de transformación debe interpolarse según los siguientes criterios.

- 1) Si el valor gaussiano está entre dos valores de la tabla de transformación entonces se usa el modelo lineal:

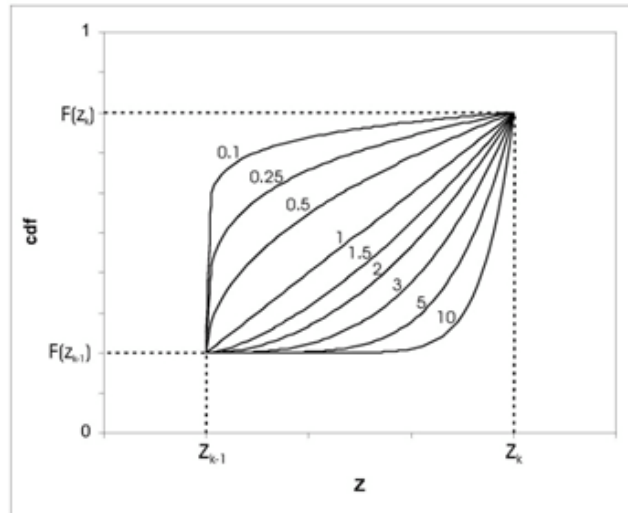


Figura 47 - Transformación de vuelta, interpolación lineal.

$$F(z) = F^*(z_{k-1}) + \left(\frac{z - z_{k-1}}{z_k - z_{k-1}} \right)^w \cdot (F^*(z_k) - F^*(z_{k-1})) \quad \forall z \in (z_{k-1}, z_k)$$

Donde el valor de w es 1 (lineal).

- 2) Si el valor gaussiano simulado es mayor que el valor gaussiano más alto de la tabla de transformación, se usa el modelo hiperbólico para extrapolar:

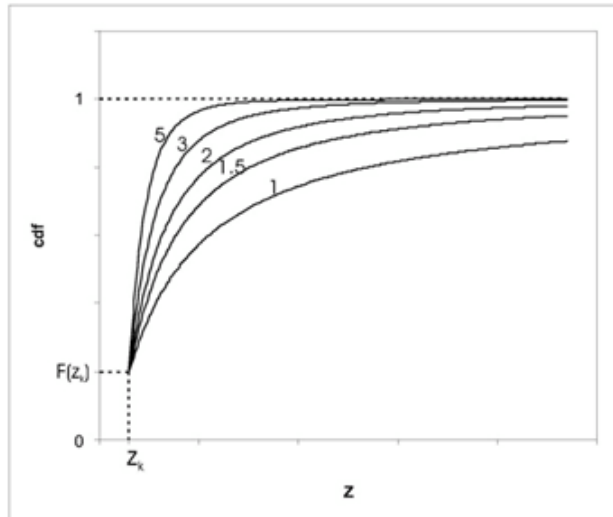


Figura 48 - Transformación de vuelta, extrapolación hiperbólica.

$$F(z) = 1 - \frac{z_k^w \cdot (1 - F^*(z_k))}{z^w}$$

El factor w recomendado en la literatura es de 1.5, que es el valor que se usó en este estudio. Se debe mencionar que este factor es muy importante ya que pequeñas diferencias en él afectan de manera importante los resultados obtenidos

Usando estos criterios se puede entonces hacer la transformación de vuelta a la variable original. En las siguientes secciones ya se estará trabajando con las variables originales, es decir, con CuT y CuS en los distintos casos de estudio.

6.7 Validación de los resultados en las variables originales.

6.7.2 Resultados de la simulación – Soporte puntual.

En las siguientes figuras se muestran los histogramas y estadísticas básicas de 3 realizaciones para las variables CuT y CuS resultantes de la transformación de vuelta.

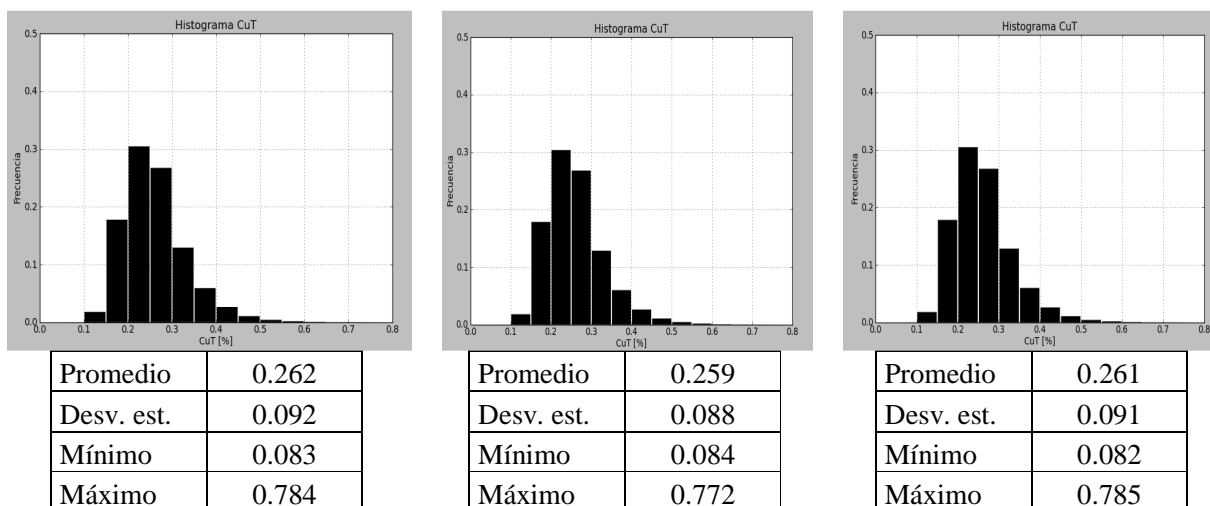


Figura 49 - Histogramas y estadísticas básicas de 3 realizaciones de la variable CuT a soporte puntual.

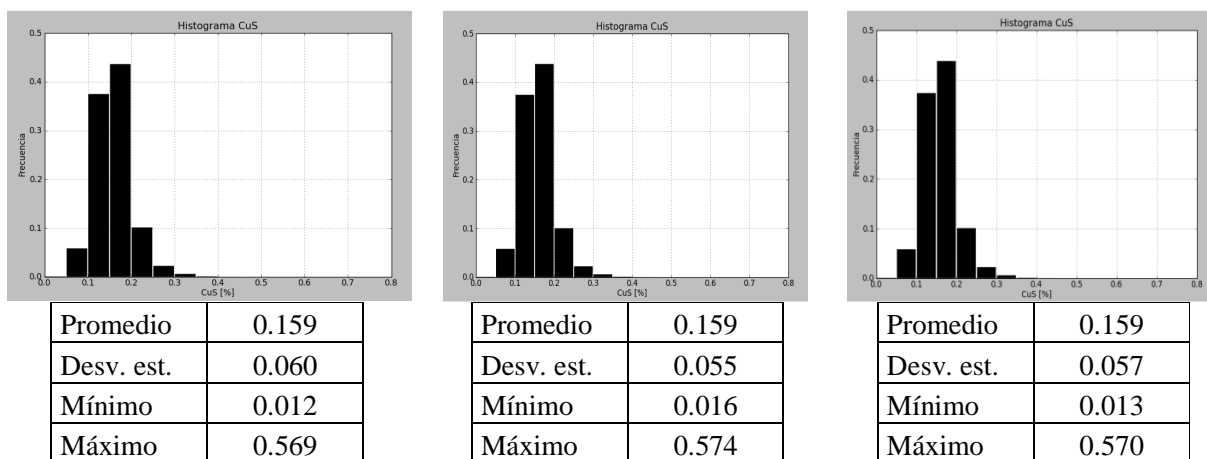


Figura 50 - Histogramas y estadísticas básicas de 3 realizaciones de la variable gaussiana CuS a soporte puntual.

En el anexo E se pueden ver las realizaciones 3 realizaciones asociadas a las bases de datos sesgadas. En la figura 51 se ven los mapas de las respectivas realizaciones para la variable CuS, donde no se aprecian grandes diferencias entre cada una de ellas.

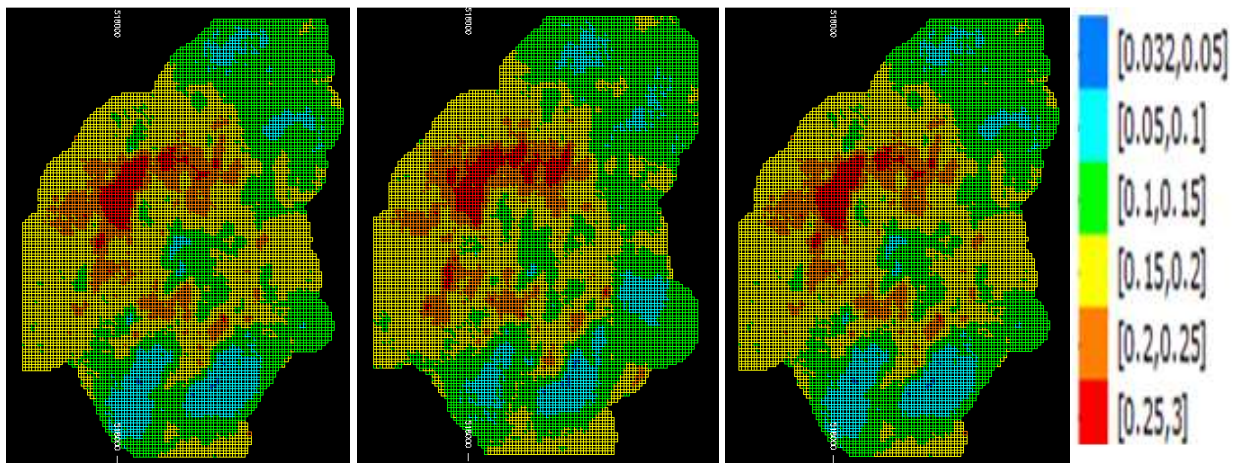


Figura 51 - Visualización grafica de 3 realizaciones, ley de CuS.

En las siguientes figuras se muestran gráficos cuantil a cuantil comparando las distribuciones de los datos resultantes de la transformación de vuelta de los valores simulados y los datos originales, es decir, la base de datos de los cerca de 15000 compositos iniciales.

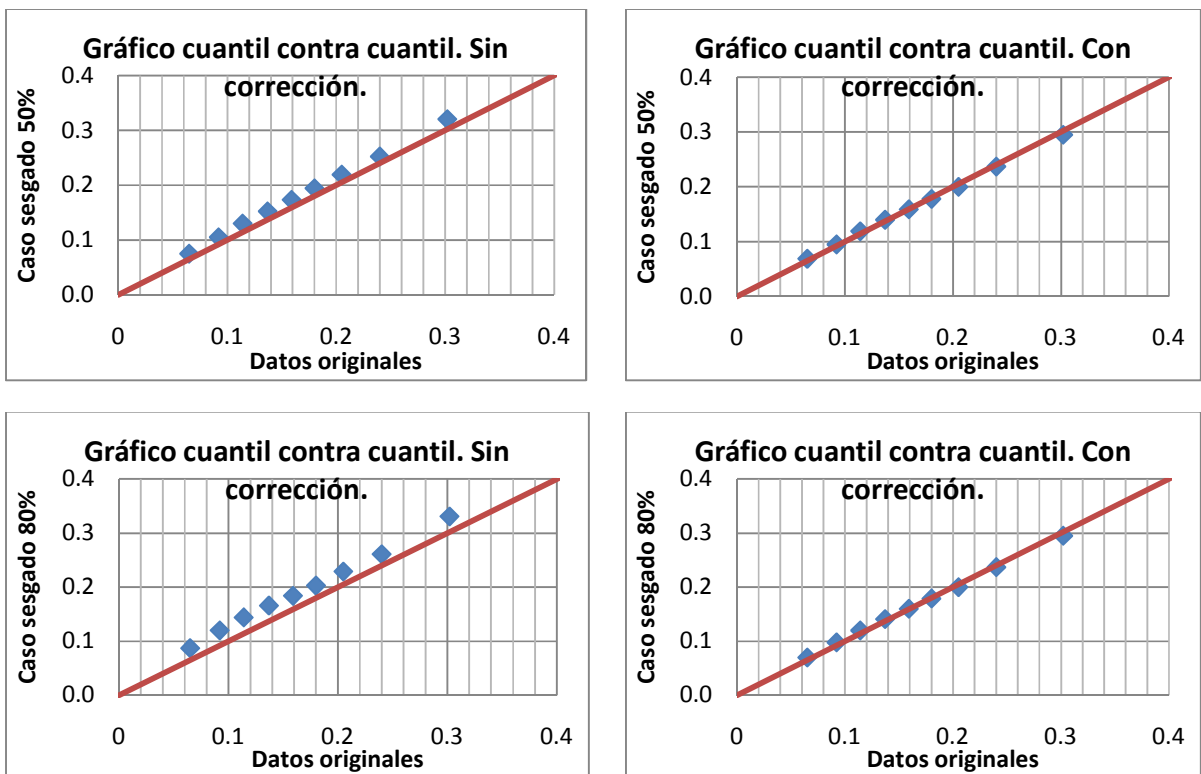


Figura 52 - Comparación entre las distribuciones de valores y simulados y los datos originales. Arriba comparando con los resultados del caso sesgado 50% y abajo comparando con el caso sesgado 80%.

Se verifica que la relación $CuT \geq CuS$ se cumpla con los datos simulados a soporte puntual para cada una de las realizaciones, de un análisis de los datos solo 20 bloques de cerca de 3.000.000 no cumplen con esta restricción. Para corregir este hecho se redefinen los valores de CuS como el valor de CuT en el mismo sitio ($CuS = CuT$). Por problemas de visualización solo se presentaran las nubes de correlación para soporte de bloques debido al menor número de datos involucrados.

6.7.2 Resultados de la simulación – Soporte de bloques.

En esta sección se presentan los mismos resultados pero a soporte de bloques. Sin embargo, solo se muestran 3 realizaciones de las variables CuT y CuS debido a que en la sección siguiente el análisis se realiza en el soporte de bloques, por lo que se presentarían muchos resultados redundantes con esta sección.

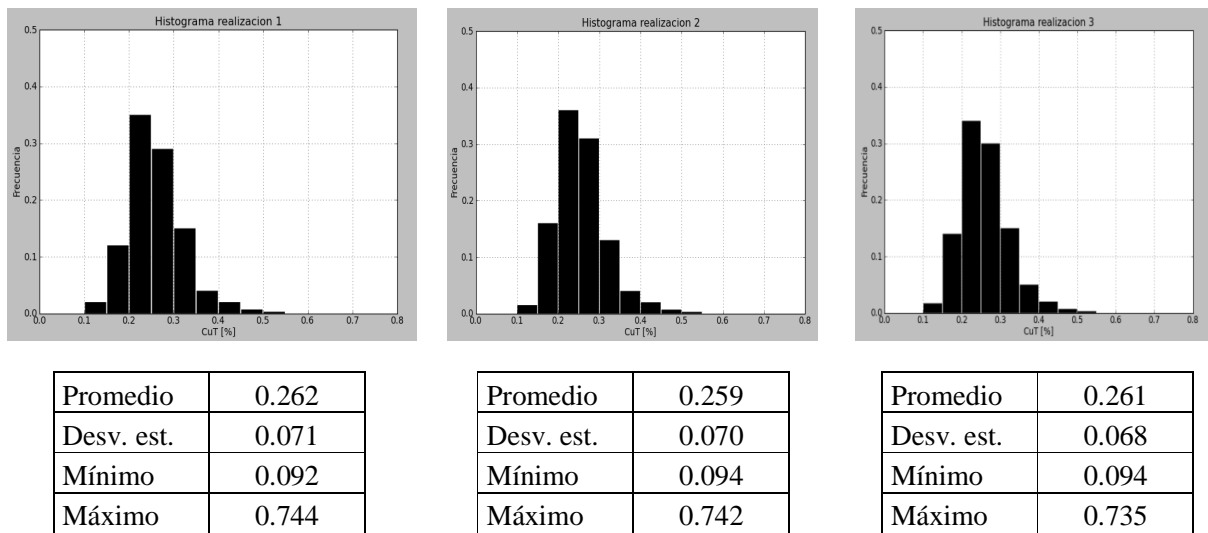


Figura 53 - Histogramas y estadísticas básicas de 3 realizaciones de la variable CuT a soporte de bloques.

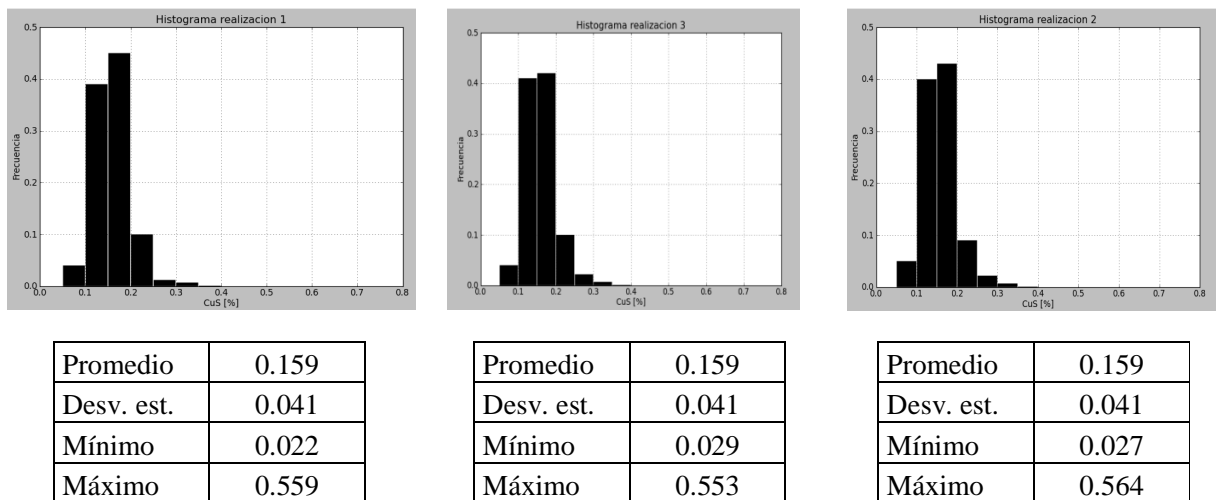


Figura 54 - Histogramas y estadísticas básicas de 3 realizaciones de la variable CuS a soporte de bloques.

En las siguientes figuras se muestra la nube de correlación entre CuT y CuS en las distintas bases de datos para el promedio de 100 realizaciones en cada caso, aunque la restricción $CuT \geq CuS$ se confirmó para cada una de las realizaciones (arriba caso base, abajo izquierda caso sesgado 50% y abajo derecha caso sesgado 80%).

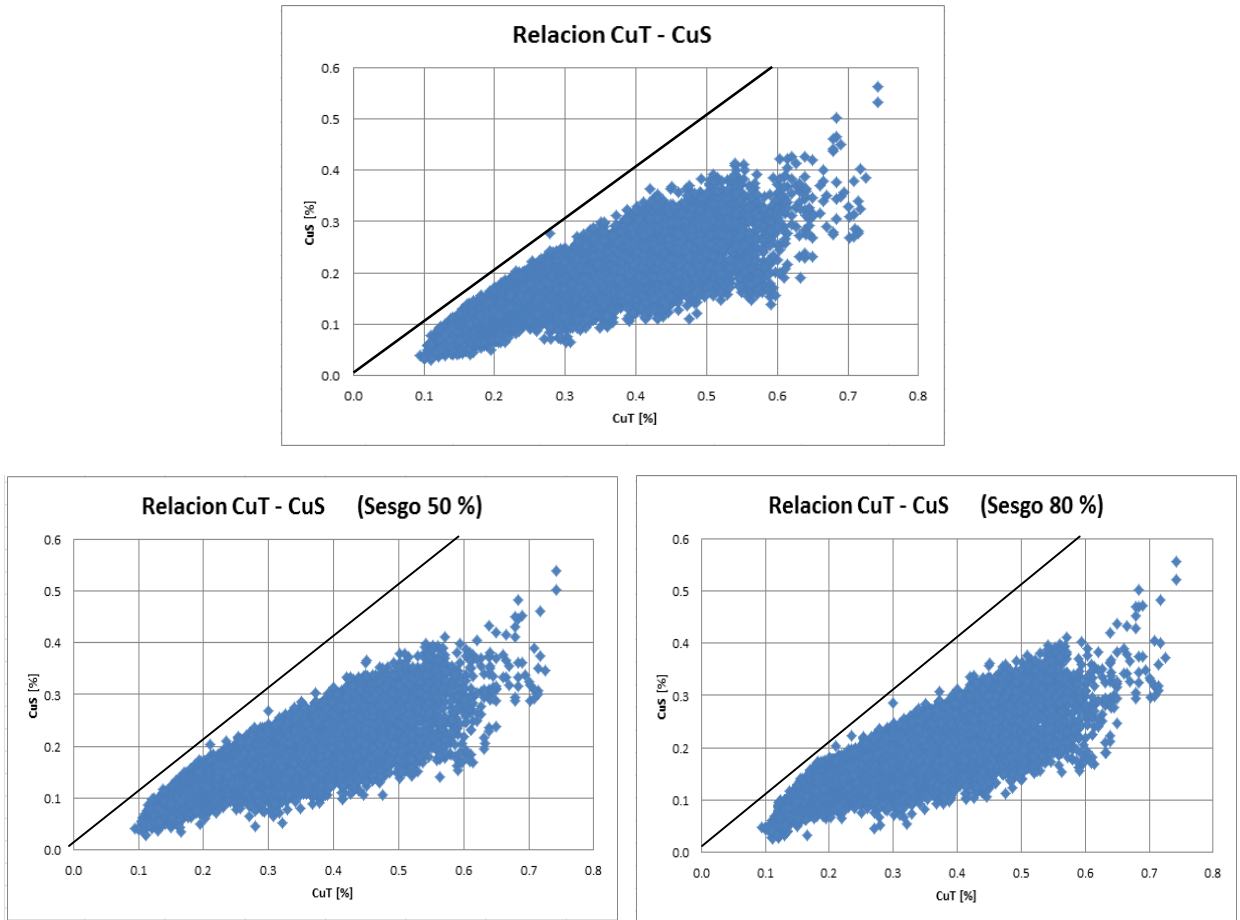


Figura 55 - Nubes de correlación CuT - CuS para los distintos casos de estudio.

7. ANÁLISIS DE LOS RESULTADOS.

7.1 Evaluación de las diferencias debido al grado de sesgo.

Después de efectuar la simulación puntual, se rebloqueó a un tamaño de 20 x 20 x 15 [m]. Asumiendo una densidad de 2.7 [ton/m³] se calcularon los tonelajes y leyes medias en cada caso, en las siguientes tablas se muestran los resultados comparativos para los tres casos a soporte de bloques y como el promedio de 100 realizaciones.

	CuT	CuS		
		Caso base	Sesgo 50%	Sesgo 80%
Media	0.260	0.159	0.157	0.158
Desv. Estándar	0.070	0.041	0.038	0.036
Mínimo	0.095	0.032	0.028	0.026
Máximo	0.742	0.562	0.539	0.557

Tabla 11 - Comparación de las estadísticas básicas reportadas de los diferentes casos de estudio.

A partir de las estadísticas básicas no se puede apreciar una gran diferencia entre cada uno de los casos.

En las siguientes figuras se muestran los histogramas de los valores simulados de la variable cobre soluble (arriba caso base, izquierda caso sesgado 50%, derecha caso sesgado 80%).

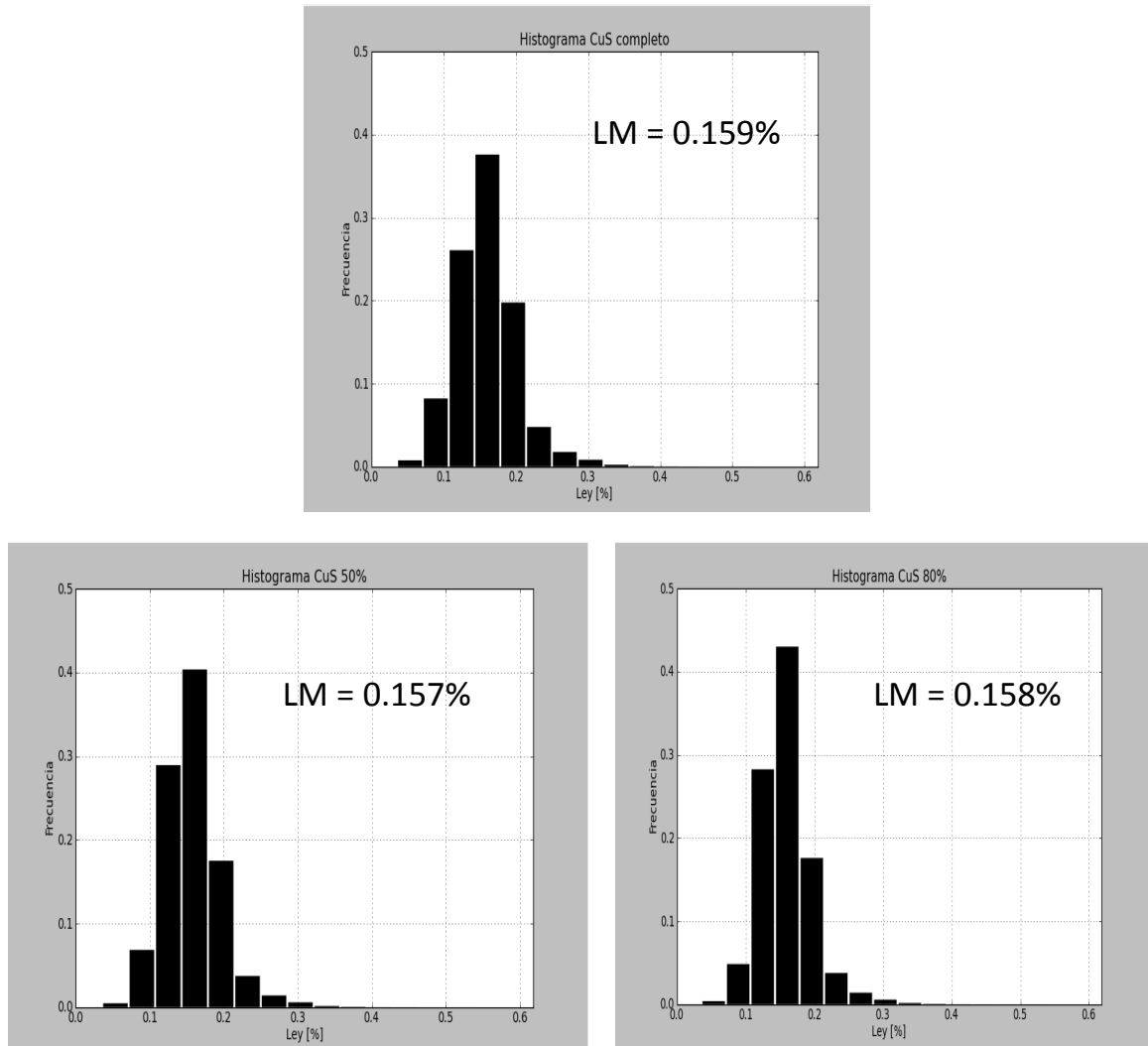


Figura 56 - Comparación de los histogramas de las estimaciones para cada uno de los casos de estudio.

Como era de esperar el algoritmo de simulación restituye la distribución original de la variable simulada y la corrección de la distribución sesgada cumple su objetivo, los valores simulados tienen una distribución bastante parecida a la distribución de los valores simulados sin sesgo (caso base).

En las siguientes figuras se comparan los gráficos cuantil contra cuantil, comparando el caso base con los casos sesgados.

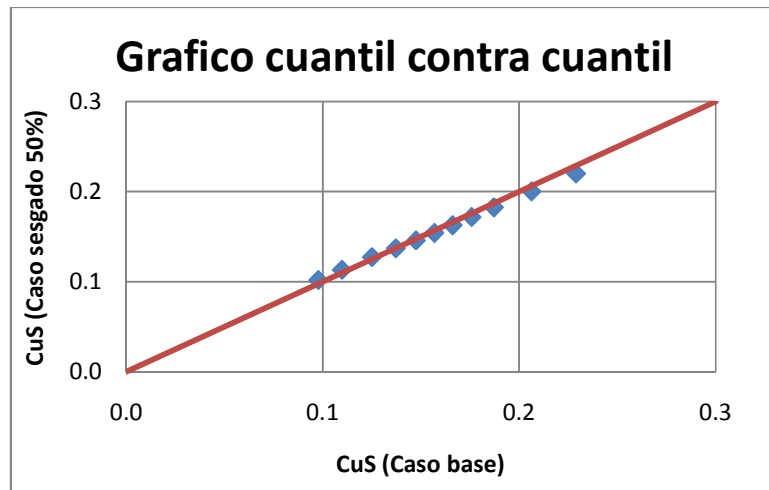


Figura 57 - Gráfico cuantil contra cuantil. Caso base y caso sesgado 50%.

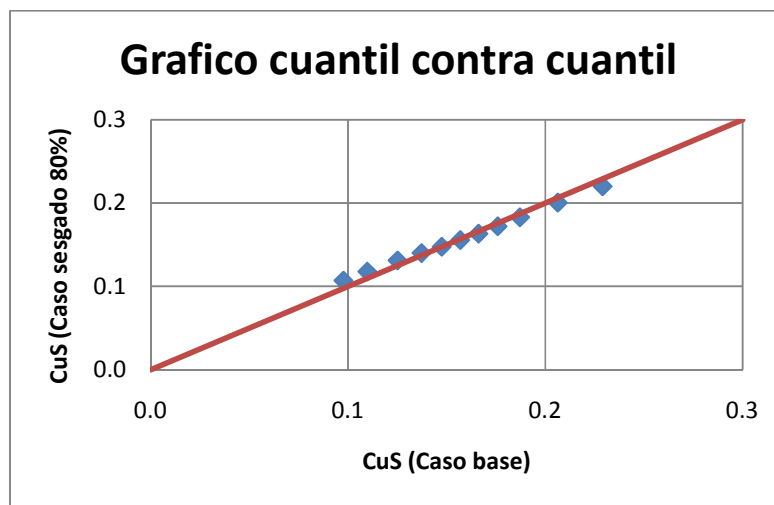


Figura 58 - Gráfico cuantil contra cuantil. Caso base y caso sesgado 80%.

De los gráficos anteriores se puede nuevamente concluir que los valores estimados de los casos sesgados tienen la misma distribución que los valores estimados del caso base, por lo tanto, se confirma que la corrección es exitosa.

7.1.1 Evaluación gráfica

Promediando las 100 simulaciones y haciendo un despliegue de estas, se pueden comparar gráficamente como se muestra en las siguientes figuras.

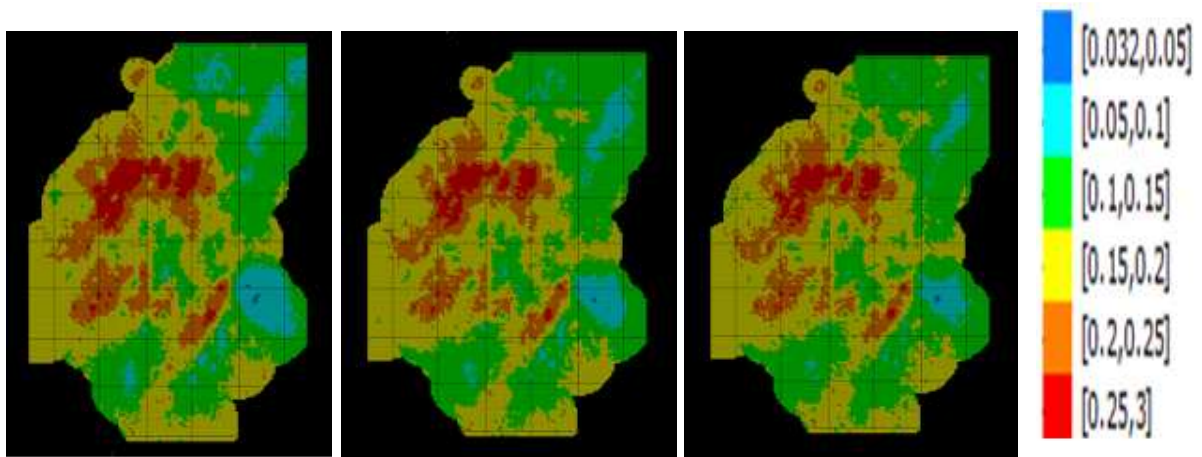


Figura 59 - Comparación gráfica, leyes de CuS. Caso base (izq), caso sesgado 50% (centro) y caso sesgado 80% (der).

De la figura anterior se aprecia que no existen mayores diferencias entre los distintos casos de estudio, lo que confirma aún más que la corrección fue exitosa. También se puede evaluar viendo la probabilidad de superar cierta ley de corte en cada uno de los casos, en la cual se aprecia lo mencionado antes, como se muestra en las siguientes figuras.

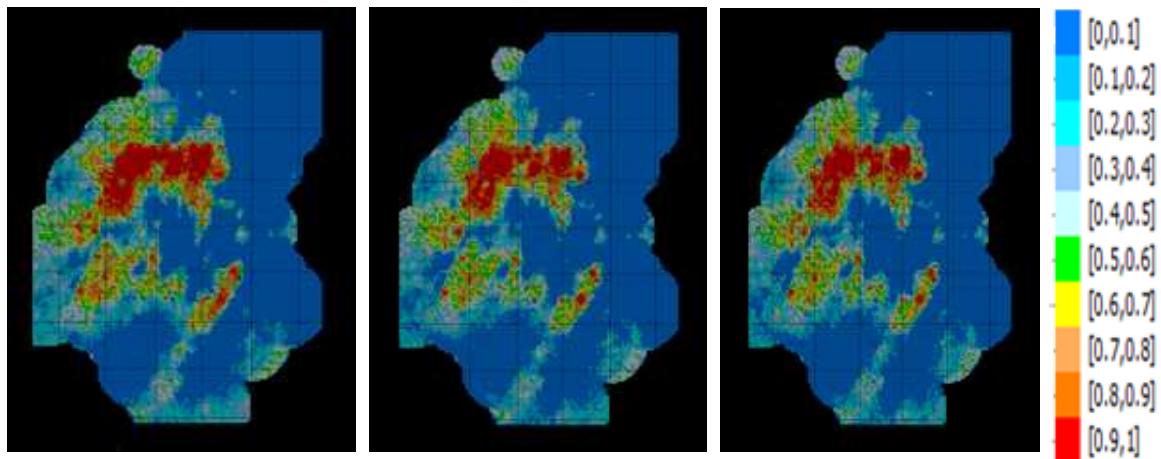


Figura 60 - Comparación gráfica, probabilidad de superar la ley de corte 0,2% CuS. Caso base (izq), caso sesgado 50% (centro) y caso sesgado 80% (der).

En las siguientes figuras se ven las curvas tonelaje ley de CuS para cada uno de los casos.

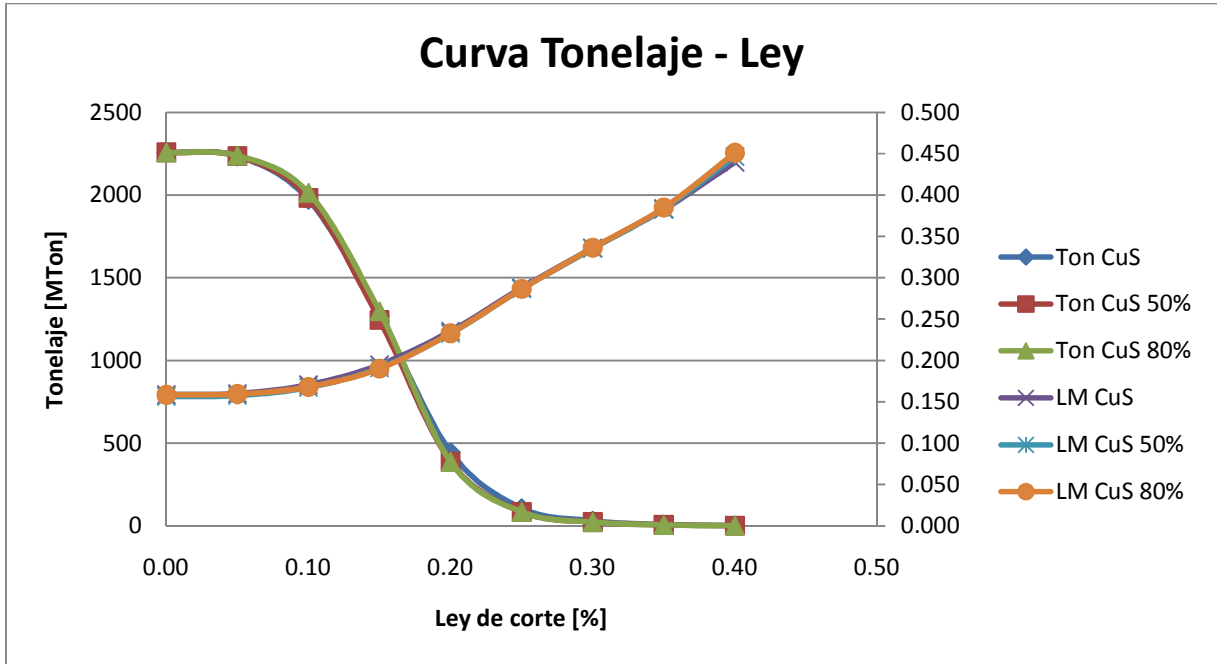


Figura 61 - Comparación de las curvas tonelaje ley para cada uno de los casos de estudio.

A primera vista no se aprecian grandes diferencias en los tonelajes y leyes medias comparando distintas leyes de corte. A modo de ejemplo se eligen 2 leyes de corte para comparar en mayor detalle, una de estas sobre la ley usada para sesgar la base de datos (0.2%) y la otra por encima de esta. La segunda fila en cada categoría corresponde a la variación respecto del caso base.

Ley de corte 0.1%				
	Unidades	CuS	CuS 50	CuS 80
Tonelaje	[Mton]	1972	1983	2014
	[%]	0	0.59	2.16
Ley media	[%]	0.170	0.167	0.168
	[%]	0	-1.74	-1.43

Tabla 12 - Comparación de los tonelajes reportados por cada uno de los casos de estudio para una ley de corte de 0.1% CuS.

Las variaciones que se aprecian entre los distintos casos de estudio son bastante pequeñas (menores a un 3% en tonelaje y menores a un 2% en la ley media). Se aprecia cierta tendencia a reportar tonelajes mayores y menores leyes medias en los casos sesgados respecto al caso base.

En la siguiente tabla se muestran los mismos resultados pero para una ley de corte superior a la ley del sesgo.

Ley de corte 0.35%				
	Unidades	CuS	CuS 50	CuS 80
Tonelaje	[Mton]	7	6	6
	[%]	0	-0.14	-0.14
Ley media	[%]	0.383	0.384	0.385
	[%]	0	0.2	0.2

Tabla 13 - Comparación de los tonelajes y leyes medias reportados por cada uno de los casos de estudio para una ley de corte de 0.35% CuS.

En este caso también se aprecian diferencias menores para los tonelajes y leyes medias reportadas, también en este caso se aprecia cierta tendencia pero al revés del caso anterior, los tonelajes de los casos sesgados son menores al caso base mientras que las leyes medias son superiores a los del caso base.

7.1.2 *Análisis de incertidumbre.*

El siguiente paso es ver la incertidumbre en la estimación tanto del tonelaje como la ley media para cierta ley de corte, como en las etapas anteriores también se usarán 2 leyes de corte, una por encima de la ley de sesgo y otra por debajo. Pero no se considerará solo el promedio de las realizaciones sino más bien se considerarán todas ellas para realizar el análisis.

Los histogramas de tonelajes para una ley de corte de 0,1% se presentan en las siguientes figuras.

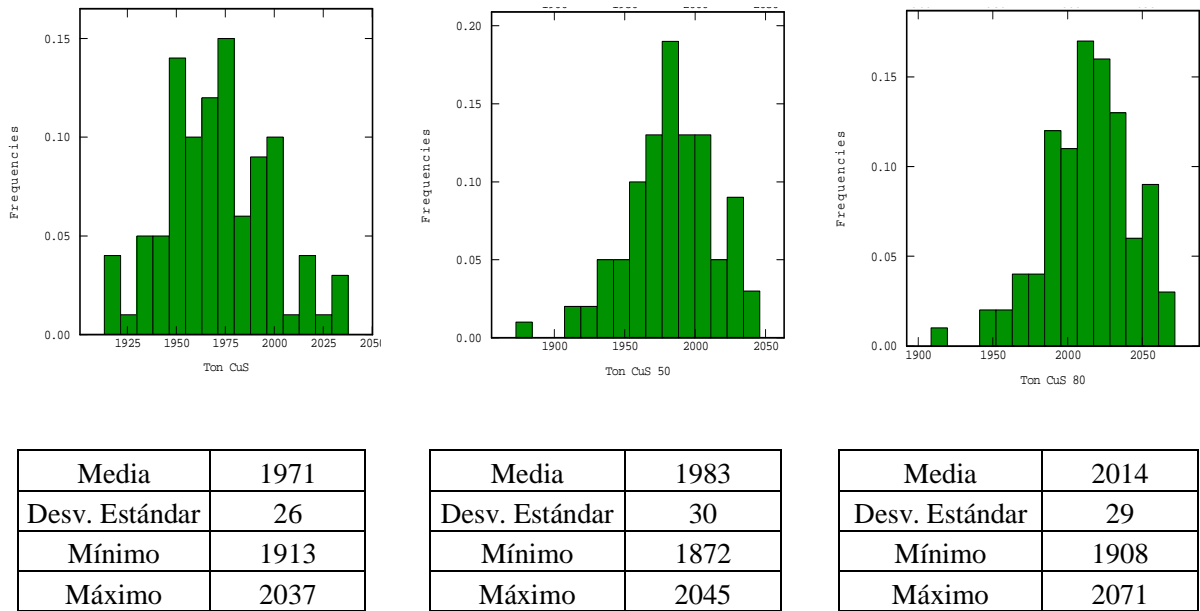


Figura 62 – Histogramas de tonelajes para una ley de corte 0.1 % (de derecha a izquierda, caso base, sesgo del 50%, sesgo del 80%).

Los histogramas de tonelajes para una ley de corte de 0,35% se presentan en las siguientes figuras.

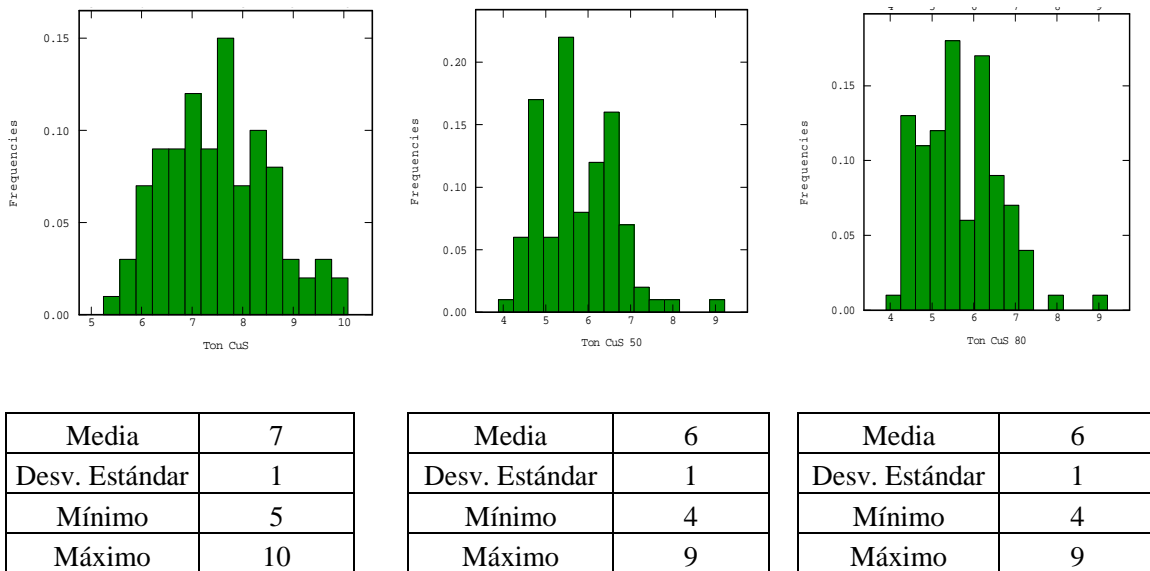


Figura 63 - Histogramas de tonelajes para una ley de corte 0.1 % (de derecha a izquierda, caso base, sesgo del 50%, sesgo del 80%).

Se confirma entonces el comportamiento identificado en la sección anterior, para una ley de corte bajo la ley del sesgo en la medida en que el sesgo es más fuerte el tonelaje reportado sobre esa ley de corte es mayor.

Por el contrario para una ley de corte sobre la ley del sesgo, los tonelajes reportados son menores en la medida en que el sesgo es más severo (respecto al caso base nuevamente), además se puede ver que la varianza de los tonelajes reportados en los casos sesgados es mayor que las del caso base.

En la siguiente sección se analiza la varianza de las realizaciones en cada uno de los casos de estudio.

7.1.3 Análisis de varianzas.

En las siguientes figuras se muestran el histograma de varianzas de cada realización para cada uno de los casos de estudio.

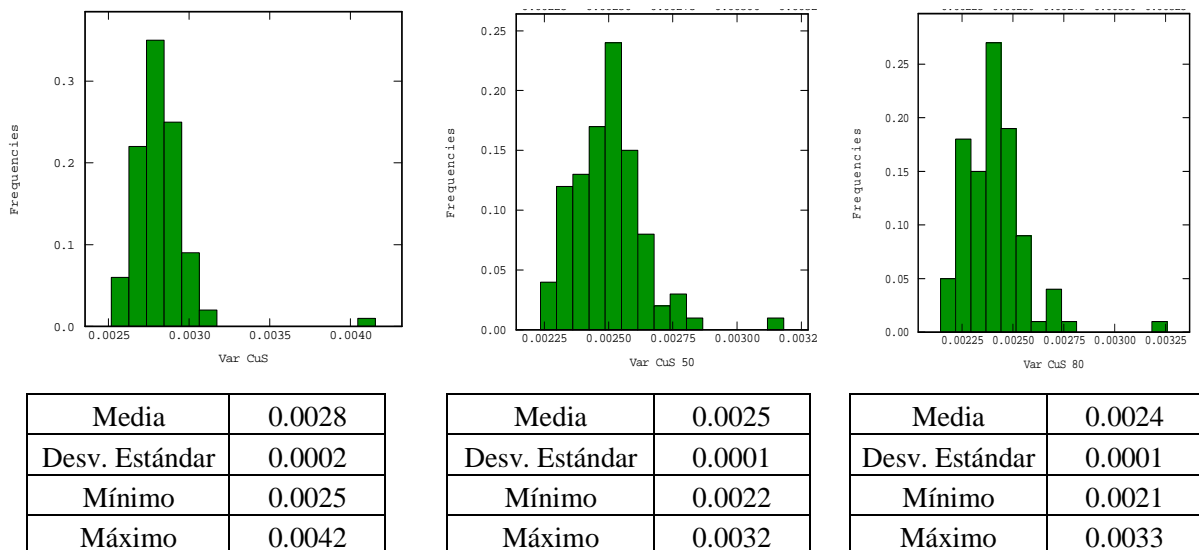


Figura 64 - Comparaciones de los histogramas de las varianzas de cada realización en cada uno de los casos de estudio.

De los gráficos de la figura anterior se puede concluir que los casos sesgados tienen una menor varianza en promedio en cada una de sus realizaciones, además el rango de varianzas esta mucho más acotado.

7.2 Comparación utilizando la distribución sesgada de CuS en la estimación.

La comparación con respecto al caso base demostró que la corrección es útil intentando restablecer el verdadero comportamiento de la variable en estudio, sin embargo, es útil también comparar el resultado de la corrección con un caso de estudio donde no se hubiese usado la corrección, es decir, si se hubiese utilizado la distribución sesgada para hacer la transformación a la variable gaussiana. El resultado de esto se puede ver en las siguientes figuras, primero para la base sesgada 50% y después para la base de datos sesgada 80%.

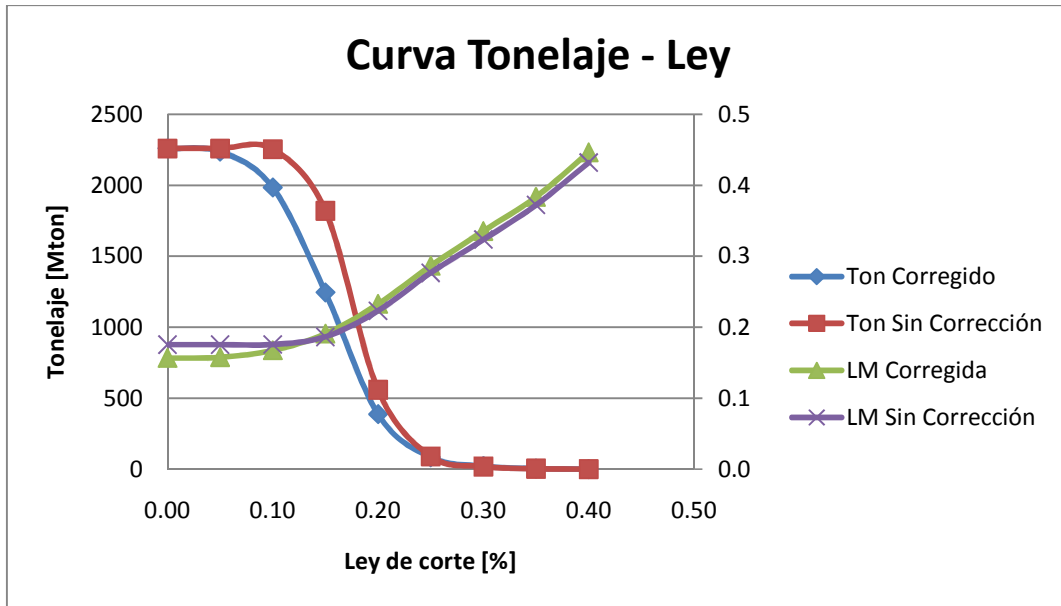


Figura 65 - Comparación de las curvas tonelaje ley, usando la corrección (caso sesgado 50%) y sin usar la corrección.

En el gráfico anterior la curva azul corresponde al tonelaje estimado usando la corrección mientras que la curva roja es el tonelaje reportado usando la distribución sesgada de CuS. Análogamente la curva verde representa la ley media obtenida mediante el uso de la corrección y la curva morada la ley media obtenida sin usar la corrección. Se nota claramente que existe una sobrestimación del yacimiento tanto en tonelaje como en ley media para cualquier ley de corte.

Para cada ley de corte las diferencias por ejemplo en tonelaje reportado pueden llegar casi al 50% entre el caso mal hecho y el caso donde se usó la corrección. En la siguiente figura se muestra la misma comparación pero usando el caso sesgado 80%.

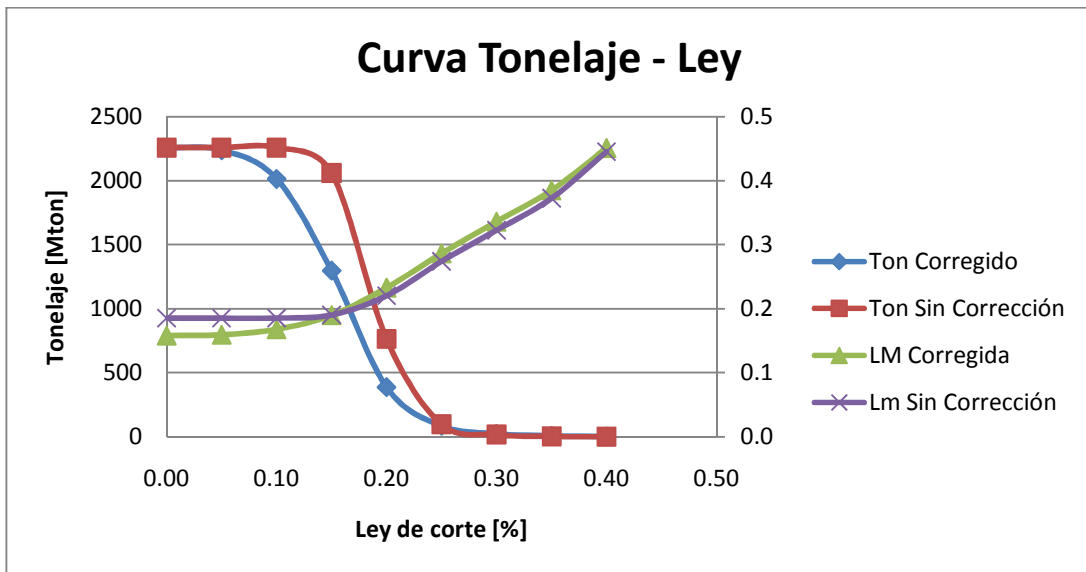


Figura 66 - Comparación de las curvas tonelaje ley, usando la corrección (caso sesgado 80%) y sin usar la corrección.

Se puede ver que lo mismo sucede para el caso de 80 % de sesgo, en este caso las diferencias son muy grandes. Por ejemplo, para una ley de corte de 0.2% el caso con corrección informa 386 [Mton] mientras que el caso sin corrección informa un tonelaje de 764 [Mton], es decir una diferencia de prácticamente un 100%.

En la siguiente figura se muestra los gráficos cuantil contra cuantil comparando las distribuciones de CuS para los casos con corrección y sin corrección.

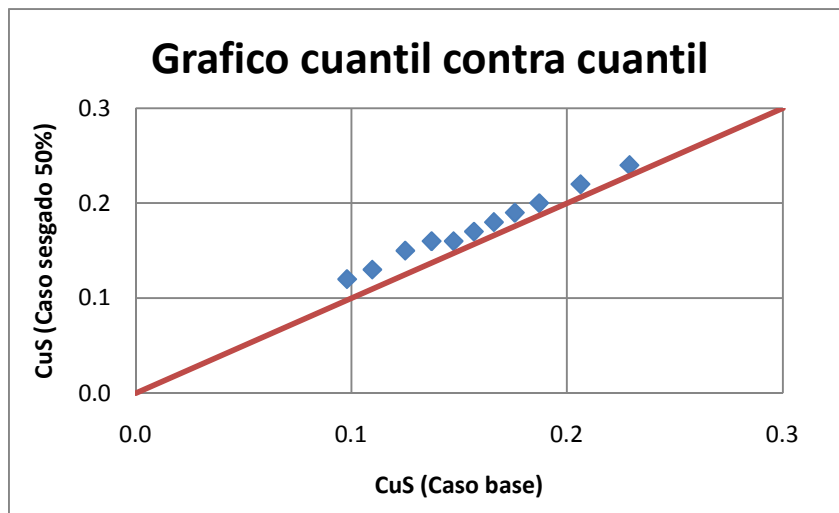


Figura 67 - Grafico cuantil contra cuantil. Caso base y caso mal hecho sesgado 50%.

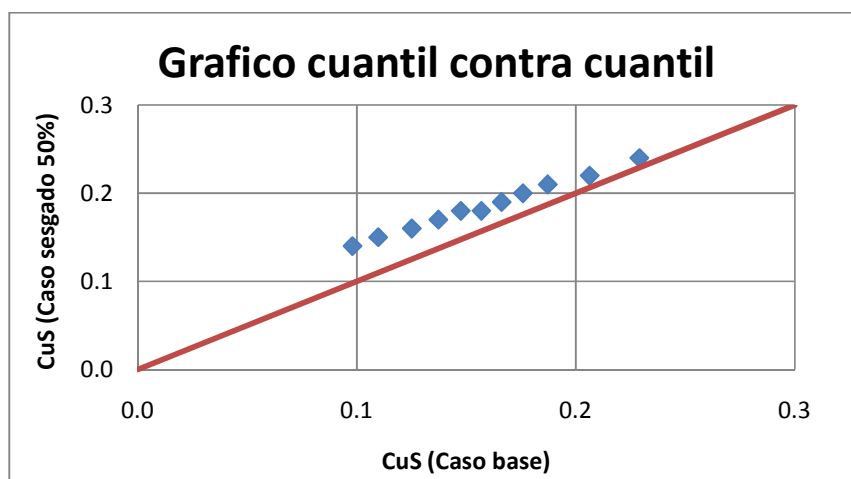


Figura 68 - Gráfico cuantil contra cuantil. Caso base y caso mal hecho sesgado 80%.

Es evidente entonces que la distribución de valores estimados en los casos donde no se corrigió la distribución de CuS es muy diferente a la distribución del caso base. Claramente la distribución está sesgada hacia los valores altos, sesgo que se hace más evidente cuando el número de datos sin información de CuS es más grande (se notan mayores diferencias entre el caso sesgado 80% y el caso base, que entre el caso 50% y el caso base). Esto también se refleja en las estadísticas básicas de los casos de estudio analizados.

	Caso base	Casos sin corrección	
		<i>Sesgo 50 %</i>	<i>Sesgo 80 %</i>
Media	0.158	0.175	0.185
Varianza	0.0017	0.0013	0.0011
Mínimo	0.03	0.06	0.07
Máximo	0.56	0.55	0.54

Tabla 14 - Estadísticas básicas de los casos sin corrección y el caso base.

Como se ve la ley media reportada es mucho mayor que la mostrada en el caso base o en los casos en los que sí se hizo la corrección.

7.3 Aplicación del suavizamiento mediante polinomios de Bernstein.

Como se mencionó anteriormente, sería recomendable poder discretizar las distribuciones de probabilidad de manera de implementar la corrección de una manera más precisa (con clases cada vez más pequeñas), en este sentido poder usar una función suavizada del histograma experimental permite poner en práctica esta idea, ya que permite discretizar tal función infinitamente. Sin embargo, como el suavizamiento no interpola los puntos de control, solo mantiene ciertas características del histograma original.

El objetivo es suavizar la distribución bivariable de CuT-CuS, a partir de ella será posible obtener las distribuciones marginales suavizadas simplemente integrando la distribución bivariable. De acuerdo al formulismo presentado anteriormente lo primero es elegir el grado de discretización de la distribución bivariable para generar los puntos de control con los cuales se hará el suavizamiento.

Para realizar el suavizamiento se eligieron 10 clases entre 0 y 1 (100 clases para la distribución bivariable) para generar los puntos de control, después de esto se realizó el suavizamiento.

En la siguiente figura se muestra la comparación entre la distribución condicional de CuT obtenida normalmente y el generado producto del suavizamiento.

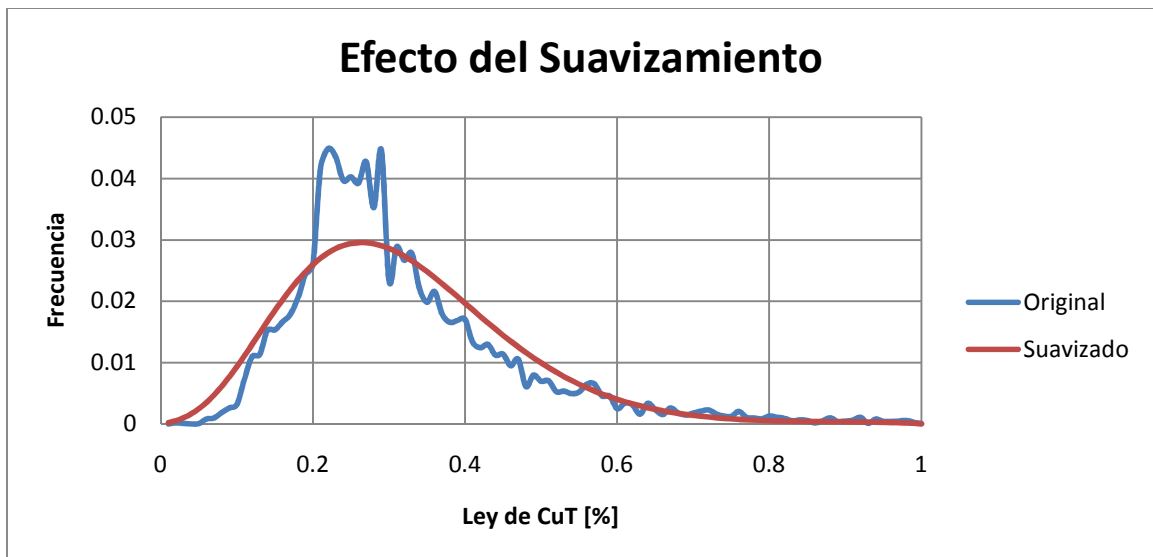


Figura 69 - Comparación entre el histograma experimental y el histograma suavizado.

Queda en evidencia que al discretizar la distribución demasiado muchas clases tienen un valor 0 como frecuencia relativa, en este caso de estudio principalmente en la zona de bajas leyes que es donde se hizo el sesgo, este hecho imposibilita la implementación de la corrección debido

a que el factor de corrección $\frac{f_{CuT}^2}{f_{CuT}^1}$ tiene un valor 0 en el denominador (recordar que f_{CuT}^1 es el valor de la frecuencia de la distribución condicional de CuT).

Aunque en forma general el suavizamiento representa la distribución “real” de la variable, localmente se producen ciertas diferencias, como en la zona de más alta probabilidad o las colas de la distribución. Este hecho se puede modificar cambiando el grado de discretización de la distribución bivariable usado para generar el suavizamiento, es decir, cambiar el número de puntos de control.

En las siguientes figuras se ve el efecto del suavizamiento usando un número diferente de puntos de control.

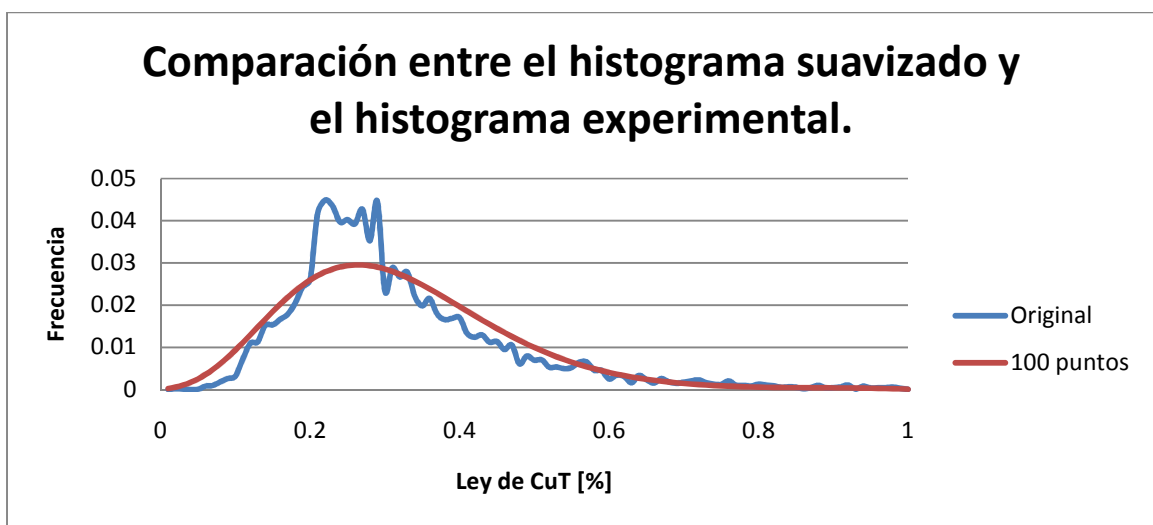


Figura 70 - Histograma suavizado, 100 puntos de control.

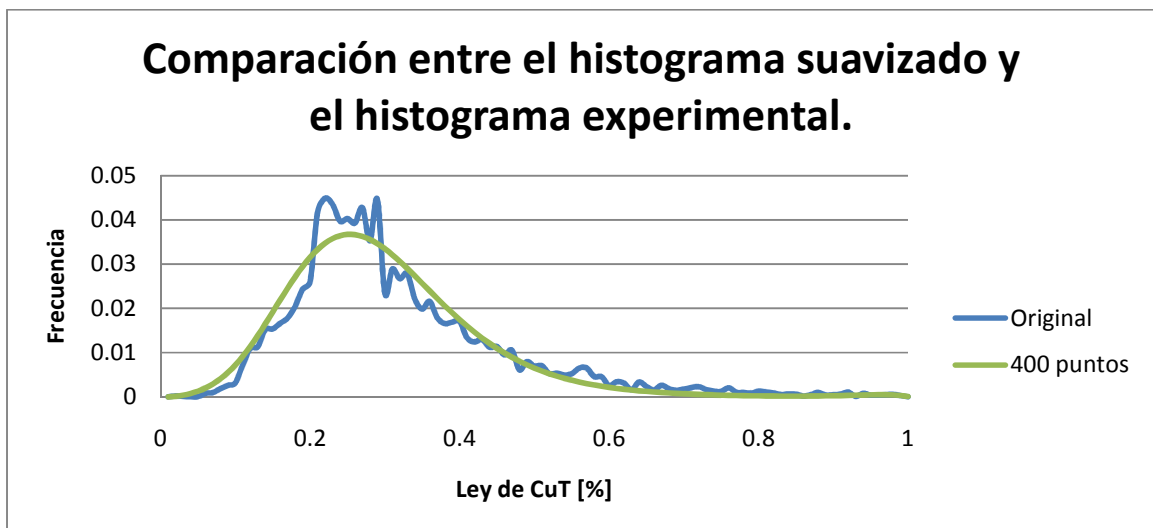


Figura 71 - Histograma suavizado, 400 puntos de control.

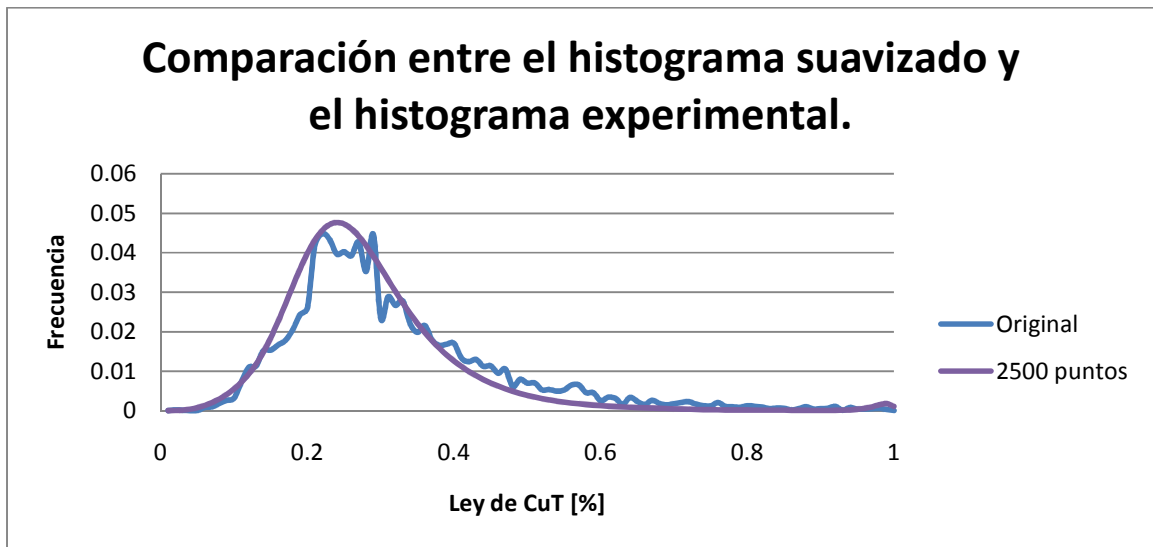


Figura 72 - Histograma suavizado, 2500 puntos de control.

Como se ve del gráfico anterior aunque se aumente el número de puntos de control con los cuales construir la función suavizada no necesariamente se obtendrá un mejor resultado. En este caso depende del usuario decidir qué curva ajusta mejor según sus propios requerimientos. Por ejemplo, la mayoría de las curvas sobreestima la frecuencia de las leyes más bajas, en cuanto a la reproducción de las leyes altas la primera curva (rojo) sobreestima la frecuencia de altas leyes, la verde ajusta mejor y la morada subestima esta frecuencia.

Para este caso de estudio se eligió la curva verde (con 400 puntos de control) como la más representativa de la distribución original, a partir de esta se implementó la corrección.

7.4 Resultados de la simulación con suavizamiento.

En la siguiente tabla se comparan los resultados del suavizamiento con respecto al caso base.

	Caso base	Casos suavizados	
		Sesgo 50 %	Sesgo 80 %
Media	0.158	0.165	0.173
Varianza	0.0017	0.0014	0.0013
Mínimo	0.03	0.04	0.04
Máximo	0.56	0.63	0.64

Tabla 15 - Estadísticas básicas, comparación entre el caso base y los casos suavizados.

Como se ve los casos con suavizamiento tienden a sobrestimar los recursos del yacimiento, no tanto como en el caso sin corrección pero sí de manera significativa. La siguiente figura reafirma esta idea, donde se comparan las curvas tonelaje ley del caso base con respecto al caso suavizado con sesgo del 50%.

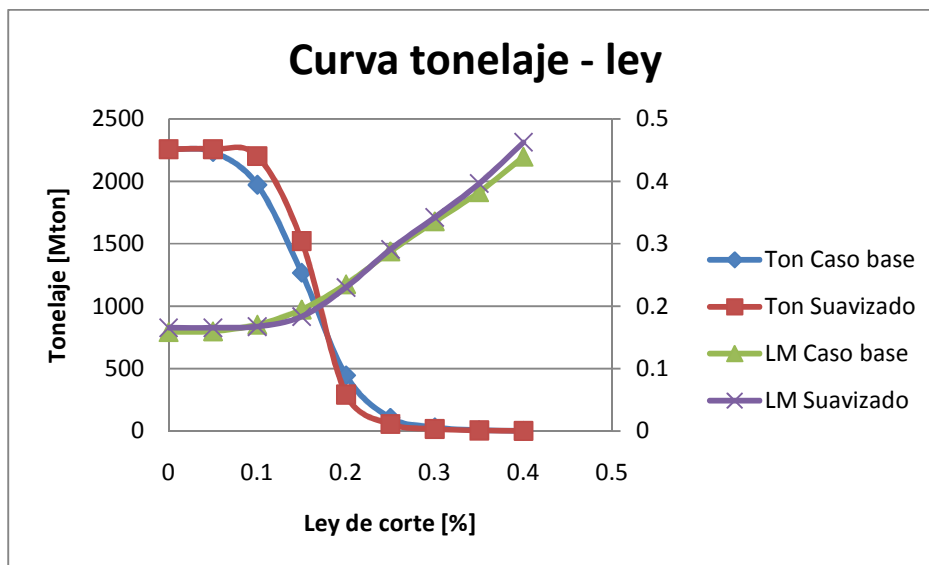


Figura 73 - Curva tonelaje - ley, comparación entre el caso base y el caso suavizado.

En este caso particular, el tonelaje informado bajo la ley del sesgo es subestimado mientras que para leyes superiores a la ley del sesgo el tonelaje es sobreestimado, sin embargo, esto depende fundamentalmente de la función de suavizamiento que se haya escogido como representativa del histograma experimental. Como esta función no ajusta perfectamente, se tienen ciertas imprecisiones en la corrección y, por ende, en la estimación.

8. DISCUSIÓN DE LOS RESULTADOS.

En términos generales la corrección cumplió con los objetivos planteados inicialmente, vale decir, generar una distribución representativa de la variable submuestreada que sirva para simular esta variable de manera correcta, ya que debido al sesgo introducido por el muestreo preferencial, los resultados obtenidos en la simulación no serían consistentes ni representativos de la realidad.

Se debe mencionar que se aprecia una gran diferencia entre simular sin usar y usando la corrección como se vio en la sección 7.2, donde se observan diferencias de hasta un 100 % en los recursos estimados. Este resultado reafirma la calidad de la solución propuesta

No obstante, la corrección no se puede aplicar si el sesgo es total bajo cierta ley. Por ejemplo, si se hubiese decidido no analizar ninguna muestra de cobre soluble si el cobre total está bajo 0.2%, entonces la corrección no se hubiese podido aplicar de ninguna forma, incluso el suavizamiento no hubiese sido una opción debido a que se está especulando acerca del comportamiento de la distribución en la zona sesgada.

Otro tema relacionado a esto es el grado de sesgo de la distribución. Como se ve, aunque la segunda base de datos tiene un fuerte sesgo bajo la ley de 0.2% solo deja sin información de CuS al 25 % de la población total. Como tema a discutir quedan las opciones de sesgar aún más la base de datos (a una ley mayor por ejemplo) y la opción de no restringir el sesgo a la parte de bajas leyes de la distribución. Sin embargo, dado la forma en que se realiza la corrección, los resultados deberían ser fundamentalmente los mismos, no importando el grado de sesgo ni en qué zona de la distribución se realicen.

Aunque la corrección resultó exitosa se debe considerar que la base de datos considerada es relativamente grande (cerca de 15000 datos), probablemente los resultados de la corrección resultarían menos favorables en bases de datos más pequeñas ya que la discretización necesaria sería potencialmente mucho menor (clases más grandes), para subsanar este hecho es necesario el suavizamiento de las distribuciones implicadas, principalmente las distribuciones condicionales a la presencia de CuS en la muestra. Sin embargo, el suavizamiento de la distribución no modela perfectamente la distribución real de las variables en estudio, lo que introduce errores e imprecisiones que afectarán más tarde en los resultados de la simulación. Debido a esto, se aprecian diferencias de hasta un 50% de los recursos estimados respecto al caso base entre los resultados usando la corrección con suavizamiento y sin él.

Por esto los resultados de la corrección usando suavizamiento dependen de que tan bien se ajuste la curva suavizada al histograma experimental, y este ajuste depende solo del criterio del usuario, según cuál sea la parte de la curva que se desee ajustar de mejor manera, por ejemplo tratar de ajustar bien las colas de la distribución. En este sentido el uso de los polinomios de Bernstein no cumplió con los objetivos propuestos.

Quizá se deba analizar otra herramienta diferente a los polinomios de Bernstein para realizar el suavizamiento del histograma experimental, quizá no un método que interpole los puntos de control pero si uno que suavice un poco menos la distribución experimental y no generen tantas diferencias entre ésta y la curva suavizada. Alternativas potenciales son el uso de funciones kernel o un ajuste polinomial.

9. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO.

Dados los resultados obtenidos en los diferentes casos de estudio se puede concluir que la corrección cumple con todos los objetivos planteados anteriormente, por lo que se convierte en una herramienta útil para generar información representativa cuando se tienen una variable submuestreada con respecto a una más importante.

No usar la corrección genera sesgos importantes en los resultados pudiendo conducir a un yacimiento sobrestimado (o subestimado) en hasta un 100% como se mostró en el presente caso de estudio, porcentaje que aumenta en la medida en que el sesgo de la distribución es mas grande.

El número de datos que se tienen en la base de datos es una variable importante en la metodología que se debería usar, con una gran cantidad de datos parece ser suficiente una discretización normal de la base de datos (15-30 clases) intentando que el ancho de cada clase sea el menor posible con la condición de que en todas ellas haya por lo menos un dato (aunque no se debería llegar a ese extremo). En el caso en que se tienen pocos datos y el histograma experimental es errático o presenta clases con una frecuencia de 0 es necesario y recomendable utilizar el suavizamiento.

El suavizamiento resuelve en cierta medida los problemas que genera tener pocos datos pero introduce imprecisiones en el ajuste de la real distribución de la variable, por ende, se recomienda su uso solo en casos extremos en donde se tienen pocos datos (por ejemplo no en el presente caso de estudio) y cuando no es posible hacer una discretización adecuada para generar la corrección, ya que aunque los resultados no son tan malos como en los casos sin usar la corrección, los resultados pueden ser significativamente diferentes al caso base.

Con el fin de subsanar este hecho y como trabajo futuro se podría implementar un algoritmo que usando los polinomios de Bernstein busque, por ejemplo, la curva que minimice el error cuadrático entre el histograma experimental y la curva suavizada, de esta manera no quedará a criterio del usuario cual curva ajusta mejor.

10. REFERENCIAS.

[1] Abbate, H. *Curvas*.

[2] Babu, G. J., Canty, A. J., Chaubey, Y. P., 2001. *Application Of Bernstein Polynomials For Smooth Estimation Of A Distribution And Density Function*.

[3] Chak, P. M., Madras, N., Smith, J. B., 2001. *Flexible Functional Forms: Bernstein Polynomials*.

[4] Deutsch, C. V., 2002. *Geostatistical Reservoir Modeling*, Oxford University Press.

[5] Emery, X, 2008. *Simulación Estocástica y Geoestadística No Lineal*. Universidad de Chile, facultad de ciencias físicas y matemáticas, departamento de ingeniería civil de minas.

[6] Leuangthong, O., Deutsch, C. V., 2003. *Stepwise Conditional Transformation For Simulation Of Multiple Variables*.

[7] Leuangthong, O., Khan, K. D., y Deutsch, C. V. 2008. *Solved Problems In Geostatistics*, pag. 28 – 34.

[8] Leuangthong, O., Ortiz, J., y Deutsch, C. V. 2005. *On The Scaling And Use Of Multivariate Distributions In Geostatistical Simulation*.

[9] Laurini, M., Pereira, P., 2009. *Conditional Stochastic Kernel Estimation By Nonparametric Methods*.

[10] Salama, A, 2002. *A Technique For Interpolation And Smoothing Of Mass Size Density Data*.

11. ANEXOS.

ANEXO A. Estudio exploratorio- Despliegue de datos.

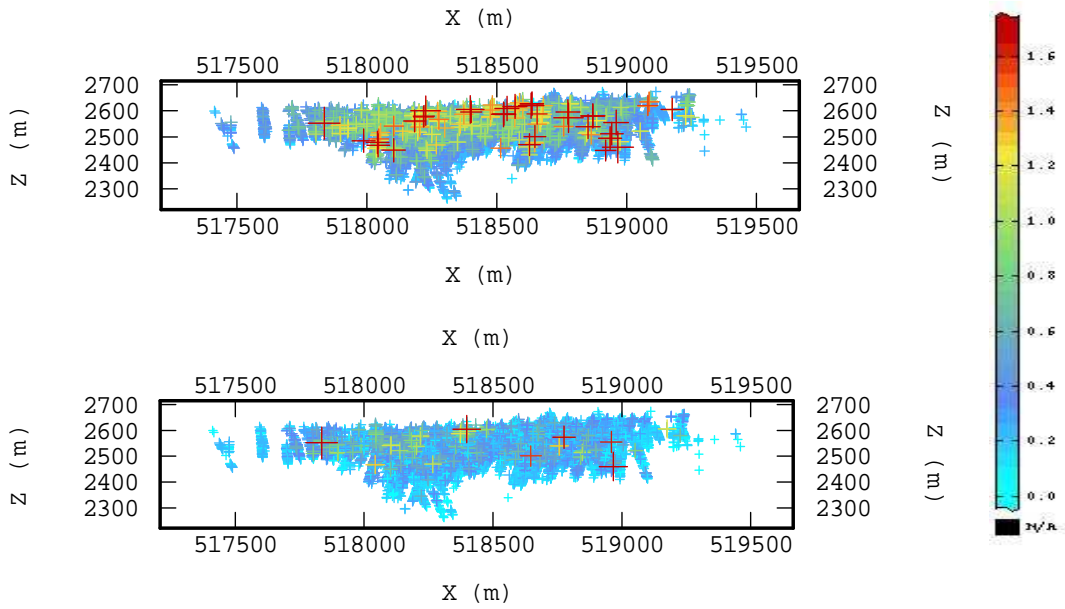


Figura 74 - Vista Este - Oeste. Ley de CuT (arriba) y CuS (abajo).

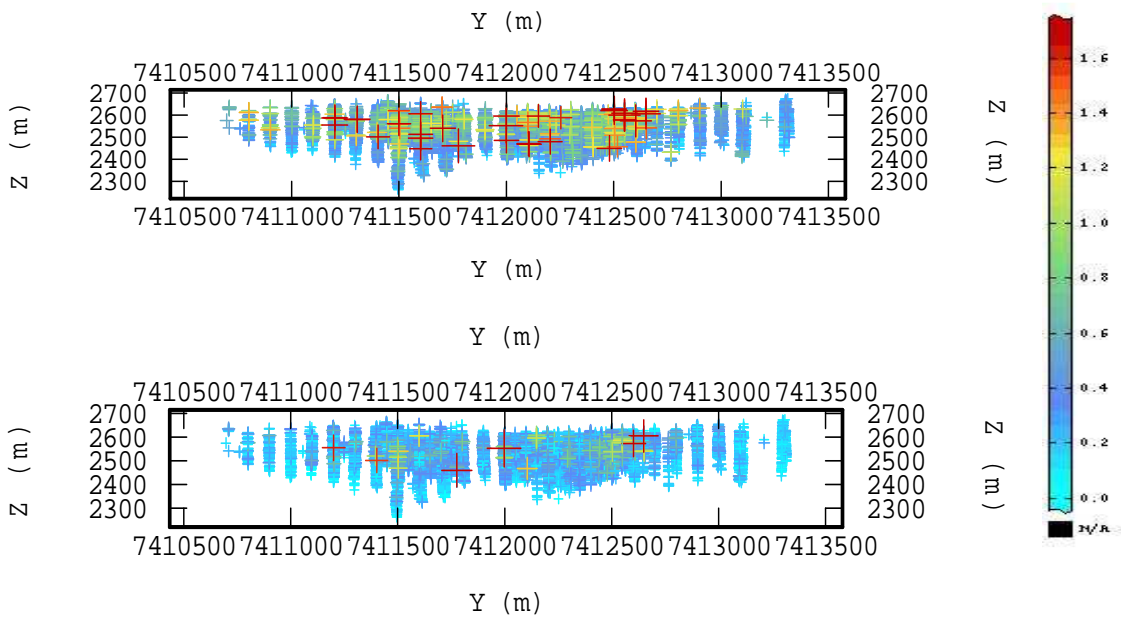


Figura 75 - Vista Norte - Sur. Ley de CuT (arriba) y CuS (abajo).

ANEXO B. Dependencia del factor de corrección con el número de clases.

En las siguientes figuras se muestra el factor de corrección que afecta a cada clase y la distribución de CuS corregida resultante, en la base de datos 80% de sesgo.

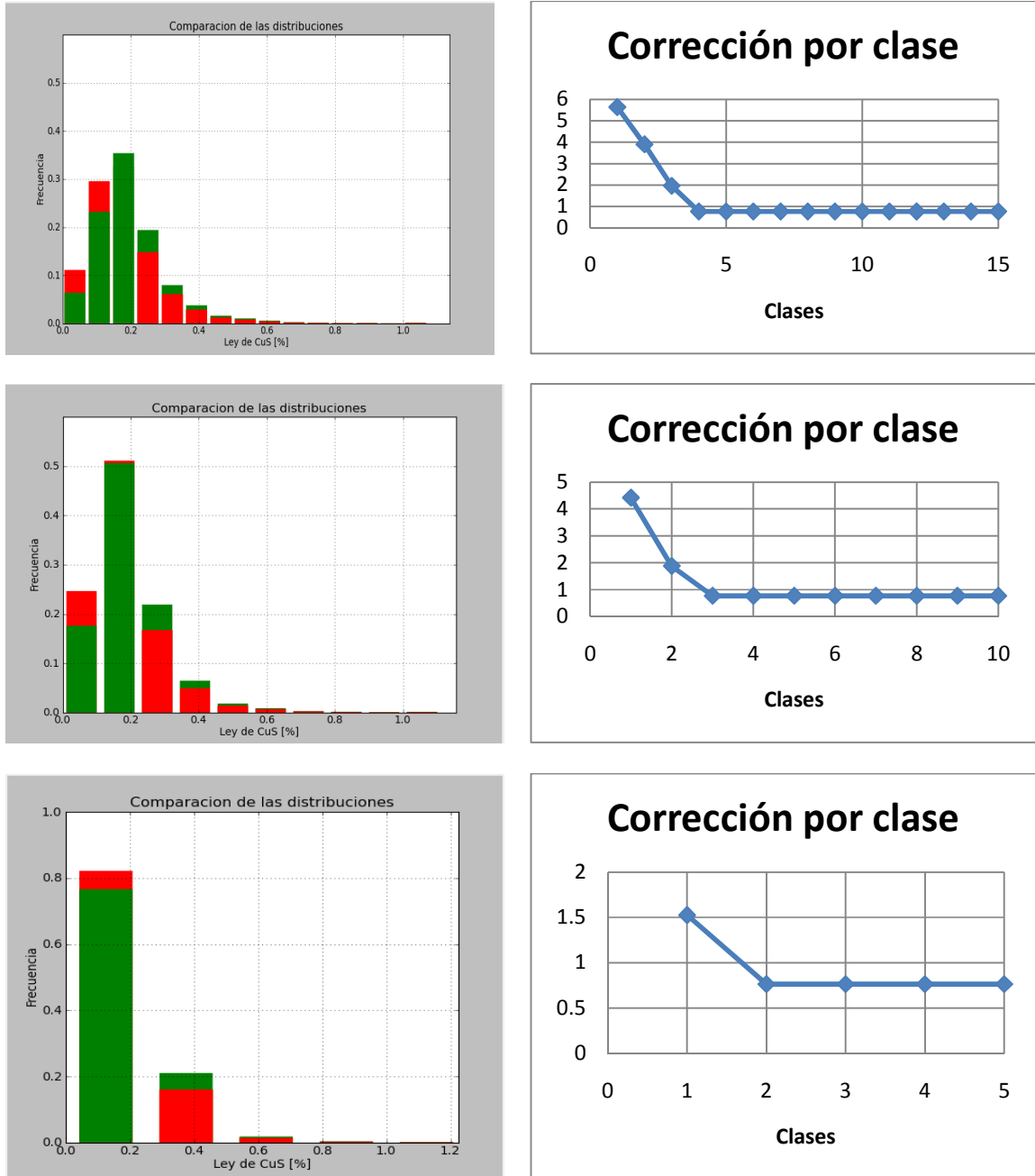


Figura 76 - Dependencia del factor de corrección del grado de discretización, caso sesgado 80%.

ANEXO C. Transformación gaussiana – Bases de datos sesgadas.

Base de datos sesgada 50% bajo 0.2% CuT

Distribución bivariable, gaussiana CuT y gaussiana CuS.

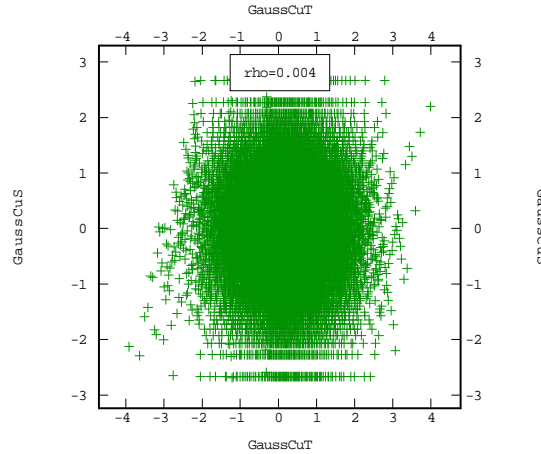
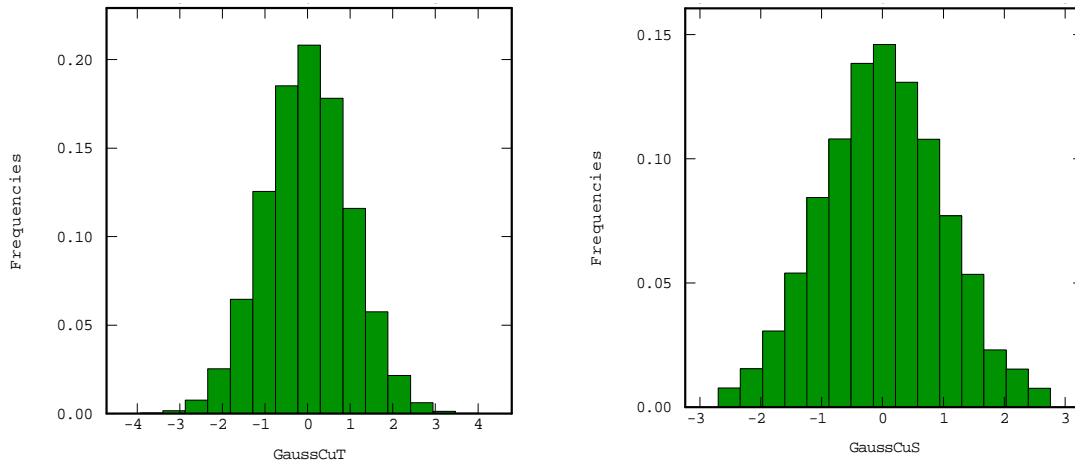


Figura 77 - Distribución bivariable, variables gaussianas, base de datos sesgada 50%.

Histogramas y estadísticas básicas de las variables gaussiana CuT (izquierda) y gaussiana CuS (derecha).



Media	0.00
Desv. Estándar	1.00
Mínimo	-3.92
Máximo	3.99

Media	0.00
Desv. Estándar	1.00
Mínimo	-2.69
Máximo	2.75

Figura 78 - Estadísticas básicas e histogramas de las distribuciones marginales, variables gaussianas, caso sesgado 50%.

Comparación de las distribuciones con una normal (0,1)

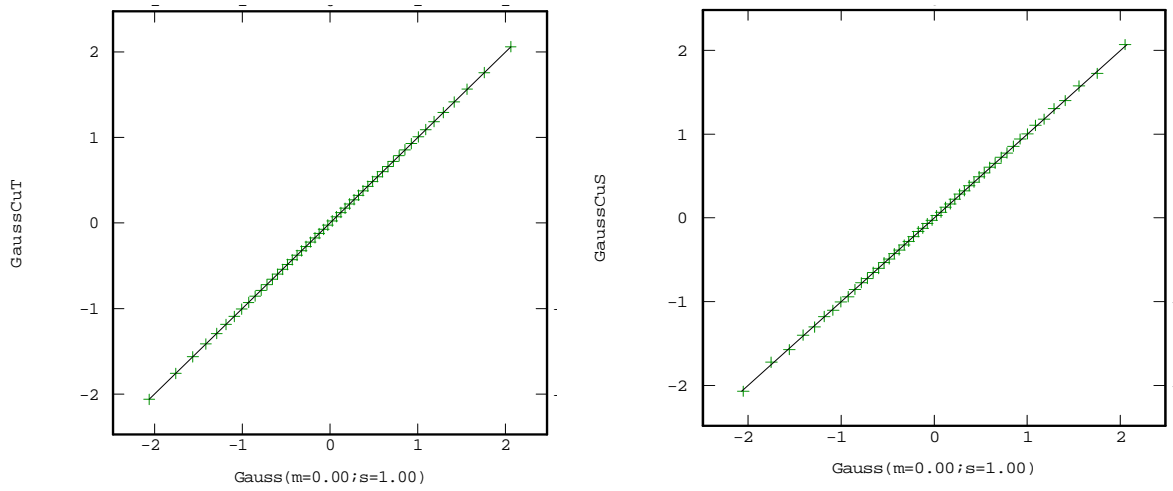


Figura 79 - Comparación de las distribuciones marginales con una normal de media 0 y varianza 1.

Base de datos sesgada 80% bajo 0.2% CuT

Distribución bivariable, gaussiana CuT y gaussiana CuS.

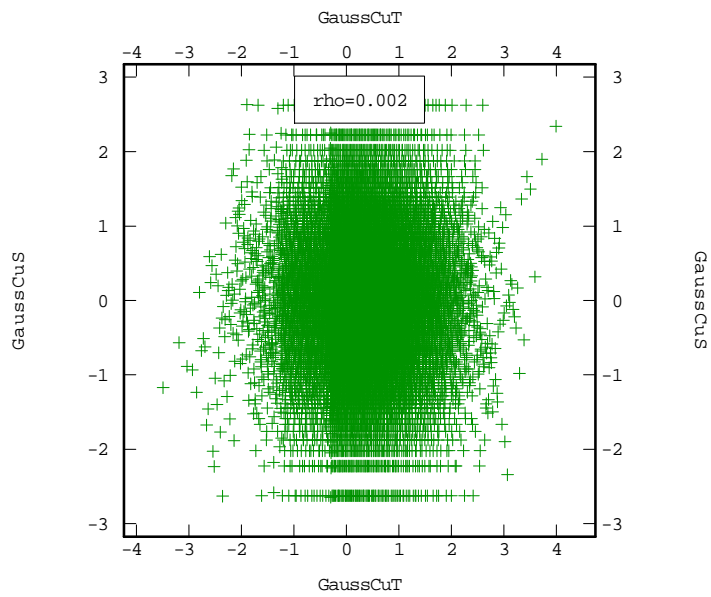
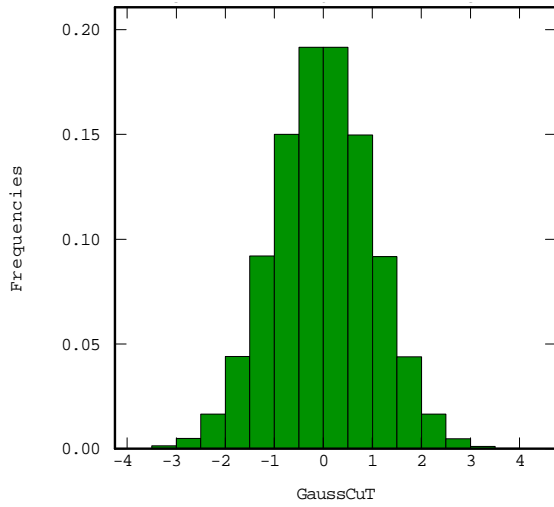
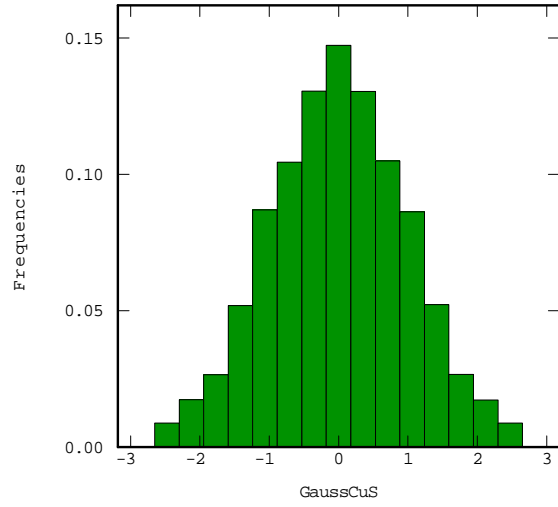


Figura 80 - Distribución bivariable, variables gaussianas, base de datos sesgada 80%.

Histogramas y estadísticas básicas de las variables gaussiana CuT (izquierda) y gaussiana CuS (derecha).



Media	0.00
Desv. Estándar	1.00
Mínimo	-3.49
Máximo	3.99



Media	0.00
Desv. Estándar	1.00
Mínimo	-2.65
Máximo	2.65

Figura 81 - Estadísticas básicas e histogramas de las distribuciones marginales, variables gaussianas, caso sesgado 80%.

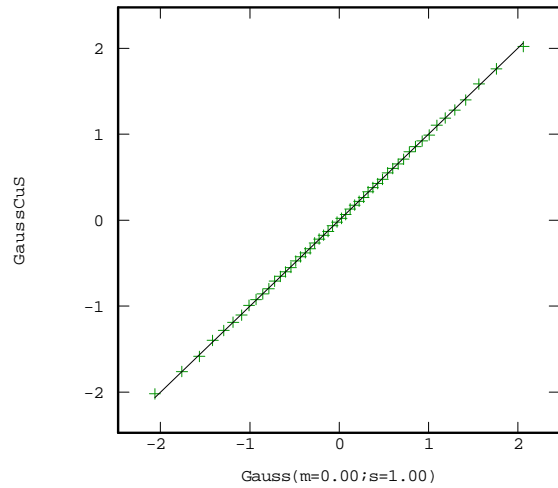
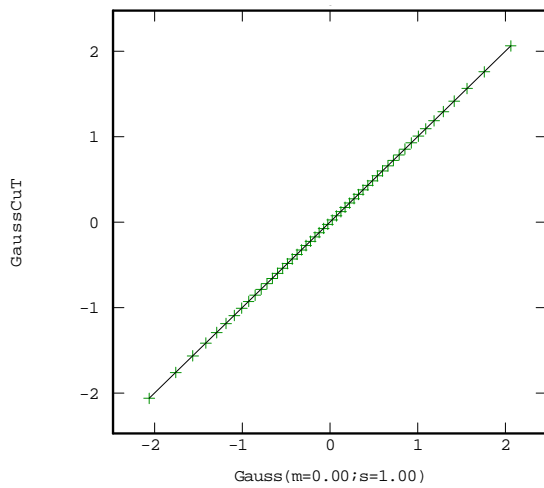


Figura 82 - Comparación de las distribuciones marginales con una normal de media 0 y varianza 1.

ANEXO D. Estudio variográfico – Variables gaussianas.

Variable gaussiana CuS sesgada 50% bajo la ley 0.2%.

Mapas variográficos.

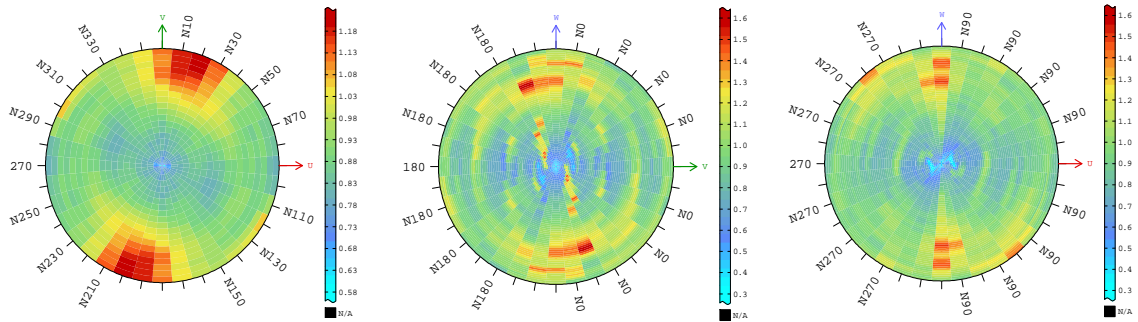


Figura 83 - Mapas variográficos, variable gaussiana CuS, caso sesgado 50%.

Variogramas experimentales y modelados.

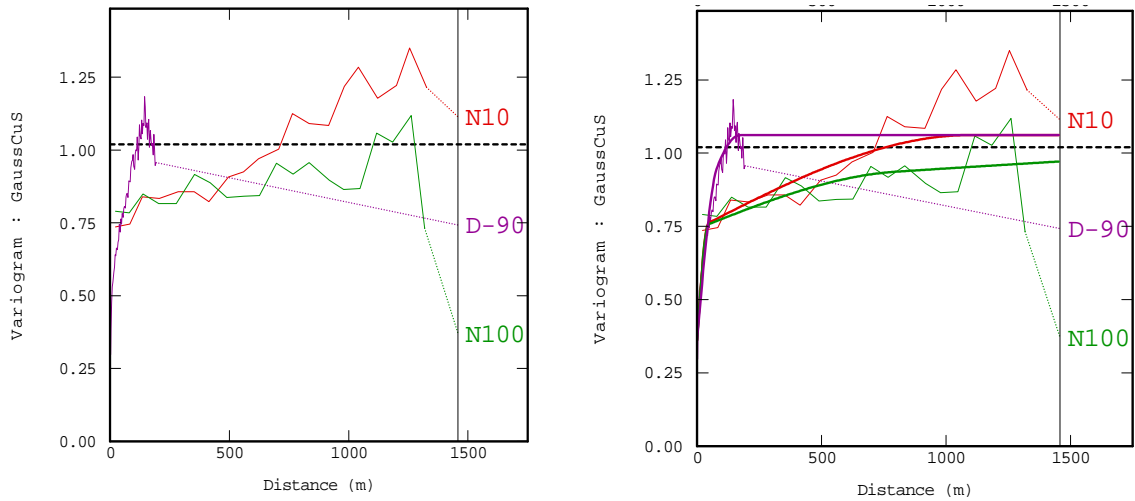


Figura 84 - Variogramas experimental y modelado, variable gaussiana CuS, caso sesgado 50%.

Donde el modelo del variograma de la variable gaussiana CuS es el siguiente:

$$\gamma(h) = 0.35 + 0.4 \cdot \text{esf}(30,30,150) + 0.14 \cdot \text{esf}(800,800,250) + 0.09 \cdot \text{esf}(1500, \infty, 300)$$

Estudio variográfico variable gaussiana CuS sesgada 80% bajo la ley 0.2%.

Mapas variográficos.

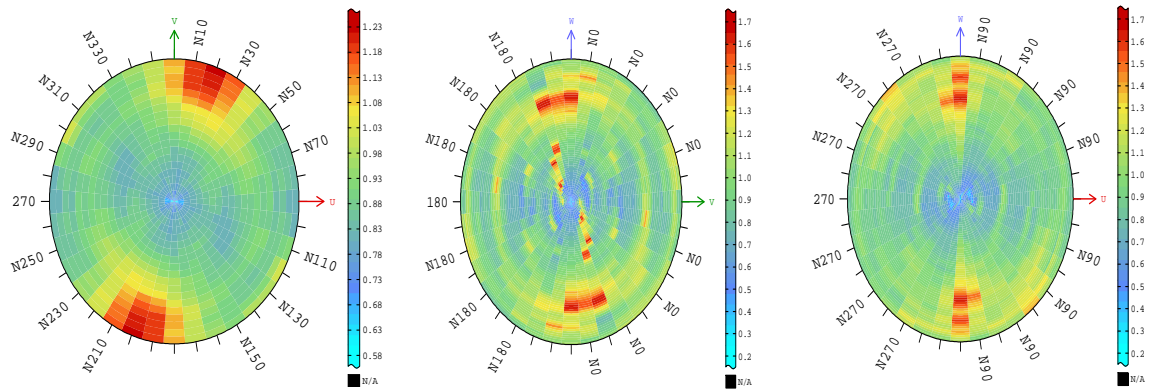


Figura 85 - Mapas variográficos, variable gaussiana CuS, caso sesgado 80%.

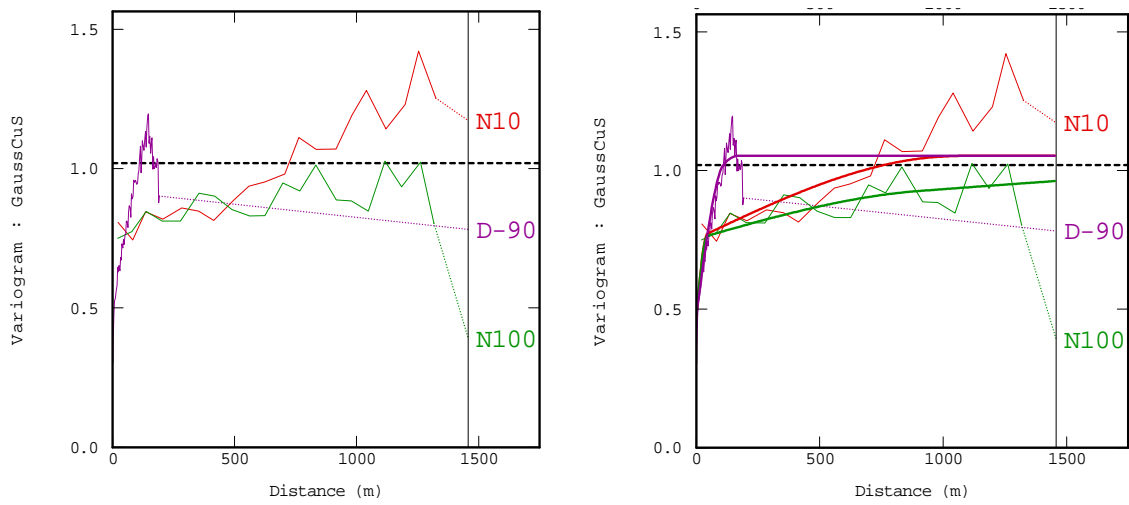


Figura 86 - Variogramas experimental y modelado, variable gaussiana CuS, caso sesgado 80%.

Donde el modelo del variograma de la variable gaussiana CuS es el siguiente:

$$\gamma(h) = 0.35 + 0.4 \cdot \text{esf}(30,30,150) + 0.14 \cdot \text{esf}(800,800,250) + 0.07 \cdot \text{esf}(1500, \infty, 300)$$

ANEXO E. Realizaciones de los casos con sesgo corregidos.

Variable CuS. Sesgo del 50%.

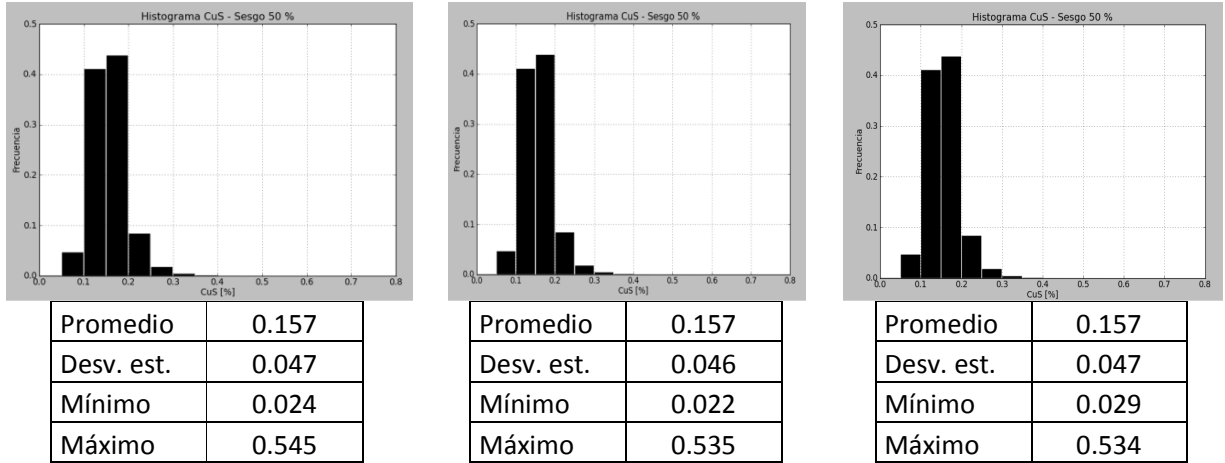


Figura 87 - Histogramas y estadísticas básicas de 3 realizaciones de la variable CuS a soporte puntual. Caso sesgado 50%.

Variable CuS. Sesgo del 80%.

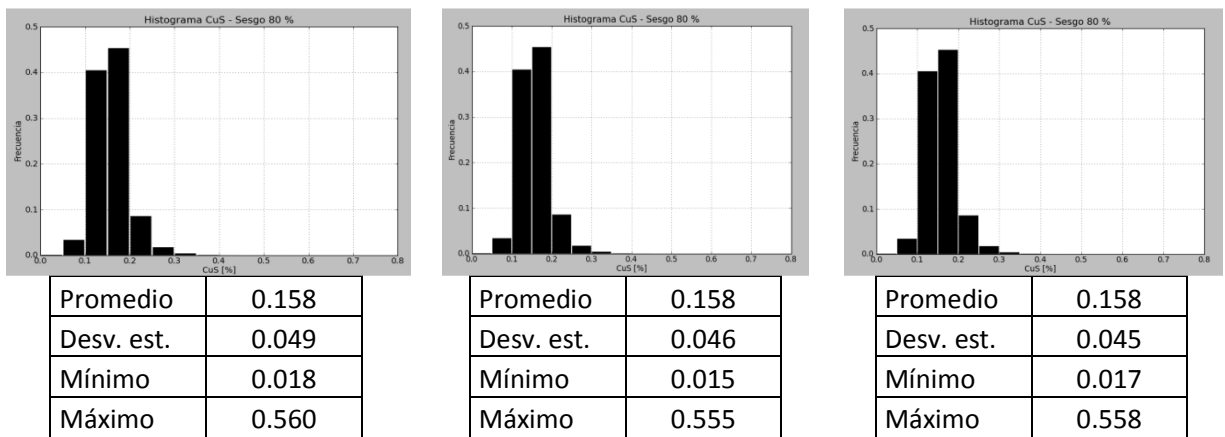
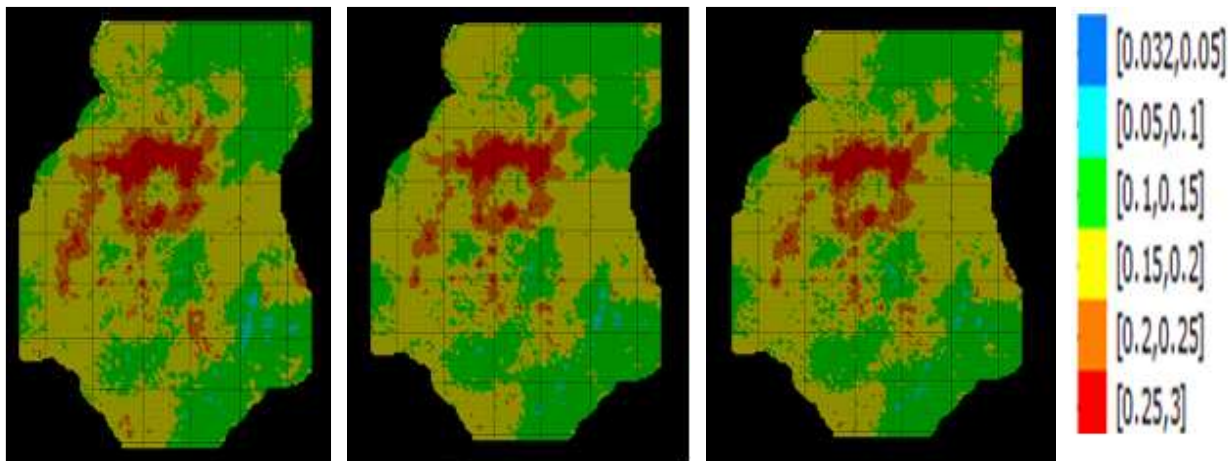


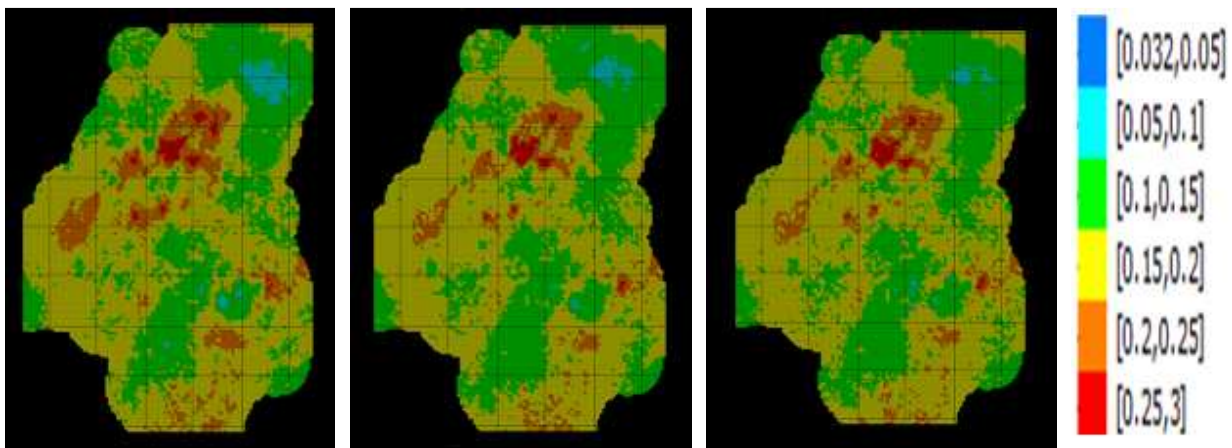
Figura 88 - Histogramas y estadísticas básicas de 3 realizaciones de la variable CuS a soporte puntual. Caso sesgado 80%.

ANEXO F. Resultados de la simulación.

Ley de CuS – Altura 2575.

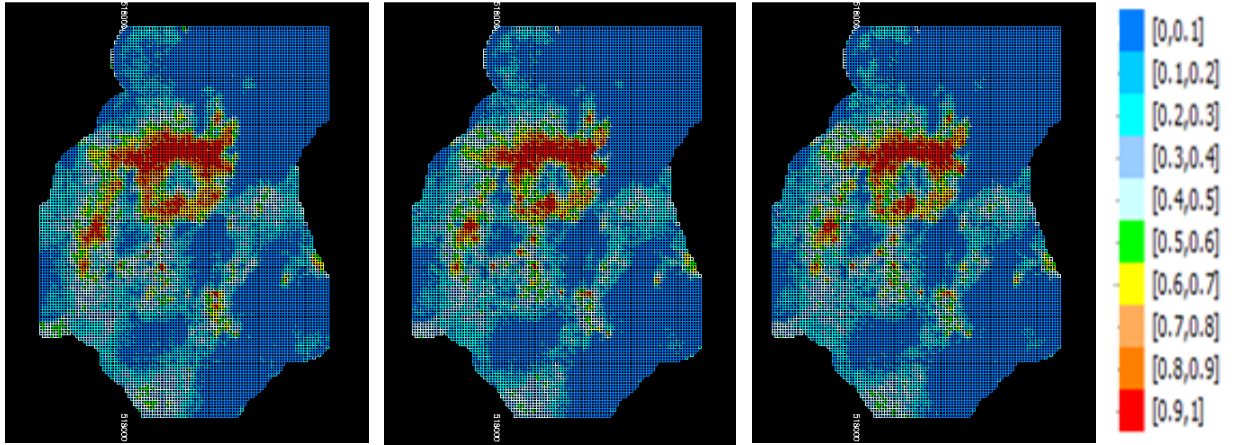


Ley de CuS – Altura 2615.

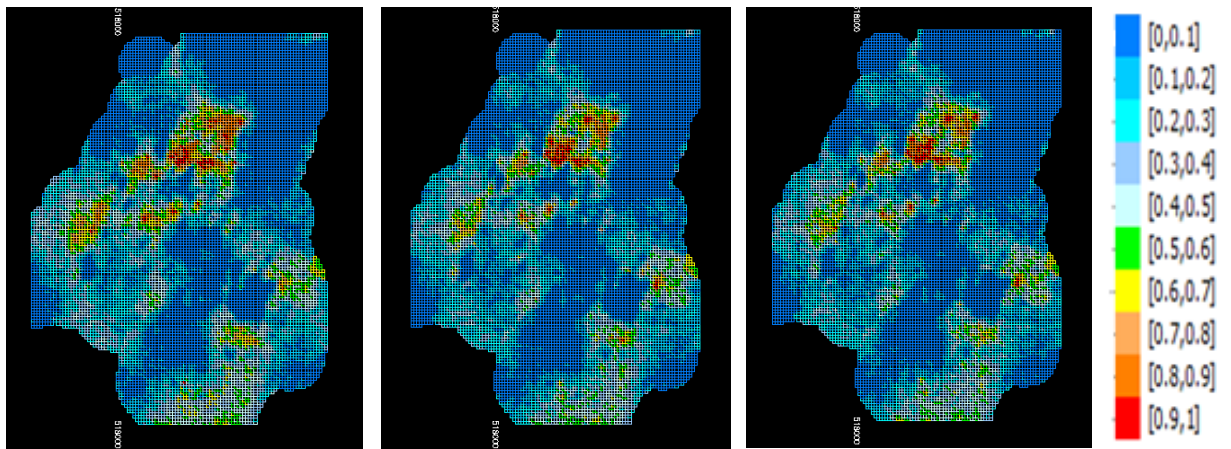


Probabilidad de superar la ley de corte de 0.2% de CuS.

Planta 2575

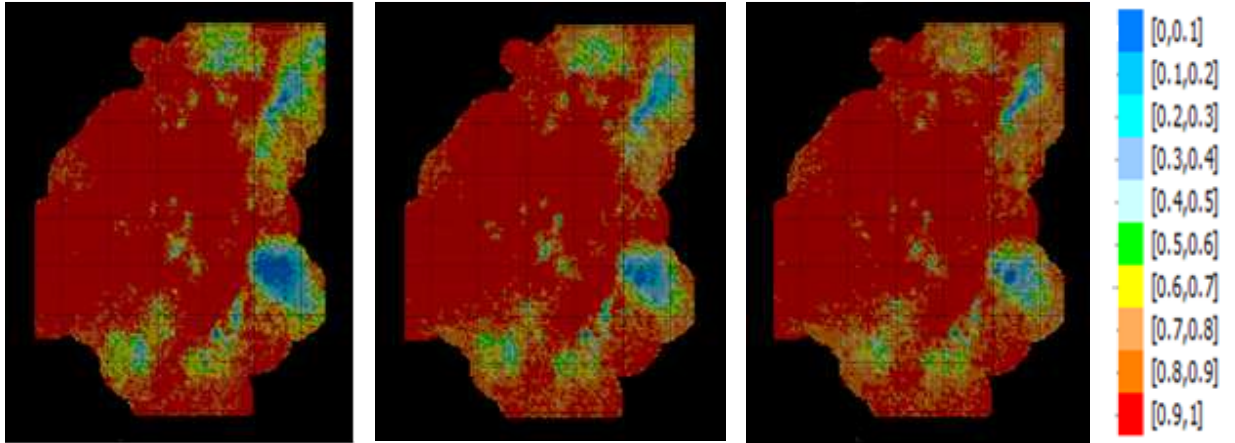


Planta 2615

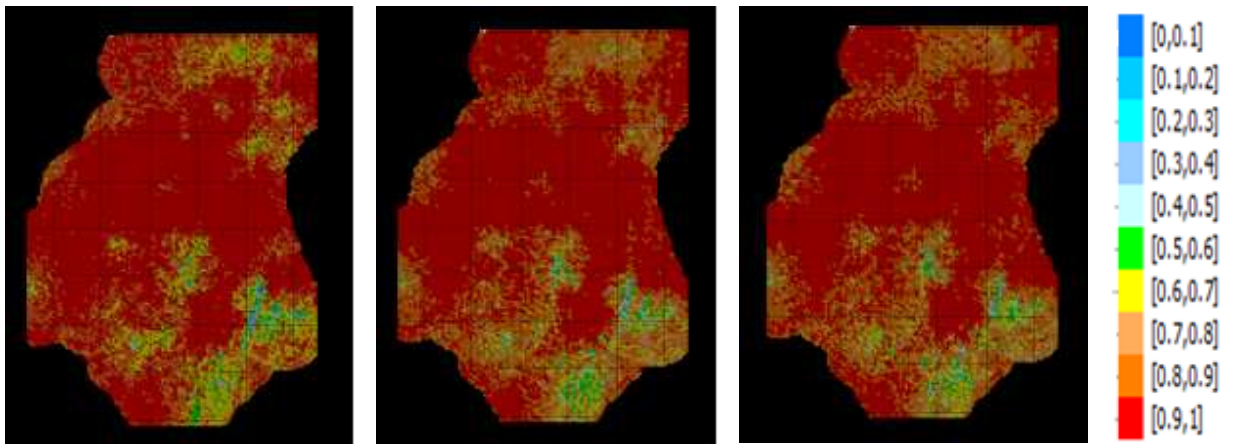


Probabilidad de superar la ley de corte de 0.1% de CuS.

Planta 2535.



Planta 2615.



ANEXO G. Comparación entre los casos de estudio.

Tabla resumen - Tonelajes y leyes medias para distintas leyes de corte, corrección normal.

Ley de corte	Tonelaje			Ley media		
	Caso base	Sesgo 50%	Sesgo 80%	Caso base	Sesgo 50%	Sesgo 80%
[%]	[Mton]	[Mton]	[Mton]	[%]	[%]	[%]
0.00	2259	2259	2259	0.159	0.157	0.158
0.05	2234	2235	2239	0.160	0.158	0.159
0.10	1972	1983	2014	0.170	0.167	0.168
0.15	1266	1245	1297	0.195	0.191	0.190
0.20	446	388	386	0.235	0.233	0.233
0.25	106	84	84	0.288	0.287	0.287
0.30	30	23	22	0.336	0.336	0.336
0.35	7	6	6	0.383	0.384	0.385
0.40	2	1	1	0.439	0.446	0.451

Tabla 16 - Comparación entre el caso base y los casos sesgados, corrección normal.

Tabla resumen - Tonelajes y leyes medias para distintas leyes de corte, sin corrección.

Ley de corte	Tonelaje			Ley media		
	Caso base	Sesgo 50%	Sesgo 80%	Caso base	Sesgo 50%	Sesgo 80%
[%]	[Mton]	[Mton]	[Mton]	[%]	[%]	[%]
0	2259	2259	2259	0.159	0.175	0.185
0.05	2234	2259	2259	0.160	0.175	0.185
0.1	1972	2254	2258	0.170	0.176	0.185
0.15	1266	1820	2060	0.195	0.186	0.190
0.2	446	558	764	0.235	0.223	0.220
0.25	106	91	98	0.288	0.277	0.274
0.3	30	19	17	0.336	0.323	0.323
0.35	7	3	3	0.383	0.372	0.372
0.4	2	1	0	0.439	0.432	0.445

Tabla 17 - Comparación entre el caso base y los casos sesgados, casos sin corrección.

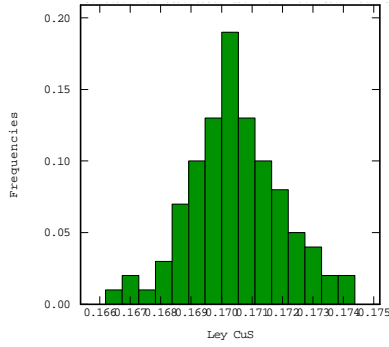
Tabla resumen - Tonelajes y leyes medias para distintas leyes de corte, usando el suavizamiento.

Ley de corte	Tonelaje			Ley media		
	Caso base	Sesgo 50%	Sesgo 80%	Caso base	Sesgo 50%	Sesgo 80%
[%]	[Mton]	[Mton]	[Mton]	[%]	[%]	[%]
0	2259	2259	2259	0.159	0.165	0.173
0.05	2234	2259	2259	0.160	0.165	0.173
0.1	1972	2202	2234	0.170	0.167	0.174
0.15	1266	1523	1708	0.195	0.183	0.186
0.2	446	293	374	0.235	0.230	0.228
0.25	106	56	67	0.288	0.291	0.291
0.3	30	17	21	0.336	0.342	0.341
0.35	7	5	6	0.383	0.396	0.394
0.4	2	2	2	0.439	0.463	0.460

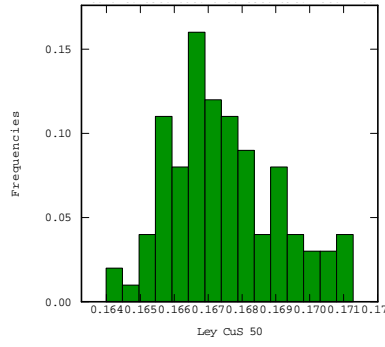
Tabla 18 - Comparación entre el caso base y los casos sesgados, usando suavizamiento.

ANEXO H. Análisis de incertidumbre - Ley media.

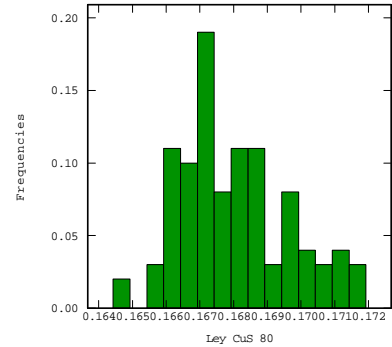
Ley de corte 0.1 % (de derecha a izquierda, caso base, sesgo del 50%, sesgo del 80%)



Media	0.17
Desv. Estándar	0.00
Mínimo	0.17
Máximo	0.17

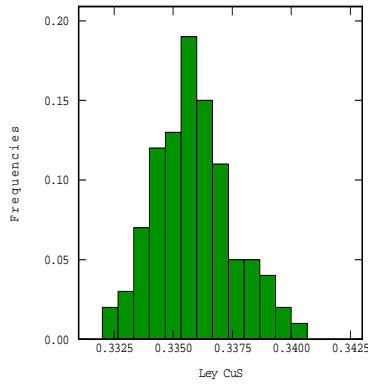


Media	0.17
Desv. Estándar	0.00
Mínimo	0.16
Máximo	0.17

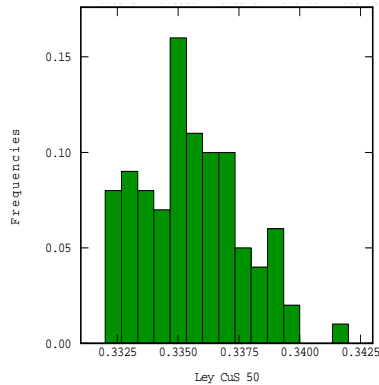


Media	0.17
Desv. Estándar	0.00
Mínimo	0.16
Máximo	0.17

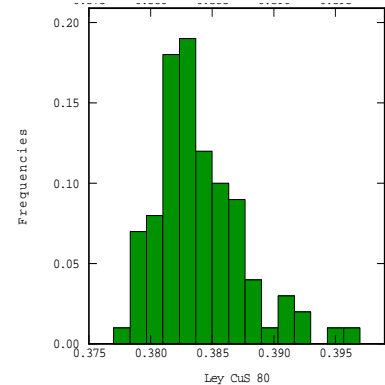
Ley de corte 0.35 % (de derecha a izquierda, caso base, sesgo del 50%, sesgo del 80%)



Media	0.34
Desv. Estándar	0.00
Mínimo	0.33
Máximo	0.34



Media	0.34
Desv. Estándar	0.00
Mínimo	0.33
Máximo	0.35



Media	0.38
Desv. Estándar	0.01
Mínimo	0.38
Máximo	0.41