A novel framework for leveraging geological environment big data to assess sustainable development goals

Wei Han,[1] Lizhe Wang,[1,*] Yuewei Wang,[1] Jun Li,[1] Jining Yan,[1] and Yinghui Shao[1]

[1]School of Computer Science, China University of Geosciences, Wuhan 430078, China

*Correspondence: lizhe.wang@gmail.com

Received: September 3, 2023; Accepted: November 7, 2024; Published Online: December 12, 2024; https://doi.org/10.59717/j.xinn-geo.2024.100122

Citation: Han W., Wang L., Wang Y., et al. (2025). A novel framework for leveraging geological environment big data to assess sustainable development goals. *The Innovation Geoscience* **3**:100122.

Geological environmental big data (GEBD) offers a significant opportunity for the comprehensive monitoring of the Earth's shallow structures, dynamic processes, and their interactions among surface spheres, promisingly contributing to the precise assessment of sustainable development goals (SDGs). To systematically harness the potential of GEBD in SDG assessment, an innovative framework has been developed to address current challenges through a four-dimensional observation network, a strategy of integrating GEBD into SDG indicators, and a knowledge-driven association-mining method. This framework potentially provides a series of positive and profound impacts, ranging from technology development and environmental conservation to enhanced public awareness.

INTRODUCTION

The geological environment comprises various geological bodies and processes within Earth's shallow lithosphere and surface space,[1] and their interactions with the atmosphere, hydrosphere, and other spheres. It directly connects to geology, which studies Earth's composition, structure, physical characteristics, and history. Geological processes shape the geological environment, while these geological environment conditions deeply influence geological processes and structures. The geological environment and human society are intrinsically linked, as it sustains human survival, industrial production, and social development through essential natural resources such as water, minerals, and land. However, over recent centuries, human activities including mining, urban development, and reservoir construction have caused unprecedented changes to the geological environment, leading to global challenges such as resource scarcity and ecological degradation.[2]

Monitoring data from the geological environment, named geological environment big data (GEBD), record information about surface rock, soil, topography, groundwater, mineral, and shallow geological structures. These data play a supportive role to access to clean water and energy, develop safe infrastructure and other applications, thereby contributing to the assessment of multiple sustainable development goals (SDGs).[3] Therefore, integrating GEBD into SDG indicators offers a promising avenue for improving the accuracy and comprehensiveness of SDG assessments. This integration provides critical insights into numerous aspects, ranging from resource availability and climate change to natural disaster assessment. Although previous studies have clarified its crucial role in SDG assessment,[4] the current research field still lacks comprehensive analysis. Key limitations regarding observation data acquisition, SDG indicator alignment, and association mining need to be systematically addressed:

(1) ***Limited observation capability***. Geological environment elements exhibit highly complex taxonomies with significant variations across multiple strata. The complexity, combined with surface coverage limitations and data acquisition constraints, makes it challenging for mainstream Earth observation systems to achieve comprehensive, continuous, and high-precision monitoring.

(2) ***Unclear SDG indicators***. The extensive volume and complexity of GEBD require systematic alignment with appropriate SDG indicators to reveal meaningful patterns. However, current indicators insufficiently incorporate geological environmental theory, limiting GEBD's potential to effectively assess sustainable development levels.

(3) ***Inconclusive influence mechanism***. The application of GEBD in SDG assessments involves multi-dimensional geological environment and human activity factors across multi-temporal scales. Identifying meaningful relationships remains challenging, particularly in establishing causality amid numerous confounding factors.

To address these challenges for unlock the potential of GEBD for SDG assessments, this perspective proposes a novel framework that integrates multi-dimensional observing techniques and advanced artificial intelligence (AI) methods.[5-6] As shown in Figure 1, the framework establishes a four-dimensional observing network through monitoring platforms equipped with various sensors and instruments to acquire comprehensive GEBD. Subsequently, it systematically incorporates GEBD into the selected SDG indicators. Finally, a knowledge-driven association-mining method identifies the intricate relationships between geological environment and human activities, enabling targeted assessments of specific SDG indicators. The following sections detail these three main components.

FOUR-DIMENSIONAL OBSERVING NETWORK

Satellite-based platforms and remote sensing (RS) sensors alone are insufficient to obtain comprehensive information about geological environment elements,[7] particularly underground features, due to geological variations, dynamic processes, surface coverage, and sensor limitations. Advanced data acquisition platforms and emerging sensor technologies offer promising solutions for monitoring Earth's surface and interior dynamics. Consequently, we propose a "space-air-ground-underground (or underwater)" (SAGU) four-dimensional observation network that integrates diverse platforms and sensors to enhance GEBD acquisition. The four dimensions are detailed as follows:

(1) ***Space***. It comprises satellite-based platforms equipped with extensive passive sensors, including high-spatial, multi-spectral, and hyper-spectral sensors, as well as active RS systems.

(2) ***Air***. The dimension consists of aircrafts, drones and other platforms equipped with various sensors. Both space and air dimensions collectively provide comprehensive coverage of surface environments and human activities. The associated datasets are widely available through established sharing platforms, such as National Aeronautics and Space Administration (NASA) Open Data Portal and European Space Agency (ESA) Earth Online. Strategic utilization of these existing resources can significantly reduce monitoring network deployment costs.

(3) ***Gound***. The ground dimension incorporates monitoring stations equipped with ground-based sensors and instruments that continuously collect high-precision data on parameters such as soil composition and biodiversity.

(4) ***Underground (or underwater)***. It deploys sensor networks (including crack meters, water quality sensors, and sonar systems) combined with borehole sampling and analysis to monitor geological features, mineral resources, and their dynamics. Major data sharing platforms, including the United States Geological Survey (USGS), China Science and Technology Resource Sharing Network, and Terrestrial Ecosystem Research Network (TERN), serve as comprehensive repositories for geological environmental data. These platforms provide extensive datasets encompassing ground and subsurface dimensions, ranging from atmospheric conditions and soil properties to wetland and oceanic parameters. However, the current fragmented and non-standardized datasets are insufficient to meet contemporary monitoring requirements.

The network construction follows a top-down principle, beginning with the identification of available space and air platforms along with historical data. The process then involves a comprehensive assessment of accessible ground and underground platforms, evaluating their instrumental capabilities and advantages. Strategic instrument selection and deployment locations are

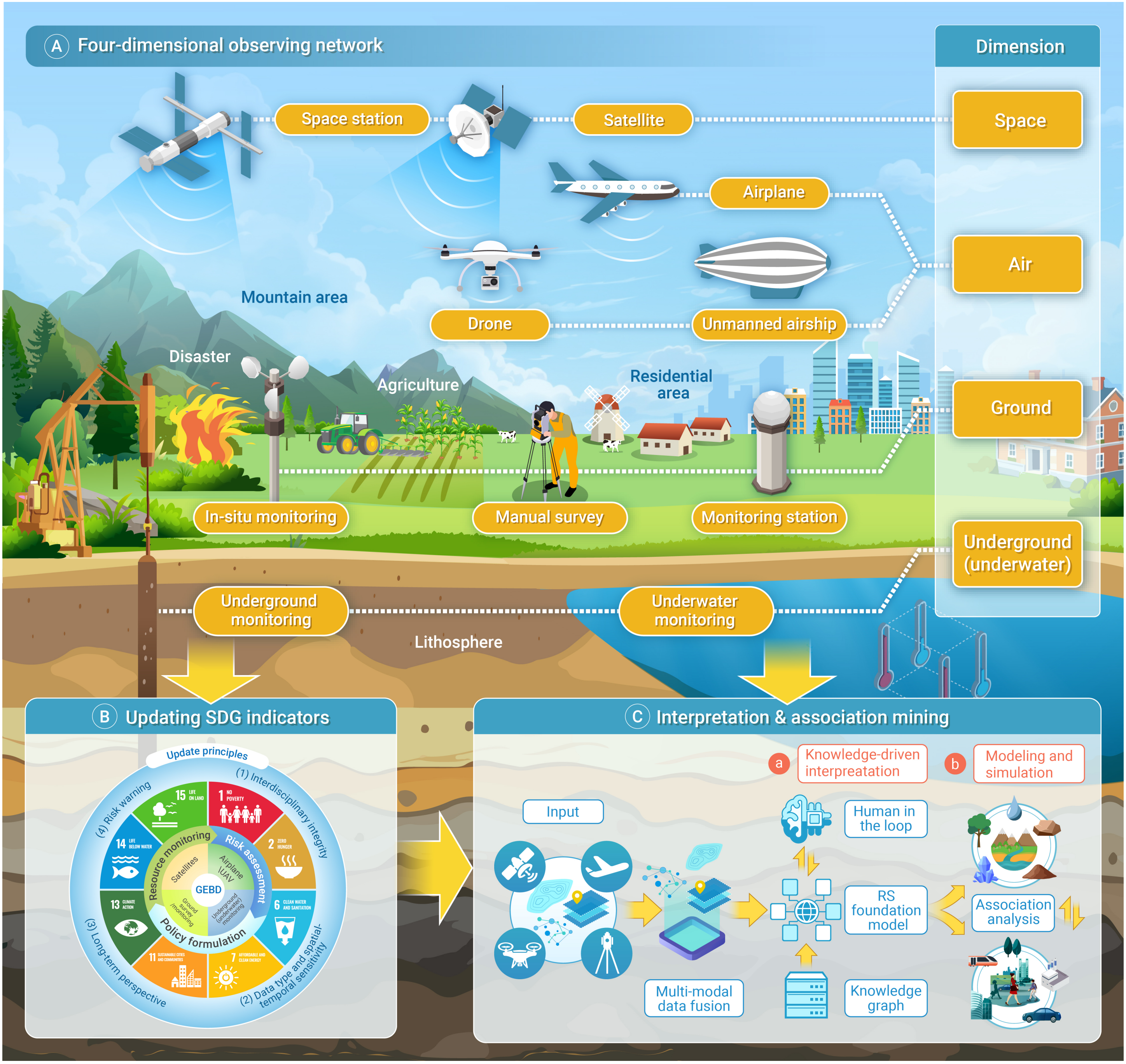

Figure 1. The overall structure of the proposed framework.

determined to address data coverage gaps in space and air dimensions, ultimately establishing the integrated four-dimensional network. Data collected across different observational dimensions provide varying modalities and resolutions, facilitating complementary information integration and cross-validation. Future development requires careful consideration of large-scale sensor deployment, network transmission protocols, data integration standards, and data security and privacy protocols.

INCORPORATING GEBD INTO SDG INDICATORS

The SDGs constitute a complex multi-disciplinary and multi-hierarchical system, making it challenging to establish clear connections with GEBD. A significant challenge lies in balancing the influence and weights of geological environmental factors against human activities while considering relevant policies. GEBD demonstrates significant potential in natural resource monitoring, climate and disaster risk assessment, and policy formulation. Based on GEBD's capabilities, we have identified eight relevant goals from the SDG framework—SDGs 1, 2, 6, 7, 11, 13, 14, and 15. This section proposes a comprehensive strategy to enhance these selected SDG indicators by integrating geological environmental elements and appropriate GEBD metrics. This section proposes a comprehensive strategy to update these selected SDG indicators by integrating geological environmental elements and appropriate GEBD types. The strategy encompasses the following aspects:

(1) ***SDG indicators and GEBD alignment***. The initial step involves identifying the relationships between geological environmental factors and selected SDG indicators and determining their relative weights. The relevance is determined based on geological principles, SDG indicator characteristics, and relevant policy frameworks.

(2) ***SDG indicators refinement***. Based on the alignment analysis, SDG indicators are refined to systematically incorporate geological environmental factors into their assessment frameworks.

(3) ***Iterative Review***. The final stage implements an iterative review process to refine and update SDG indicators. This systematic approach ensures that

the indicators accurately reflect geological environmental changes and incorporate scientific advances.

For example, SDG indicator 11.5 measures the population exposure ratio for climate extremes,[8] evaluating the combined effects of climate change, population growth, and land use change. The assessment begins with analyzing the relevance of geological environmental factors to this indicator, identifying key elements from shallow and surface geological environments, such as elevation, topography, engineering geological conditions, and groundwater characteristics. This perspective then incorporates these geological environmental elements to enhance the calculation of the exposure ratio E, which can be expressed as follows:

$$E = F \times P \times (\alpha LU + \beta GE) \tag{1}$$

where F represents the frequency of extreme climate events, P represents the population density and LU represents the land use type. $GE = \sum_{i=1}^{N} Z_i$ comprises various geological environmental elements Z_i. The coefficients α and β represent the weights of LU and GE, respectively. Historical GEBD enables analysis of coefficient variations across distinct periods (such as 2000-2010, and 2010-2020), facilitating the identification of key drivers affecting population exposure patterns. Furthermore, the updated population exposure ratio will be aligned with the Shared Socioeconomic Pathways (SSPs) developed under the Coupled Model Intercomparison Project Phase 6 (CMIP6) through an iterative review process. This alignment involves refining the types and weights of GE and LU elements to enhance the precision of future exposure ratio predictions at the local scale, thereby supporting more accurate assessments of SDG 11.5.

Moreover, the development of novel indicators should not only adhere to the fundamental principles of scientific validity, measurability, and comprehensiveness, but also incorporate the following essential principles: (1) ***Interdisciplinary integrity***: It requires integrated coupling among geological environmental science, sociology, and governance to achieve comprehensive assessment outcomes. (2) ***Data type and scale sensitivity***: Selection of data sources and spatial-temporal scales should align with specific SDG indicators, regional contexts, and geological environmental principles. (3) ***Long-term perspective***: Incorporating GEBD's temporal depth into existing SDG indicators enhances future planning effectiveness. Balanced weighting between current and long-term sustainability is essential. (4) ***Risk warning***: The updated indicators should incorporate potential unsustainable scenarios identified through GEBD analysis to provide early warnings of environmental risks. By implementing these methods and principles, we can establish robust evidence-based frameworks for SDG assessments.

INTERPRETATION AND ASSOCIATION MINING

The third part focuses on extracting geological element distributions and analyzing complex relationships among geological environment, human activities, and SDG assessments through a novel knowledge-driven approach combining interpretation methods and multi-element association mining.

The interpretation component utilizes a "point-surface" fusion approach to amalgamate multi-modal GEBD through optimal transport theory and specialized AI models.[9] High-precision "point-shape" monitoring data facilitates the enhancement and calibration of "surface-shape" remote sensing images. Drawing from geoscience knowledge graphs, expert knowledge is transformed into knowledge rules and graphical representations, serving as incremental guidance for the interpretation process. These knowledge representations and fused data features are then integrated with a pre-trained remote sensing vision-language foundation model to accurately identify geological environmental elements. Additionally, a human-in-the-loop strategy ensures the reliability of interpretation results.[10]

In association mining, two primary subjects encompass the geological environment and human activities, each characterized by an intricate web of factors and interconnections. To effectively structure this complexity, complex network techniques are utilized, wherein discrete nodes symbolize the individual elements and edges encapsulate observed relationships or interactions. Subsequently, we further enhance the temporal and spatial analysis with nonlinear dynamics and Monte Carlo methods. Taking nonlinear dynamics as an example, in the complex network $G(V,E)$, geological environment factors are defined as $X_{GE} = \{x_1, \ldots, x_N\}$ and the factors in the human activity layer are represented by $X_{HU} = \{x_{N+1}, \ldots, x_M\}$. Then the geological environment element x_{GE}^t at time t can be computed as,

$$\frac{dS(x_{GE}^t)}{dt} = A_{inner}F_{inner}(X_{GE}) + A_{inter}F_{inter}(X_{HU})$$

where A_{inner} and F_{inner} respectively denote the connectivity and interaction intensity among geological environment elements. Meanwhile, A_{inter} and F_{inter} represent the connectivity and intensity of human activities' effects on geological elements. These techniques facilitate simulation of multi-temporal evolution processes in geological environment and human activities, enabling comprehensive assessment of their potential impacts on sustainable development.

The final phase focuses on selecting representative case studies for practical localized evaluations of sustainable development levels, such as water resources and land degradation monitoring for SDGs 6.1 and 15.3 assessment. The selection process is guided by complex network centrality measures and spatio-temporal analysis outcomes, ensuring focus on the most significant scenarios. The SDG assessments for these cases are then conducted using the identified relationships and updated SDG indicators.

CONCLUSION

This perspective proposes a novel framework to explore data-driven methods for GEBD in supporting precise SDG assessments and practices. Advanced technologies and services will enhance the accuracy and efficiency of geological environmental monitoring. The SDG indicators will be further developed to incorporate geological environment dimensions, enabling more effective natural resource management and utilization. Moreover, this framework will improve public awareness of the fundamental role of geological environments in SDG assessment, thereby promoting environmentally responsible social behavior.

REFERENCES

1. Han W., et al. (2023). A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities, ISPRS J. Photogramm. *Remote Sens.* **202**:87-113. DOI:10.1016/j.isprsjprs.2023.05.032
2. United Nations. (2022). The sustainable development goals: Report 2022. https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf
3. United Nations. (2015). Transforming our world: The 2030 Agenda for sustainable. https://documents.un.org/doc/undoc/gen/n15/291/89/pdf/n1529189.pdf
4. Gill J C., (2017). Geology and the sustainable development goals, *Episodes* **4**:70-76. DOI:10.18814/epiiugs/2017/v40i1/017010
5. Luo L., et al., (2023). Space technology: A powerful tool for safeguarding world heritage, *The Innovation* **4**:100420. DOI:10.1016/j.xinn.2023.100420
6. LeCun Y., et al., (2015). Deep learning, *Nature* **521**:436-444. DOI:10.1038/nature14539
7. Guo H., et al., (2022). Measuring and evaluating SDG indicators with Big Earth Data, *Science Bulletin* **67**:1792-1801. DOI:10.1016/j.scib.2022.07.015
8. Gao J., et al., (2022). Urban land patterns can moderate population exposures to climate extremes over the 21st century, *Nature Communications* **14**:6536. DOI:10.1038/s41467-023-42084-x
9. Zhao T., et al., (2024). Artificial intelligence for geoscience: Progress, challenges and perspectives, *The Innovation* **5**:100691. DOI:10.1016/i.xinn.2024.100691
10. Mosqueira-Rey E., et al., (2023). Human-in-the-loop machine learning: A state of the art, *Artificial Intelligence Review* **56**:3005-3054. DOI:10.1007/s10462-022-10246-W

FUNDING AND ACKNOWLEDGMENTS

The work was supported by the National Natural Science Foundation of China under Grant 41925007 and 42201415; the Open Fund Project of Hubei Key Laboratory of Intelligent Geo-Information Processing under Grant KLIGIP-2023-B05. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

W. Han conceptualized and designed the overall framework, drafted and revised the manuscript; L. Wang contributed to the overall framework and participated in the manuscript review; Y. Wang, J. Li, J. Yan and Y. Shao contributed to the manuscript writing and review process. All authors contributed to and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.