Towards the Intelligent Era of Spatial Analysis and Modeling

Di Zhu*
dizhu@umn.edu
University of Minnesota, Twin Cities
Minnesota, USA

Song Gao
song.gao@wisc.edu
University of Wisconsin, Madison
Wisconsin, USA

Guofeng Cao
guofeng.cao@colorado.edu
University of Colorado, Boulder
Colorado, USA

ABSTRACT

Geographic phenomena are considered complex due to the heterogeneous nature of spatial dependencies. It is impossible to specify a universal law described in statistical or physical languages that can perfectly characterize a real-world geographic process and explain how it forms certain observed patterns. Traditional spatial analytics based on strict statistical principles, strong assumptions, or classic computation workflows are facing great challenges and opportunities when embracing the explosive growth of geospatial data and recent technical innovations. Here, we highlight the promises of Intelligent Spatial Analytics (ISA), a new set of spatial analytical approaches based on spatially explicit deep neural networks with more flexible data representation, modules for complex spatial dependence, weaker model prior assumptions, and hence the enhanced ability to predict/explain unknowns. Three essential topics in spatial analysis, i.e., geostatistics, spatial econometrics, and flow analytics are elaborated as examples in the vision of ISA. We also discuss challenging issues of ISA as an invitation to explore deeper linkages between machine/deep learning and spatial analysis at the frontier of Geospatial Artificial Intelligence.

CCS CONCEPTS

• **Applied computing** → *Environmental sciences*; • **Information systems** → *Geographic information systems*; • **Computing methodologies** → *Model verification and validation.*

KEYWORDS

GIS, GeoAI, spatial analysis, intelligent spatial analytics, neural networks

ACM Reference Format:
Di Zhu, Song Gao, and Guofeng Cao. 2022. Towards the Intelligent Era of Spatial Analysis and Modeling. In *The 5th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI '22) (GeoAI '22), November 1, 2022, Seattle, WA, USA.* ACM, Seattle, WA, USA, 4 pages. https://doi.org/10.1145/3557918.3565863

*Contact author

GeoAI '22, November 1, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9532-8/22/11…$15.00
https://doi.org/10.1145/3557918.3565863

1 BACKGROUND

Spatial analytics, such as spatial interpolation methods, spatial regression models and point pattern analysis, provide both qualitative and quantitative approaches to understanding complex spatial patterns formed by many real-world geographic processes. Classic operations of such spatial analysis and modeling (SAM) in Geographic Information Systems are well established with standard data structures and computation workflows, answering questions such as: "*what are the spatial autocorrelation measures for a rasterized air pollution pattern*?" "*how the socioeconomic factors are associated with the crime event points?*", and "*which is the best location to open up a restaurant given the street network layout?*". Despite the rich history of SAM in the fields of Geographic Information Science (GIScience) and regional studies, we are seeing explosive growth and diversity of geospatial data coming from various sources together with methodological and technical innovations in relevant computational disciplines posing great opportunities but also challenges to the current paradigm of spatial analytics.

The uniqueness of geospatial problems lies in the fact of spatial heterogeneity, i.e., it is impossible to find a universal law described in statistical or physical languages that can perfectly replicate a geospatial process all over the geographical surface [4]. Taking spatial regression as an example, it's common to accept that the association coefficient between household average income and the crime rate can change in space, while it is not yet considered in classic SAM that the spatial regression function itself could be of different mathematical nature across study areas or scales. For predictive scenarios such as the spatial interpolation of air temperature, stationary statistical assumptions and ad-hoc mathematical models are commonly adopted to simplify the variation structure of the spatial process, in order to conclude at the unknown locations based on data observed at samples. Such mismatches, i.e., gaps between the complex nature of geospatial phenomena to be understood and the classic statistical computation paradigm of SAM, are widely-aware but yet not adequately emphasized in the community.

To get the most out of the explosive growth and diversity of geospatial data and technical innovations, traditional spatial analytics face a major task in the coming years: deriving new methods and models that can learn more from the data than current approaches can, while still respecting our evolving understanding of the geospatial nature's complexity. In particular, recent advances in statistical modeling and artificial intelligence (AI) can be employed to the development of a future essential toolkit for spatial analytics. AI techniques promise to make the modeling of complex systems more manageable, because of the use of more general symbolic representations arguing that various logical languages provide the basis for construction such intelligent models [12]. Taken together, it is not easy to define what can be considered intelligence in spatial analytics, but the undebatable renovation comes with AI is already happening in related fields such as remote sensing [1] and urban studies [9, 14]. AI techniques are now adopted in many geographical applications, while their use lag behind some other disciplines mainly because (deep) neural networks are considered a "black box" and are believed to be extremely complicated compared to traditional SAM. It is not well-introduced to geographers on how to take advantage of AI for new designs of spatial analytics. There is also a lack of research emphasis on bringing such methodological innovation into spatial analysis.

In this paper, we identify the promises of Intelligent Spatial Analytics (ISA), which extends traditional SAM with more flexibility on representing diverse spatial data, less model prior assumptions, modules to learn complex spatial dependence, and higher accuracy in approaching unknowns. Three methodological branches that are essential in spatial analytics are elaborated as examples in the vision of ISA, i.e., geostatistics, spatial econometrics, and flow analytics. Research potentials are further discussed with respect to issues such as transferability, interpretability, and biases.

2 ENABLING TECHNOLOGIES

The widespread proliferation of information and communication techniques (ICTs) has fostered a plethora of big geo-data characterizing the spatiotemporal information within our human-environment systems. A challenge coming with the abundance of data is the increasing data complexity. Our ability to collect and create geospatial data far outpaces our ability to process and digest it, let alone understand it sensibly. We are seeing issues of contemporary spatial analytics in handling the data diversity when dealing with interdisciplinary geospatial problems before providing a satisfying solution. Spatial features derived from diverse sources may be represented in different data types, combining raster, vector, network, flow and other irregular structures. Also, there could be multiplex geospatial knowledge to be learned that can not be simplified by ad hoc modeling functions based on inaccurate domain knowledge.

As a result of computational innovation, AI is flourishing in many fields. Deep learning (DL) methods have been proven their superior abilities to approach numerous complex problems that are almost impossible to solve in the past, such as speech recognition, image understanding, and language translation [10]. For Geography, Openshaw's influential book, named Artificial Intelligence in Geography, provided one of the first discussions to strongly support the use of AI in geography [12]. One year later, computational neural network was proposed as a pioneer prototype that combined spatial analytics with neural networks [2]. Around five years ago, geospatial artificial intelligence (GeoAI) emerged as an interdisciplinary area, where AI techniques are further developed and utilized for geographic knowledge discovery [6].

In this context, a series of GeoAI workshops have been organized at ACM SIGSPATIAL since 2017, recognized as the premier conference at the intersection of geospatial data analysis and computer science [5]. Despite many GeoAI studies published, most efforts are still bonded to the end-to-end DL frameworks, focusing on adjusting geospatial data into well-established DL models such as convolutional neural networks (CNN) and recurrent neural networks (RNN) to achieve better performances on downstream applications such as land feature classification, image geolocalization, traffic prediction, to name a few [3]. That is, the comprehensive linkages between SAM methods and DL models have seldom been examined so far. There is usually insufficient context provided on the logic to adopt certain neural network architecture and a lack of investigation on how new GeoAI models would contribute to the current family of spatial analytics.

3 INTELLIGENT SPATIAL ANALYTICS

Spatial analytics have been evolving through the integration of better theories, following strict statistical principles, fixed domain knowledge, and classic computation paradigm. In the complex human-environment systems, contemporary questions raised are even harder to answer, while SAM methods are not intelligent enough to cope with the exploding data, handle the geospatial complexity, take advantage of the computational innovation that is happening, or support new geographic knowledge discovery. The focus of this paper is to propose the concept of Intelligent Spatial Analytics (ISA). We acknowledge that ambiguity may persist in understanding what can be considered as intelligence in spatial analysis and uncertainty may even increase when embracing the transformation. However, we assume there are some general trends towards the intelligent era of SAM. We invite future works to explore the combination of ISA and current spatial analytics in fields such as geostatistics, spatial econometrics, spatial interaction modeling, spatial optimization, pattern classification/clustering, spatial simulation, human mobility, etc.

3.1 Motivation

We suggest ISA to be a new set of AI-powered computational approaches designed specifically for SAM. Conceptually, ISA methods and models are defined based on spatially explicit deep neural networks that fulfill the invariance, representation, formulation and outcome tests in the context of GeoAI [6]. Besides that, there are four basic motivations to be considered:

- **Flexible data representation**: Usage of various neural network techniques to handling diverse data structures, e.g., regular CNN for rasterized spatial data while graph convolutional networks (GCN) for vectorized spatial data.
- **Modules for spatial dependence**: Explicit consideration to model complex spatial dependence concepts such as the notion of neighborhood, variation structure, autocorrelation, distance decay, scale, etc.
- **Weak model assumption**: Loose the strict statistical assumptions and not to over-specify prior model definition on the spatial process.
- **Ability to predict and explain unknowns**: Go beyond just summarizing patterns into better supporting geospatial knowledge discovery and contributing to explainable AI.

In the following sections, we select three methodological branches in spatial analysis as examples to show how traditional spatial analytics can be further enriched in the vision of ISA.

3.2 Geostatistics

The key of geostatistics is to derive statistical relationships describing how the values of a target spatial variable are related to the information collected at samples. Traditional Kriging-based interpolation devises semivariogram models using prior functions such as Gaussian, Exponential and Spherical to approximate the variation structure γ w.r.t. lag distances h. Then the fitted model f_γ^* provides the weights knowledge for interpolation. In view of ISA, spatially explicit neural networks can be developed to formalize the spatial interpolation process without such strong prior model assumptions. For example, a generative adversarial neural networks (GAN) can be adopted to approximate the spatial conditioned probability distribution $p_{model}(\hat{Z}|x_j, \forall j \in 1, \cdots, k)$ that best describes data observations $p_{data}(\hat{Z}|x_j, \forall j \in 1, \cdots, k)$ through an adversarial game between the generator G and the discriminator D [21], as shown in Figure 1. By structuring each location as a neuron in the deep learning model, learnable convolutional weights can imitate the data-borrowing process from nearby samples, offering a more flexible notion on the spatial dependence. Besides, auxiliary learning is found to be able to further amplify spatial knowledge such as the autoregressive structures and local patterns for more accurate geostatistical modeling [8].

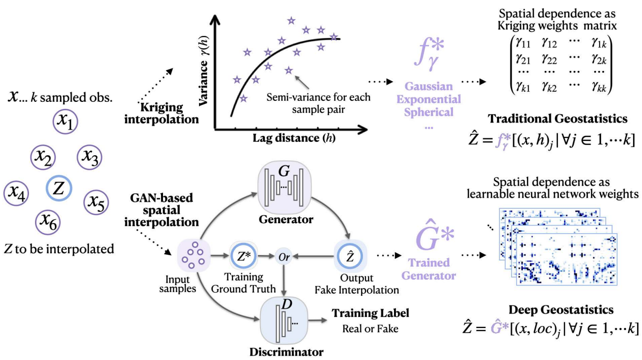

Figure 1: Traditional Kriging-based v.s. Generative adversarial learning based spatial interpolation

3.3 Spatial econometrics

For scenarios where the multivariate distributional data is available, spatial regression models are commonly used to seek the statistical associations between observed variables. Current endeavors rooted in spatial econometrics use prior model assumptions such as the linearity of regression equations, the normality of distributions, and ad hoc spatial dependence structure in weights matrix. These assumptions overlook the non-linear nature of spatial associations and enforce the specification of spatial lag effects such as the auto-regressive, cross-regressive and linear coefficients to be within the predefined weights matrix and linear regression. As shown in Figure 2, GCN as a variant of the CNN framework, is naturally suitable to build a conceptual mapping between the graph structure and the spatial weights matrix for irregularly distributed spatial units, and is capable of capturing complex spatial lagged effects via the multi-layer graph convolutional filters across feature channels. Encouraging evidences have been founded recently that GCN can replicate the workflow of spatial regression, thanks to its propagation mechanisms, spatial locality learning nature and the semi-supervised training strategy [22]. This new spatial regression logic has been adopted in several latest urban studies to understand complex association relationships based on irregular geographic units such as places [23] and street segments [16].

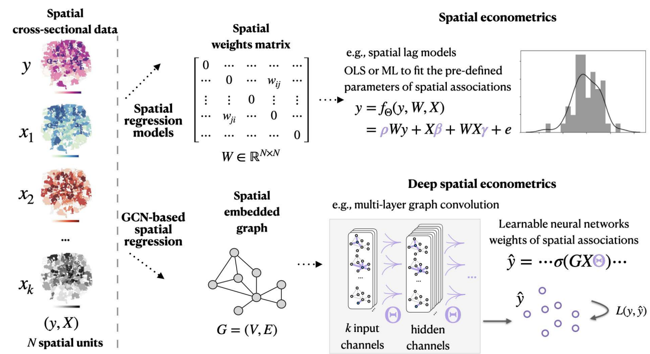

Figure 2: Traditional spatial lag models v.s. Graph convolutional neural networks in spatial regression

3.4 Flow analytics

Flow analytics aim at modeling the spatial flow intensity between locations (e.g., the number of people move from one location to another) given the demographics and geographic characteristics (e.g., population and distance). Traditional spatial interaction models such as the gravity model and radiation model assume the flow between two locations increases with the locations' populations while decreases with spatial displacement factors like distance or intervening opportunities. These models have relatively fixed inputs and formal expressions, unable to capture the structure of real flows, and also ignore the rich features that are essential to account for the complex geographical landscape. In light of intelligent flow analytics, deep gravity model [15] extends gravity model into a shallow neural network with added hidden layers to introduce nonlinearities and additional geographical features such as land uses and POI types. Latest GCN-based deep learning models such as SI-GCN [19] and ConvGCN-RF [20] also report accuracy improvement in predicting unobserved spatial flows with the help of distributed embeddings to integrate diverse geographic features and the spatially-informed graph convolutional modules to approximate the greater variability of real spatial flows [14].

4 POTENTIALS AND CHALLENGES

In this section, we outline several challenging issues that need to be addressed to advance the development of ISA.

- **Transferability and Generalization**. GeoAI model's transferability and generalization across space are weak due to the spatial heterogeneity [4]. To address this issue, geospatial knowledge-informed models have been developed and are generalizable for both natural and man-made features [11]. Spatial-heterogeneity-aware deep learning architectures have shown promising results in spatial prediction tasks [17]. ISA should take the advantages of generalization capability across geographic scales from traditional SAM and the automatic spatiotemporal feature extraction capability from deep neural networks. ISA models for small regions should be transferable to large regions and vice versa.

- **Interpretability and Explainability**. To open the AI model "black boxes", great efforts have been made in the AI community to increase the interpretability and explainability of deep learning models, such as the layer-wise relevance propagation (LRP) to assess the feature importance in classification tasks and the attention mechanisms to explain the relevant context in neural networks [11]. Future development of ISA should keep the interpretability and explainability in mind, especially how to incorporate spatial principles (e.g, spatial dependence) and geographic knowledge to advance explainable ISA models. One direction might be the incorporation of spatiotemporal-LRP and attention weights in assessing the relevance of geographic contexts in spatially explicit neural network models [18]. Moreover, the inclusion of causal inference capabilities including association, intervention and counterfactuals from econometrics would further enhance the intelligence of spatial analytics [13].
- **Biases and Ethics**. There may exist multiple types of biases in the ISA computational frameworks such as data bias, model bias, and inductive bias. ISA may become reliant on undesired sensitive features (e.g., human socioeconomic status), and result in unfair decision making. Those biases should be mitigated throughout the data-model-action loop to ensure equitable outcomes. Emerging ethical principles including transparency, justice, non-maleficence, responsibility and privacy should also be considered when developing ISA for social good [7]. For example, while human mobility and digital contact tracing are important for geospatial modeling of virus spread, privacy concerns in location tracking have generated barriers for data usage and method replicability.

5 CONCLUSIONS

In this paper, we call for attentions on Intelligent Spatial Analytics (ISA), a set of new computational approaches designed specifically for spatial analysis and modeling based on spatially explicit deep neural networks. ISA methods and models have the motivations to extend traditional spatial analytics with more flexible data representation, intelligent module for spatial dependence, weaker model prior assumptions, and higher accuracy for predictions and explanations. Geostatistics, spatial econometrics and flow analytics are presented as examples of how major methodological branches in spatial analysis can be enriched with the ISA motivations. Researchers should also be aware of challenging issues such as model transferability, interpretability, bias and ethics to promote the further breadth and widespread of GeoAI in spatial analytics.

ACKNOWLEDGMENTS

Di Zhu is supported by the New Faculty Set-up Funding, University of Minnesota (1000-10964-20042-5672018) and the Faculty Interactive Research Program from Center for Urban and Regional Affairs (1801-10964-21584-5672018); Song Gao is supported by the National Science Foundation (no. 2112606); Guofeng Cao is supported by the National Science Foundation (no. 2026331).

REFERENCES

[1] Ying Chen, Xu Ouyang, and Gady Agam. 2019. ChangeNet: Learning to detect changes in satellite images. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. 24–31.

[2] Manfred M Fischer. 1998. Computational neural networks: a new paradigm for spatial analysis. *Environment and Planning A* 30, 10 (1998), 1873–1891.

[3] Song Gao. 2020. A review of recent researches and reflections on geospatial artificial intelligence. *Geomatics and Information Science of Wuhan University* 45, 12 (2020), 1865–1874.

[4] Michael F Goodchild and Wenwen Li. 2021. Replication across space and time must be weak in the social and environmental sciences. *Proceedings of the National Academy of Sciences* 118, 35 (2021).

[5] Yingjie Hu, Song Gao, Dalton Lunga, Wenwen Li, Shawn Newsam, and Budhendra Bhaduri. 2019. GeoAI at ACM SIGSPATIAL: progress, challenges, and future directions. *Sigspatial Special* 11, 2 (2019), 5–15.

[6] Krzysztof Janowicz, Song Gao, Grant McKenzie, Yingjie Hu, and Budhendra Bhaduri. 2020. GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science* 34, 4 (2020), 625–636.

[7] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9 (2019), 389–399.

[8] Konstantin Klemmer and Daniel B Neill. 2021. Auxiliary-task learning for geographic data with autoregressive embeddings. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*. 141–144.

[9] Stephen Law and Mateo Neira. 2019. An unsupervised approach to geographical knowledge discovery using street level and street network images. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. 56–65.

[10] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.

[11] Wenwen Li, Chia-Yu Hsu, and Maosheng Hu. 2021. Tobler's First Law in GeoAI: A spatially explicit deep learning model for terrain feature detection under weak supervision. *Annals of the American Association of Geographers* 111, 7 (2021), 1887–1905.

[12] Stan Openshaw and Christine Openshaw. 1997. *Artificial intelligence in geography*. John Wiley & Sons, Inc.

[13] Judea Pearl. 2019. The seven tools of causal inference, with reflections on machine learning. *Commun. ACM* 62, 3 (2019), 54–60.

[14] Nastaran Pourebrahim, Selima Sultana, Jean-Claude Thill, and Somya Mohanty. 2018. Enhancing trip distribution prediction with twitter data: comparison of neural network and gravity models. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. 5–8.

[15] Filippo Simini, Gianni Barlacchi, Massimilano Luca, and Luca Pappalardo. 2021. A Deep Gravity model for mobility flows generation. *Nature communications* 12, 1 (2021), 1–13.

[16] Longzhu Xiao, Siuming Lo, Jiangping Zhou, Jixiang Liu, and Linchuan Yang. 2021. Predicting vibrancy of metro station areas considering spatial relationships through graph convolutional neural networks: The case of Shenzhen, China. *Environment and Planning B: Urban Analytics and City Science* 48, 8 (2021), 2363–2384.

[17] Yiqun Xie, Erhu He, Xiaowei Jia, Han Bao, Xun Zhou, Rahul Ghosh, and Praveen Ravirathinam. 2021. A statistically-guided deep network transformation and moderation framework for data with spatial heterogeneity. In *2021 IEEE International Conference on Data Mining (ICDM)*. IEEE, 767–776.

[18] Bo Yan, Krzysztof Janowicz, Gengchen Mai, and Song Gao. 2017. From ITDL to Place2Vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In *Proceedings of the 25th ACM SIGSPATIAL international conference on advances in geographic information systems*. 1–10.

[19] Xin Yao, Yong Gao, Di Zhu, Ed Manley, Jiaoe Wang, and Yu Liu. 2020. Spatial origin-destination flow imputation using graph convolutional networks. *IEEE Transactions on Intelligent Transportation Systems* 22, 12 (2020), 7474–7484.

[20] Ganmin Yin, Zhou Huang, Yi Bao, Han Wang, Linna Li, Xiaolei Ma, and Yi Zhang. 2022. ConvGCN-RF: A hybrid learning model for commuting flow prediction considering geographical semantics and neighborhood effects. *GeoInformatica* (2022), 1–21.

[21] Di Zhu, Ximeng Cheng, Fan Zhang, Xin Yao, Yong Gao, and Yu Liu. 2020. Spatial interpolation using conditional generative adversarial neural networks. *International Journal of Geographical Information Science* 34, 4 (2020), 735–758.

[22] Di Zhu, Yu Liu, Xin Yao, and Manfred M Fischer. 2021. Spatial regression graph convolutional neural networks: A deep learning paradigm for spatial multivariate distributions. *GeoInformatica* (2021), 1–32.

[23] Di Zhu, Fan Zhang, Shengyin Wang, Yaoli Wang, Ximeng Cheng, Zhou Huang, and Yu Liu. 2020. Understanding place characteristics in geographic contexts through graph convolutional neural networks. *Annals of the American Association of Geographers* 110, 2 (2020), 408–420.