

Optimización de parámetros geoestadísticos para interpolación de lluvia para uso en el sector agropecuario.

Palabras clave: geoestadística, lluvia, optimización, hiperparámetros.

Introducción

La actividad agropecuaria es altamente dependiente del clima y estado del tiempo, principalmente de la variable lluvia. A su vez, dentro del sector agropecuario, la agricultura es un sector clima dependiente, potencialmente afectado por las consecuencias del clima, siendo la variabilidad climática interanual, la que presenta la principal incidencia relativa comparada con la variabilidad decadal y la de largo plazo (Tiscornia et al., 2016).

Tener mapas interpolados de lluvia es de gran importancia para la toma de decisiones y gestión de riesgos entre otras posibilidades. El objetivo del siguiente trabajo es evaluar diferentes técnicas de optimización de hiperparámetros como grid search, random search y optimización bayesiana para obtener imágenes interpoladas raster de lluvia con el menor error posible y hacerlo de manera automática.

Materiales y métodos

Para el siguiente trabajo se utilizó información de la red pluviométrica del Instituto Nacional de Meteorología (INUMET) de Uruguay para el mes de marzo de 2020. La red pluviométrica de INUMET cuenta con casi 350 pluviómetros distribuidos en todo el país. Del mes de marzo se evaluaron 3 días con lluvias: 11, 15 y 17 de marzo. De esos días se usaron solo los pluviómetros que contaran con datos. En la tabla 1 se ven la cantidad de pluviómetros usados cada día.

Fecha	11/03/2020	15/03/2020	17/03/2020
Pluviómetros	227	223	226

Tabla 1. Cantidad de pluviómetros usados en cada fecha evaluada.

Los paquetes R utilizados en este trabajo fueron gstat, mlrMBO, randomsearch y raster. Es importante aclarar que para el adecuado funcionamiento del paquete mlrMBO es necesario tener instalado el paquete rgenoud.

Para la generación de imágenes raster de precipitación se utilizó la técnica de interpolación conocida como kriging ordinario (Bivand, Pebesma, & Gómez-Rubio, 2013). En la técnica de kriging a partir de puntos georreferenciados con información observada para una variable de interés, en nuestro caso lluvia, se genera una grilla uniforme en donde cada celda cuenta con un valor obtenido por interpolación a partir de datos reales. Esta grilla uniforme tiene valores no solo para los puntos donde se conoce el valor de la variable, sino también para aquellos puntos donde no se tiene observaciones de la variable, es decir los valores son predichos.

Para poder realizar la interpolación es necesario definir valores a usar por determinados parámetros geoestadísticos como nugget, psill y range. Según los valores que se usen, el error que tenga el raster final. La búsqueda manual de los parámetros que generen el raster con el menor error de predicción es una tarea ardua, y prácticamente imposible por que las combinaciones de parámetros son infinitas.

En R la interpolación de lluvia devuelve 2 grillas raster, una con el valor predicho de lluvia para cada celda, y otra con la varianza de la predicción para cada celda.

El error de predicción del raster se evalúa a través del desvío estándar de lluvia definido como la raíz del promedio de la varianza de lluvia de las celdas del raster final. El objetivo final es generar un raster de lluvia con el menor desvío estándar.

En R se evaluaron las técnicas de optimización de hiperparámetros conocidas como grid search (GS), random search (RS), y optimización bayesiana (OB) (Feurer & Hutter, 2019). Estas técnicas de optimización buscan encontrar de manera automática aquellos parámetros geoestadísticos que minimicen una función objetivo, en este estudio el desvío estándar.

Las técnicas de optimización requieren que uno defina de antemano para cada variable a ajustar, el rango de búsqueda, es decir entre que valores mínimos y máximos puede estar cada variable, cuando es numérica, y que valores puede tomar si es discreta. En la tabla 2 se muestra el rango de búsqueda para cada parámetro geoestadístico.

Parametro	Psill	Range	Nugget
Rango	1 a 100	1 a 200	1 a 100

Tabla 2. Rango de búsqueda para cada parámetro geoestadístico.

Resultados y discusión

A continuación, a modo de ejemplo se presentan los resultados obtenidos para las diferentes técnicas de optimización solo para el día 15/03/2020. En la tabla 3 se ven los parámetros que dieron como resultado el desvío estándar más bajo en la etapa de interpolación para cada técnica de optimización.

OPTIMIZACIÓN	CORRIDAS	NUGGET	PSILL	RANGE	DESVÍO ESTÁNDAR	TIEMPO DE EJECUCIÓN (segundos)
BAYESIANA	125	1.00	1.02	196.30	1.19	306
GRID SEARCH		23.20	18.70	196.80	5.61	50
RANDOM SEARCH		1.48	5.66	159.89	1.90	52

Tabla 3. Mejores corridas según técnica de optimización para el día 15/03/2020.

Como se ve en la tabla 3, las técnicas evaluadas obtuvieron valores diferentes de desvío estándar en la etapa de interpolación, siendo optimización bayesiana la que tuvo mejor desempeño con el desvío más bajo, seguida por random search. En cuanto al tiempo de ejecución optimización bayesiana es la que lleva más tiempo demorando 6 veces más en su ejecución en comparación a las otras técnicas.

Conclusiones

Es posible usando diferentes técnicas de optimización de hiperparámetros obtener un raster de interpolación de lluvia con valores de desvío estándar para lluvia diferentes a partir de diferentes valores de parámetros geoestadísticos. Optimización bayesiana fue la técnica que obtuvo el desvío estándar más bajo. Sin embargo, hay que tener en cuenta que el resultado mostrado es sobre una sola fecha, pudiendo en otras obtener resultados distintos. A futuro es necesario seguir explorando y profundizando el conocimiento de estas técnicas para obtener mapas de lluvia con menor incertidumbre para uso en el sector agropecuario.

Bibliografía

- Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). Interpolation and Geostatistics. En *Applied Spatial Data Analysis with R* (pp. 213-261). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-7618-4_8
- Feurer, M., & Hutter, F. (2019). Hyperparameter Optimization. En F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), *Automated Machine Learning: Methods, Systems, Challenges* (pp. 3-33). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-05318-5_1
- Tiscornia, G., Cal, A., & Giménez, A. (2016). Análisis y caracterización de la variabilidad climática en algunas regiones de Uruguay. *RIA. Revista de investigaciones agropecuarias*, 42(1), 66-71.