Machine Learning Meets Big Spatial Data

Ibrahim Sabek
Dept. of Computer Science and Engineering
University of Minnesota
Minnesota, USA
sabek@cs.umn.edu

Mohamed F. Mokbel*
Qatar Computing Research Institute
Hamad bin Khalifa University
Doha, Qatar
mmokbel@hbku.edu.qa

ABSTRACT

The proliferation in amounts of generated data has propelled the rise of scalable machine learning solutions to efficiently analyze and extract useful insights from such data. Meanwhile, spatial data has become ubiquitous, e.g., GPS data, with increasingly sheer sizes in recent years. The applications of big spatial data span a wide spectrum of interests including tracking infectious disease, climate change simulation, drug addiction, among others. Consequently, major research efforts are exerted to support efficient analysis and intelligence inside these applications by either providing spatial extensions to existing machine learning solutions or building new solutions from scratch. In this 90-minutes tutorial, we comprehensively review the state-of-the-art work in the intersection of machine learning and big spatial data. We cover existing research efforts and challenges in three major areas of machine learning, namely, data analysis, deep learning and statistical inference, as well as two advanced spatial machine learning tasks, namely, spatial features extraction and spatial sampling. We also highlight open problems and challenges for future research in this area.

PVLDB Reference Format:
Ibrahim Sabek, and Mohamed F. Mokbel. Machine Learning Meets Big Spatial Data. *PVLDB*, 12(12): 1982-1985, 2019.
DOI: https://doi.org/10.14778/3352063.3352115

Keywords

Big Spatial Data, Machine Learning, Scalability

1. INTRODUCTION

There has been a recent wide deployment of machine learning (ML) solutions, with their different areas (e.g., data analysis, deep learning), in various big data applications, including public health [19], information extraction [47], data cleaning [37], among others. Meanwhile, spatial applications have witnessed unprecedented explosion in the amounts of generated and collected data. For example, space telescopes generate up to 150 GB weekly spatial data, medical devices produce spatial images (X-rays) at a rate of 50 PB per year, while a NASA archive of satellite earth images has more than 500 TB. To efficiently process such tremendous amounts of spatial data, researchers and developers worldwide have proposed either spatial extensions to existing machine learning systems (e.g., Azure Geo AI [3]) or new end-to-end solutions (e.g., ESRI ArcGIS [11]). Such extensions and new solutions have motivated a wide variety of applications in biology [50], environmental science [51], climatology [14], among others.

In this tutorial, we aim to provide a comprehensive review of existing machine learning systems and approaches that efficiently support big spatial data. In particular, we focus on explaining the main ideas, architectures, strengths and weaknesses of existing systems and approaches. We also highlight the strong bond between spatial data management and spatial machine learning workflows, discuss the related technical challenges, and outline the open research opportunities. Previous SIGMOD tutorials have focused on the techniques and challenges in machine learning for big data in general [7, 25]. Another previous VLDB tutorial focused on big spatial data management [10]. Unlike these tutorials, our tutorial aims to combine the two worlds of *scalable machine learning* and *big spatial data* together, which is beyond just applying techniques from one area to another.

2. TUTORIAL OUTLINE

Figure 1 gives the **90-minutes** tutorial outline, composed of six parts. The first part motivates the need for machine learning systems to support big spatial data, and provides the basic background on these two worlds (Section 2.1). The second, third, and fourth parts delve into the ongoing efforts and challenges of supporting big spatial data in three major areas of machine learning, namely, *data analysis*, *deep learning*, and *statistical inference*, respectively (Sections 2.2-2.4). The fifth part reviews advanced spatial machine learning pipeline in terms of two common tasks, namely, *spatial features extraction*, and *spatial sampling* (Section 2.5). The sixth part concludes the tutorial by discussing several open research problems (Section 2.6).

[*] Also affiliated with University of Minnesota, MN, USA.

[1] This work is partially supported by the National Science Foundation, USA, under Grants IIS-1525953 and CNS-1512877.

Proceedings of the VLDB Endowment, Vol. 12, No. 12
ISSN 2150-8097.
DOI: https://doi.org/10.14778/3352063.3352115

2.1 Part 1: Spatial Data and ML Synergy

This part advocates for the need to develop machine learning systems and techniques for big spatial data that go beyond simple extensions of existing work for general data. We start by describing some motivating applications, introducing the world of big spatial data, and discussing its machine learning related concepts. We then quickly review the landscape of spatial machine learning systems, algorithms, applications, and needs. Then, we give a brief introduction about three major machine learning areas, namely, spatial data analysis, spatial deep learning, and spatial statistical inference, which will be heavily discussed in the next parts.

2.2 Part 2: Data Analysis Solutions

This part covers the big spatial data analysis systems and approaches from three aspects: (1) The research efforts of adding spatial support in existing big data analysis systems and tools, which are either: (a) in the form of add-ons libraries and tools that enable processing spatial data with classical operations (e.g., clustering, classification). Examples include spatial extensions to Spark core (e.g., Simba [57], Magellan [29], GeoSpark [60], GeoMesa [20], UlTraMan [9]) to enable using Spark MLib [30] with spatial data, ESRI spatial data analysis extensions for Hive [12], and PostGIS [34] that can be used along with MADLib [19] to support spatial analytics for PostgreSQL [35], or (b) in the form of built-in native support of spatial analysis operations (e.g., hot spot detection, spatial co-location) inside existing data analysis engines. (2) The research efforts of providing full-fledged big spatial data analysis systems and tools. In such systems, all execution steps in any data analysis operation are optimized for efficient and scalable processing of spatial data. We will classify existing work based on the underlying architecture, which could be either (a) *in-memory systems* (e.g., CrimeStat [27], GDAL [55], GeoDa [2], PySAL [38]), (b) *RDBMS-based systems* (e.g., ESRI ArcGIS [11]), or (c) *cloud-based services* (e.g., IBM PAIRS [23]). For all these systems and services, we will give motivational case studies, and a brief on their supported spatial analysis operations and running time efficiency. (3) The research efforts for the scalability of five stand alone (without much of system support) common big spatial analysis operations, namely, spatial outlier detection [46, 63], spatial classification [8, 16, 22], spatial clustering [13, 31, 54, 53, 61], hotspot detection [4], and spatial co-location [36].

2.3 Part 3: Deep Learning Solutions

This part covers the interplay between spatial data management and analysis techniques and deep learning approaches. We start by highlighting the role of spatial data management techniques in improving the performance of various deep learning tasks when applying on big spatial data. For example, Quad-tree partitioning [15] is used for: (a) balancing the convolution computation in Convolutional Neural Networks (CNN) for object detection applications [21] and (b) efficient automatic features extraction and matrix factorization operations inside deep learning models [59]. Meanwhile, k-nearest neighbor operations are used to efficiently build specific neural network architectures from big spatial datasets [6, 33]. Then, we discuss the role of deep learning in efficiently supporting numerous large-scale spatial prediction queries (e.g., aggregate prediction [56], forecasting queries [28]), and other spatial analysis tasks (e.g., geospatial object detection [58], outdoors localization [48]).

- **Part 1: Spatial Data and ML Synergy (10 minutes)**
  - Importance of ML with big spatial data
  - Quick review of spatial ML landscape
- **Part 2: Data Analysis (DA) Solutions (25 minutes)**
  - Spatial support in big data analysis systems
  - End-to-end big spatial data analysis systems
  - Scalability of common spatial analysis operations
- **Part 3: Deep Learning (DL) Solutions (20 minutes)**
  - Spatial DB techniques to improve DL approaches
  - DL approaches to improve spatial prediction queries
- **Part 4: Statistical Inference Solutions (15 minutes)**
  - Review of spatial Bayesian inference concepts
  - Approaches for supporting scalable spatial inference
- **Part 5: Advanced ML Pipeline Tasks (10 minutes)**
  - Spatial features extraction techniques
  - Spatial sampling techniques
- **Part 6: Future Opportunities (10 minutes)**
  - Inference model maintenance
  - Spatial ML models optimizer
  - Spatially-optimized deep learning frameworks

Figure 1: Tutorial Outline (90 minutes)

2.4 Part 4: Statistical Inference Solutions

This part covers the scalable Bayesian inference solutions that are designed to analyze big spatial data. We will start by a brief review for the basic statistical concepts of spatial Bayesian modeling and their big data applications. We will then discuss existing spatial inference systems and approaches, categorized into: (a) *in-memory* solutions, where the input dataset of the inference model is first spatially partitioned into a grid. Then, each partition is analyzed using a Bayesian spatial process model (e.g., [17]). Finally, an approximate posterior inference for the entire dataset is obtained by optimally combining the individual posterior distributions from each partition [17, 44, 49]. (b) *RDBMS-based* solutions, where the assumption of fitting the whole model data in memory is no longer valid. Hence, RDBMSs are exploited to support scalable spatial inference computation. Although, there are many RDBMS-based inference systems (e.g., [5, 32, 47]), they do not provide specialized support for spatial data and operations. As a result, recent research efforts started to support the spatial inference in these systems through either implementing on-top user-defined functions, e.g., TurboReg [40] and Flash [42, 39] on top of DeepDive [47], and [43] on top of Alchemy [1], or providing built-in modules, e.g., Sya [41] inside DeepDive [47].

2.5 Part 5: Advanced ML Pipeline Tasks

This part covers more advanced spatial machine learning tasks. In particular, we focus on two main tasks that are used extensively in different spatial machine learning pipelines: (1) *Spatial features extraction.* Many spatial machine learning algorithms require the features extraction from raw spatial data as a pre-processing step, which is very time-consuming [26]. In response, the database community has offered system solutions to scale up the performance of such task. For example, SkewReduce [26] is a Hadoop-based system that expresses and executes the spatial features extraction task in a scalable manner while avoiding skewness issues, TELEIOS [24] is a scalable data exploration system that provides a built-in support for spatial features extraction, while DeepDive [47] is an RDBMS-based information extraction system that can be exploited to extract spatial features through user-defined functions as shown in a recent work [40]. (2) *Spatial sampling.* Due to the massive amounts of spatial data that are available for training any spatial machine learning algorithm, spatial sampling becomes a critical task to efficiently select a set of representative data objects while taking the spatial distribution into account. Existing sampling techniques over big spatial data can be either incremental (i.e., generated samples are refined over many iterations) [52] or satisfying certain locality constraints (e.g., zooming level) [18, 45].

2.6 Part 6: Future Opportunities

This part discusses several future opportunities and open research challenges in the intersection of machine learning with big spatial data and applications. In particular, we will focus on the following three points: (1) *Inference model maintenance*: Materialized views have been heavily used to support incremental maintenance over inference models and their predictions [47]. However, typical materialization techniques are not efficient to handle spatial data. A promising direction is to exploit recent materialization techniques for multi-dimensional data [62] to support incremental spatial inference. (2) *Spatial ML models optimizer*: There is a major opportunity in borrowing ideas from spatial database query optimizers (e.g., cost models, and operations re-ordering) to efficiently select among different spatial machine learning models. (3) *Spatially-optimized deep learning frameworks*: Unlike spatial data analysis frameworks (e.g., [11, 23]), all current deep learning approaches for big spatial data applications are stand alone efforts, without any system support. A future direction is to distill the commonalities from all these approaches and bring them into end-to-end full-fledged system frameworks.

3. TARGET AUDIENCE

This tutorial targets researchers, developers, and practitioners, who are interested in large-scale machine learning and big spatial data. No prior knowledge is required to understand the systems and approaches in the tutorial. The tutorial will also be very beneficial for graduate students as it will help in identifying various topics and challenges for PhD topics. Practitioners will get to know the state-of-the-art systems for enriching their machine learning systems and tools with spatial data support. This tutorial will act as an invitation to the database community to join arms for satisfying the emerging needs of big spatial data analysis and machine learning applications.

4. RELEVANCE TO VLDB

Research in the areas of spatial data and scalable machine learning has been always active in the database community in general, and in the VLDB community in particular. With the proliferation of proposed systems and approaches in these areas, it becomes inevitable to present a tutorial that surveys the current state-of-the-art techniques and suggests future research directions for the community. Many of the research efforts covered in this tutorial were recently published in major database conferences including ICDE, VLDB, and SIGMOD [4, 5, 9, 18, 19, 24, 31, 32, 41, 42, 45, 47, 52, 53, 54, 57, 61, 62].

5. BIOGRAPHICAL SKETCHES

Ibrahim Sabek is a PhD candidate at the department of Computer Science and Engineering, University of Minnesota. He received his M.Sc. degree at the same department in 2017. His research interests lie in the intersection area between big spatial data management, spatial computing, and scalable machine learning systems. Ibrahim has been awarded the University of Minnesota Doctoral Dissertation Fellowship in 2019 for this dissertation focus on scalable machine learning for big spatial data and applications. His research work has been nominated for the Best Paper Award of ACM SIGSPATIAL 2018, and has been qualified to the final stage of ACM SIGMOD Student Research Competition (SRC) 2017. During his PhD, he has collaborated with NEC Labs America, and Microsoft Research (MSR) in Redmond. Ibrahim has published many papers in top research venues, including ACM TSAS, IEEE ICDE, ACM SIGSPATIAL, IEEE TMC, and demonstrated his work at ACM SIGMOD. For more information, please visit: http://www.cs.umn.edu/∼sabek.

Mohamed F. Mokbel is the Chief Scientist of Qatar Computing Research Institute and a Professor at University of Minnesota. His current research interests focus on systems and machine learning techniques for big spatial data and applications. His research work has been recognized by the VLDB 10-years Best Paper Award, four conference Best Paper Awards, and the NSF CAREER Award. Mohamed has delivered six tutorials in VLDB/SIGMOD/ICDE/EDBT conferences, in addition to tutorials in other communities' first-tier venues, including IEEE ICDM and ACM CCS. None of these tutorials overlaps with this tutorial proposal. Mohamed is the past elected Chair of ACM SIGPATIAL, current Editor-in-Chief for Distributed and Parallel Databases Journal, and on the editorial board of ACM Books, ACM TODS, VLDB Journal, ACM TSAS, and GoeInformatica journals. He has also served as PC Vice Chair of ACM SIGMOD and PC Co-Chair for ACM SIGSPATIAL and IEEE MDM. For more information, please visit: www.cs.umn.edu/∼mokbel.

6. REFERENCES

[1] Alchemy. https://alchemy.cs.washington.edu/.

[2] L. Anselin et al. GeoDa: An Introduction to Spatial Data Analysis. *Journal of Geographical Analysis*, 38(1):5–22, 2006.

[3] Azure Geo AI. https://azure.microsoft.com/en-us/blog/microsoft-and-esri-launch-geospatial-ai-on-azure/.

[4] S. Bhadange, A. Arora, and A. Bhattacharya. GARUDA: A System for Large-scale Mining of Statistically Significant Connected Subgraphs. *PVLDB*, 9(13):1449–1452, 2016.

[5] Z. Cai, Z. Vagena, L. Perez, S. Arumugam, P. J. Haas, and C. Jermaine. Simulation of Database-valued Markov Chains Using SimSQL. In *SIGMOD*, pages 637–648, 2013.

[6] C.-R. Chen and U. T. Kartini. K-Nearest Neighbor Neural Network Models for Very Short-Term Global Solar Irradiance Forecasting Based on Meteorological Data. *Journal of Energies*, 10(2):186–203, 2017.

[7] T. Condie, P. Mineiro, N. Polyzoti, and M. Weimer. Machine Learning for Big Data (Tutorial). In *SIGMOD*, pages 939–942, 2013.

[8] E. Diday. Spatial Classification. *Journal of Discrete Applied Mathematics*, 156(8):1271–1294, 2008.

[9] X. Ding, L. Chen, Y. Gao, C. S. Jensen, and H. Bao. UlTraMan: A Unified Platform for Big Trajectory Data Management and Analytics. *PVLDB*, 11(7):787–799, 2018.

[10] A. Eldawy and M. F. Mokbel. The Era of Big Spatial Data (Tutorial). *PVLDB*, 10(12):1992–1995, 2017.

[11] ESRI ArcGIS. `https://www.esri.com/en-us/arcgis/about-arcgis/overview`.

[12] ESRI Tools for Hive. `https://github.com/Esri/spatial-framework-for-hadoop`.

[13] M. Ester, H. Kriegel, J. Sander, and X. Xu. A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *SIGKDD*, pages 226–231, 1996.

[14] J. H. Faghmous and V. Kumar. *Spatio-temporal Data Mining for Climate Data: Advances, Challenges, and Opportunities*, pages 83–116. Springer, 2014.

[15] R. Finkel and J. Bentley. Quad Trees a Data Structure for Retrieval on Composite Keys. *Acta Informatica*, 1974.

[16] R. Frank, M. Ester, and A. Knobbe. A Multi-relational Approach to Spatial Classification. In *SIGKDD*, pages 309–318, 2009.

[17] R. Guhaniyogi and S. Banerjee. Meta-Kriging: Scalable Bayesian Modeling and Inference for Massive Spatial Datasets. *Journal of Technometrics*, 60(4):430–444, 2018.

[18] T. Guo, K. Feng, G. Cong, and Z. Bao. Efficient Selection of Geospatial Data on Maps for Interactive and Visualized Exploration. In *SIGMOD*, pages 567–582, 2018.

[19] J. M. Hellerstein, C. R'e, F. Schoppmann, D. Z. Wang, E. Fratkin, A. Gorajek, K. S. Ng, C. Welton, X. Feng, K. Li, and A. Kumar. The MADlib Analytics Library: or MAD Skills, the SQL. *PVLDB*, 5(12):1700–1711, 2012.

[20] J. Hughes et al. GeoMesa: A Distributed Architecture for Spatio-temporal Fusion. In *SPIE Defense+Security*, 2015.

[21] P. K. Jayaraman et al. Quadtree Convolutional Neural Networks. In *ECCV*, pages 546–561, 2018.

[22] Z. Jiang and S. Shekhar. *Spatial Big Data Science: Classification Techniques for Earth Observation Imagery*. Springer Publishing Company, 1st edition, 2017.

[23] L. J. Klein et al. PAIRS: A Scalable Geo-spatial Data Analytics Platform. In *IEEE Big Data*, pages 1290–1298, 2015.

[24] M. Koubarakis, M. Datcu, C. Kontoes, U. Giammatteo, S. Manegold, and E. Klien. TELEIOS: A Database-powered Virtual Earth Observatory. *PVLDB*, 5(12):2010–2013, 2012.

[25] A. Kumar, M. Boehm, and J. Yang. Data Management in Machine Learning: Challenges, Techniques, and Systems (Tutorial). In *SIGMOD*, pages 1717–1722, 2017.

[26] Y. Kwon et al. Skew-resistant Parallel Processing of Feature-extracting Scientific User-defined Functions. In *SoCC*, pages 75–86, 2010.

[27] N. Levine. *CrimeStat: A Spatial Statistical Program for the Analysis of Crime Incidents*, pages 381–388. Springer, 2017.

[28] Y. Lin et al. Exploiting Spatiotemporal Patterns for Accurate Air Quality Forecasting Using Deep Learning. In *SIGSPATIAL*, pages 359–368, 2018.

[29] Magellan: Geospatial analytics using spark. `https://github.com/harsha2010/magellan`.

[30] X. Meng et al. MLlib: Machine Learning in Apache Spark. *Journal of Machine Learning Research*, 17(1), 2016.

[31] R. T. Ng and J. Han. Efficient and Effective Clustering Methods for Spatial Data Mining. In *VLDB*, pages 144–155, 1994.

[32] F. Niu, C. Ré, A. Doan, and J. Shavlik. Tuffy: Scaling Up Statistical Inference in Markov Logic Networks Using an RDBMS. *PVLDB*, 4(6):373–384, 2011.

[33] T. Plötz and S. Roth. Neural Nearest Neighbors Networks. In *NIPS*, pages 1087–1098, 2018.

[34] PostGIS. `http://postgis.net/`.

[35] PostgreSQL. `https://www.postgresql.org/`, 2019.

[36] F. Qian, Q. He, and J. He. Mining Spatial Co-location Patterns with Dynamic Neighborhood Constraint. In *European Conference on ML and Knowledge Discovery in Databases*, pages 238–253, 2009.

[37] T. Rekatsinas, X. Chu, I. F. Ilyas, and C. Ré. HoloClean: Holistic Data Repairs with Probabilistic Inference. *PVLDB*, 10(11):1190–1201, 2017.

[38] S. Rey et al. *PySAL: A Python Library of Spatial Analytical Methods*, pages 175–193. Springer, 2010.

[39] I. Sabek. Adopting Markov Logic Networks for Big Spatial Data and Applications. In *VLDB PhD Workshop*, 2019.

[40] I. Sabek, M. Musleh, and M. Mokbel. TurboReg: A Framework for Scaling Up Spatial Logistic Regression Models. In *SIGSPATIAL*, pages 129–138, 2018.

[41] I. Sabek, M. Musleh, and M. F. Mokbel. A Demonstration of Sya: A Spatial Probabilistic Knowledge Base Construction System. In *SIGMOD*, pages 1689–1692, 2018.

[42] I. Sabek, M. Musleh, and M. F. Mokbel. Flash in Action: Scalable Spatial Data Analysis Using Markov Logic Networks. *PVLDB*, 12(12):1834–1837, 2019.

[43] N. A. Sakhanenko and D. J. Galas. Markov Logic Networks in the Analysis of Genetic Data. *Journal of Computational Biology*, 17(11):1491–1508, Nov. 2010.

[44] Y.-L. K. Samo and S. Roberts. Scalable Nonparametric Bayesian Inference on Point Processes with Gaussian Processes. In *ICML*, pages 2227–2236, 2015.

[45] A. D. Sarma, H. Lee, H. Gonzalez, J. Madhavan, and A. Halevy. Efficient Spatial Sampling of Large Geographical Tables. In *SIGMOD*, pages 193–204, 2012.

[46] S. Shekhar, C.-T. Lu, and P. Zhang. Detecting Graph-based Spatial Outliers: Algorithms and Applications (a Summary of Results). In *SIGKDD*, pages 371–376, 2001.

[47] J. Shin, S. Wu, F. Wang, C. D. Sa, C. Zhang, and C. Ré. Incremental Knowledge Base Construction Using DeepDive. *PVLDB*, 8(11):1310–1321, 2015.

[48] A. Shokry, M. Torki, and M. Youssef. DeepLoc: A Ubiquitous Accurate and Low-overhead Outdoor Cellular Localization System. In *SIGSPATIAL*, pages 339–348, 2018.

[49] C. R. Stephens, V. Snchez-Cordero, and C. G. Salazar. Bayesian Inference of Ecological Interactions from Spatial Data. *Journal of Entropy*, 19(12), 2017.

[50] R. Tibshirani and P. Wang. Spatial Smoothing and Hot Spot Detection for CGH Data Using the Fused Lasso. *Biostatistics*, 9(1):18–29, Jan. 2008.

[51] T. VoPham et al. Emerging Trends in Geospatial Artificial Intelligence (geoAI): Potential Applications for Environmental Epidemiology. *Environmental Health*, 2018.

[52] L. Wang, R. Christensen, F. Li, and K. Yi. Spatial Online Sampling and Aggregation. *PVLDB*, 9(3):84–95, 2015.

[53] W. Wang, J. Yang, and R. R. Muntz. STING: A Statistical Information Grid Approach to Spatial Data Mining. In *VLDB*, pages 186–195, 1997.

[54] W. Wang, J. Yang, and R. R. Muntz. STING+: An Approach to Active Spatial Data Mining. In *ICDE*, pages 116–125, 1999.

[55] F. Warmerdam. *The Geospatial Data Abstraction Library*, pages 87–104. Springer Berlin Heidelberg, 2008.

[56] H. Wei et al. Residual Convolutional LSTM for Tweet Count Prediction. In *WWW*, pages 1309–1316, 2018.

[57] D. Xie, F. Li, B. Yao, G. Li, L. Zhou, and M. Guo. Simba: Efficient In-Memory Spatial Analytics. In *SIGMOD*, pages 1071–1085, 2016.

[58] Y. Xie et al. An Unsupervised Augmentation Framework for Deep Learning Based Geospatial Object Detection: A Summary of Results. In *SIGSPATIAL*, pages 349–358, 2018.

[59] H. Yin, W. Wang, H. Wang, L. Chen, and X. Zhou. Spatial-Aware Hierarchical Collaborative Deep Learning for POI Recommendation. *TKDE*, 29(11):2537–2551, 2017.

[60] J. Yu, Z. Zhang, and M. Sarwat. Spatial Data Management in Apache Spark: The GeoSpark Perspective and Beyond. *Journal of GeoInformatica*, pages 1–44, 2018.

[61] T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH: An Efficient Data Clustering Method for Very Large Databases. In *SIGMOD*, pages 103–114, 1996.

[62] W. Zhao, F. Rusu, B. Dong, K. Wu, and P. Nugent. Incremental View Maintenance over Array Data. In *SIGMOD*, pages 139–154, 2017.

[63] G. Zheng, S. L. Brantley, T. Lauvaux, and Z. Li. Contextual Spatial Outlier Detection with Metric Learning. In *SIGKDD*, pages 2161–2170, 2017.