

# Declaración de investigación: Aprendizaje espacial de pocas muestras (few-shot) para problemas con datos georreferenciados escasos

## 1 Introducción y motivación

El presente trabajo se centra en el aprendizaje con pocas muestras aplicado a tareas espaciales en las que solo se dispone de datos georreferenciados limitados. Este tema es relevante porque el análisis espacial se ha vuelto esencial en campos como la geología, la ecología, la epidemiología y la planificación urbana, y actualmente su uso se ha manifestado en prácticamente todas las industrias [Pierdicca et al., 2022]. En efecto, la disponibilidad masiva de datos geoespaciales (por ejemplo imágenes satelitales, registros de sensores, etc.) contrasta con la dificultad de obtener etiquetas o mediciones de campo. Como señalan expertos, la etiquetación manual de imágenes georreferenciadas es costosa y *frequently discourages the adoption of geospatial deep learning* [Zhu et al., 2017]. Este esfuerzo titánico disuade a instituciones y empresas de explotar plenamente el aprendizaje profundo en contextos geográficos.

Además, el análisis espacial requiere datos de posición precisos: son de vital importancia los datos de posición, distancia y la interacción con el espacio [Cressie, 1993, Anselin, 1995] para cualquier estudio cuantitativo de fenómenos geográficos. Sin embargo, en zonas remotas o poco muestreadas estos datos son escasos o de baja calidad [Yuan et al., 2023, Zhang et al., 2024], lo que limita la efectividad de métodos tradicionales. En la teledetección, por ejemplo, se reportan desafíos como la *inadequate field validation data in remote areas* [Ma et al., 2023], evidenciando la falta de muestras en regiones de interés.

El interés por este problema surge de la brecha existente en la literatura y la importancia práctica del mismo. Mientras las técnicas convencionales –tales como modelos de procesos gaussianos (kriging) y redes neuronales profundas– requieren grandes volúmenes de datos etiquetados [Rasmussen and Williams, 2006, Stein, 1999], los enfoques de few-shot learning prometen superar esta limitación. En particular, pocos trabajos han explorado cómo adaptar aprendizaje de pocas muestras al ámbito espacial: las investigaciones recientes muestran que los métodos que funcionan con datos masivos *struggle in data-scarce scenarios* [Lu et al., 2020, Song et al., 2022], lo que subraya la necesidad

de explorar nuevas soluciones. Así, estudiar cómo aplicar few-shot learning en problemas georreferenciados puede aportar conocimiento novedoso y relevante tanto académica como profesionalmente. Este trabajo busca contribuir a comprender cómo estos modelos pueden generalizar en tareas espaciales con datos limitados.

## 2 Planteamiento del problema

Los métodos actuales de predicción espacial suelen basarse en enfoques clásicos o supervisados. Por ejemplo, en estadística geoespacial es común usar procesos gaussianos (kriging) para interpolar valores en el espacio, pero esto requiere muchas observaciones para lograr alto rendimiento [Stein, 1999]. De manera similar, las redes neuronales profundas en teledetección o análisis de imágenes de satélite exigen bases de datos extensas y etiquetadas. Sin embargo, en aplicaciones reales a menudo solo contamos con pocos puntos muestreados; esta disparidad limita la precisión de los métodos tradicionales.

Por otro lado, el campo de few-shot learning ha avanzado principalmente en visión por computador, donde los modelos aprenden a reconocer nuevas clases con muy pocos ejemplos mediante meta-aprendizaje [Finn et al., 2017]. En contraste, existe un vacío de estudios específicos para datos georreferenciados. Aunque hay trabajos pioneros (por ejemplo, modelos recientes adaptan el aprendizaje de pocas muestras a regresión espacial [Li et al., 2020]), todavía faltan análisis exhaustivos que comparen estos enfoques con técnicas convencionales. En concreto, no está bien documentado cómo las propiedades espaciales (correlación con la distancia, anisotropía, etc.) afectan la eficacia de algoritmos de few-shot. Además, los estudios disponibles sugieren que modelos entrenados con datos masivos logran buen desempeño con grandes conjuntos etiquetados, pero tienen problemas de generalización en escenarios con pocos datos [Vinyals et al., 2016, Snell et al., 2017]. Este vacío justifica realizar un estudio que explore detalladamente la aplicabilidad de métodos de aprendizaje de pocas muestras en problemas de predicción espacial con datos escasos, con el fin de valorar su desempeño y límites en comparación con métodos tradicionales.

## 3 Objetivos y preguntas de investigación

**Objetivo general:** Investigar la viabilidad y el desempeño de modelos de aprendizaje de pocas muestras (few-shot learning) para tareas de regresión o clasificación espacial sobre datos georreferenciados cuando la cantidad de datos de entrenamiento es muy limitada.

**Objetivos específicos:**

- Revisar la literatura sobre enfoques de few-shot learning y métodos estadísticos espaciales, identificando sus fortalezas y limitaciones en entornos con datos escasos.

- Diseñar o adaptar un modelo de few-shot learning (por ejemplo, usando meta-aprendizaje o redes siamesas) para predecir valores espaciales con pocos datos.
- Implementar experimentos cuantitativos utilizando conjuntos de datos georreferenciados (reales o sintéticos) con muestras limitadas, evaluando la precisión del modelo propuesto frente a técnicas convencionales (e.g., kriging o redes entrenadas con pocos datos).

#### Preguntas de investigación:

- ¿Cómo afecta la aplicación de técnicas de aprendizaje de pocas muestras a la precisión de la predicción espacial con datos georreferenciados limitados?
- ¿Qué métodos específicos (p. ej., meta-aprendizaje, máquinas de soporte vectorial ajustadas, procesos gaussianos adaptados) resultan más efectivos en escenarios de datos escasos?
- ¿Cómo influyen las características espaciales (distribución, autocorrelación) en el rendimiento de los modelos few-shot en comparación con modelos tradicionales?

## 4 Metodología propuesta

La investigación se desarrollará bajo un enfoque mixto. En primer lugar, se efectuará una revisión bibliográfica exhaustiva de estudios en GeoAI, teledetección y meta-aprendizaje aplicados a datos espaciales. Esta etapa cualitativa permitirá contextualizar el problema y elegir técnicas adecuadas. En paralelo, se adoptará un enfoque cuantitativo: se diseñará e implementará un modelo de few-shot learning para datos georreferenciados. Por ejemplo, aprovechando el concepto de meta-aprendizaje, donde el modelo aprende a aprender identificando similitudes entre ejemplos [Sung et al., 2018], se entrenará una red neuronal que genere representaciones de tareas espaciales con pocos puntos de apoyo.

En la práctica, se planifica:

- **Recolección de datos:** Selección de conjuntos georreferenciados relevantes (por ejemplo, mediciones ambientales o imágenes satelitales con etiquetas puntuales) con número limitado de muestras etiquetadas, como la Encuesta Nacional Agropecuaria (ENA) [Instituto Nacional de Estadística e Informática (INEI), 2022].
- **Desarrollo de algoritmos:** Implementación de un algoritmo few-shot (por ejemplo, Prototypical Networks o redes siamesas adaptadas) y, si procede, integración con técnicas espaciales (p. ej. redes con capas de convolución espacial o procesos gaussianos adaptativos [Rasmussen and Williams, 2006]).

- **Experimentación:** Entrenar y evaluar el modelo en múltiples escenarios de prueba con pocos datos, comparándolo con métodos tradicionales (kriging, regresión lineal, etc.). Se medirán métricas de desempeño (error de predicción, etc.) para cuantificar la ganancia del enfoque few-shot.

## 5 Impacto y proyección

Se espera que esta investigación contribuya académicamente a llenar un vacío en el campo de GeoAI y aprendizaje automático. Al estudiar aplicabilidad y límites de modelos few-shot en tareas espaciales con datos limitados, el trabajo brindará evidencia empírica sobre su eficacia, generando discusión sobre cómo integrar meta-aprendizaje en problemas georreferenciados. Además, podría derivar recomendaciones prácticas para analistas GIS y especialistas en teledetección, al mostrar qué métodos funcionan mejor cuando la recolección de datos es difícil o costosa.

En el ámbito práctico, los resultados podrían facilitar el monitoreo y la modelación en zonas remotas (donde validar datos en terreno es complicado [Ma et al., 2023]), así como en proyectos de planificación urbana o medioambiental con muestras escasas. Por último, en el contexto académico del curso, esta investigación integrará conocimientos de inteligencia artificial y análisis geoespacial, fortaleciendo la formación del equipo al aplicar técnicas avanzadas (meta-aprendizaje, estadística espacial) a un problema real. Constituirá así un primer paso que permitirá explorar en el futuro mejoras y extensiones, como la generalización de modelos a diferentes regiones o la inclusión de datos temporales.

## References

- [Anselin, 1995] Anselin, L. (1995). Local indicators of spatial association — lisa. *Geographical Analysis*, 27(2):93–115.
- [Cressie, 1993] Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, revised edition.
- [Finn et al., 2017] Finn, C., Abbeel, P., and Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 1126–1135.
- [Instituto Nacional de Estadística e Informática (INEI), 2022] Instituto Nacional de Estadística e Informática (INEI) (2022). Encuesta nacional agropecuaria (ena) — microdatos y documentos metodológicos. Ficha técnica, cuestionarios y microdatos disponibles en el portal de INEI / datos abiertos.

- [Li et al., 2020] Li, H., Cui, Z., Zhu, Z., Chen, L., Zhu, J., Huang, H., and Tao, C. (2020). Rs-metanet: Deep meta metric learning for few-shot remote sensing scene classification. *arXiv preprint*.
- [Lu et al., 2020] Lu, J., Gong, P., Ye, J., Zhang, J., and Zhang, C. (2020). A survey on machine learning from few samples. *arXiv preprint*.
- [Ma et al., 2023] Ma, Y. et al. (2023). Dictionary learning for few-shot remote sensing scene classification. *Remote Sensing*, 15(3):773.
- [Pierdicca et al., 2022] Pierdicca, R. et al. (2022). Geoai: a review of artificial intelligence approaches for the analysis of geomatics data. *Geoscience and Remote Sensing / Copernicus (GI)*.
- [Rasmussen and Williams, 2006] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.
- [Snell et al., 2017] Snell, J., Swersky, K., and Zemel, R. S. (2017). Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- [Song et al., 2022] Song, Y., Wang, T., Mondal, S. K., and Sahoo, J. P. (2022). A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *arXiv preprint*.
- [Stein, 1999] Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.
- [Sung et al., 2018] Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H. S., and Hospedales, T. (2018). Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1199–1208.
- [Vinyals et al., 2016] Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., and Wierstra, D. (2016). Matching networks for one shot learning. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- [Yuan et al., 2023] Yuan, Z. et al. (2023). Few-shot remote sensing image scene classification. *Remote Sensing*, 15(3):831.
- [Zhang et al., 2024] Zhang, J. et al. (2024). Few-shot object detection for remote sensing imagery. *Remote Sensing*, 16(19):3630.

[Zhu et al., 2017] Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., and Fraundorfer, F. (2017). Deep learning in remote sensing: A review. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36.