**Análisis espacial de potencial inicial de gas en Virginia**

**Esteban Vargas y Miguel Coto**

**Resumen**

Con datos del Sistema de Datos de Petróleo y Gas de la Encuesta Geológica y Económica del oeste de Virginia sobre flujos abiertos finales de gas de las rocas del Alto Devónico en un campo en Virginia Occidental con valores del potencial inicial medidos en mil pies cúbicos por día (Mcfpd) se realizó un análisis univariado utilizando análisis exploratorio y métodos geoestadísticos clásicos como el calculo de semivariograma y modelado del mismo así como estimaciones y mapeo sobre el área usando técnicas de kriging con el fin de determinar el patrón espacial de la variable de interés, la mejor representación gráfica de la continuidad de la variable y cuál es el grado de incertidumbre de estas estimaciones. Los principales resultados arrojan que el mejor método de estimación para semivariograma fue el esférico y que si bien alrededor de donde había más puntos muestreados se consiguieron buenas estimaciones con variancia bastante baja, el pobre muestreo del área resultó en correlaciones espaciales bajas en una gran zona con la implicación de alta variabilidad y estimaciones muy cercanas a la media.

**Palabras clave**

Geoestadística, kriging, gas, interpolación.

**Introducción**

El Sistema de Datos de Petróleo y Gas de la Encuesta Geológica y Económica del oeste de Virginia es una estructura de datos relacional que consta de 8 tipos de datos, codificados en el código del condado y el número de permiso de los pozos. Estos archivos incluyen ubicaciones, terminaciones, pagos, estratigrafía, producción, enchufe, catálogo de registro mecánico, entre otros. Este sistema respalda las necesidades de investigación e información de la agencia y está disponible para el público en general en aplicaciones basadas en la web u otros.

Este sistema se diseñó originalmente en 1966 y se ha rediseñado varias veces desde entonces, generalmente junto con las principales actualizaciones del sistema de hardware y software. El portal administra registros de más de 145,000 pozos de petróleo y gas perforados en Virginia Occidental desde hace unos 150 años. este sistema contiene datos de pozos de petróleo y gas perforados en 53 de los 55 condados de West Virginia; no se han perforado pozos de petróleo o gas en el extremo oriental del estado del estado, en los condados de Berkeley y Jefferson.

Como parte del proceso de recolección, la mayoría de los datos son extraídos o interpretados por los geólogos a partir de los registros de los perforadores y las placas enviadas por los operadores a la autoridad reguladora estatal. Otros datos se obtienen directamente de los operadores o de otras fuentes, como publicaciones de encuestas, registros de geólogos, mapas de propiedades, registros mecánicos, muestras y núcleos. Las coordenadas del pozo se almacenan en una forma modificada de latitud / longitud medida en centésimas de milla al sur de la latitud de referencia y al oeste de la longitud de referencia. Estas coordenadas se convierten en la entrada de datos a través de un cálculo a las coordenadas de Universal Transverse Mercator (UTM) para la zona 17 UTM y los grados decimales de latitud y longitud.

Algunos de los atributos que se recolectan son: ubicación del pozo; datos de terminación, propiedad de superficie / minerales y operador de pozos; profundidades y formaciones de pagos y espectáculos de hidrocarburos, almacenamiento, observación, inyección, etc. y profundidad y tipo de agua encontrada; tapas y espesores de taponamiento estratigráfico; agua encontrada en la secuencia de perforación, entre otros.

Las unidades de medida según el material que se mide son, por lo general: elevaciones, en pies sobre el nivel medio del mar; profundidades - en pies debajo del dato superficial; volúmenes de gas - en Mcf (mil pies cúbicos); volúmenes de petróleo - en barriles; presiones - en psi (libras por pulgada cuadrada); tiempo de tratamiento - en horas.

Los métodos geoestadísticos que se utilizan en el sistema se han llevado a cabo en WVGES durante 20 años para calcular estimaciones de recursos naturales calificadas por riesgo. Se utilizan variografía, *kriging* y simulación condicional en áreas como petróleo y gas, mediante la medición del potencial inicial y producción acumulada de petróleo y gas.

La geoestadística es de gran utilidad para describir la continuidad espacial de cualquier fenómeno natural. Es posible conocer la forma en que varía cualquier variable continua en el espacio (patrón espacial) a una o varias escalas seleccionadas, con un nivel de detalle que permite cuantificar la variación espacial de la variable en distintas direcciones del espacio. Este método utiliza funciones para modelar esta variación espacial, y estas funciones son utilizadas posteriormente para interpolar en el espacio el valor de la variable en sitios no muestreados. La fortaleza de la geoestadística es que esta interpolación (*conocida como kriging*) es considerada una estimación muy robusta ya que se basa en la función continua que explica el comportamiento de la variable en las distintas direcciones del espacio, y que en contraste con otros métodos de interpolación (como por ejemplo interpolar un punto usando los valores de los puntos que le rodean ponderados por la distancia que los separa) permite asociar la variabilidad de la estima (conocido como *grado de incertidumbre*). A partir de esto se pueden responder a las siguientes preguntas: ¿Cuál es el patrón espacial de mis variables de interés? ¿A qué escala se repite este patrón espacial? ¿Existe covariación espacial entre las distintas variables de interés? ¿Cuál es la mejor representación gráfica de la continuidad de mi variable? ¿Cuál es el grado de incertidumbre de estas estimaciones?

Para el presente proyecto se tomaron datos relacionados con flujos abiertos finales de gas de las rocas del Alto Devónico en un campo en Virginia Occidental. Los valores del potencial inicial, variable observada, se dan en mil pies cúbicos por día (Mcfpd)., con el final de analizar la distribución espacial del indicador y aplicar una serie de técnicas espaciales de interpolación para conocer cual método realiza mejores estimaciones.

**Metodología**

La técnica para utilizar con los datos de los potenciales iniciales *“Initial Potentials”* de gas corresponde a un análisis espacial de datos georreferenciados el cual se detalla a continuación.

El uso de la palabra "*kriging*" se ha convertido en sinónimo de "predicción óptima" en el espacio, utilizando observaciones tomadas en ubicaciones cercanas conocidas. Ord (1983), describe en la Enciclopedia de Ciencias Estadísticas, al “*kriging”* como "un método de interpolación para procesos espaciales aleatorios" con predictores tanto lineales como no lineales. Según Hemyari y Nofziger (1987), “*Kriging*” es una forma de promedio ponderado en el que los pesos se eligen de manera que el error asociado con el predictor sea menor que para cualquier otra suma lineal. Los pesos dependen de la ubicación de los puntos utilizados en la predicción y sobre la covarianza, lo cual se ve reflejado en el semivariograma.

El análisis de “*kriging”* es un método geoestadístico de interpolación que ha probado ser útil y popular en muchos campos (Burgess y Webster, 1980). Este método provee, a partir de una muestra de puntos, ya sean regular o irregularmente distribuidos, valores **estimados** de aquellos sitios donde no hay información, sin sesgo y con una varianza mínima conocida.

Para la interpolación es necesario conocer previamente ciertos aspectos básicos acerca de los datos de elevación a interpolar: 1) el semivariograma, 2) el efecto nugget 3) la anisotropía y 4) la presencia o no de un componente estructural o “drift” (Burgess y Webster, 1980).

Mediante el ajuste de un semivariograma experimental, también es posible determinar si existe un *efecto nugget* en los datos. Si bien por definición en cualquier modelo de semivariograma, cuando existe un *efecto nugget* en los datos, el límite de (h) cuando h tiende a cero es diferente de cero. Para una variable continua, el *efecto nugget* se puede presentar como producto de errores de medición y variaciones sobre distancias mucho más pequeñas que los intervalos de muestreo utilizados (Burgess y Webster, 1980). +

El ajuste mediante los semivariogramas permite extraer una serie de parámetros que son los que van a ser usados para la interpolación geostadística (kriging) y que definen el grado y escala de variación espacial. Estos parámetros son el rango (A0), el nugget (C0), el sill (C0+C), y la proporción de la varianza explicada por el espacio (C/C0+C), a menudo expresada en porcentaje. El rango (A0) es la distancia a la que la semivarianza deja de aumentar, por tanto, indica la distancia a partir de la cual las muestras son espacialmente independientes unas de otras, y representa el tamaño o mancha que representa la variable (Paramá, 2006). El nugget (C0) es la varianza no explicada por el modelo, y se calcula como la intersección con el eje Y. La máxima semivarianza encontrada entre pares de puntos se conoce como sill y nos da el grado de variación espacial; el grado de incertidumbre a la hora de interpolar puntos en el espacio. En otras palabras, un alto cociente nos indica una variable espacialmente muy predecible.

Como ya se comentó, los datos provienen de una muestra de puntos de flujos abiertos finales de gas de las rocas del Alto Devónico en un campo en Virginia Occidental. Los valores del potencial inicial, variable observada, se dan en mil pies cúbicos por día (Mcfpd). La muestra consta de 122 puntos georreferenciados mediante proyección UTM en kilómetros con referencia a la zona 17.

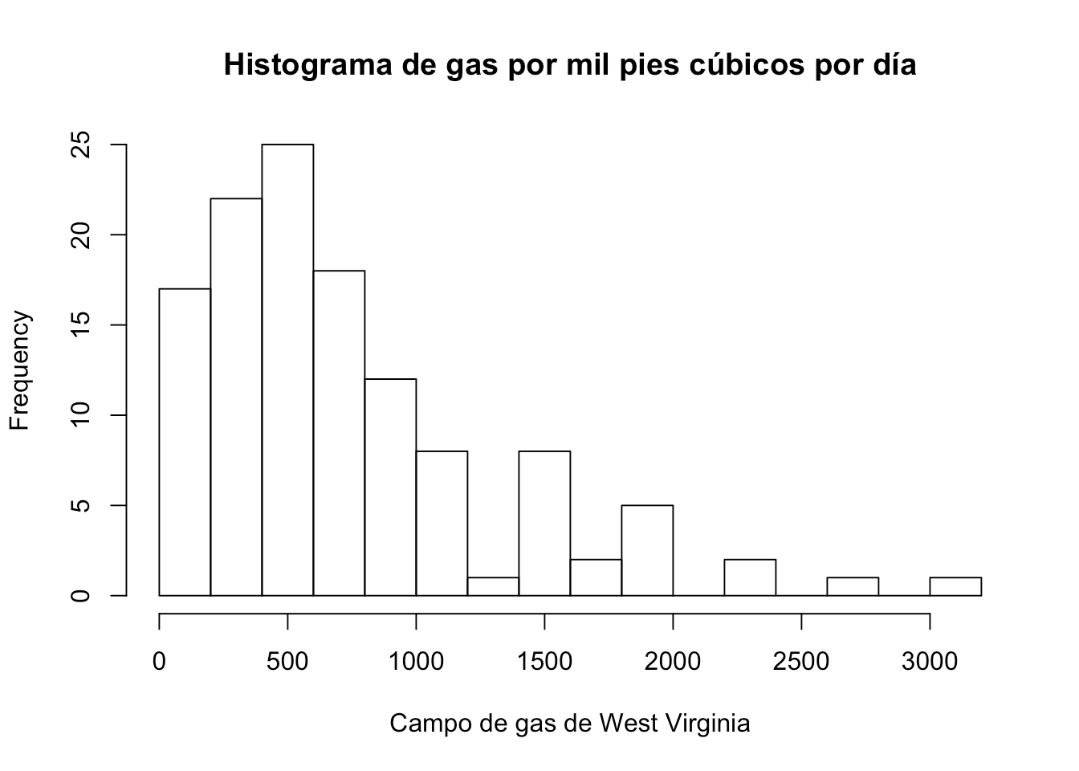
Para el análisis se hizo exploración de los datos, en este caso de la variable de medición de gas (objetivo), además se estimo el variograma mediante 3 métodos distintos (esférico, exponencial y de Matern), se determinó el efecto nugget, así como si existía anisotropía o no para luego realizar estimaciones a través de kriging a la grilla completa para estimar los valores esperados y variancia de las estimaciones de valores iniciales para posibles puntos de flujo abierto de gas. Así, el objetivo principal del presente análisis es determinar cuál es el patrón espacial de la variable de interés, cuál es la mejor representación gráfica de la continuidad de la variable través de los métodos de kriging usando el modelo de variograma y cuál es el grado de incertidumbre de estas estimaciones.

**Resultados**

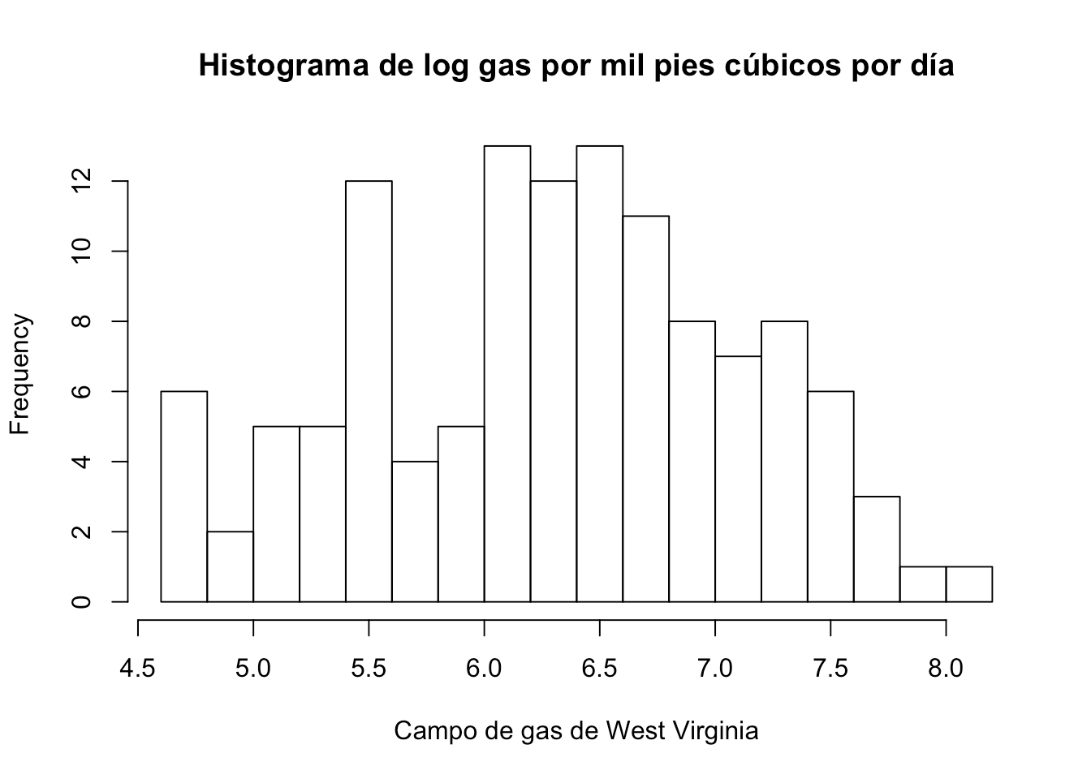
**Visualización de datos**

Como primera medida, se procedió a obtener resultados descriptivos del potencial inicial de gas. En la figura 1 se muestra el histograma, el cual refleja que los datos están sesgados hacia uno de los extremos de la distribución, por lo que se procedió a utilizar los valores transformados logarítmicamente, como se muestra en el gráfico 2.

**Gráfico 1.** Histograma de potencial inicial de gas (pies cúbicos por día)

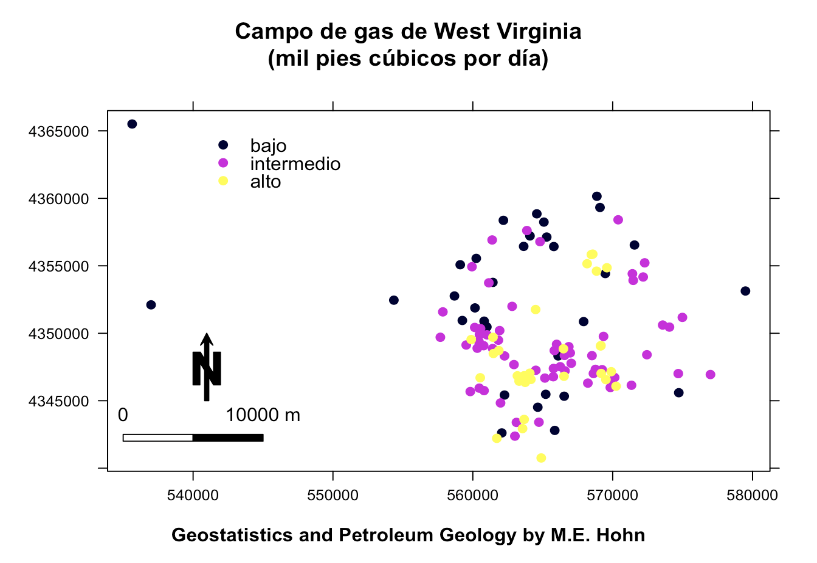
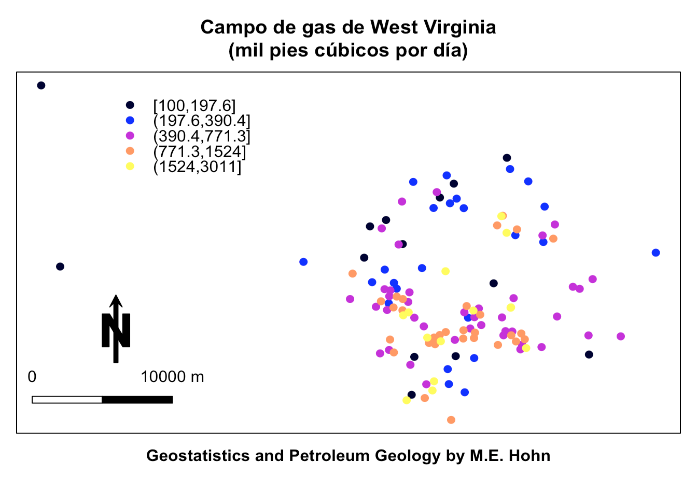


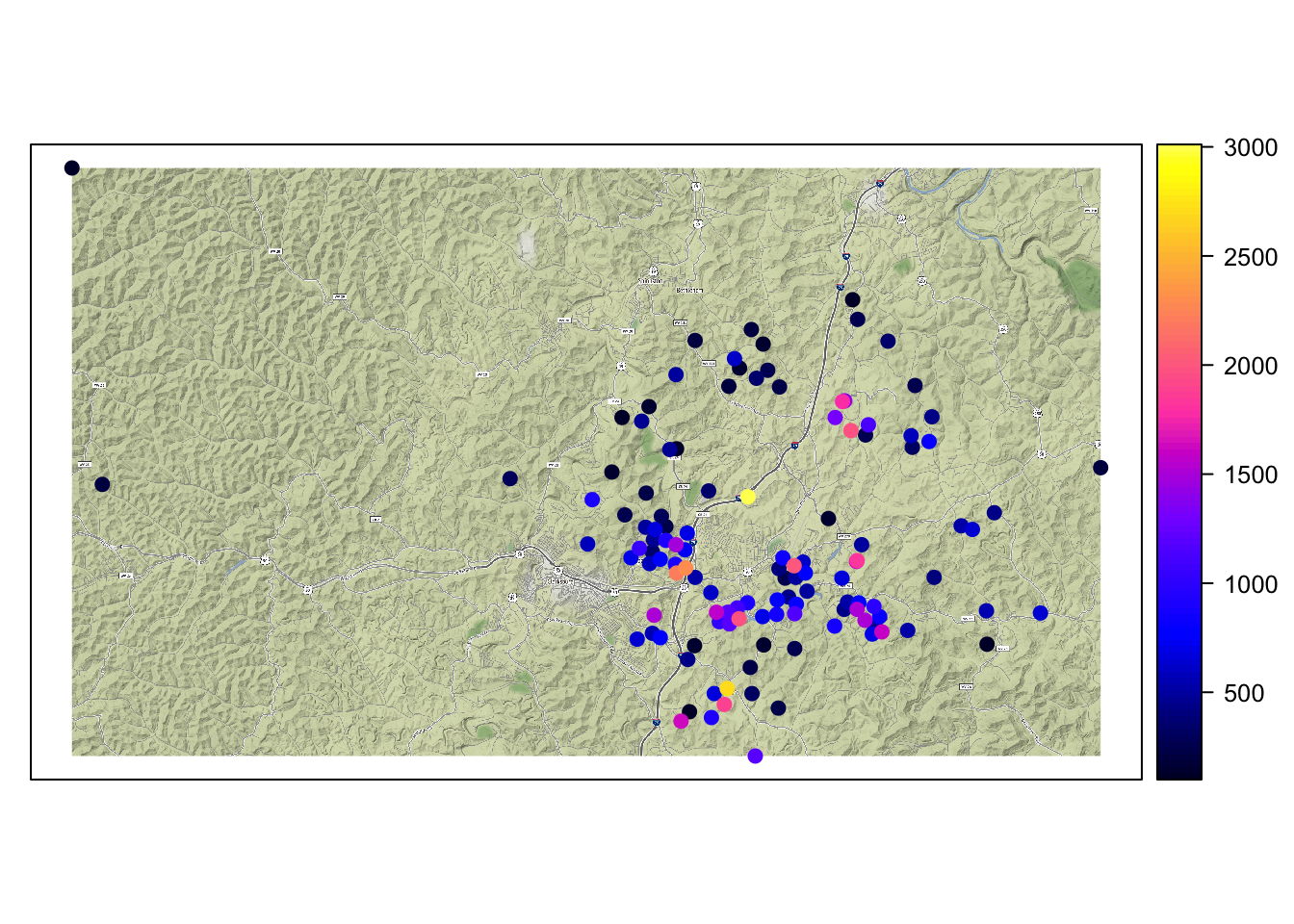
**Gráfico 2.** Histograma del logaritmo del potencial inicial de gas (pies cúbicos por día)



En el gráfico 2 se muestra la variable transformada a logaritmo; esto permite obtener una distribución más equiparada de los datos a lo largo de la escala. Cabe destacar que este tipo de transformaciones apropiadas de los datos, pueden mejorar la apariencia del semivariograma, y consecuentemente el análisis.

**Gráfico 3.** Distribución espacial del potencial inicial de gas (pies cúbicos por día)



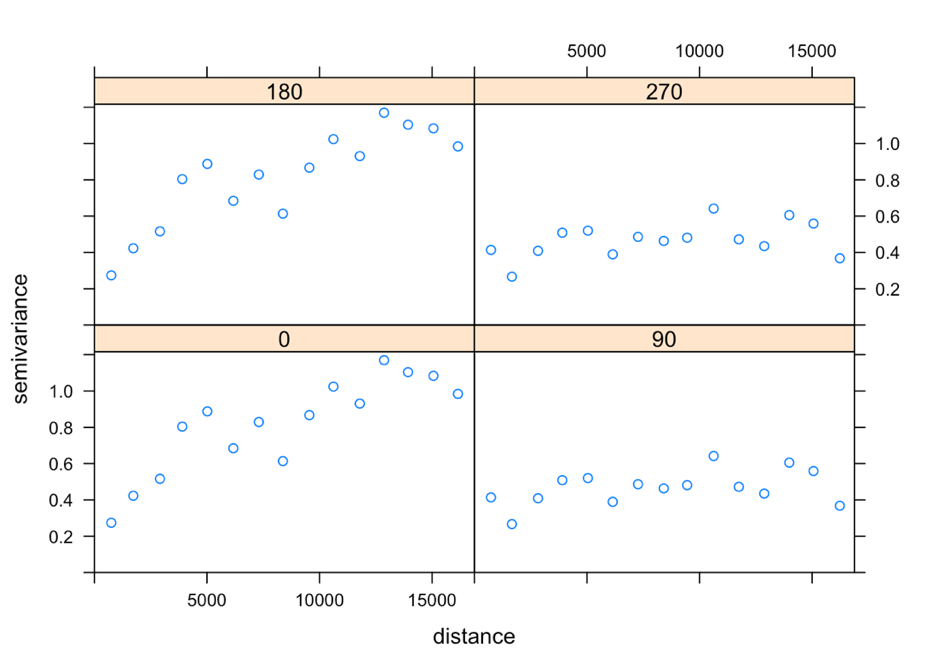
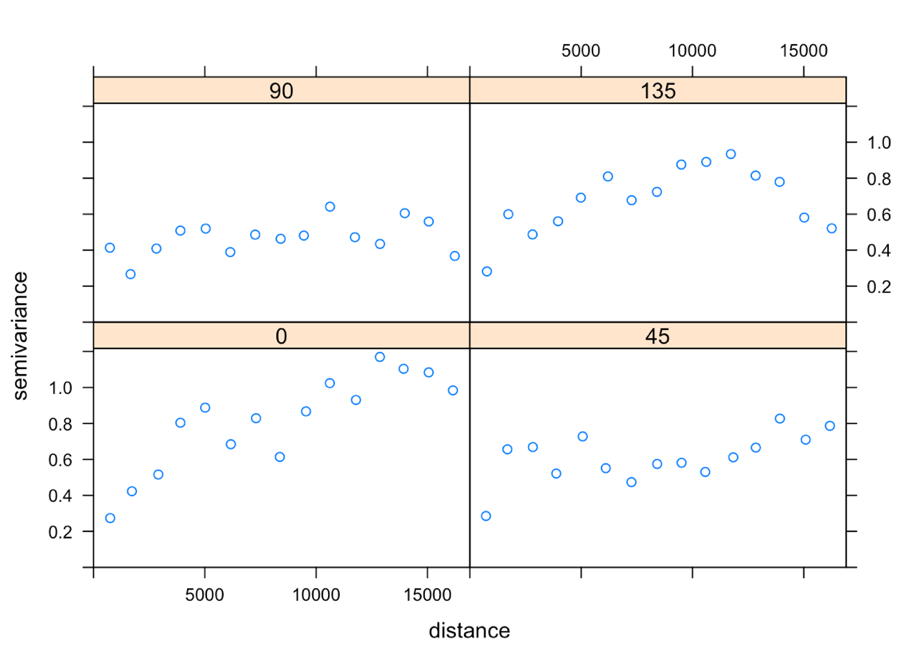


En el gráfico 3, se muestran los datos georreferenciados, en el primer gráfico se tienen los valores de potencial se dan en cinco grupos, donde se observa que el grupo de potencial más bajo presenta dos valores bastante alejados del resto. Esto puede ser de utilidad a la hora de realizar la interpolación. En el segundo gráfico, se observa la distribución espacial en tres grupos: bajo medio y alto; donde se observan datos, en su mayoría, de nivel medio. Además, en el tercer grafico (usando mapas de google) se puede observar que la distribución de puntos con valores mas altos se encuentra alrededor de la carretera que corre de norte a sur y los puntos con menores valores parecen seguir una forma mas o menos circular alrededor de los puntos altos.

**Isotropía**

Para conocer si la variación del valor del potencial con el espacio es igual en todas las direcciones, se utiliza el semivariograma en diferentes direcciones. Si esto ocurre decimos que la variable tiene un comportamiento isotrópico, caso contrario, la variación espacial es diferente en las distintas direcciones del espacio (anisotropía).

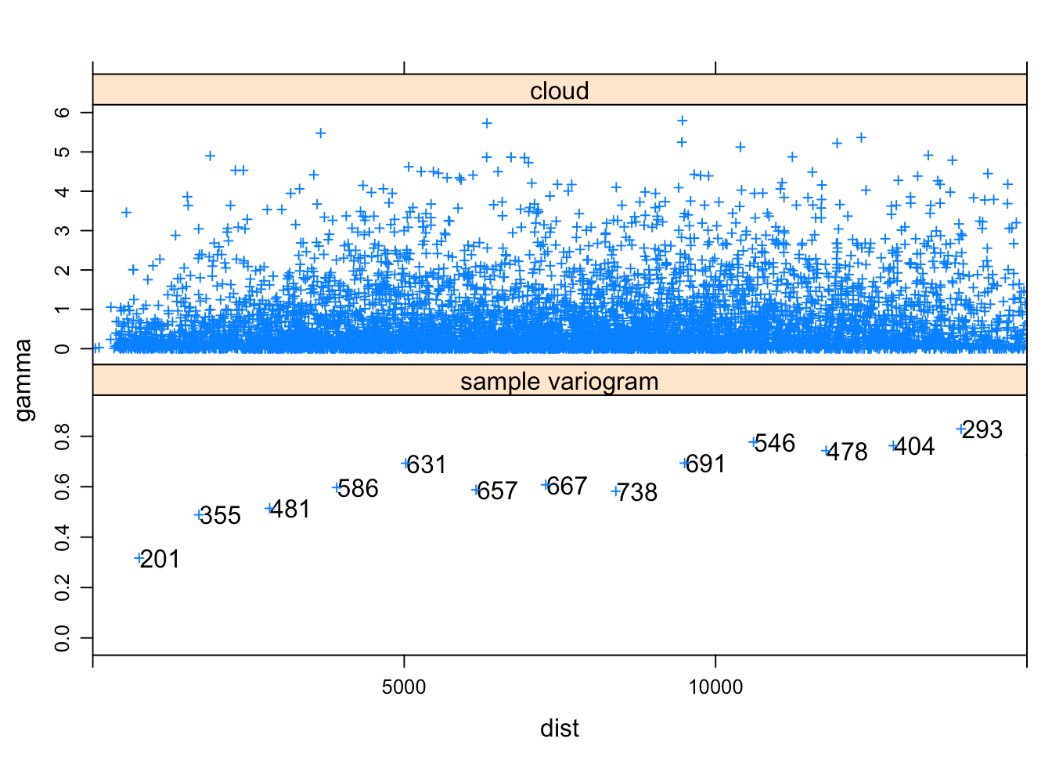
**Gráfico 4**. Semivariogramas direccionales



Al observar los semivariogramas no se obtienen sospechas de que exista anisotropía pues al menos en 3 de las direcciones se observa claramente como la curva ajusta de manera similar con los puntos del semivariograma, por lo cual no es necesario considerar varias direcciones del espacio (semivariogramas direccionales). Por otro lado, la presencia de los puntos alejados de muestreo puede influir en los variogramas por dirección.

El semivariograma del gráfico 4 proporciona bastante información del comportamiento espacial de la variable. Sin embargo, es necesario ajustar una función para cuantificar el grado y escala de variación espacial. Existen numerosos modelos que se utilizan en geoestadística, siendo los más comúnmente usados el modelo esférico, el modelo exponencial y el modelo matern. Estos modelos se pueden observar en el gráfico 5.

**Gráfico 5**. Modelajes de los semivariogramas



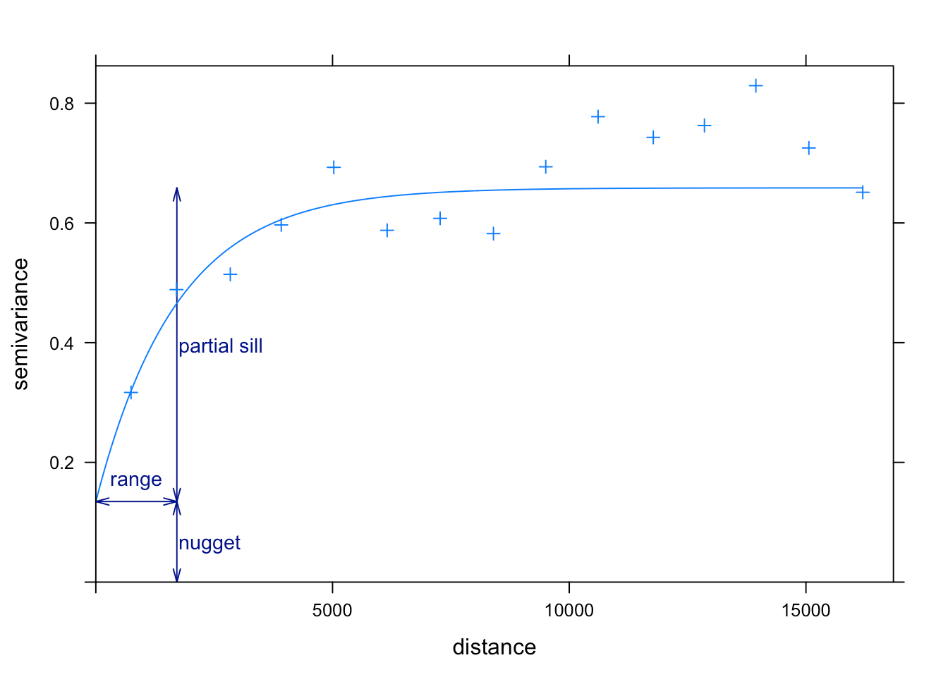
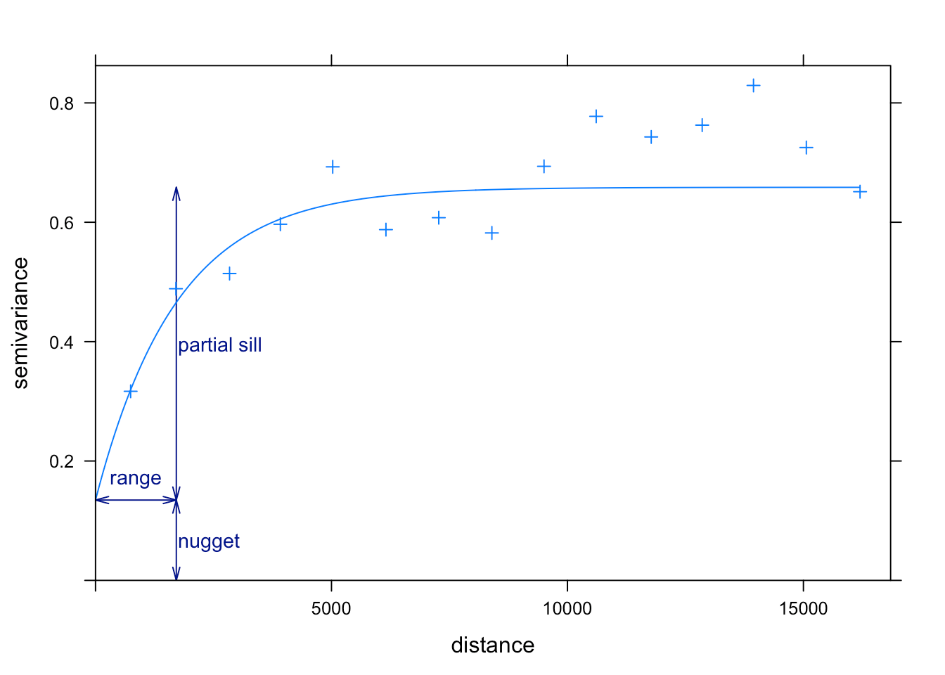
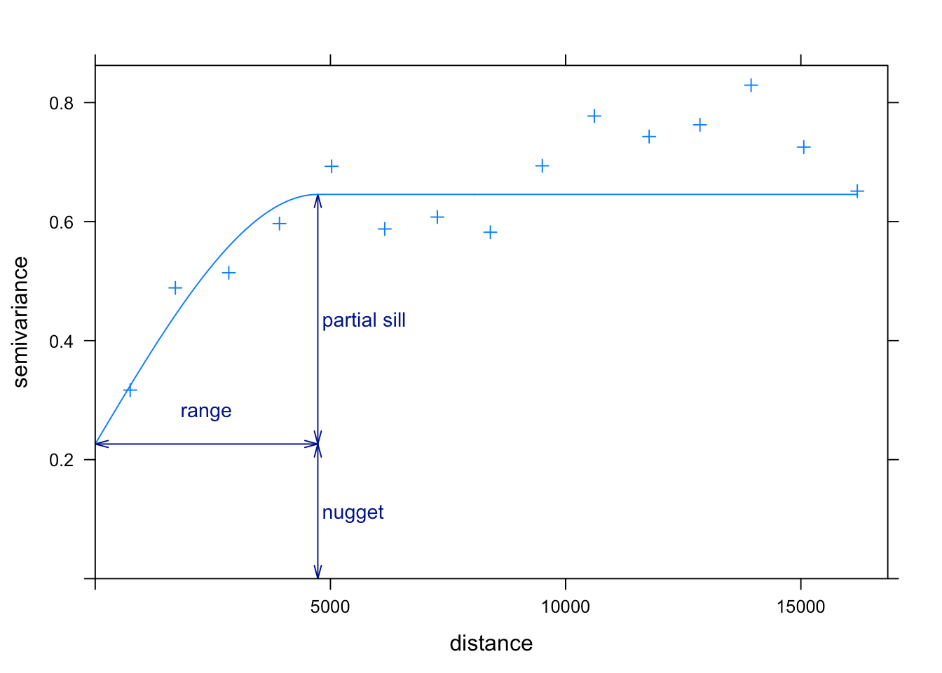
En el variograma general en conjunto con la nube de puntos se aprecia que hay una alta porción de variabilidad no explicada por el modelo pero que a partir aproximadamente de 5000 metros la variancia se estabiliza, es decir, el comportamiento de la variable analizada se considera independiente con respecto a su distribución espacial.

**Gráfico 6**. Modelajes de los semivariogramas

Esférico

Matern

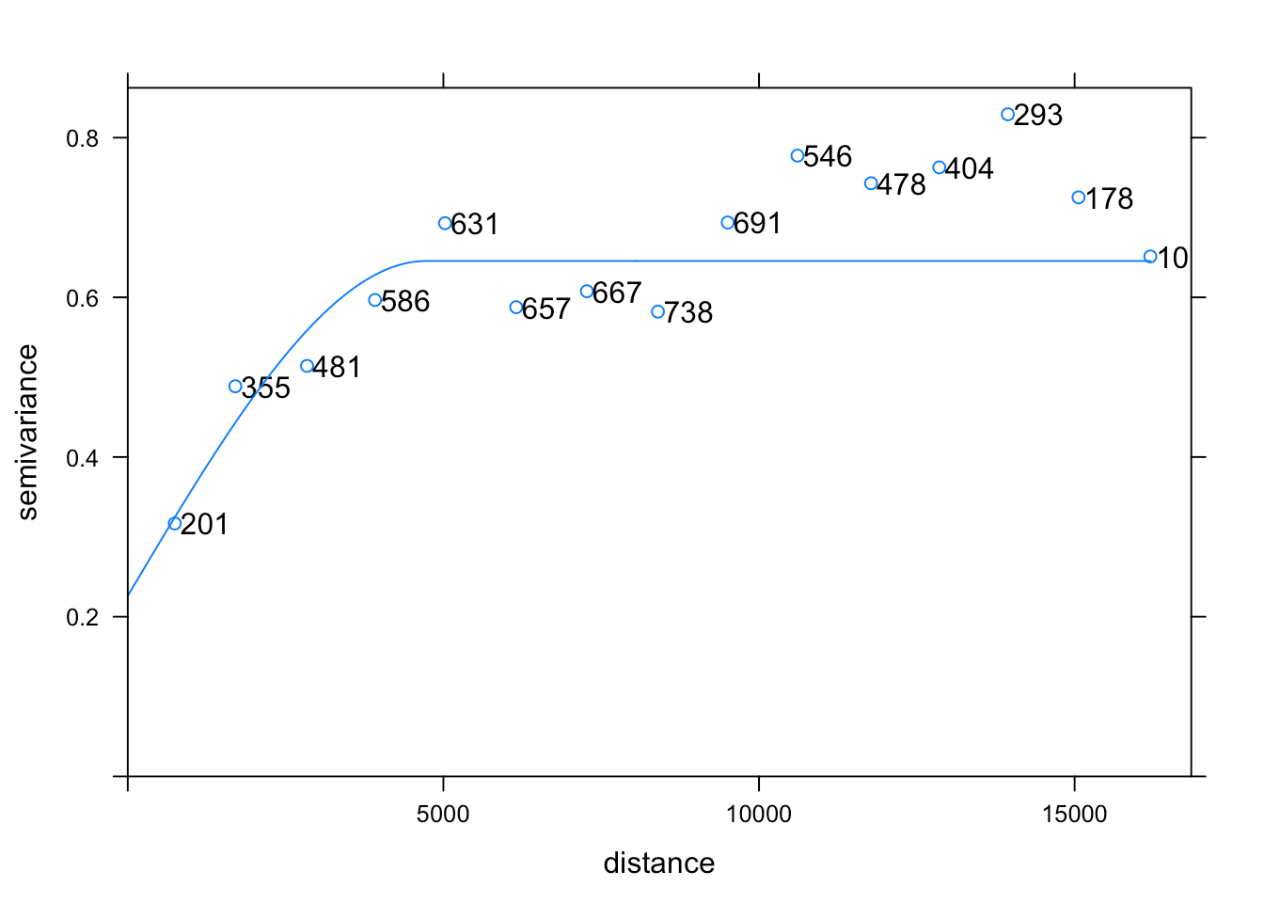
Exponencial



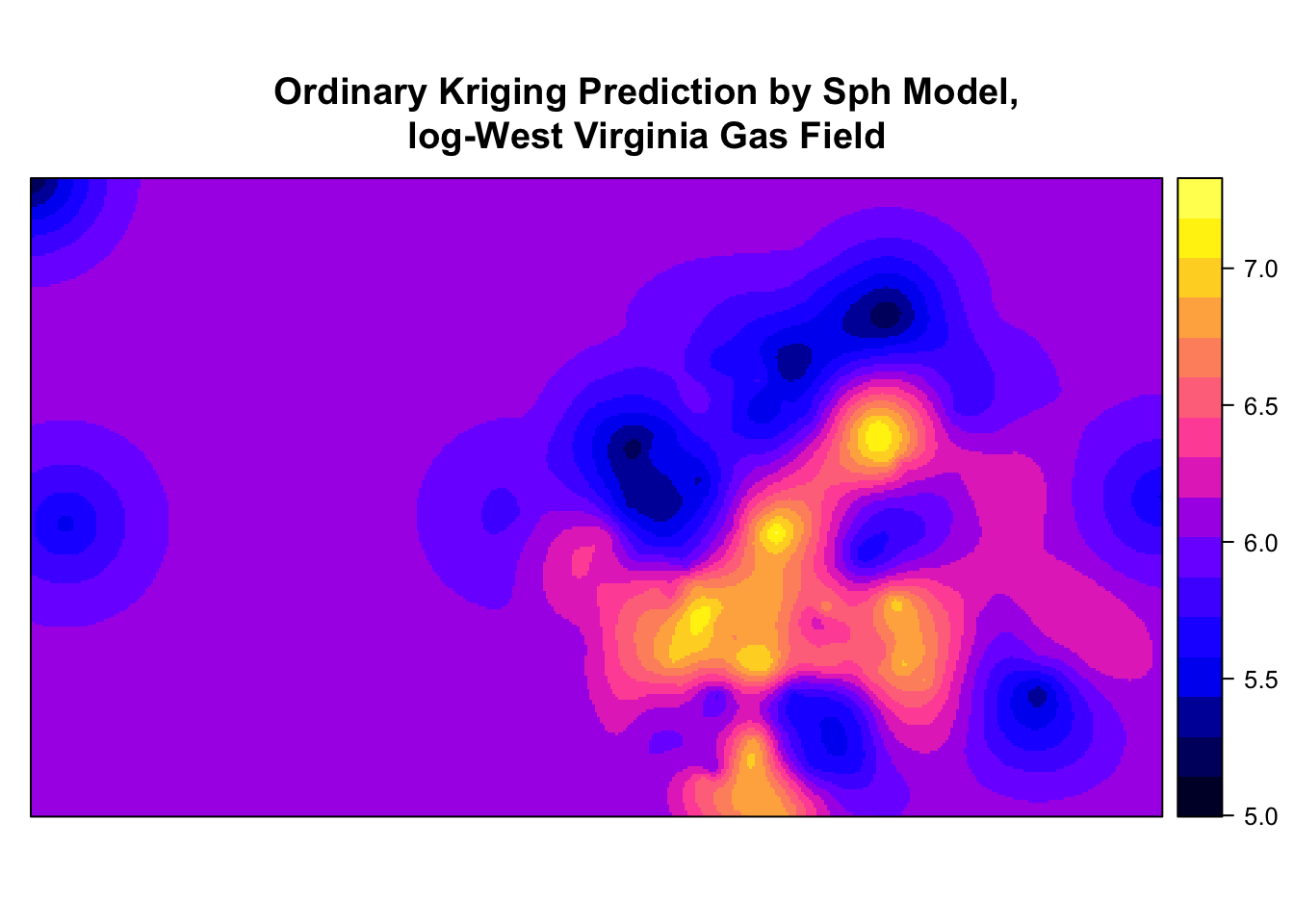
De acuerdo con los modelos utilizados se puede observar que el ajuste mediante el modelo esférico es el que más difiere con respecto a los demás; el matern y exponencial. La varianza no explicada por el modelo es mayor en el modelo esférico, mientras que en los demás es más bajo y similar. Un resultado similar se observa con la distancia a partir de la cual las muestras son espacialmente independientes unas de otras; en el modelo esférico esta distancia es de aproximadamente 5000 metros, mientras que en los modelos matern y exponencial, la distancia es de 2000 metros aproximadamente.

Por lo tanto, aunque el efecto nugget es mayor en el modelo esférico se aprecia que el rango en los otros 2 es muy bajo. Además, el sill o efecto de incertidumbre es mas alto en el modelo esférico comparado con el exponencial y el Matern con lo cual se puede hablar de hay menor incertidumbre en el modelo esférico y por ende es mas fácil predecir a la hora de hacer kriging.

**Gráfico 7**. Semivariograma Final

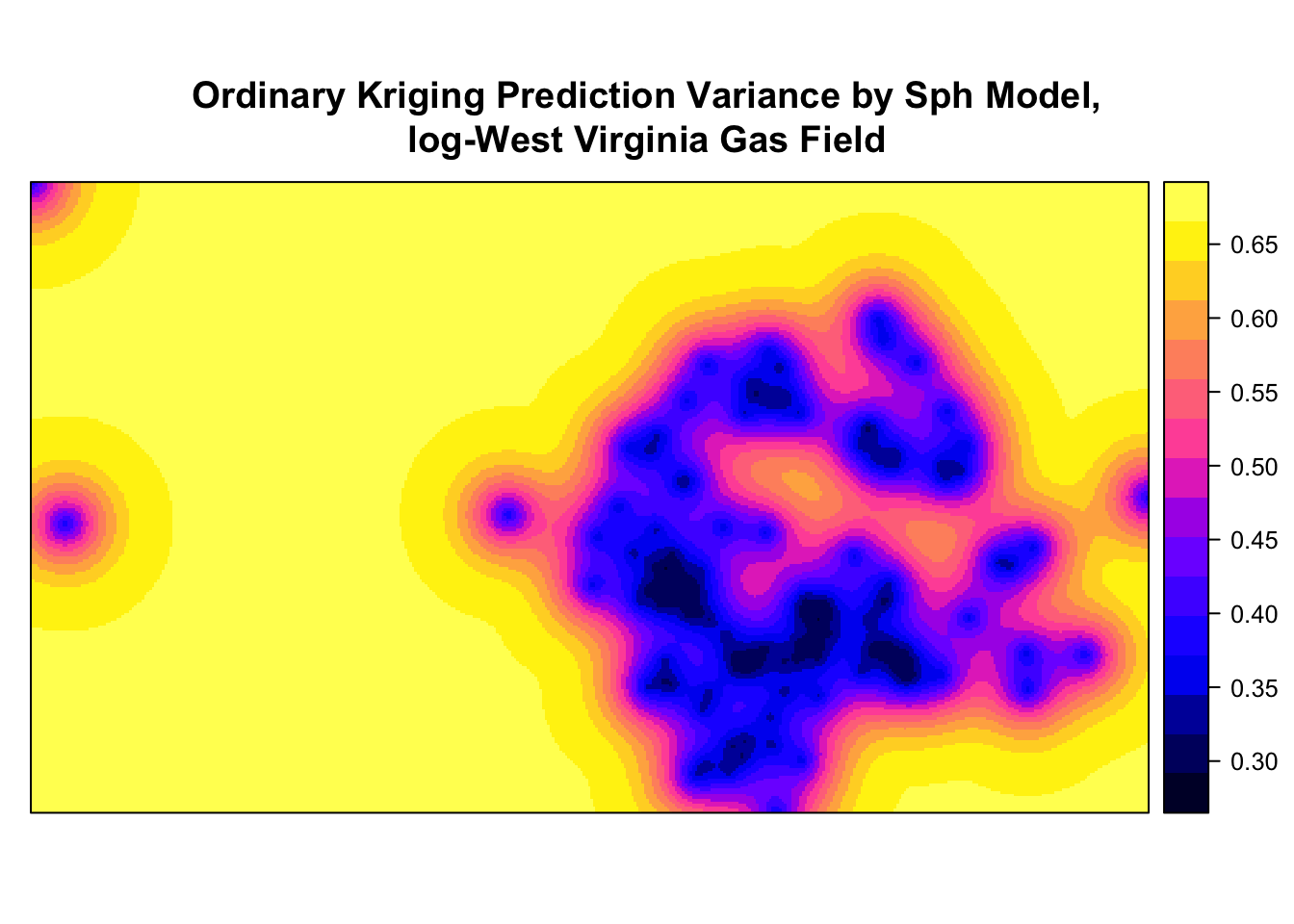


El ajuste mediante el semivariograma final permite extraer una serie de parámetros que son los que van a ser usados para la interpolación geoestadística (kriging). El rango (A0) o distancia a la que la semivarianza deja de aumentar, es de aproximadamente 5000 metros, a partir de esta distancia las muestras son espacialmente independientes unas de otras. La varianza no explicada por el modelo o nugget es de 0,23 aproximadamente. El parámetro sill, grado de variación espacial o de incertidumbre a la hora de interpolar puntos en el espacio es de 0,42.

**Gráfico 8**. Predicción de *kriging* por medio del modelo *esférico*

En el grafico 8 se puede apreciar la predicción mediante kriging del logaritmo de los valores iniciados de gas en el área analizada. Se puede observar que la concentración de valores altos se encuentra cerca de la zona intermedia del mapa en donde se pasa la carretera antes mencionada con varios focos claros de muy alta concentración de gas. Por otro lado, existen puntos claros en donde la concentración es baja con evidentes concentraciones de valores medios y bajos alrededor en la misma forma como se determinó en la exploración de datos, alrededor en forma circular a los puntos de concentración altos. Finalmente, debido a los puntos con valores bajos y la lejanía de estos, en medio se producen estimaciones cercanas a la media.

**Gráfico 9**. varianza de predicción de *kriging* por medio del modelo *esférico*



Tal y como se esperaba, la variancia alrededor de los puntos muestreados es muy baja y se puede hablar de muy buenas estimaciones con variancia minina a unos 10 o 15 kilómetros con respecto al centro de donde se encuentran los puntos. Por el contrario, fuera de este rango de 15 kilómetros, la variabilidad aumenta de forma significativa lo que se puede explicar por la gran distancia que existe entre las concentración de puntos muestreados en el centro y los 2 puntos al borde izquierdo del mapa.

**Conclusiones**

A partir del análisis realizado fue posible obtener un modelo para ajustar el método de interpolación de kriging utilizado. El cual permitió estimar las zonas en donde no se tenían mediciones de potencial inicial de gas y así conocer la intensidad del indicador de potencial inicial de gas en el campo del oeste de Virginia.

También fue posible observar en cuales zonas se obtuvo mejor capacidad de predicción; la varianza más pequeña se centró en el centro derecha del mapa, donde se observaron varianzas alrededor de 0,3 y 0,4.

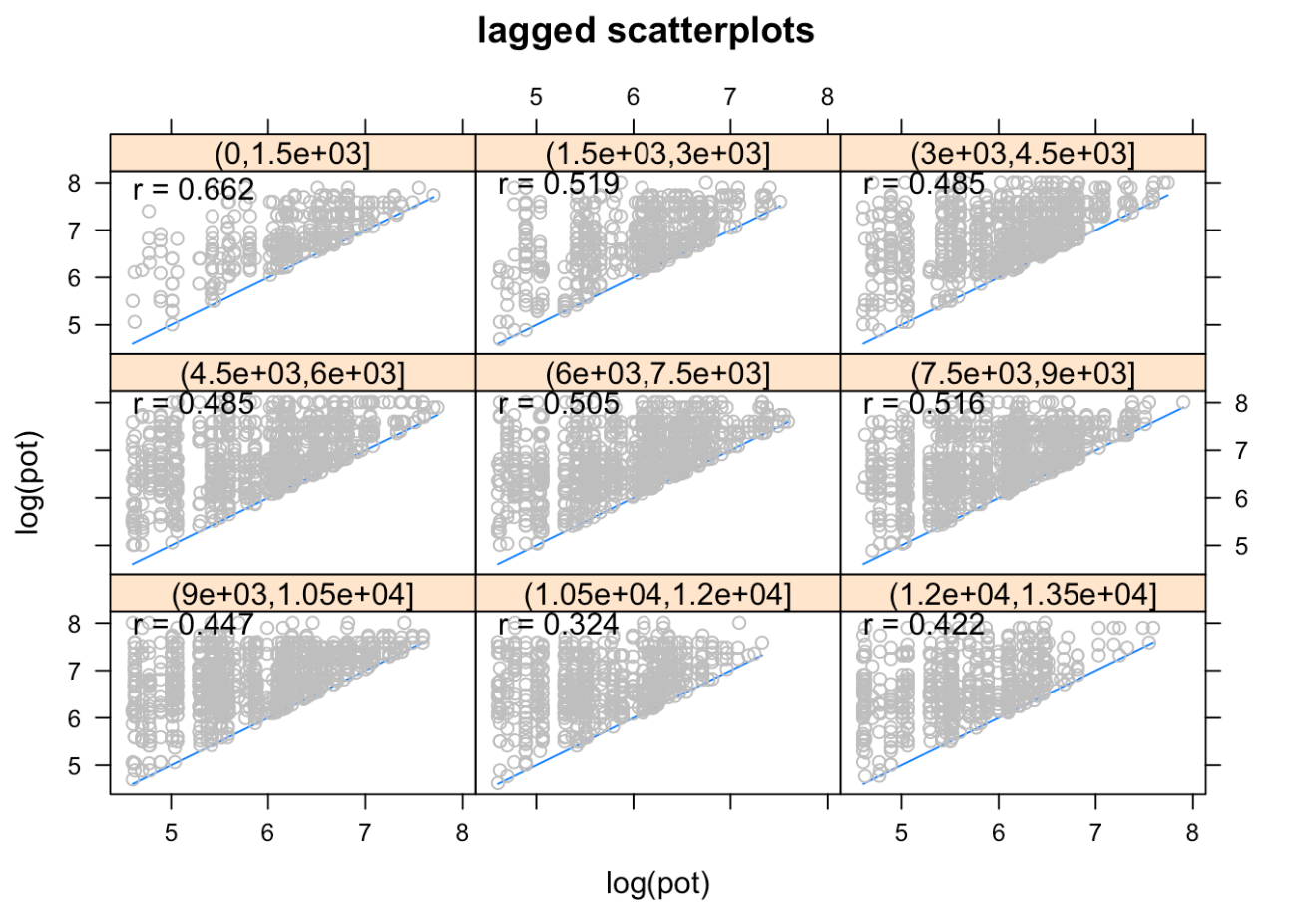
El mejor modelo para realizar la interpolación de kriging fue es esférico pues presentó menor incertidumbre y mas rango. Con este, se realizó la interpolación hacia toda la zona del mapa en donde al rededor de donde había mayor numero de puntos muestreados se obtuvieron valores apropiados con respecto a lo observado empíricamente con estimaciones cuya variancia era baja y en un rango de unos 15 kilómetros aproximadamente, pero como había tanta distancia entre ese grupo de observaciones y las 2 mas al oeste las estimaciones en este espacio no fueron las mejores.

Como observación y propuesta final, se sugiere que para futuros análisis de este tipo se realicen también mediciones que apoyen la estimación en regiones mas alejadas pues la enorme área que quedó sin muestrear no obtuvo las mejores estimaciones en términos de modelos geoestadísticos.

**Referencias**

1. Burgess, T.M. & Webster, R. 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. Journal of Soil Science, 31, 315-331.: Commentary on the impact of Burgess & Webster (1980a) by R.M. Lark, G.B.M. Heuvelink and T.F.A. Bishop
2. Cressie, N. Math Geology (1990). The origins of kriging. Kluwer Academic Publishers-Plenum Publishers 22(3), 239-252.
3. Hemyari, P., and Nofziger, D. L., 1987, Analytical solution for punctual kriging in one dimension: Soil Sci. Soc. Am. J., 51, p. 268–269.
4. Ord, J. K., 1983, Kriging, entry,in Encyclopedia of statistical sciences, vol. 4, S. Kotz and N. Johnson (Eds.): Wiley, New York, p. 411–413.
5. Paramá, R. 2006. Heterogeneidad Espacial de Nutrientes del Suelo en Ecosistemas Terrestres. Tesis Doctoral. Universidad de Vigo.

**Anexos**

**Anexo 1**. Gráficos de dispersión de los pares de observaciones

**Anexo 2.** Prueba para constatar que la señal no es “ruido”

