Towards Algorithmic Fairness in Space-Time: Filling in Black Holes

Anonymous Author(s)
Affiliation
Address
email

Abstract

New technologies and the availability of geospatial data have drawn attention to spatio-temporal biases present in society. For example: the COVID-19 pandemic highlighted disparities in the availability of broadband service and its role in the digital divide; the environmental justice movement in the United States has raised awareness to health implications for minority populations stemming from historical redlining practices; and studies have found varying quality and coverage in the collection and sharing of open-source geospatial data. Despite the extensive literature on machine learning (ML) fairness, few algorithmic strategies have been proposed to mitigate such biases. In this paper we highlight the unique challenges for quantifying and addressing bias in spatio-temporal data, through the lens of use cases presented in the scientific literature and media. We envision a roadmap of ML strategies that need to be developed or adapted to quantify and overcome these challenges—including transfer learning, active learning, and reinforcement learning techniques. Further, we discuss the potential role of ML in providing guidance to policy makers on issues related to spatial fairness.

1 Motivation

As machine learning (ML) based decision making systems become integrated in our daily lives, demographic fairness and equity concerns continue to surface across diverse applications [29]. Many of these applications are spatio-temporal in nature—such as deployment of public or private services (broadband, 5G, transportation, healthcare, education), or user request prioritization (same-day delivery, rideshare). In spite of the obvious importance, discourse on spatio-temporal applications is largely absent in the extensive literature on algorithmic fairness. Given that the geographical aspect of racial equity and social justice is by now common knowledge—through use cases such as location-based gaming [12], response to COVID-19 [13], and distribution of bikeshare stations [2]—it is important to develop general methodology for rigorous measurement and mitigation of fairness issues in spatio-temporal data.

Due to the complex dependency patterns that spatio-temporal data exhibits between observations (Section 2), algorithmic fairness techniques designed for non-spatial datasets do not directly carry over into the spatial domain. In this paper, we address this challenge, by outlining the research opportunities and charting the path for future work in developing bias detection and mitigation methods tailored for spatio-temporal data (Section 3). While there is some recent work that propose fairness-aware methods in the context of urban mobility [44, 45], we lay out the broader vision of how researchers can approach methodological work in the near future that generalizes to a range of use cases and addresses the complex interplay between sources of spatio-temporal demographic bias. Further, we discuss how the results of technical approaches can inform effective policy changes in this context (Section 4).

Submitted to 2022 Trustworthy and Socially Responsible Machine Learning (TSRML 2022). Do not distribute.

2 Sources of Bias

Based on a literature review of spatio-temporal use cases, we have identified five main sources of demographic bias in this domain.

Geographical differences: Geographical differences in residential and/or work areas may cause bias. A large number of developing countries today are located in tropical areas with extreme climatic conditions. Poverty can drastically aggravate the impact on communities living in such areas [11], with land-locked tropical countries typically being among the poorest. Sustainable efforts to countering climate changes need also to ensure the simultaneous upliftment of poverty-stricken areas [39]. Another example of such disproportionate impacts is floods in the United States. The population distribution in floodplains is marked by a disproportionately large number of people from minority communities [33], causing a racial/ethnic bias in determining who gets affected. For example, [18] states that "Of the seven ZIP codes that suffered the costliest flood damage from Katrina, four of them had populations that were at least 75% Black, government records show". Moreover, flooding outside regular floodzones also disproportionately impacts racial and ethnic minorities according to a recent study [40].

Mental maps Mental or prejudice maps form another common source of bias. A combination of familiarity bias with spatial homophily in residential clusters [6], along with a historical and continued disproportionate wealth distribution among racial and ethnic minorities [5] means that such communities can be disproportionately affected in a sharing economy. For example, in TaskRabbit, UberX, and most other sharing economy services, low-SES (socio-economic status) individuals have a harder time satisfying microentrepreneur enrollment requirements (e.g., microentrepreneurs must have a bank account in many cases), which likely reduces microentrepreneur participation in low-SES neighborhoods and thereby diminishes the effectiveness of the overall platforms in these neighborhoods. Disproportionate population distribution in low-SES neighbourhoods in turn implies a disproportionate impact on minority communities. Some of these mental maps may be further aggravated by confirmation biases through predictive policing [25] because of differences in crime reporting across population segments, equivalent treatment of unreported crimes with reported ones, and the presence of bias in the initial training data. This further leads to feedback loops [17].

Spatial clustering: Bias can also be induced because of existing residential clustering patterns, where people of different demographic profiles and socioeconomic segments tend to cluster into different spatial regions over time. This has led to uneven access to infrastructure and services, e.g., broadband internet [41], and public transportation [4], and Amazon same-day delivery [21]. Some of these infrastructure and services have spatial constraints on their span and coverage that further worsen the problem of uneven access, e.g., 5G cellular network stations [1] need to be placed close to each other to be within range for communication.

Spatio-temporal data glitches Another form of bias occurs from non-random missingness in data. As an example, consider the following. Poverty and wealth information is a factor in deciding government policies. However, if mobile phones are used as the crowdsourcing mechanism [36] to collect that data, it may cause bias in the missingness of the data [13], which may be further aggravated by inferences drawn using the data. Even though almost everyone has a phone, the distribution of the types of phones is not uniform across various population subgroups. For example, it is well-known that iPhone users are more likely to belong to higher income groups than android users. If data collected predominantly through iPhones are used to inform government policies, that would mean a differential impact on lower-income groups, and due to correlation between ethnicity and income, minority groups. Other causes of biased missingness could be lack of access to resources for reporting [7] or context-unaware reuse of external databases. One such example is given by the Pokémon Go example [12]. Data collected as portal locations generated by users of the augmented reality mobile game Ingress were used to determine Gyms and Pokestops for the 2016 game Pokemon Go, and it was later shown that the stops are disproportionately concentrated in White-majority neighborhoods primarily due to the demographic make-up of Ingress players [9].

Spatio-temporal feedback loop: We may make decisions today based on biased outcomes resulting from historical practice or demand, forming feedback loops and perpetuating the bias. For example, electric car charging station [16] roll-out prioritizes regions with compatible technologies, products, demand, and willingness to pay. This leads some spatial regions to have less support for electric cars, which reduces demand further for electric cars and subsequently demand for charging

stations for these regions. This feedback loop continues to widen the divide across different regions for electric car ownership and its associated environmental impacts. Similarly, "black citizens are about half as likely to live in neighborhoods with access to Amazon same-day delivery as white residents." [21] Pricing algorithms for ridehailing services such as Uber charge more for neighborhoods with larger non-white populations and higher poverty levels [34]. Historical practices and policies may have also caused such feedback loops. For example, "the use of the [home] sales comparison approach has allowed historical racialized appraisals to influence contemporary values." [20] Similarly, "[p]eople who live in neighborhoods that were once subjected to a discriminatory lending practice called redlining are today more likely to experience shorter life spans" [19], and poorer cardiovascular health [30]. More recently, discussions around environmental racism in the U.S. and its connection to historical segregation and redlining practices highlight the impact of historical practices and policies. [23] states that "racial/ethnic air pollution exposure disparities persist in part because the underlying sociological, economic, and policy drivers typically evolve on generational time scales."

The five identified sources of bias above often share sources of origin and more than one may impact any particular use case. This can create a confounding effect. Understanding which sources of bias are at play could be useful when identifying mitigation strategies. As pointed out earlier, while geography can play a major role in terms of which areas are affected, social and economic policies and situations can often exacerbate the impact among the poor. Geographical isolation, social exclusion and limited access to education often results in high-levels of poverty among ethnic minorities across the world. Additionally, even when mitigation strategies are implemented, residues of inappropriate racial historical practices such as expropriation of land from Indigenous people, racially exclusive housing covenants, and redlining (in the case of the United States) or racial whitening [15] and restriction of voting to literates (in the case of Brazil) continue to affect minority groups differentially. For example, redlining practices in the United States meant low investments in minority-rich areas, leading to the lack of infrastructure, proper living conditions or access to education. All of these issues have contributed significantly to the five sources of bias identified above.

3 Research Opportunities

There are several challenges to bias detection and mitigation in spatial and spatio-temporal settings. First, spatial and spatio-temporal data often exhibit autocorrelation, i.e., neighboring areas and times tend to be similar due to reasons such as residential clustering. As a result, observations may not be independent-and-identically distributed (i.i.d.), as commonly assumed in existing bias detection and mitigation techniques. Second, certain services have additional requirements or constraints in where they should be placed; for example bike sharing stations need to be placed within biking distance to each other. These constraints need to be accounted for or accommodated by any bias detection and mitigation strategies. Third, the confounding effect discussed in the previous section creates challenges for identifying and accounting for the source(s) of bias in spatio-temporal settings. To tackle these challenges, we lay out a roadmap of research opportunities in the areas of spatio-temporal bias measurement and mitigation. We discuss three strategies for bias mitigation—gathering additional data or using existing data effectively, algorithmic 'debiasing' methods, and spatial experimentation—that are relevant to the different sources of bias introduced in Section 2, (see Table 1).

3.1 Measurement

Currently, several bias metrics have been proposed in the ML fairness literature [31]; however, these metrics assume independence among data samples. This is often an invalid assumption in spatio-temporal settings and can impact the statistical distribution of a bias metric. For example, [27] demonstrated that ignoring spatial clustering when computing common bias metrics can result in higher false-positive rates than would be expected with independent data samples, and proposed adjusting bias metrics using a spatial filtering approach. Spatial Durbin models have also been used to account for spatial clustering in location-based gaming [12] and the sharing economy [42], where the bias detection problem could be represented by a regression framework. In related work on fair clustering (see [8] and references therein), distance-based bias metrics are commonly used and have been applied in fairness studies on resource center locations. These works focus on detecting

Table 1: Promising data-driven mitigation strategies for each source of spatio-temporal bias.

Source of bias	Mitigation stratrgies
Geographical differences	Data gathering
Mental maps	Spatial experimentation
Spatial clustering	Algorithmic methods
Spatio-temporal data glitches	Data gathering
Spatio-temporal feedback loop	Spatial experimentation

a single source of bias, whereas in reality there is often a combination of biases at play. Thus, we identify an interesting research opportunity and two specific research directions.

Research Opportunity 1. New algorithmic methods to quantify and detect bias while accounting for the interplay between different sources of spatio-temporal bias.

Research Direction 1a. New algorithmic methods to quantify and detect bias while accounting for the impact of underlying human mobility patterns.

For example, resource allocation problems may need to consider human mobility patterns that impact the availability of the resource itself—e.g., Uber drivers—and/or the distribution of individuals with access to the resource—e.g., cell-phone users who move in-and-out of areas with 5G coverage. Bias metrics need to account for patterns in such trajectories, as well as any confounding from mental maps, spatial clustering, or feedback loops that reinforce biased behavior.

Research Direction 1b. New algorithmic methods to detect censored feedback tied to varying socio-economic status of geographical areas.

In 2012, the city of Boston launched a smart-phone app to automatically identify potholes by detecting a change in the phone's vertical acceleration and mapping the location using the phone's GPS. By design, pothole reporting was limited to smartphone users, which were predominantly high-income users at that time. Combined with bias in driving behaviors, this resulted in an under-reporting of potholes in lower-income neighborhoods [14]. This example highlights the need for metrics and techniques to automatically detect data glitches due to censored feedback and design choices connected to socio-economic status. One option is to combine data with other cross-referenced sources with connected information. For example, the crowd-sourced data for identifying potholes in Boston could have been combined with census or other location data to reveal the lack of uniform coverage across different areas.

3.2 Data Gathering

As mentioned in Section 2, bias may arise in the data gathering stage through missing data or skewness of data distributions. These data glitches may also be confounded with other sources of bias such as spatial clustering or a spatio-temporal feedback loop. We propose the following research opportunities and two distinct potential research directions within this space.

Research Opportunity 2. New, generalized, or adapted algorithmic methods to mitigate bias caused by data sparsity and quality issues in some spatial regions.

Research Direction 2a. Algorithmic approaches to collect more data for under- and misrepresented spatial regions.

To collect more data, active learning strategies [3] can be used to identify from which instances, individuals, regions, or segments to collect data next in the presence of data paucity and resource limitations. Given a fixed budget, we may want to focus on obtaining data from where data is most sparse, uncertain, dissimilar, or biased, and we may need to design additional incentives or targeting to encourage data collection. ML models can also be used in addition to help infer more data or data labels, e.g., semi-supervised learning, cross-referencing with other sources, programmatic label consolidation [37], or data augmentation [24].

Research Direction 2b. Algorithmic approaches to increase efficiency of the existing data in the underrepresented spatial regions.

To increase the efficiency of existing data, transfer learning [47] can be used; models can be trained on data from regions or sources with enough well-represented data, and transferred or adapted to regions or sources without. Semi-supervised or weakly supervised models can make use of unlabeled data to learn the underlying data distribution and space.

3.3 Algorithmic Mitigation

The sizeable recent literature on algorithmic fairness—for non-spatial but related data problems (e.g. social networks)—presents opportunities to design demographic bias mitigation methods for spatial or spatio-temporal problems. To this end, we identify a broad research opportunity with directions for future investigation.

Research Opportunity 3. New, generalized, or adapted algorithmic methods to mitigate bias through modifications in different phases of the ML pipeline.

Research Direction 3a. Algorithmic approaches to obtain equitable outputs from ML models trained on biased spatial data.

For supervised predictive models, spatio-temporal dependency structure among samples can augment the input dataset (pre-processing), learning objective (in-processing), or model outputs (post-processing). As a pre-processing strategy, input dataset(s) can be replaced with a lower dimensional representation that accounts for fairness constraints *and* underlying spatial autocorrelation [45]. Additonal pre-processing strategies may be explored in the spatio-temporal context. In-processing strategies, such as adversarial debiasing [46] and regularization methods [28], that try to minimize the association between sensitive attributes and predicted outcomes can be generalized to include underlying spatial factors—for example, as a separate regularization term. Post-processing mitigation techniques that aim to minimally change model outputs in a 'fair' way while preserving performance may be adapted to take into account use case-specific spatial constraints.

Research Direction 3b. Fair algorithmic methods tailored for geospatial network analysis.

Many domains of geospatial problems—such as transportation and mobility networks or cell phone networks—can be seen as spatial networks, i.e. networks that have nodes and edges embedded in physical space. Relevant directions of future work in our context include adapting fair versions of unsupervised network analysis methods, such as clustering and community detection [10, 43]. The networked nature of discrimination and social (in)equity themselves are well-studied in a non-ML context (e.g. see references in [43]). Such guidance should inform stages in ML workflows such as model choices and data collection strategies when analyzing spatio-temporal data. This is especially important in the face of complexities discussed earlier, such as data glitches and confounding effects.

3.4 Spatial Experimentation

As discussed in Section 2, different forms of bias may stem from common sources. In some cases, A/B testing could be used to guide design choices to mitigate underlying biases. For example, awareness campaigns could be evaluated to help address biases [42] in sharing economies by removing false perceptions propagated through mental maps. However, in general, context-unaware usage of data as in [12], or models unaware of spatial associations further perpetuate feedback loops aggravating the biases [16]. In light of this, we identify the following research opportunity and subsequent research directions.

Research Opportunity 4. New, generalized, or adapted spatial experimentation methods for bias mitigation.

Research Direction 4a. Exploration-exploitation approaches to break feedback loops that further bias.

Reinforcement learning techniques could help balance the exploration/exploitation trade-off. Existing fairness work on the Multi-armed bandit problem [35] and its connections to corporate social responsibility [38] could be adapted to account for the interplay between demographic sub-populations and geographical areas. This would mitigate feedback loops to avoid early adopter biases when expanding existing infrastructure/deploying new products.

Research Direction 4b. Context-aware usage of models to overcome existing biases.

Existing literature on context-aware ML models [32] use temporal contexts to determine the best ML model. These works could be extended to consider spatio-temporal contexts to mitigate bias. As an example, assessing the ability for wireless versus wireline broadband technologies to satisfy certain performance criteria is an important task in mitigating broadband deserts [41]. To forecast broadband needs in different neighborhoods, unsupervised learning methods could be used to learn contexts that capture variations in geographical characteristics and socio-economic traits across neighborhoods over time. Then a context-aware ML approach could experiment and determine the best demand forecasting model for each context. The resulting model could be used to determine where different broadband technologies would meet broadband needs.

4 Discussions

This paper proposes a roadmap for algorithmic methods to detect and mitigate bias in spatio-temporal data settings. We identify five sources of bias that often co-occur and create challenges in spatio-temporal settings, and highlight four main research opportunities to address these challenges.

Translating the research opportunities on algorithmic fairness in space-time identified earlier in this paper to adequately address topics such as climate change, broadband, 5G, transportation, healthcare, and education requires effective policy changes. Given the enormity of the task at hand, a key question that policy makers around the world face is how to prioritize these opportunities.

The importance of data-driven policy making has been widely recognized. As a direct example of how ML practitioners and data scientists can assist policy makers, the RAND corporation recently released an online tool to "...help policymakers and community residents understand the links between historical discriminatory practices and current environmental inequities and identify hot spots of environmental burdens that can be targeted for environmental improvement efforts" [26]. This tool provides a map-based visualization of historically redlined areas compared to different environmental metrics, and box plots summarizing the distributional differences by Home Owners' Loan Corporation (HOLC) zone. To ensure fair policy, the data used for this purpose needs to be comprehensive, unbiased and widely shared. However, as we have seen, available data gathering mechanisms and hence the collected data are often limited. Research directions 2a and 2b will be critical to help policy makers improve the quality and quantity of data available for data-driven policy making.

To this end, the algorithmic techniques identified in the paper to quantify and overcome bias in spatio-temporal data provide a solid technical foundation. Policymakers would be well advised to make effective use of these algorithmic techniques, in conjunction with seeking input from the various constituencies that have been historically discriminated against, to create policy in a prioritized manner to maximize their impact. In particular, usage of data or ML models for policies that are context-unaware or unmindful of the spatial associations could lead to undesirable feedback loops. Socioeconomic policies targeted at neighbourhood improvement have the potential to effectively mitigate these biases going forward [22].

References

[1] Breland A. Experts worry 5g could widen digital divide in cities. *The Hill*, 2018.

[2] Kabra A. et al. Bike-share systems: accessibility and availability. In *Chicago Booth Research Paper*, 2018.

[3] C. Aggarwal et al. Active learning: a survey. In *Data classification: algorithms and applications*. Chapman and Hall/CRC, 2014.

[4] McKenzie B. Transit access and population change. *U.S Census Bureau SEHSD Working Paper*, 2015.

[5] N. Bhutta, A. Chang, et al. Disparities in wealth by race and ethnicity in the 2019 survey of consumer finances. *Economic Research*, 2020.

[6] B. Bishop and R. G. Cushing. *The big sort: Why the clustering of like-minded America is tearing us apart.* 2008.

[7] J. Blumenstock, G. Cadamuro, and R. On. Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076, 2015.

[8] B. Brubach, D.Chakrabarty, et al. Fairness in clustering: A tutorial at aaai 2022.

[9] K. Capps. Why it sucks to play 'pokemon go' if you don't live in a white neighborhood. *Bloomberg*, 2016.

[10] F. Chierichetti, R. Kumar, et al. Fair clustering through fairlets. In *NIPS '17*, volume 30, 2017.

[11] U.S. Global Leadership Coalition. Climate change and the developing world: A disproportionate impact, 2021.

[12] A. Colley, J. Thebault-Spieker, et al. The geography of pokémon go: beneficial and problematic effects on places and movement. In *CHI '17*, 2017.

[13] A. Coston, N. Guha, et al. Leveraging administrative data for bias audits: Assessing disparate coverage with mobility data for covid-19 policy. In *FAccT '21*, 2021.

[14] K. Crawford. The hidden biases in big data. *Harvard Business review*, 2013.

[15] dos Santos SA. Historical roots of the "whitening" of brazil. *Latin American Perspectives.*, 29(1):61–82, 2002.

[16] W. Englund. Without access to charging stations, black and hispanic communities may be left behind in the era of electric vehicles. *Washington Post*, 2021.

[17] D. Ensign et al. Runaway feedback loops in predictive policing. In *FAT* '18*, 2018.

[18] Thomas Frank. Flooding disproportionately harms black neighborhoods. *Scientific American*, 2020.

[19] M. Godoy. In u.s. cities, the health effects of past housing discrimination are plain to see. *NPR*, 2020.

[20] J. Howell and E. Korver-Glenn. The increasing effect of neighborhood racial composition on housing values, 1980–2015. *Social Problems*, 2021.

[21] D. Ingold and S. Soper. Amazon doesn't consider the race of its customers. should it? *Bloomberg*, 2016.

[22] Paul Jargowsky. Are minority neighborhoods a disaster? *The Century Foundation*, 2016.

[23] H. M. Lane, R. Morello-Frosch, J. D. Marshall, and J. S. Apte. Historical redlining is associated with present-day air pollution disparities in u.s. cities. *Environmental Science & Technology Letters*, 9(4):345–350, 2022.

[24] Y. Li, Y. Luo, et al. Context-aware attention-based data augmentation for POI recommendation. In *ICDEW '19*, 2019.

[25] K. Lum and W. Isaac. To predict and serve? *Significance*, 13:14–19, 2016.

[26] J. Madrigano et al. Environmental racism: A tool for exploring the enduring legacy of redlining on urban environments. 2022.

[27] S. Majumdar, C. Flynn, and R. Mitra. Detecting bias in the presence of spatial autocorrelation. In *AFCR '21*, volume 171, pages 6–18, 2021.

[28] J. Mary, C. Calauzènes, and N. El Karoui. Fairness-aware learning for continuous attributes and treatments. In *ICML '19*, pages 4382–4391, 2019.

[29] N. Mehrabi, F. Morstatter, et al. A survey on bias and fairness in machine learning. 54(6), 2022.

[30] M. Mujahid et al. Historical redlining and cardiovascular health: the multi-ethnic study of atherosclerosis. *PNAS*, 2021.

[31] A. Narayanan. Translation tutorial: 21 fairness definitions and their politics. In *FAT* '18*, 2018.

[32] N. Nascimento, P.Alencar, et al. A context-aware machine learning-based approach. In *CASCON '18*, 2018.

[33] O.Wing, W.Lehman, et al. Inequitable patterns of us flood risk in the anthropocene. *Nature Climate Change*, 12:156–162, 2022.

[34] A. Pandey and A. Caliskan. Disparate impact of artificial intelligence bias in ridehailing economy's price discrimination algorithms. *AAAI/ACM conference on Artificial Intelligence, Ethics, and Society*, 2021.

[35] V. Patil, G. Ghalme, et al. Achieving fairness in the stochastic multi-armed bandit problem. In *AAAI '20*, 2020.

[36] G. Quattrone, L. Capra, and P. De Meo. Thereś no such thing as the perfect map: Quantifying bias in spatial crowd-sourcing datasets. In *CSCW '15*, 2015.

[37] A. Ratner et al. Data programming: creating large training sets, quickly. In *NIPS '16*, 2016.

[38] T. Ron et al. Corporate social responsibility via multi-armed bandits. In *FAccT '21*, 2021.

[39] A. Shapiro et al. Congo basin peatlands have trapped years' worth of carbon. how can they be protected? *Climate*, 2022.

[40] Kevin T. Smiley. Flooding outside flood zones more common in minority neighborhoods. *Finance & Commerce*, 2020.

[41] J. Stankey. A game plan to – finally – connect every american to broadband, 2020.

[42] J. Thebault-Spieker et al. Toward a geographic understanding of the sharing economy: Systemic biases in uberx and taskrabbit. *ACM Trans. Comput.-Hum. Interact.*, 2017.

[43] S. Venkatasubramanian, C. Scheidegger, et al. Fairness in networks: Social capital, information access, and interventions. In *KDD '21*, 2021.

[44] A. Yan and B. Howe. Fairness-Aware Demand Prediction for New Mobility. In *AAAI '20*, 2020.

[45] A. Yan and B. Howe. Equitensors: Learning fair integrations of heterogeneous urban data. In *SIGMOD/PODS '21*, 2021.

[46] B. Zhang, B. Lemoine, and M. Mitchell. Mitigating unwanted biases with adversarial learning. In *AIES '18*, 2018.

[47] F. Zhang et al. A comprehensive survey on transfer learning. In *Proceedings of the IEEE*, 2021.