Natural Language to DGGS-Aware Methane Insights with a Multi-LLM-Agent Framework

Mingke Erin Li[1], Steve H.L. Liang[2]

[1]Department of Geomatics Engineering, Schulich School of Engineering, University of Calgary, Calgary, T2N 1N4, Canada, mingke.li@ucalgary.ca
[2]Department of Geomatics Engineering, Schulich School of Engineering, University of Calgary, Calgary, T2N 1N4, Canada, liangs@ucalgary.ca

ABSTRACT

Fragmented spatial data and inconsistent semantics hinder environmental analytics and confound large language models. We present a Discrete Global Grid System powered framework that harmonizes gridded methane inventories and enables explainable, traceable natural language analytics. Demonstrating the methane use case, we standardize spatial and semantic references using DGGS as a global index, then allow the interaction with a modular multi-agent system. Agents parse intent, resolve locations to DGGS cells, and generate validated SQL that powers summaries and maps with cell-level citations. The result is fast, reproducible methane analysis with clear provenance.

1. Introduction

Spatial data needed for environmental decision-making is scattered across platforms and arrives with incompatible coordinate systems, formats, resolutions, time snapshots, and even sector-specific vocabularies. This fragmentation creates long preprocessing cycles and introduces avoidable errors, which slow action before, during, and after events (Thompson et al., 2023). It also confounds large language models, which many users now rely on to ask spatial questions (Wang et al., 2024; Zhang et al., 2024). When inputs are inconsistent, models hesitate or invent details, so the central requirement is not better prompting but standardization that makes answers explainable and traceable.

We address this requirement by using a Discrete Global Grid System (DGGS) as the standard framework for data and analytics. A DGGS provides a global, hierarchical, cell-based spatial index with persistent identifiers, enabling consistent spatial referencing across time, scale, and data sources (Li & Stefanakis, 2020; Sahr et al., 2003). In this study, we adopt rearranged Hierarchical Equal Area isoLatitude Pixelization (rHEALPix; shown in Figure 1), which provides an equal-area global index with fixed cell locations that support consistent change analysis and a hierarchical structure that enables both zooming and aggregation (Gibb, 2016). Coupled with a standard schema for features and attributes, the grid anchors each value to a stable identifier. When users ask spatial questions, the system connects answers to specific cells and the sources, turning opaque outputs into auditable results.

Volume DOI: 10.5281/zenodo.20598496

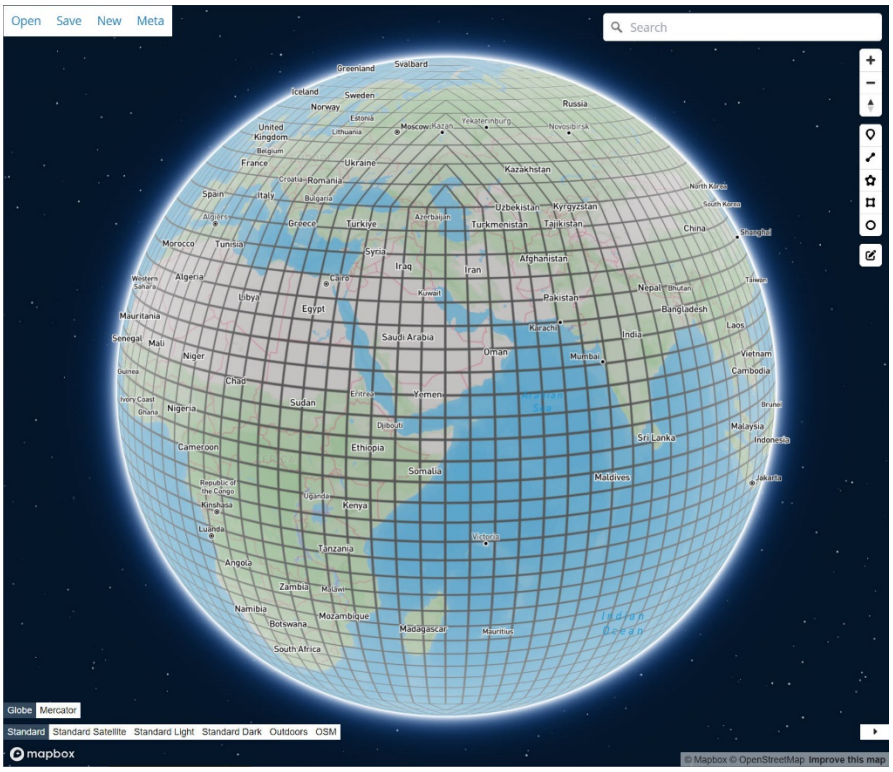

Figure 1: Overview of rHEALPix DGGS at level 3.

Several alternative spatial indexing systems are widely used in industry, including Uber's H3 and Google's S2 (Uber, 2017; Veach et al., 2017). While these systems are effective for tiling and indexing, they are not equal area, which complicates quantitative comparison and aggregation for environmental variables such as emissions. For this reason, they are not adopted in this study. A detailed rationale for selecting rHEALPix, including its equal area properties and suitability for environmental analysis, is provided in our previous work (Li & Liang, 2026).

The research problem is to transform multi-source environmental data into a form that large language models can query reliably while keeping answers fully traceable to their spatial and data origins, demonstrated through a methane emissions use case. We show how a DGGS can harmonize heterogeneous methane inventories into a single analytical substrate, how a multi-agent language system can reason over space, time, and sector semantics while grounded in indexed cells, and how the pair supports end-to-end methane analytics from national aggregation to basin-level hotspot detection with explicit provenance.

2. Methods & Data

2.1 Data sources

This study synthesizes 13 open-source, bottom-up gridded methane inventories published in peer-reviewed journals or official repositories. Coverage spans global products (Crippa et al., 2023; Scarpelli, Jacob, Grossman, et al., 2022), a continental product for Europe (Kuenen et al., 2022), national products for the United States, Canada, Mexico, China, Switzerland, India, and Australia (Gong & Shi, 2021; Hiller et al., 2014; Maasakkers et al., 2023; Sadavarte et al., 2022; Scarpelli, Jacob, Moran, et al., 2022; Scarpelli et al., 2020), and a high-resolution subnational

Volume DOI: 10.5281/zenodo.20598496

product for New York State (Loman et al., 2025). We emphasize energy-related methane from coal mining and from oil and gas systems, given their dominant role in anthropogenic budgets and their direct relevance for mitigation.

2.2 Data harmonization pipeline

The collected inventories differ in native resolution, coordinate reference systems, sector schemes, formats, and reporting units. For backend development, we standardized spatial reference, units, and semantics by converting all sources to the rHEALPix DGGS. Figure 2 illustrates the workflow from the original inventories to the DGGS outputs. Inputs in NetCDF, CSV, and GeoTIFF are first rendered as intermediate rasters, then redistributed to rHEALPix cells by area-weighted intersection. The Python library 'uraster' was developed to accomplish this task efficiently (Liao et al., 2026). DGGS levels are selected to best match the native resolution of each source inventory, thereby preserving the original spatial information content and avoiding artificial refinement during harmonization.

Aggregation across scales is performed at query time rather than during data preparation. Cell values are summarized on the fly through the DGGS hierarchy according to the spatial extent of the user query, which is typically much coarser than the native resolution of the source inventories. Because DGGS cells retain explicit hierarchy levels, the effective resolution and associated spatial uncertainty remain transparent to users.

Because original inventories report in mixed units, we convert all values to Mg a^{-1} per cell to ensure comparability across the DGGS. Sector semantics are harmonized with the Intergovernmental Panel on Climate Change (IPCC) 2006 hierarchy. The resulting harmonized DGGS methane dataset is available at https://doi.org/10.5281/zenodo.17362125, supporting reproducibility and reuse.

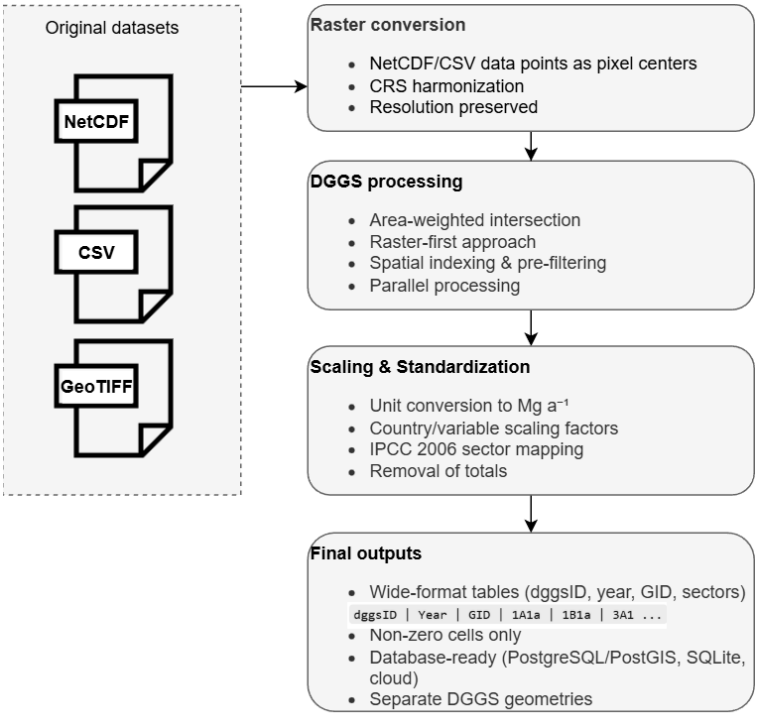

Figure 2: Workflow of converting original gridded methane inventories to rHEALPix DGGS.

2.3 Multi-agent query system

For frontend development, we implement a modular system that converts natural language questions into DGGS-aware analytics and returns concise answers. The system combines multiple LLM calls with deterministic steps, where agents communicate through LangGraph, providing a directed workflow with explicit state at each step (Figure 3).

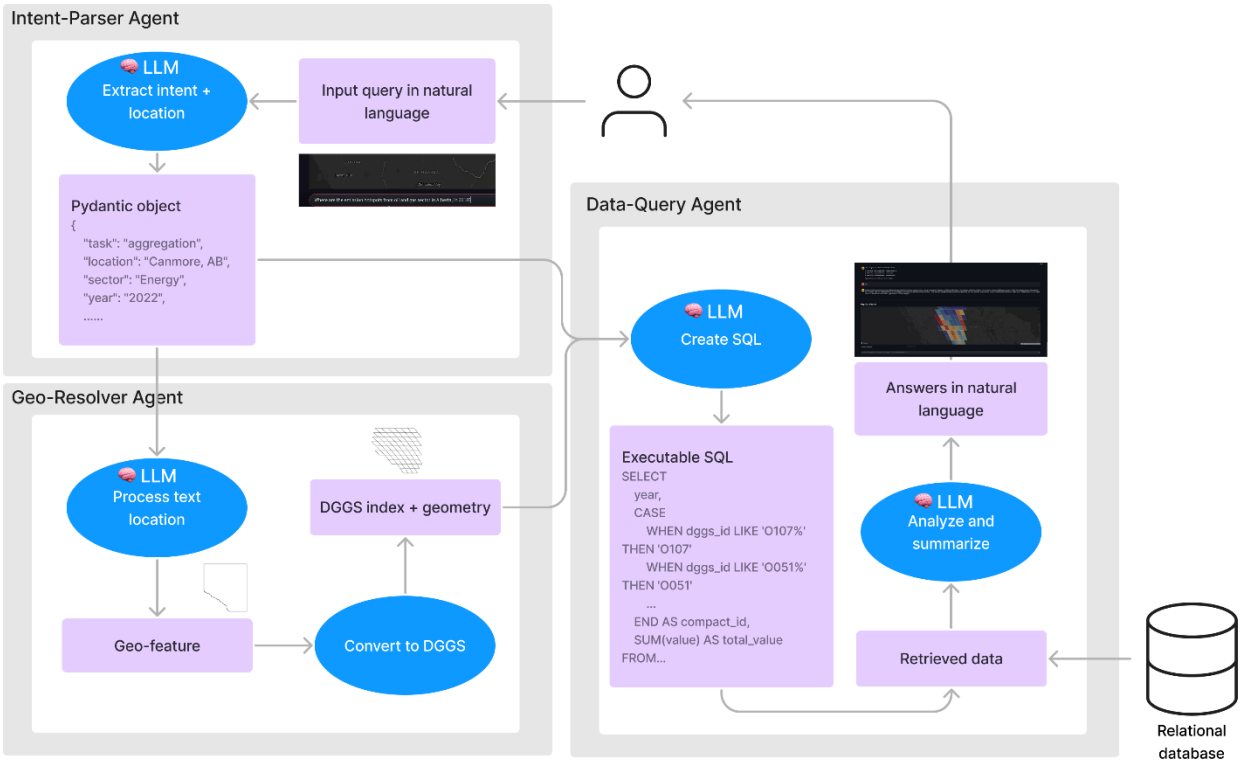

Figure 3: Architecture of the multi-agent query system.

2.3.1 Intent-parser agent

The intent-parser agent converts an input question into a structured request. A compact schema captures the key fields that downstream components require, such as task type, location text, and year (Figure 3). The agent runs a small LLM with a constrained output format, producing a complete, validated request object.

2.3.2 Geo-resolver agent

The geo-resolver agent maps the location text to DGGS cells at an appropriate resolution. The agent first resolves the text to a geo-feature using a gazetteer and administrative boundary services, then converts it to DGGS grids using DGGAL (St-Louis, 2025).

2.3.3 Data-query agent

The data-query agent turns the structured request and DGGS cells into executable SQL, executes the query, and composes a concise answer with citations. It uses an LLM to populate parameterized SQL templates that cover aggregation, comparison, trend analysis, and extremum

Volume DOI: 10.5281/zenodo.20598496

analysis. After the database returns results, the agent uses an LLM to summarize findings in clear language that reflects the original question.

3. Results

3.1 Harmonized methane inventory

The harmonized inventories provide a single, consistent foundation across data sources, years, and sector schemes. We publish a compact tabular structure with explicit fields for DGGS cell, year, dataset, sector code, and units, which supports transparent joins and reproducible aggregation (Li & Liang, 2025). Users can aggregate quickly at national and global scales, or zoom to fine-resolution tasks such as basin-level assessments or hotspot detection. This balance of consistency, efficiency, and portability enables reliable comparison and fast analysis across the full inventory stack.

3.2 Natural language to DGGS analytics

Figures 4 through 8 showcase the multi-agent system that converts plain language into DGGS-based analytics and maps. The interface supports four task types, including aggregation, comparison, trend analysis, and extremum analysis. For spatial questions, users can choose a preferred data source, then drill into sector hierarchies to select specific subsectors that align with the question's intent. Visualizations are rendered directly from DGGS cell outputs, with geometries generated on the fly. Interactive grids reveal DGGS identifiers and cell values on hover, which makes every number traceable to a spatial unit and a dataset.

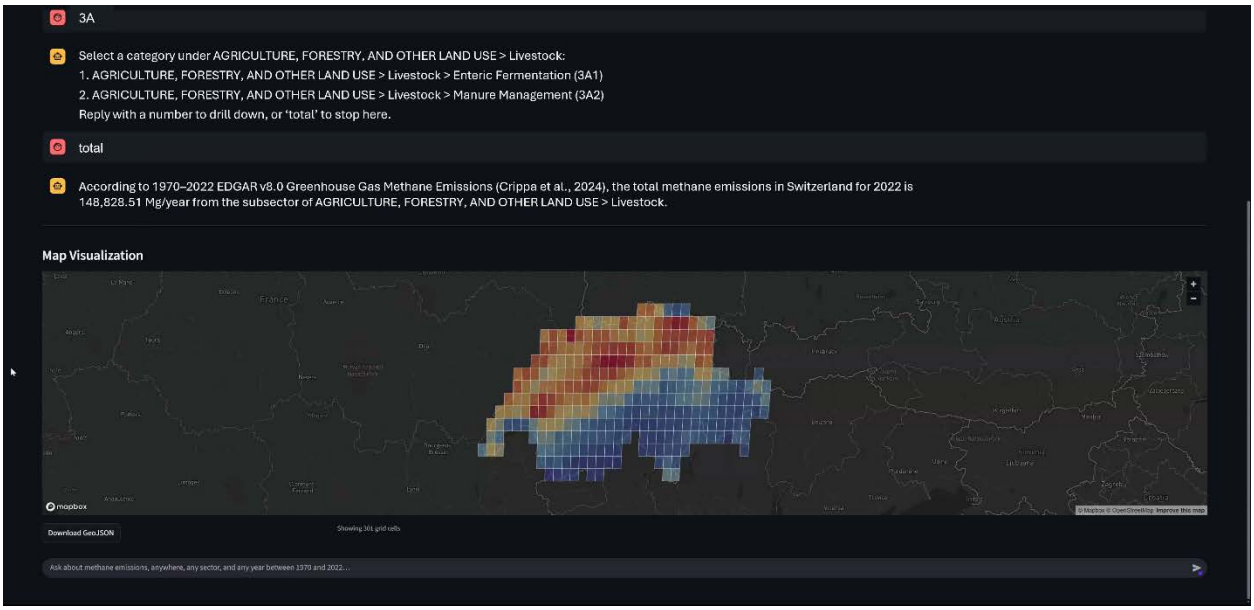

Figure 4: Interface of the multi-agent query system, showing answer and visualization to the user question "Show me the total methane emissions from the agriculture sector in Switzerland in the year 2022".

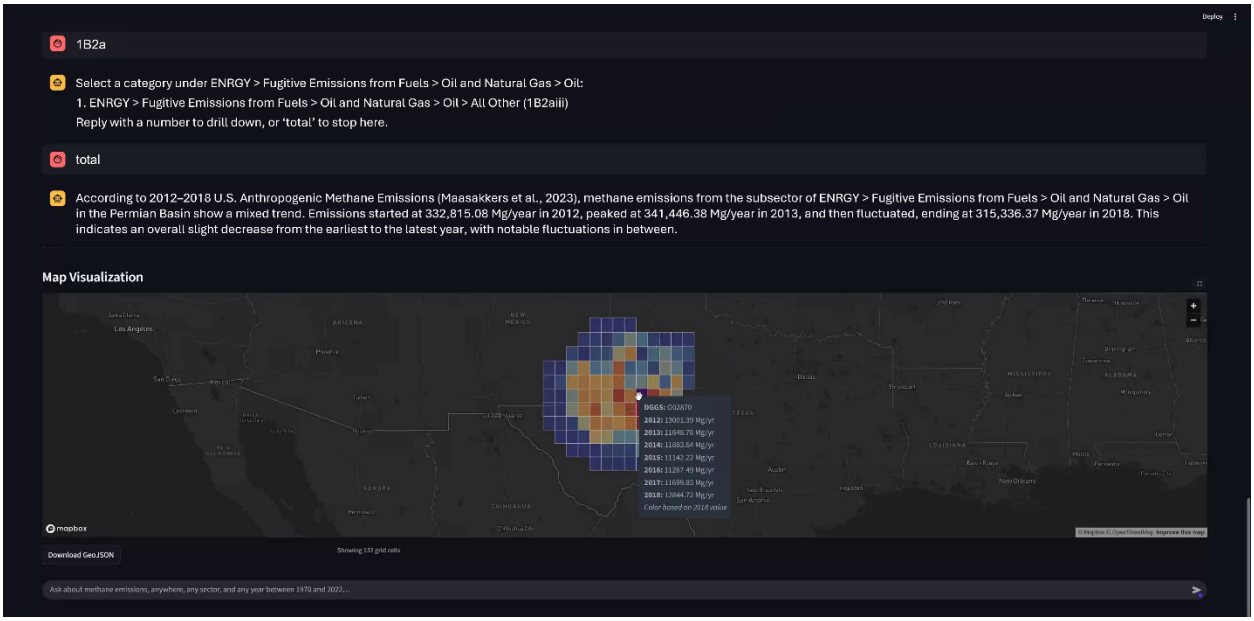

Figure 5: Interface of the multi-agent query system, showing answer and visualization to the user question "Compare the emissions from the energy sector between 2017 and 2018 in Texas".

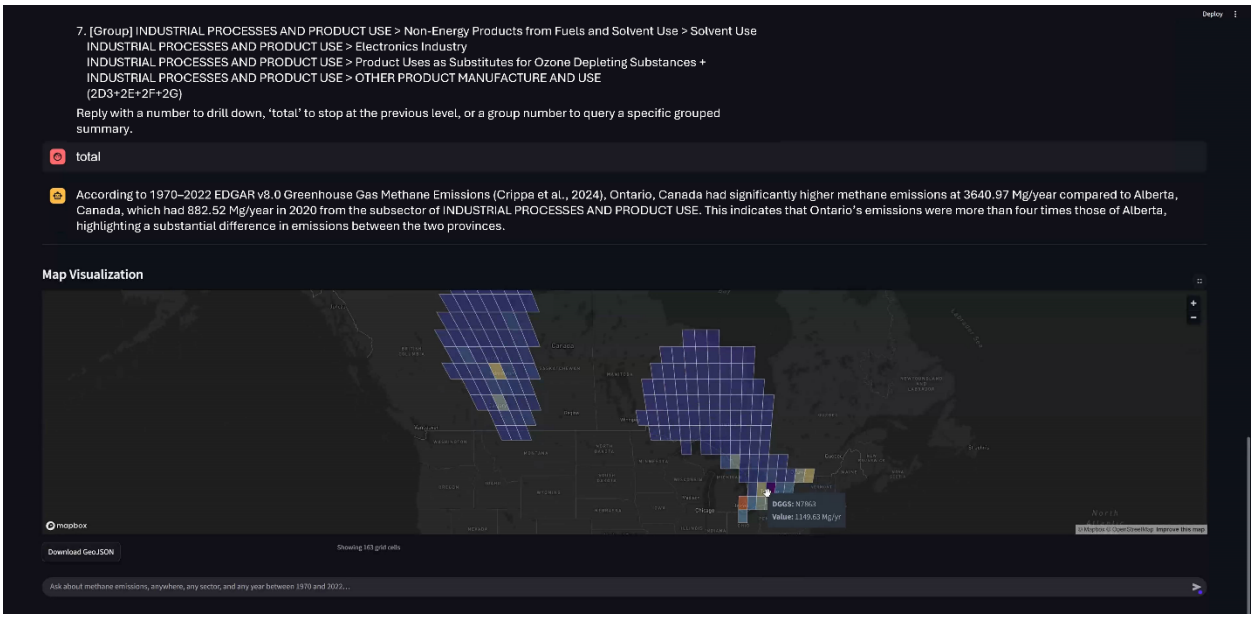

Figure 6: Interface of the multi-agent query system, showing answer and visualization to the user question "Compare the emissions from the industry processing sector in Alberta and Ontario in the year 2020".

Volume DOI: 10.5281/zenodo.20598496

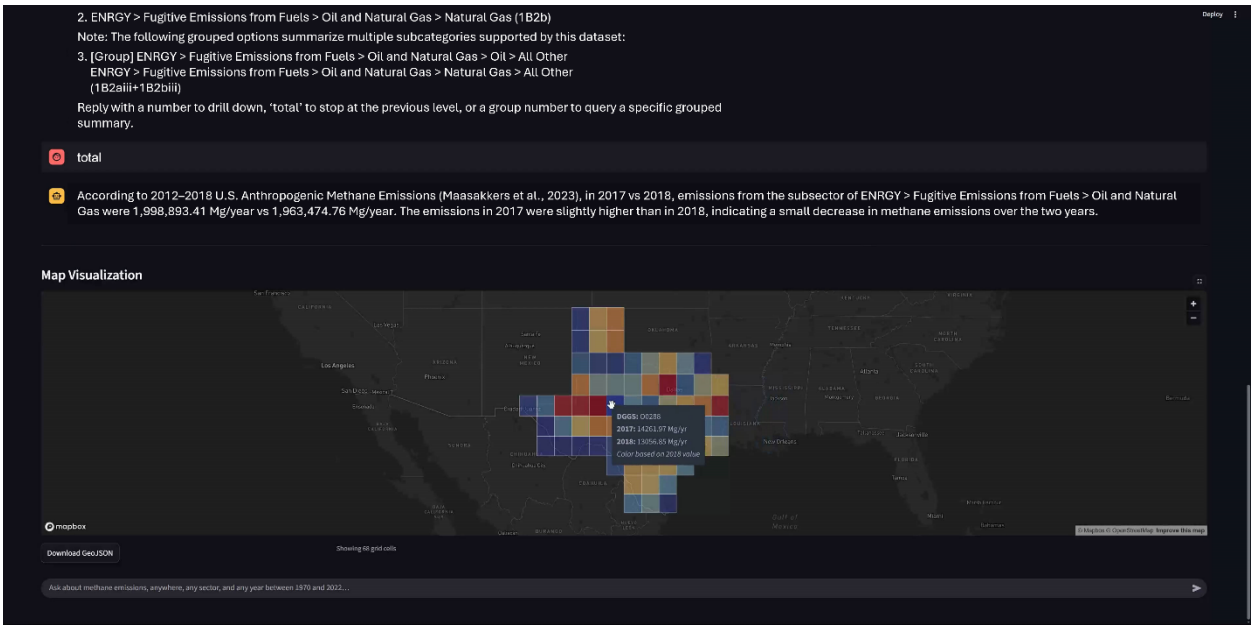

Figure 7: Interface of the multi-agent query system, showing answer and visualization to the user question "Are emissions increasing in Permian Basin from 2012 to 2018?".

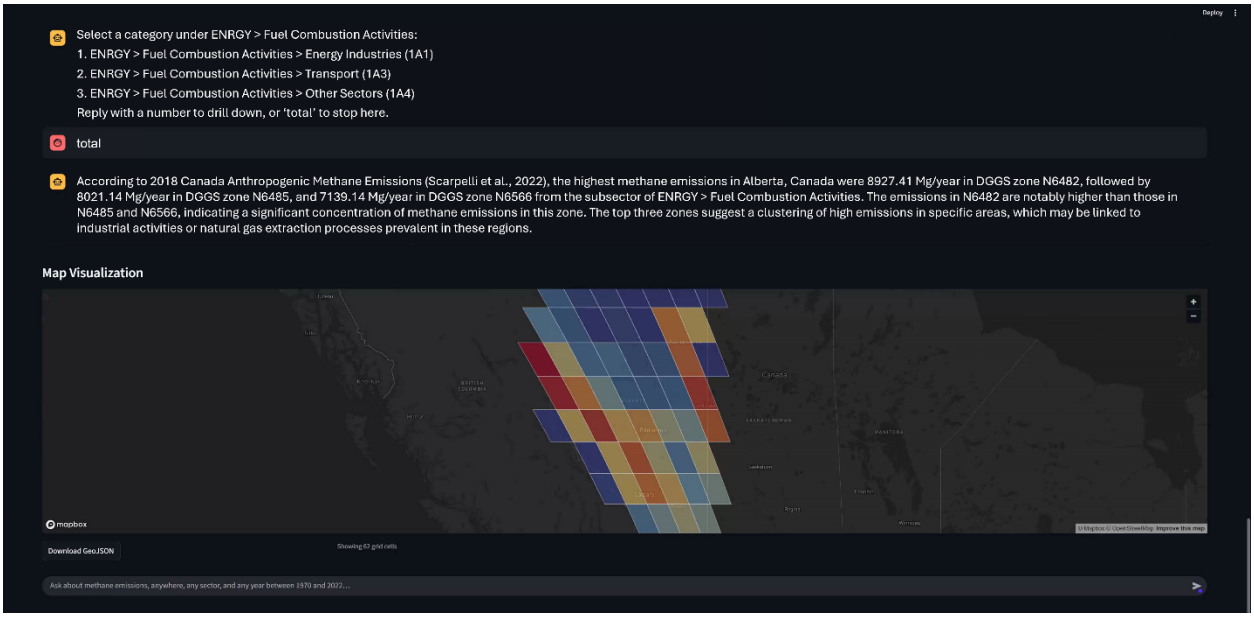

Figure 8: Interface of the multi-agent query system, showing answer and visualization to the user question "Where are the emission hotspots from the oil and gas sector in Alberta in 2018?".

Figure 4 illustrates an aggregation task, where total methane emissions from the agriculture sector in Switzerland are computed for a single year and visualized consistently across DGGS cells. Figure 5 demonstrates temporal comparison by contrasting energy sector emissions in Texas between two years, while Figure 6 shows a spatial comparison across jurisdictions by comparing industrial processing emissions in Alberta and Ontario for the same year. Trend analysis is illustrated in Figure 7, where the system evaluates whether emissions in the Permian Basin

increased over a multi-year period. Figure 8 demonstrates extremum analysis by identifying and visualizing emission hotspots from the oil and gas sector in Alberta.

These examples illustrate a complete loop from question to cited answer to interpretable map, with clear provenance and immediate feedback for further exploration.

4. Discussion & Conclusion

This work shows how DGGS provides a stable spatial and semantic reference frame for AI agents, turning black-box answers into explainable, traceable results. Every value in the harmonized database carries a DGGS cell identifier, so each statement in an answer can be linked to specific cells and to the original datasets. Maps expose these links directly through interactive grids that reveal cell IDs and values on hover, while the text response cites datasets and DGGS cells. Analysts can inspect the origin of a number, confirm which sources were used, and reproduce the exact computation, which supports auditability and trust.

In this implementation, DGGS results are visualized through an embedded Mapbox interface, with cell geometries generated dynamically at query time. Although DGGS cells must be projected for display, the analytical meaning of the results remains independent of the chosen map projection, as DGGS cells have nearly equal areas and consistent shapes. This avoids the projection-dependent distortions common in traditional map-based visualizations and supports stable visual interpretation across different cartographic contexts.

The approach is general. DGGS supplies the standard spatial index, and agents operate over extensible schemas and controlled vocabularies, which makes the framework transferable beyond methane to other greenhouse gases, pollutants, and climate hazards. Future work will focus on systematic evaluation of system performance, including query accuracy, execution efficiency, and response consistency across spatial scales. We also plan to assess uncertainty propagation from source inventories through harmonization, aggregation, and language-based summarization. In addition, comparative studies with existing natural language geospatial tools and prompt-driven analytics will be conducted to better quantify the added value of DGGS-grounded, multi-agent reasoning. Finally, more advanced agent patterns, such as planner executor loops and optional human-in-the-loop review, will be explored to support more complex analytical workflows.

Acknowledgments

This research was funded by the Rogers Internet of Things Research Chair. We are grateful to the members of the UbiSensing & AI Lab at the University of Calgary for their valuable feedback and collaborative discussions during the development of this study.

References

Crippa, M., Guizzardi, D., Pagani, F., Banja, M., Muntean, M., Schaaf, E., Becker, W. E., Monforti-Ferrario, F., Quadrelli, R., Risquez Martin, A., Taghavi-Moharamli, P., Köykkä, J., Grassi, G., Rossi, S., Melo, J., Oom, D., Branco, A., San-Miguel, J., & Vignati, E. (2023). *GHG emissions of all world countries* (JRC134504). P. O. o. t. E. Union.

Gibb, R. (2016, October 5–9.). *The rHEALPix discrete global grid system* IOP Conference Series: Earth and Environmental Science, 9th Symposium of the International Society for Digital Earth (ISDE), Halifax, Canada.

Gong, S., & Shi, Y. (2021). Evaluation of comprehensive monthly-gridded methane emissions from natural and anthropogenic sources in China. *Sci Total Environ*, *784*, 147116. https://doi.org/10.1016/j.scitotenv.2021.147116

Hiller, R. V., Bretscher, D., DelSontro, T., Diem, T., Eugster, W., Henneberger, R., Hobi, S., Hodson, E., Imer, D., Kreuzer, M., Künzle, T., Merbold, L., Niklaus, P. A., Rihm, B., Schellenberger, A., Schroth, M. H., Schubert, C. J., Siegrist, H., Stieger, J.,…Brunner, D. (2014). Anthropogenic and natural methane fluxes in Switzerland synthesized within a spatially explicit inventory. *Biogeosciences*, *11*(7), 1941–1959. https://doi.org/10.5194/bg-11-1941-2014

Kuenen, J., Dellaert, S., Visschedijk, A., Jalkanen, J.-P., Super, I., & Denier van der Gon, H. (2022). CAMS-REG-v4: a state-of-the-art high-resolution European emission inventory for air quality modelling. *Earth System Science Data*, *14*(2), 491–515. https://doi.org/10.5194/essd-14-491-2022

Li, M., & Liang, S. (2025). *Harmonized Global to Regional Gridded Methane Inventories on rHEALPix DGGS* (Version V1.0) Zenodo. https://doi.org/10.5281/zenodo.17362125

Li, M., & Stefanakis, E. (2020). Geospatial operations of discrete global grid systems—a comparison with traditional GIS. *Journal of Geovisualization and Spatial Analysis*, *4*, 26. https://doi.org/10.1007/s41651-020-00066-3

Li, M. E., & Liang, S. H. L. (2026). Enabling a Digital Earth for Methane Emissions Management with Equal-Area Discrete Global Grids. *International Journal of Digital Earth*, *19*(1), 2607210. https://doi.org/10.1080/17538947.2025.2607210

Liao, C., Li, M., & Little, B. (2026). *uraster: Structured Raster to Unstructured Mesh*. https://doi.org/10.5281/zenodo.18249075

Loman, M. L., Murray, L. T., Leibensperger, E. M., & Maasakkers, J. D. (2025). A High-Resolution Inventory of Anthropogenic Methane Emissions in New York State. *Environ Sci Technol*, *59*(32), 16933–16946. https://doi.org/10.1021/acs.est.5c07245

Maasakkers, J. D., McDuffie, E. E., Sulprizio, M. P., Chen, C., Schultz, M., Brunelle, L., Thrush, R., Steller, J., Sherry, C., Jacob, D. J., Jeong, S., Irving, B., & Weitz, M. (2023). A Gridded Inventory of Annual 2012-2018 U.S. Anthropogenic Methane Emissions. *Environ Sci Technol*, *57*(43), 16276–16288. https://doi.org/10.1021/acs.est.3c05138

Sadavarte, P., Pandey, S., Maasakkers, J. D., van der Gon, H. D., Houweling, S., & Aben, I. (2022). A high-resolution gridded inventory of coal mine methane emissions for India and Australia. *Elementa: Science of the Anthropocene*, *10*(1), 00056. https://doi.org/10.1525/elementa.2021.00056

Sahr, K., White, D., & Kimerling, A. J. (2003). Geodesic discrete global grid systems. *Cartography and Geographic Information Science*, *30*(2), 121–134. https://doi.org/10.1559/152304003100011090

Scarpelli, T. R., Jacob, D. J., Grossman, S., Lu, X., Qu, Z., Sulprizio, M. P., Zhang, Y., Reuland, F., Gordon, D., & Worden, J. R. (2022). Updated Global Fuel Exploitation Inventory (GFEI) for methane emissions from the oil, gas, and coal sectors: evaluation with inversions of atmospheric methane observations. *Atmospheric Chemistry and Physics*, *22*(5), 3235–3249. https://doi.org/10.5194/acp-22-3235-2022

Scarpelli, T. R., Jacob, D. J., Moran, M., Reuland, F., & Gordon, D. (2022). A gridded inventory of Canada's anthropogenic methane emissions. *Environmental Research Letters*, *17*(1), 014007. https://doi.org/10.1088/1748-9326/ac40b1

Scarpelli, T. R., Jacob, D. J., Villasana, C. A. O., Ramírez Hernández, I. F., Cárdenas Moreno, P. R., Cortés Alfaro, E. A., García García, M. Á., & Zavala-Araiza, D. (2020). A gridded inventory of anthropogenic methane emissions from Mexico based on Mexico's national inventory of greenhouse gases and compounds. *Environmental Research Letters*, *15*(10), 105015. https://doi.org/10.1088/1748-9326/abb42b

St-Louis, J. (2025). *DGGAL, the Discrete Global Grid Abstraction Library*. Retrieved September 30 from https://github.com/ecere/dggal

Thompson, P. T., Ojha, S., Powell, C. D., Pennell, K. G., & Moseley, H. N. B. (2023). A proposed FAIR approach for disseminating geospatial information system maps. *Sci Data*, *10*(1), 389. https://doi.org/10.1038/s41597-023-02281-1

Uber. (2017). *H3: a hexagonal hierarchical geospatial indexing system*. Retrieved 25 November from https://github.com/uber/h3

Veach, E., Rosenstock, J., Engle, E., Snedegar, R., Basch, J., & Manshreck, T. (2017). *S2 geometry library: computational geometry and spatial indexing on the sphere*. Retrieved 5 June from https://s2geometry.io

Wang, S., Hu, T., Xiao, H., Li, Y., Zhang, C., Ning, H., Zhu, R., Li, Z., & Ye, X. (2024). GPT, large language models (LLMs) and generative artificial intelligence (GAI) models in geospatial science: a systematic review. *International Journal of Digital Earth*, *17*(1), 2353122. https://doi.org/10.1080/17538947.2024.2353122

Zhang, Y., Wei, C., He, Z., & Yu, W. (2024). GeoGPT: An assistant for understanding and processing geospatial tasks. *International Journal of Applied Earth Observation and Geoinformation*, *131*, 103976. https://doi.org/10.1016/j.jag.2024.103976