International Journal of Applied Earth Observation and Geoinformation 149 (2026) 105239

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Advancing sustainable geospatial analytics and geoinformatics through repeatable, reproducible, and expandable (RRE) framework and design

Siqin Wang [a,b,c,*], Xiao Huang [d], Filip Biljecki [e,f], Francisco Rowe [g], Veruska Muccione [h,i]

[a] Spatial Sciences Institute, University of Southern California, Los Angeles, CA, United States
[b] School of Science, RMIT University, Melbourne, Victoria, Australia
[c] School of the Environment, University of Queensland, Brisbane, Australia
[d] Department of Environmental Sciences, Emory University, Atlanta, GA, United States
[e] Department of Architecture, National University of Singapore, Singapore
[f] Department of Real Estate, National University of Singapore, Singapore
[g] Geographic Data Science Lab, University of Liverpool, Liverpool, United Kingdom
[h] Department of Geography, University of Zurich, Zurich, Switzerland
[i] Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

A R T I C L E I N F O

Keywords:
Sustainable geospatial analytics
Repeatable framework
Reproducibility and replicability
FAIR principles
Scientific workflows
Geoinformatics

A B S T R A C T

In the era of data-intensive science, the complexity and volume of geospatial data have grown exponentially. Compared to traditional data sources, non-traditional sources are more complex and structured, necessitating sophisticated methods and a series of decisions to transform raw data inputs into usable and actionable data products. This Special Issue, "Sustainable geospatial analytics and geoinformatics with repeatable, reproducible, and expandable (RRE) framework and design," brings together a collection of seven pioneering papers that address the critical need for consistency and transparency in geospatial research. These studies explore diverse domains, including explainable machine learning, disaster risk assessment, urban ecological health, infectious disease control and scientific workflow management. Collectively, they advocate for the adoption of an RRE framework to ensure that results can be verified and reproducible across different environments and expanded with new data or methodologies. By integrating visual programming, service-oriented strategies, as well as Findable, Accessible, Interoperable, and Reusable (FAIR) principles, the featured research lowers technical barriers for non-experts while enhancing the robustness of complex models. This editorial synthesizes the contributions of these papers, highlighting how they foster a sustainable and collaborative geospatial knowledge ecosystem. This collection serves as a roadmap for the next generation of geoinformatics, where transparency and flexibility are foundational to addressing global environmental and social challenges.

1. Introduction

Like many others, the fields of geospatial analytics and geoinformatics are currently navigating a "reproducibility crisis," where the lack of process transparency often hinders the verification and reuse of scientific findings. Modern geoprocessing often involves intricate chains of data manipulation, parameter tuning and computational modeling that are difficult for peer scientists to reconstruct. To address these hurdles, the concept of Repeatability, Reproducibility and Expandability (RRE) have emerged as fundamental pillars for sustainable scientific development (Arribas-Bel et al 2021; Nelson et al., 2025). Repeatability ensures that the same researchers can achieve consistent results using the same data and methods. Reproducibility allows different researchers to obtain identical conclusions using the original data and methods. Expandability provides flexibility to adapt existing workflows to new data, variables, or geographic regions. These principles are essential for transitioning from isolated studies to a cumulative, structured body of geographic knowledge. This Special Issue showcases how the RRE framework can be applied across various geospatial domains from monitoring urban vegetation productivity to optimizing pandemic mass testing to create scientific tools that are not only powerful but also trustworthy and adaptable.

* Corresponding author at: Spatial Sciences Institute, University of Southern California, Los Angeles, CA, United States.
E-mail addresses: siqinwan@usc.edu (S. Wang), xiao.huang2@emory.edu (X. Huang), Singapore.filip@nus.edu.sg (F. Biljecki), F.Rowe-Gonzalez@liverpool.ac.uk (F. Rowe), veruska.muccione@geo.uzh.ch (V. Muccione).

https://doi.org/10.1016/j.jag.2026.105239
Received 31 January 2026; Received in revised form 7 March 2026; Accepted 9 March 2026

2. Synthesis of RRE frameworks in geospatial analytics

To provide a coherent understanding of the advancements presented in this Special Issue, the following synthesis is organized into three progressive themes: computational infrastructure and standardized workflows, advanced analytical modeling for complex spatial phenomena, and sustainable domain-specific applications. We first examine the structural foundations, such as FAIR-oriented management systems, visual programming, and action-based tracking that facilitate the technical execution of RRE principles. Building upon this scaffolding, we analyze methodological innovations in explainable machine learning, compound hazard assessment and mobility-based sampling, which ensure that complex geoprocessing remains transparent and adaptable. Finally, we demonstrate how these frameworks are applied to critical sustainability challenges, including urban carbon sequestration, pandemic control and disaster resilience, thereby bridging the gap between theoretical geoinformatics and real-world policy impact.

2.1. Foundations of reproducible infrastructure

A primary hurdle in geospatial research is the "technical barrier" that prevents peer scientists from accurately recreating complex computational environments. The papers in this collection address this by shifting focus from isolated code to comprehensive, service-oriented workflows.

Hu et al. (2025) lead this shift by introducing a FAIR-oriented geospatial Workflow Management System (WfMS). Their logic dictates that for a workflow to be truly reproducible, every component, including data, tools, and processes, must be Findable, Accessible, Interoperable, and Reusable (FAIR). By integrating the Common Workflow Language (CWL) and Docker containers, they ensure that workflows are platform-independent and that software dependencies are encapsulated, eliminating the "it works on my machine" syndrome. This architectural rigor is mirrored by Zhu et al. (2023) who propose a strategy to reproduce computational processes by tracking researchers' actions online. Instead of just sharing static results, they decompose research into node-link ensembles, where action nodes represent specific resource interactions (e.g., data preprocessing or model selection) and logical links represent the procedural sequence. This allows peer scientists to re-execute exact dataflows through the OpenGMS platform, ensuring consistency in output even when models are heterogeneous.

Complementing these backend systems is the drive toward democratizing geospatial science through visual programming platforms. Liu et al. (2024) demonstrate how the "Geospatial Analytics Extension for KNIME" can elevate the RRE framework by replacing textual coding with intuitive, node-based graphical interfaces. Their 4E approach "Examine Innovation, Engineer Workflow, Establish Nodes, and Embed Structure" provides a systematic pathway for transforming individual research innovations into a shared "Geospatial Knowledge Tree". This logic ensures that complex spatial accessibility models, such as the Generalized Two-Step Floating Catchment Area (G2SFCA), can be standardized and expanded by researchers who may lack an extensive programming background.

2.2. Methodological innovation for complex geospatial analysis

Beyond infrastructure, the RRE framework requires analytical methods that remain robust and interpretable when applied to non-linear and high-dimensional data. Liu (2024) addresses the interpretability of GeoAI by proposing the XGeoML ensemble framework. The core logic here is that while traditional models like GWR capture spatial heterogeneity, they often fail to decode non-linear interactions between geographical features and explanatory variables. By merging local spatial weighting with Explainable AI (XAI) tools like SHAP and LIME, XGeoML allows researchers to not only predict outcomes but also verify the spatial mechanisms behind them. This focus on parameter uncertainty and model comparison is essential for the "Reproducible" pillar of RRE, as it forces researchers to cross-validate results across different ML algorithms.

Similarly, Zhou et al. (2025) tackles the complexity of compound natural hazards through the ComHazAsTC-RRE model. Their logic emphasizes that traditional single-hazard assessments lead to an underestimation of risk during events like tropical cyclones. By utilizing a C-Vine Copula structure, the framework accurately models the integrated dependencies of wind, rain, and storm surges. To maintain expandability, the model is built entirely on globally accessible datasets (like GSOD and GTSM) and open-source code, allowing the framework to be applied to different coastal regions worldwide.

In the realm of public health, Zhang et al. (2024) introduce a repeatable framework for optimizing infectious disease detection. The logic focuses on using human mobility and Point-of-Interest (POI) data to inform spatial sampling. By developing metrics like Case Flow Intensity (CFI) and Case Transmission Intensity (CTI), the study proves that targeting high-risk communities based on movement patterns is significantly more efficient than simple random sampling. This framework is inherently expandable, offering a scalable solution for containing future outbreaks by prioritizing resources in densely populated urban environments.

2.3. Sustainable domain applications

The ultimate goal of the RRE design is to support sustainable development through consistent and scalable geoinformation. Chen et al. (2024) demonstrate this by providing a multi-source data-driven estimation of urban Net Primary Productivity (NPP) in Wuhan. Their improved CASA model logic accounts for the unique fragmentation of urban landscapes by excluding impervious surfaces like roads and buildings. By integrating Sentinel-2 remote sensing with high-resolution meteorological data, they provide a repeatable method for monitoring carbon sequestration. This research directly supports carbon neutrality strategies, illustrating how RRE-compliant models can provide the technical backing needed for long-term ecological planning.

Across all seven papers, the logic remains consistent: sustainability in geoinformatics is achieved through transparency. Whether it is simulating flood runoff in the Wangjiaba basin using reproducible service-based models or evaluating healthcare accessibility in Oklahoma through FAIR-compliant workflows (Hu et al., 2025), these studies prove that a standardized, expandable design is the key to transitioning from fragmented data to cumulative, reliable geographic knowledge.

3. Challenges and future directions

Despite the significant advancements highlighted in this Special Issue, key fundamental challenges persist in the pursuit of a fully sustainable and RRE-driven geospatial ecosystem. One of the most prominent obstacles is the technical and knowledge barrier that restricts accessibility for researchers and practitioners, versed in this domain and interdisciplinary audiences. The literature consistently highlights that the rapid expansion of geospatial data and emergence of complex "black box" machine learning models often require steep learning curves and advanced coding skills (Wang et al., 2024; Liu et al., 2025). Even when tools are shared, the rebuilding of environmental dependencies, such as specific operating systems, library versions, and software configurations, remains a daunting task that frequently leads to reproducibility failures. Furthermore, the rapid pace of geospatial knowledge expansion creates a synthesis dilemma, where the sheer volume of fragmented tools and models makes it difficult to structure knowledge into an intuitive and accessible knowledge collection for the broader scientific community.

A second major challenge concerns data quality, uncertainty and computational scalability. The accuracy of sophisticated models, such as the CASA model for urban productivity or XGeoML for spatial heterogeneity presented in our Special Issue, is heavily dependent on high-quality, high-resolution input data. However, researchers often encounter data gaps and sparsity, particularly in less-developed or high-altitude regions where meteorological stations are limited; in areas frequently obscured by cloud cover in satellite imagery; or in regions impacted by conflict or natural disasters which have restricted access but where timely information is essential to inform humanitarian efforts (Iradukunda et al. 2025; Pietrostefani, et al. 2025). Additionally, as models move toward higher-dimensional dependencies, they encounter the "curse of dimensionality" and massive computational demands. Current systems often lack the scalability to support these computation-intensive workflows on a global or real-time scale, calling for more robust infrastructure (Rowe, 2023).

Finally, the field of geospatial sciences has been challenged by achieving interoperability, standardization and best practices (Nelson et al., 2025; Gu et al. 2025). While many powerful platforms exist, they often rely on proprietary frameworks or specific standards that are not easily shared across different systems. This fragmentation hinders the seamless exchange of model resources and prevents the construction of complex, cross-disciplinary workflows. Even in open-source environments, there is often a lack of process transparency. Documenting the exact sequence of a researcher's actions and interactions with data remains inconsistent, making it difficult for peer scientists to verify and rely on previous findings with absolute confidence.

To facilitate new research ideas and address these hurdles, specific future directions can be further explored. First, the integration of generative AI (GenAI) and autonomous geospatial science presents a transformative opportunity to democratize geoinformatics (Wang et al., 2024). Future research should explore using large language models (LLMs) to automate the development, execution and documentation of geospatial workflows (Li and Ning, 2023). By allowing users to design complex spatial analyses through natural language or intuitive interfaces, the barrier to entry can be significantly lowered, while AI-generated documentation can ensure that every step of a study is citable and transparent. This shift toward autonomous Geographic Information System (GIS) could serve as a vital bridge between high-level theory and practical, repeatable execution.

Second, there is a critical need to transition toward cloud-native architectures and containerization to ensure scalability and environment independence. Future RRE frameworks should encapsulate software dependencies, ensuring that workflows can be executed seamlessly across diverse computing environments, from personal desktops to high-performance clusters. Expanding current infrastructures into distributed computing clusters will allow for the parallel execution of massive tasks, such as high-resolution carbon monitoring or global-scale pandemic modeling, making sustainable geoinformatics truly scalable.

Third, the field should advance towards holistic error checking and multi-source validation tools. As research increasingly relies on the fusion of remote sensing, ground observations, and mobility data, understanding how errors propagate through integrated models is essential (Huang et al., 2024). Developing automated tools that track error origins and conduct validation studies between initial and reproduced outputs will enhance the scientific rigor of RRE frameworks. By prioritizing these directions, geoinformatics can transition from a collection of isolated studies to a collaborative, transparent, and resilient knowledge ecosystem capable of addressing global sustainability challenges.

4. Conclusion

The papers within this Special Issue collectively signal a paradigm shift toward more sustainable and transparent geoinformatics. By aligning geospatial analytics with the Repeatable, Reproducible, and Expandable (RRE) framework, researchers are ensuring that their work is not merely a static publication but a living, adaptable resource. The integration of FAIR principle and service-oriented architecture provides the necessary scaffolding to build a collaborative scientific community where data and models can be verified and improved by peers around the world. As we move forward, the focus must remain on reducing technical barriers and standardizing workflows to ensure that the powerful tools of GeoAI and spatial simulation are available to address the urgent environmental and social challenges of our time. The studies presented here prove that when we prioritize transparency, we do not just make science more reliable, but we make it more resilient and sharable.

CRediT authorship contribution statement

Siqin Wang: Writing – original draft, Resources, Conceptualization. **Xiao Huang:** Writing – review & editing, Conceptualization. **Filip Biljecki:** Writing – review & editing, Conceptualization. **Francisco Rowe:** Writing – review & editing, Conceptualization. **Veruska Muccione:** Writing – review & editing, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Arribas-Bel, D., Green, M., Rowe, F., Singleton, A., 2021. Open data products-a framework for creating valuable analysis ready data. J. Geogr. Syst. 23 (4), 497–514.

Chen, J., Shao, Z., Huang, X., Hu, B., 2024. Multi-source data-driven estimation of urban net primary productivity: a case study of Wuhan. Int. J. Appl. Earth Obs. Geoinf. 127, 103638.

Gu, Y., Quintana, M., Liang, X., Ito, K., Yap, W., Biljecki, F., 2025. Designing effective image-based surveys for urban visual perception. Landsc. Urban Plan. 260, 105368.

Hu, T., Liu, T., Jarugumalli, V.S.D., Cheng, S., Deng, C., 2025. FAIR principles in workflows: a GIScience workflow management system for reproducible and replicable studies. Int. J. Appl. Earth Obs. Geoinf. 138, 104477.

Huang, X., Wang, S., Yang, D., Hu, T., Chen, M., Zhang, M., Hohl, A., 2024. Crowdsourcing geospatial data for earth and human observations: a review. J. Remote Sens. 4, 0105.

Iradukunda, R., Rowe, F., Pietrostefani, E., 2025. Estimating internal displacement in Ukraine from high-frequency GPS mobile phone data. Humanities Soc. Sci. Commun. 12 (1), 1863.

Nelson, T., Frazier, A.E., Kedron, P., Dodge, S., Zhao, B., Goodchild, M., Wilson, J., 2025. A research agenda for GIScience in a time of disruptions. Int. J. Geogr. Inf. Sci. 39 (1), 1–24.

Li, Z., Ning, H., 2023. Autonomous GIS: the next-generation AI-powered GIS. Int. J. Digital Earth 16 (2), 4668–4686.

Liu, L., 2024. An ensemble framework for explainable geospatial machine learning models. Int. J. Appl. Earth Obs. Geoinf. 132, 104036.

Liu, L., Huang, X., Wang, S., Fu, X., 2025. Sustainable GeoAI in Human Geography: Reproducible, Replicable, and Expandable. In: *GeoAI and Human Geography: the Dawn of a New Spatial Intelligence Era*. Springer Nature Switzerland, Cham, pp. 327–344.

Liu, L., Wang, F., Fu, X., Kötter, T., Sturm, K., Guan, W.W., Bao, S., 2024. Elevating the RRE framework for geospatial analysis with visual programming platforms: an exploration with geospatial analytics extension for KNIME. Int. J. Appl. Earth Obs. Geoinf. 130, 103948.

Pietrostefani, E., Mason, M., Iradukunda, R., Tran-Jones, H., Loktieva, I., Rowe, F., 2025. Dynamic estimates of Displacement in disaster Regions: a Policy-driven framework triangulating data. UN Publications, International Organization for Migration (IOM), Geneva.

Rowe, F., 2023. Big data. In Concise Encyclopedia of Human Geography (pp. 42-47). In *Concise Encyclopedia of Human Geography*. Eds: Edited Lees, L and Demeritt, D. Edward Elgar Publishing.

Wang, S., Huang, X., Liu, P., Zhang, M., Biljecki, F., Hu, T., Bao, S., 2024a. Mapping the landscape and roadmap of geospatial artificial intelligence (GeoAI) in quantitative human geography: an extensive systematic review. Int. J. Appl. Earth Obs. Geoinf. 128, 103734.

Wang, S., Hu, T., Xiao, H., Li, Y., Zhang, C., Ning, H., Ye, X., 2024b. GPT, large language models (LLMs) and generative artificial intelligence (GAI) models in geospatial science: a systematic review. Int. J. Digital Earth 17 (1), 2353122.

Zhang, D., Ge, Y., Wang, J., Liu, H., Zhang, W.B., Wu, X., Lai, S., 2024. Optimizing the detection of emerging infections using mobility-based spatial sampling. Int. J. Appl. Earth Obs. Geoinf. 131, 103949.

Zhou, Z., Yang, S., Wang, S., Liu, X., Hu, F., Wu, Y., Chen, Y., 2025. ComHazAsTC-RRE: compound hazard assessment of tropical cyclones within repeatable, reproducible, and expandable framework. Int. J. Appl. Earth Obs. Geoinf. 136, 104314.

Zhu, Z., Chen, M., Sun, L., Qian, Z., He, Y., Ma, Z., Lue, G., 2023. Reproducing computational processes in service-based geo-simulation experiments. Int. J. Appl. Earth Obs. Geoinf. 124, 103520.