

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) para el mes de marzo de 2020. La red pluviométrica de INUMET cuenta con casi 350 pluviómetros distribuidos en el territorio nacional. 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.

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 width, cutoff, range, nugget, psill, y el modelo de variograma (esférico, gaussiano, exponencial, etc). 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.

El error de predicción del raster se evalúa a través de la raíz del error cuadrático medio (rmse) entre valores observados (pluviómetros) y valores predichos por el raster. El objetivo final es generar un raster de lluvia con el menor rmse posible.

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 rmse.

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	Width	Cutoff	Psill	Range	Nugget	Modelo
Rango	1 a 100	1 a 1000	1 a 3000	1 a 1000	1 a 500	Exponencial, Esférico, Gausiano, Lineal

Tabla 2. Rango de búsqueda para cada parámetro geoestadístico.

Los datos de lluvia para cada día fueron divididos en entrenamiento y validación en 80 y 20% respectivamente. Los algoritmos de optimización se aplicaron sobre los datos de entrenamiento.

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 rmse más bajo en la etapa de entrenamiento para cada técnica de optimización para el día 15/03/2020.

OPTIMIZACIÓN	MODELO	WIDTH	CUTOFF	PSILL	RANGE	NUGGET	Entrenamiento	Validación
							RMSE (mm)	RMSE (mm)
GRID SEARCH	Lineal	2.2	757	1327.3	692.1	10.3	4.82	7.53
RANDOM SEARCH	Esférico	1.2	883.5	161.9	535.5	45.1	4.39	7.25
BAYESIANA	Gausiano	4.2	601.6	1497.6	967.0	49.3	4.52	9.51

Tabla 3. Corridas con valores más bajos de rmse en entrenamiento para el día 15/03/2020.

Como se ve en la tabla 3, las técnicas evaluadas obtuvieron valores similares de rmse en la etapa de entrenamiento, siendo random search lo que tuvo mejor desempeño, con modelos y valores geoestadísticos diferentes. En la etapa de validación, random search se desempeñó mejor que las otras técnicas. Optimización bayesiana, a pesar de ser la técnica más moderna, se desempeñó peor, con un rmse sensiblemente más alto. Es importante aclarar que, de las técnicas evaluadas, optimización bayesiana tiene un tiempo de ejecución mucho más alto que grid search y random search. En la etapa de entrenamiento los peores valores de rmse obtenidos para grid search, random search, y optimización bayesiana fueron 7.0, 9.3, 7.3, respectivamente.

Conclusiones

Es posible usando diferentes técnicas de optimización de hiperparámetros obtener un raster de interpolación de lluvia con valores de rmse similares a partir de diferentes valores de parámetros geoestadísticos, aunque algunas técnicas se desempeñaron mejor que otras en la etapa de validación, como el caso de random search. Sin embargo, hay que tener en cuenta que el resultado mostrado es sobre una sola fecha, pudiendo en otras obtener resultados distintos. De todas maneras, cualquiera de las 3 técnicas permite minimizar el error de la predicción de valores en torno a 7 a 9 mm a valores inferiores a 5 mm en entrenamiento. 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.