UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Bridging community knowledge and big data towards environmental justice

Permalink

https://escholarship.org/uc/item/4wr3z6pd

Author

Karasaki, Seigi

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Bridging community knowledge and big data towards environmental justice

by

Seigi Karasaki

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Energy and Resources Group

and the Designated Emphasis

in

Development Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Isha Ray, Co-chair
Professor Rachel Morello-Frosch, Co-chair
Professor Charisma Acey
Professor Duncan Callaway

Spring 2024

Bridging community knowledge and big data towards environmental justice

Copyright 2024
by
Seigi Karasaki

Abstract

Bridging community knowledge and big data towards environmental justice

by

Seigi Karasaki

Doctor of Philosophy in Energy and Resources Group

and the Designated Emphasis in

Development Engineering

University of California, Berkeley

Professor Isha Ray, Co-chair

Professor Rachel Morello-Frosch, Co-chair

Environmental burdens have repeatedly been shown to concentrate in low-income and/or communities of color, and disagreement persists on how these inequalities should be understood and addressed. Navigating these disagreements has been a long-standing struggle for environmental justice (EJ) communities and researchers as they interact with polluters and the state. My dissertation is about environmental justice, three of the "main players" (the state, communities, and researchers), and how two emerging research methods ("big data" and community-engaged research) are shaping how these players interact and leverage science to inform and advocate for policy change. My motivation is to develop new research tools and methods to characterize uneven burdens of environmental health hazards more responsibly. I bridge participatory research methods and machine learning to explore the benefits of and challenges to bringing together community knowledge and big data in environmental justice research. I structure my dissertation around two research questions: (1) what might researchers' "due diligence" look like when applying machine learning models to environmental justice questions; and (2) how do modern tools in the environmental justice toolbox - namely, big data analyses, predictive modeling, and interactive mapping – assist or obstruct the success of hyper-local environmental justice priorities and agendas?

In recent years, the popularity of "big data" and machine learning algorithms has spread from disciplines such as computer science and medicine to research on environmental monitoring and justice. These methods – and in particular, their predictive power – offer resource-efficient approaches to environmental justice-motivated questions, especially given the re-

source constraints of grassroots activists and their government counterparts. They come with caveats, however, as researchers in other fields have demonstrated how algorithms can further entrench structural inequalities. Learning about such findings pushed me to think critically about the implications of applying similar black box predictive models to environmental justice contexts. To address question (1) above, I present a case of machine learning for environmental justice: predicting water quality violations in community water systems across California. I then reflectively de-construct this model and its findings to evaluate potential implications if this algorithmic approach were to be implemented. The result is a well-performing (highly accurate) model for water quality violations in California, paired with a set of recommendations for future environmental justice research utilizing machine learning. In short, my recommendations are to transparently report and discuss the (sociodemographic) characteristics of communities that the model accurately and inaccurately predicts for, and to assess the effects of different input and output variable choices for potentially biasing the model. I conclude with a statistical-based approach that researchers might take to reduce the bias in their outcomes.

To address question (2), I bring big-picture algorithmic approaches in conversation with community engaged research through a practice brief on participatory science. The case for community engagement in research is based on the premise that communities possess knowledge of local environmental health risks that have historically been overlooked by regulatory agencies and academics. Over the past three years, I have been involved in two projects with EJ partners (Toxic Tides and the Drinking Water Tool) that blend community participation with modern tools in the EJ toolbox - large scale datasets, predictive models, and mapping. I am both proud of and grateful for the experience of working together with environmental justice groups on these projects. The practice brief – chapter six of this dissertation – summarizes the ways in which our state-EJ organization-academic collaboration helped strengthen our science and ultimately advance our environmental justice goals.

My dissertation follows the "three paper" format, and is comprised of three manuscripts - one published, and two in peer-review (at the time of filing). Chapter one is the introduction chapter, in which I go over the narrative arc of my dissertation. Chapter two is a critical review of primary data-studies on environmental justice and drinking water. This was my first first-author publication, and has played a big role in how I think about and understand secondary data. In chapter three of my dissertation, I set the stage for my research through a perspective piece on the emerging and growing use of machine learning methods in environmental justice research. (I have also submitted this piece to a journal, but it is only being looked at by the editors, and will not go through a formal peer-review process.) In chapter four, I present the findings of a quantitative analysis using machine learning to predict drinking water quality in California. In chapter five, I offer a reflexive take on my machine learning approach, outlining how different upstream decisions resulted in demographic shifts on who the model was right or wrong about. Finally, in chapter six, I conclude with a practice brief reflecting on the work done by the Water Equity Science Shop

– a research collaborative comprised of an environmental justice community-based organization (the Community Water Center), research groups at UC Berkeley and UCLA, and the Cal EPA's Office of Environmental Health Hazard Assessment – as a practical contribution to theoretical discourse on community-engaged research and environmental justice.

In summary, I explore the opportunities and tensions of two popular approaches to environmental justice research: large scale predictive analysis and community-engagement. Both are powerful and compelling tools for furthering environmental justice research and empowering environmental justice communities, but they come with their caveats, and due diligence is required on the part of the researcher. My hope is that my research helps make these caveats both more legible and manageable to future researchers and their environmental justice collaborators. The longer-term motivation of my work is to inform efforts to connect state resources to on-the-ground community needs.

Contents

Contents **i**

List of Figures **iii**

List of Tables **v**

1 Introduction **1**
1.1 Coming into ERG . . . 1
1.2 Why environmental justice? . . . 2
1.3 Why participatory research? . . . 3
1.4 Why big data and machine learning? . . . 4
1.5 Dissertation organization . . . 5
References . . . 6

2 Environmental justice and drinking water: A critical review of primary data studies **9**
2.1 Introduction . . . 9
2.2 Background . . . 10
2.3 Methods . . . 14
2.4 Results . . . 15
2.5 Discussion . . . 27
2.6 Conclusion . . . 32
References . . . 34

3 Environmental Justice is at a crossroads with machine learning **52**
References . . . 55

4 Using machine learning to predict drinking water quality in California **58**
4.1 Background . . . 58
4.2 A case study for machine learning: drinking water quality in California . . . 59
4.3 Methods . . . 60
4.4 Results . . . 66
References . . . 67

5 Dissecting an algorithmic approach to environmental justice **72**
5.1 Dissecting the model to explore justice implications of modeling decisions . . 72
5.2 Discussion 76
5.3 Future recommendations 77
References 78

6 On strengthening the rigor, relevance, reach, and reflexivity of science **80**
6.1 Background 80
6.2 Findings 81
6.3 Practice Recommendations 84
6.4 Funding 85
References 85

7 Conclusion **88**

A First Appendix: for chapter 2 **90**

B Second Appendix: for chapters 4-5 **110**

List of Figures

2.1 Literature search flow chart (additional details in Figures S1 and S2). 49
2.2 Distribution of drinking water studies (blue) and author affiliations (red). 50
2.3 Diagram of the multidirectional relationships between different types of justice. . 51

3.1 Research articles with terms "environmental justice" and "machine learning" in two research databases. 53

4.1 Process for deriving label and feature values from EDT data for a given water system and contaminant. 62
4.2 A flowchart of our data subsetting process, using arsenic as an example. 64
4.3 Example of how metric choice can affect evaluations of model performance (contaminant: arsenic). 65
4.4 Example of how metric choice can affect evaluations of model performance (contaminant: arsenic). 67

5.1 Distribution of demographic indicators across the four confusion matrix quadrants, predicting for arsenic (MCL) using elastic net. 74

6.1 Water Equity Science Shop. Collaboration outputs are listed next to their primary agents. 81
6.2 Timeline of the Water Equity Science Shop's (WESS) development, the evolution of its Drinking Water Tool (DWT), and the broader landscape of California drinking water policy. 82

A.1 Screening & selection flow chart, phase I (2017) 105
A.2 Screening & selection flow chart, phase II (2020) 106

B.1 Time series plot of EDT sample counts over time. 116
B.2 Process of determining the "most recent sample" for systems with MR violations. 117
B.3 Model performance for all models and contaminants, MCL and half MCL (F2 score). 118
B.4 Confusion matrices for test sets and MR sets using Elastic Net (label = MCL), all contaminants. 119

B.5 Confusion matrices for test sets and MR sets using Elastic Net (label = $\frac{1}{2}$ MCL), arsenic. 120
B.6 Example of a resource prioritization system using classification probabilities. . . 120
B.7 Distribution of demographics indicators for the four other contaminants across classification outcomes using Elastic Net. 121
B.7 Distribution of demographics indicators for the four other contaminants across classification outcomes using Elastic Net. 122

List of Tables

2.1 Initial search terms and Boolean operators for review of EJ and drinking water. 15
2.2 A summary of data collection methods used across reviews. 16
2.3 How different forms of justice reinforce one another or are co-produced. 21

A.1 PRISMA 2020 main checklist 90
A.2 PRISMA 2020 abstract checklist 99
A.3 Studies' community characteristics (general demographic). 101
A.4 Studies' community characteristics (water system-related). 103
A.5 Additional information on data extraction. 107

B.1 List of response and feature variables' data sources. 111
B.3 Census variables and their codes (2017-2021 5-year American Community Survey).112
B.5 Land-use variables 113
B.6 Distribution of MCL and half-MCL occurrences in training and testing sets. . . 114
B.8 Model input features organized into broad categories. 114
B.9 Mann-Whitney U test results for sample counts and model outcomes (false negatives and true positives) for arsenic 115

Acknowledgments

I am so lucky and privileged to have shared these years in the company of so many intellectually and emotionally brilliant people. Coming into ERG, I knew that I was in for a period of growth, but there was no way I could have anticipated the depth and breadth of mental, intellectual, and emotional self-discovery this department would provide. My time here has been transformative. I found myself, here.

I am the person I am today because of my dissertation committee members. Thank you, Isha, for having and pushing my back, and for your tireless compassion and patience in witnessing my growth. Thank you, Rachel, for your constant encouragement to trust and be confident in who I am. Thank you, Duncan, for letting me learn from your laser-sharp intuition and attention to detail. Thank you, Charisma, for being so generous and gracious with your time, even for a mentee who appeared out of nowhere; I hope to emulate a fraction of the care and thoughtfulness you bring into your practice of mentorship.

My dissertation was only possible because of all the collaborators I've had the privilege to work with: my friends and colleagues in the Water Equity Science Shop, the Community Water Center, the Asian Pacific Environmental Network, members of the Toxic Tides Advisory Council, and the California EPA Office of Health Hazard Assessment. I also could not have done it without my generous funding support, from inside and outside the university – notably from the Fulbright Program, InFEWS NSF, the UC Berkeley Human Rights Center, FLAS, and the Lau Graduate Fellowship in Climate Equity. I am truly grateful to have found an academic home in Water Group+. Thank you for welcoming me into your family. I was also lucky to have found a second and third home in s/he lab and CEJ lab in my latter PhD years.

The ERG community made these nine years truly special. To my cohort: we've scattered to the wind, but it was with you I learned what (and how much) it meant to be in this department. To my many ERGie roommates over the years, in Berkeley and in Bellingham: thank you for adding such depth and color into my life. I deeply appreciate the friendships I made wading and basking in everything non-school: climbing, throwing (on the wheel), shooting (pool), reading (PtCH book club!), sewing, eating, and drinking. Thank you Kay, Amanda, Candace, and Megan, for all the love and energy you put into maintaining order in our everyday. For everyone else, in cohorts above and below: I would love more than anything to write a personal note to each of you, but I really, really just need to get this dissertation filed.

Lastly, to the rest who've seen me from way up close: I am immensely grateful for my parents and my brother who have supported me in their own special and irreplaceable ways, each honed over time; Annie for her bottomless support and love, and for being my rock through this all; Aukeem and Lorenzo, for sharing a will to change and unlearn; Pete, for all the nights we've spent on ledges and Hardin Flat Road; JT and Joe, for our friendship tricycle we've ridden in the city and on singletrack; my friends outside of ERG who fill my

cup and inspire me to be better; and to my extended family, who have taken me in as their own. I love you all so deeply. Thank you.

Chapter 1

Introduction

This dissertation is on how two environmental justice research methods – data science and community-engaged research – bring the state, affected communities, and researchers together in the scientific enterprise. In this Introduction, I present the narrative arc of my PhD dissertation. I begin with a brief note on my background prior to entering the Energy and Resources Group (herein ERG). I follow this with my motivation for studying environmental justice (EJ), and why I became interested in (and have chosen to center in my own projects) participatory methods and machine learning as emerging aspects of environmental justice research. I conclude with an organizational overview of my chapters and their contributions.

1.1 Coming into ERG

Prior to coming to Berkeley, I enrolled in a master's program at the University of Tokyo. The department's name roughly translated into "International Studies" (国際協力学専攻; kokusai kyōryoku gaku senkō) within the "Graduate School of Frontier Sciences" (新領域創成科学研究科; shin-ryōiki sōseikagaku kenkyūka). Following a year of living in the Hui Autonomous Region (on the northern tip of the Gobi desert in China), I was interested in the relationship between water access, water management institutions, and state interventions. Specifically, I was skeptical about some research findings claiming a "steady spread" of water management reform in northern China, and wanted to take a closer look at the extent to which these institutional reforms were actually taking place. I reached out to a researcher on the paper above, who assured me I would have access to their dataset.

Halfway through my program, the contact abruptly retracted their offer to share data, and I was left with the decision to postpone my graduation by a year or to pivot to a research topic with less fieldwork. It was in this scramble that I happened upon Professor Isha Ray's work with then-student Carolina Balazs on drinking water access and EJ in the San Joaquin Valley. As someone born and raised in California, I was stunned and embarrassed; how had

I been so unaware of injustices happening just a couple hours from where I was born? As I learned more about what was happening in the Valley, I realized a strong, visceral desire to have both my research and professional identity be tied to the issues in California ("at home") than anywhere else. Nearly a decade later, I'm both happy and proud to say that I have been able to identify, define, and develop the intellectual identity that I was looking for when I arrived at ERG. Drinking water was my entryway into environmental justice work and is what I am most familiar with, but I am deeply aware that it is just one of many injustices that fenceline communities face every day. This dissertation reflects my journey to nurture an ability to listen to community voices and to translate their lived experiences into models, maps, and research designs, all to push the needle towards a more environmentally just future.

1.2 Why environmental justice?

The motivating force behind my research is the existence of uneven burdens of and exposures to environmental health hazards. I chose to situate my work within the academic discipline of EJ because I believe its explicit orientation towards social justice is conducive for transparent and accountable community-based research. At its core, EJ is grounded in the principle that all people deserve access to a healthy and safe environment where they "live, work, and play" (Alston 1991). The environmental justice movement emerged and gained momentum in the 1980s in protests of concentrated siting of hazardous waste facilities in under-served communities (Bullard 1990). The academic arm of the movement has since been occupied with developing an evaluative framework to identify the drivers of environmental injustice to inform policies meant to address them (Cutter 1995; Pellow 2000; Schlosberg 2004). Originally at the margins of environmentalism, EJ has gradually become more mainstream, reflected in its formal integration into multiple levels of government and policy (Taylor 2000).

Although environmental burdens have repeatedly been shown to concentrate in low-income and/or communities of color, there is disagreement on how these inequalities should be addressed, or whether unequal patterns are symptomatic of unjust or racist drivers and policies (Holifield 2001). Navigating and resolving these disagreements has been a long-standing struggle for EJ communities, coalitions, and researchers. As someone interested in studying the structural mechanisms of injustice, I wanted to balance my work between the questions of how and why injustices develop and persist, and what should be done to address them. This led me to focus primarily on two methods of data collection and analysis that often operate on opposite ends of the spatial spectrum in EJ research: participatory research and predictive machine learning analysis.

1.3 Why participatory research?

In what I see as an effort to address disagreements between communities, coalitions, and researchers, participatory research has risen in popularity as a research framework in EJ and public health work for its emphasis on both elevating and integrating the affected community's voice into the research process. The EJ praxis of approaching and maintaining community-researcher relationships draws from the literature on participatory research – an umbrella term for a collection of approaches that actively incorporate stakeholders into the research design to "democratize the knowledge creation process" (Cargo and Mercer 2008). All forms of participatory research converge on how they challenge the norms of research design, from reimagining participation to turn research subjects into partners, to de-elevating researcher positionality as the "educated expert," to redistributing power through more equitable and transparent research collaborations. This emphasis on researcher positionality resonated with my own pursuit to develop a cohesive, personal research philosophy.

In EJ research, the case for community involvement is based on the premise that EJ communities possess knowledge of local environmental health risks that have historically been overlooked by regulatory agencies, public health practitioners, and academics. Corburn (2003) was one of the earlier voices to argue for incorporating local knowledge into EJ research frameworks to augment professional knowledge. While there are many reasons to fold participatory methods into research frameworks, it is not without its challenges. Examples include the selection of the research question (how community-driven can a project be if the research question came from the outside?), differential timelines and priorities (e.g., finite funding horizons, or the academic push to publishing peer-reviewed papers over on-the-ground action), or when study findings do not resonate with local knowledge or validate on-the-ground experiences (Minkler 2004). For all its strengths, participatory methods do not guarantee "good" or "positive" research outcomes; it requires much thoughtfulness and care to be implemented well.

There were a few areas of theoretical "fuzziness" which I was particularly interested in exploring through my time at ERG. For one: even if it is not difficult to accept that community involvement in the research process is often desirable, the question of how to determine to what extent or with what timing collaboration should take place is not clear-cut. Good participatory research practice calls upon researchers to be mindful of community capacity, the time and processes required to establish trusting, sustainable, and productive relationships between communities and researchers, and the need to properly compensate community members for their time and labor. Much of this dialogue focuses on how a hypothetical research project should involve community partners, but few, if any, participatory research discuss what type of research should or need not involve community collaborators. Community involvement in participatory research is a continuum and not a binary, as it can advance organizing goals under different forms and levels of engagement. The relevance of these questions are thrown into sharp relief when community priorities are held up alongside machine learning, a research method often employed with the specific purpose of extrapolat-

ing into the "unknown" - an "unknown" which can overlap with local experiences that are undocumented or "invisible" to the data (Borderon et al. 2021).

This brings me to my interest in "Big Data" and predictive analysis using machine learning.

1.4 Why big data and machine learning?

Both "Big Data" and "machine learning" (ML) have exploded in popularity and usage in the past few decades, and are firmly rooted in a variety of disciplines ranging from computer science to medicine (Marx 2013; Sagiroglu and Sinanc 2013). More recently, ML's popularity has spread to environmental-related research, with applications including but not limited to environmental monitoring and projections (Hino, Benami, and Brooks 2018; Rolnick et al. 2023). One of the strengths of these new methods is their enhanced ability to interpolate or extrapolate information. It can also be a resource-efficient way of pushing forward EJ-motivated work, especially given the resource constraints of on-the-ground activists. This can be seen as an extension of "traditional" EJ work that relied on statistical methods and spatial analyses to estimate the strength of patterns in burden distribution.

I became particularly interested in two points of inquiry regarding how ML is being used in EJ research. The first point is the application of ML to augment EJ research. As mentioned earlier, an advantage of these tools is in their resource efficiency, particularly in areas that simply do not have access to the necessary time, money, or technologies to monitor environmental indicators of concern (Lobo, Laraway, and Gadgil 2022; Mulhern et al. 2022). This emphasis on efficiency and increased information accessibility in the context of constrained resources raises questions about how these tools might (or whether they can) interface with participatory research designs. For one, it complicates the question of at what point community voices can or should be integrated into the research process. It also represents a departure from traditional applications of secondary data and statistics in EJ analysis. The difference is that these past exercises were focused primarily on quantifying relationships between known and measurable environmental conditions (e.g., toxic facility locations or pollution levels) and community characteristics or adverse health outcomes, both of which can be ground-truthed (Morello-Frosch et al. 2005; Sadd et al. 2014). Much of ML's allure, however, lies in its predictive application for current conditions that are difficult or cost-prohibitive to ground-truth, or future conditions that are impossible to ground-truth *now* by definition. When local knowledge cannot be used to ground-truth research findings, what are the implications for community voice and representation?

The second point of inquiry is on the topic of algorithmic bias and fairness: specifically, the need to more carefully analyze the consequence of applying "black box" predictive models to EJ research. There is extensive research in other fields such as medicine on how structural racism and oppression can be encoded in machine learning algorithms (Benjamin 2019; Robinson, Renson, and Naimi 2020). I was particularly inspired by work that specifically

evaluates the socio-demographic (e.g., racial) biases and consequences of algorithm-guided policy, as Obermeyer et al (2019) showed with their analysis of a health system's predictive calculations of future patient health and cost. Given the mission of EJ and the relative disadvantage of communities affected by its policy prescriptions, it is imperative that algorithms applied in EJ-motivated work be also assessed for their potential socio-demographic impacts.

1.5 Dissertation organization

This dissertation follows the "three paper format" of bringing together three separate manuscripts, with one published and two in the process of peer review (at the time of graduation). Their shared throughlines are drinking water and environmental justice, and each engage with the topics of participatory research and/or machine learning to varying extents. While each paper is quite different – a literature review, an application of machine learning for environmental justice, and a practice brief (akin to a policy brief) co-authored by community and government partners – each reflects a particular moment in my growth and development as a scholar, and together summarize the journey I took in ERG. I am thankful for an indebted to my coauthors for all three papers; I learned much along the way and grew tremendously through our collaboration.

This introduction and narrative arc make up the first chapter of my dissertation. In chapter two, I present a literature review on environmental justice studies that use primary data to investigate issues on drinking water. Presented as a critical review drawing on systematic review methods, we organized our results on the following questions: (1) what sampling and data collection methods are used; (2) how is (un)just access to water defined and measured; (3) what forms of environmental injustice are discussed; (4) how are affected communities resisting or coping; and (5) what, if any, mechanisms of redress are advocated? This review's focus on primary data and environmental justice helped lay the foundation for my understanding of the strengths and weaknesses of secondary data studies (which big data studies are by definition). Looking back: as my first first-author paper, it was a difficult but valuable experience.

In chapter three, I set the stage for my research concerning the growing use of machine learning methods in environmental justice research through a short perspective piece. After a brief description of what I see as both the promise and peril of machine learning for environmental justice, I propose four practice recommendations for EJ scholars using ML methods going forward. I submitted this piece as a complement to a longer and more applied paper (chapters four and five). My motivation to write this perspective was to present a distilled and easy-to-digest paper on why and how care should be taken when ML methods are applied to answer EJ questions.

Chapters four and five are two halves of one paper. In chapter four, I present the findings of a quantitative analysis using machine learning to predict drinking water quality

in California. In chapter five, I offer a reflexive take on this machine learning approach, outlining how different upstream decisions resulted in demographic shifts on who the model was right or wrong about. Chapter four presents an example of how ML can be used to extend EJ; chapter five outlines the ways in which such modeling exercises can have unintended consequences with regards to equity and presents some practical ways in which researchers can check for bias and unfairness in their own projects.

In chapter six, I conclude with a practice brief reflecting on the work done by the Water Equity Science Shop – a research collaborative comprised of an environmental justice community-based organization (the Community Water Center), research groups at UC Berkeley and UCLA, and the Cal EPA's Office of Environmental Health Hazard Assessment – as a practical contribution to theoretical discourse on community-engaged research and environmental justice, grounded in a long and collaborative partnership. In this practice brief, we reflect on how our collaboration strengthened the "4Rs" (rigor, relevance, reach, and reflexivity) of our science.

In summary, I was able to explore the opportunities and tensions of two popular approaches to environmental justice research through my time at ERG: large scale predictive analysis and community-engagement. Both are powerful and compelling tools for furthering environmental justice research and empowering environmental justice communities, but they come with their caveats, and due diligence is required on the part of the researcher. My hope is that my research helps make these caveats both more legible and manageable to future researchers and their environmental justice collaborators.

References

Alston, Dana (1991). "Transforming a Movement". In: *Race, Poverty & the Environment* 2.3, pp. 1–29. ISSN: 1532-2874. URL: https://www.jstor.org/stable/41555193 (visited on 03/22/2022).

Benjamin, Ruha (Oct. 25, 2019). "Assessing risk, automating racism". In: *Science* 366.6464, pp. 421–422. DOI: 10.1126/science.aaz3873. URL: https://www.science.org/doi/full/10.1126/science.aaz3873 (visited on 03/17/2022).

Borderon, Marion et al. (Dec. 6, 2021). "The risks of invisibilization of populations and places in environment-migration research". In: *Humanities and Social Sciences Communications* 8.1, pp. 1–11. ISSN: 2662-9992. DOI: 10.1057/s41599-021-00999-0. URL: https://www.nature.com/articles/s41599-021-00999-0 (visited on 02/19/2022).

Bullard, Robert D. (Oct. 16, 1990). *Dumping In Dixie: Race, Class, And Environmental Quality*. Avalon Publishing. 204 pp. ISBN: 978-0-8133-7954-8.

Cargo, Margaret and Shawna L. Mercer (2008). "The Value and Challenges of Participatory Research: Strengthening Its Practice". In: *Annual Review of Public Health* 29.1, pp. 325–350. DOI: 10.1146/annurev.publhealth.29.091307.083824. URL: https://doi.org/10.1146/annurev.publhealth.29.091307.083824 (visited on 11/05/2021).

Corburn, Jason (June 1, 2003). "Bringing Local Knowledge into Environmental Decision Making: Improving Urban Planning for Communities at Risk". In: *Journal of Planning Education and Research* 22.4, pp. 420–433. ISSN: 0739-456X. DOI: 10.1177/0739456X03022004008. URL: https://doi.org/10.1177/0739456X03022004008 (visited on 10/21/2021).

Cutter, Susan L. (Mar. 1, 1995). "Race, class and environmental justice". In: *Progress in Human Geography* 19.1, pp. 111–122. ISSN: 0309-1325. DOI: 10.1177/030913259501900111. URL: https://doi.org/10.1177/030913259501900111 (visited on 05/06/2022).

Hino, M., E. Benami, and N. Brooks (Oct. 2018). "Machine learning for environmental monitoring". In: *Nature Sustainability* 1.10, pp. 583–588. ISSN: 2398-9629. DOI: 10.1038/s41893-018-0142-9. URL: https://www.nature.com/articles/s41893-018-0142-9 (visited on 01/20/2022).

Holifield, Ryan (Feb. 1, 2001). "Defining Environmental Justice and Environmental Racism". In: *Urban Geography* 22.1, pp. 78–90. ISSN: 0272-3638. DOI: 10.2747/0272-3638.22.1.78. URL: https://doi.org/10.2747/0272-3638.22.1.78 (visited on 04/09/2024).

Lobo, G. P., J. Laraway, and A. J. Gadgil (Jan. 10, 2022). "Identifying schools at high-risk for elevated lead in drinking water using only publicly available data". In: *Science of The Total Environment* 803, p. 150046. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2021.150046. URL: https://www.sciencedirect.com/science/article/pii/S0048969721051214 (visited on 01/14/2022).

Marx, Vivien (June 2013). "The big challenges of big data". In: *Nature* 498.7453, pp. 255–260. ISSN: 1476-4687. DOI: 10.1038/498255a. URL: https://www.nature.com/articles/498255a (visited on 05/05/2022).

Minkler, Meredith (Dec. 1, 2004). "Ethical Challenges for the "Outside" Researcher in Community-Based Participatory Research". In: *Health Education & Behavior* 31.6, pp. 684–697. ISSN: 1090-1981. DOI: 10.1177/1090198104269566. URL: https://doi.org/10.1177/1090198104269566 (visited on 10/21/2021).

Morello-Frosch, Rachel et al. (2005). "Citizens, science and data judo: Leveraging secondary data analysis to build a community-academic collaborative for environmental justice in Southern California." In: *Methods in community-based participatory research for health*. San Francisco, CA: Jossey-Bass, pp. 371–393.

Mulhern, Riley et al. (Mar. 1, 2022). "A new approach to a legacy concern: Evaluating machine-learned Bayesian networks to predict childhood lead exposure risk from community water systems". In: *Environmental Research* 204, p. 112146. ISSN: 0013-9351. DOI: 10.1016/j.envres.2021.112146. URL: https://www.sciencedirect.com/science/article/pii/S001393512101447X (visited on 01/17/2022).

Obermeyer, Ziad et al. (Oct. 25, 2019). "Dissecting racial bias in an algorithm used to manage the health of populations". In: *Science* 366.6464, pp. 447–453. DOI: 10.1126/science.aax2342. URL: https://www.science.org/doi/full/10.1126/science.aax2342 (visited on 03/18/2022).

Pellow, David N. (Jan. 1, 2000). "Environmental Inequality Formation: Toward a Theory of Environmental Injustice". In: *American Behavioral Scientist* 43.4, pp. 581–601. ISSN:

0002-7642. DOI: 10.1177/0002764200043004004. URL: https://doi.org/10.1177/0002764200043004004 (visited on 04/03/2022).

Robinson, Whitney R, Audrey Renson, and Ashley I Naimi (Apr. 1, 2020). "Teaching yourself about structural racism will improve your machine learning". In: *Biostatistics* 21.2, pp. 339–344. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxz040. URL: https://doi.org/10.1093/biostatistics/kxz040 (visited on 05/06/2022).

Rolnick, David et al. (Mar. 31, 2023). "Tackling Climate Change with Machine Learning". In: *ACM Computing Surveys* 55.2, pp. 1–96. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3485128. URL: https://dl.acm.org/doi/10.1145/3485128 (visited on 05/02/2022).

Sadd, James et al. (June 1, 2014). "The Truth, the Whole Truth, and Nothing but the Ground-Truth: Methods to Advance Environmental Justice and Researcher–Community Partnerships". In: *Health Education & Behavior* 41.3. Publisher: SAGE Publications Inc, pp. 281–290. ISSN: 1090-1981. DOI: 10.1177/1090198113511816. URL: https://doi.org/10.1177/1090198113511816 (visited on 05/03/2022).

Sagiroglu, Seref and Duygu Sinanc (May 2013). "Big data: A review". In: *2013 International Conference on Collaboration Technologies and Systems (CTS)*. 2013 International Conference on Collaboration Technologies and Systems (CTS), pp. 42–47. DOI: 10.1109/CTS.2013.6567202.

Schlosberg, David (Sept. 1, 2004). "Reconceiving Environmental Justice: Global Movements And Political Theories". In: *Environmental Politics* 13.3, pp. 517–540. ISSN: 0964-4016. DOI: 10.1080/0964401042000229025. URL: https://doi.org/10.1080/09644010420002290 25 (visited on 09/01/2021).

Taylor, Dorceta E. (Jan. 1, 2000). "The Rise of the Environmental Justice Paradigm: Injustice Framing and the Social Construction of Environmental Discourses". In: *American Behavioral Scientist* 43.4, pp. 508–580. ISSN: 0002-7642. DOI: 10.1177/0002764200043004003. URL: https://doi.org/10.1177/0002764200043004003 (visited on 09/02/2021).

Chapter 2

Environmental justice and drinking water: A critical review of primary data studies[1]

2.1 Introduction

Safe, affordable, and accessible drinking water is integral to health, dignity, and human rights. Globally, about 2 billion people lack access to safe and affordable drinking water (UNICEF and WHO 2019). Recent media attention on lead contamination in Flint and wells running dry in California has elevated public awareness of drinking water injustices in the US (Felton, Gill, and Kendall 2021; Graham 2016). Even today, at least 500,000 US households struggle with incomplete plumbing, and over a thousand community water systems were recently out of compliance with the Safe Drinking Water Act (Mueller and Gasteyer 2021). These violations are even more prevalent in rural, under-resourced areas (Patel et al. 2020). There is a substantial body of academic work using an environmental justice (EJ) frame to understand and evaluate disparities in drinking water access. Several review papers have synthesized theories and practices of EJ writ large (e.g., Agyeman et al. 2016; Mohai, Pellow, and Roberts 2009), but none, to our knowledge, focus specifically on drinking water. There are also many insightful reviews on various metrics and dimensions of drinking water access such as affordability (e.g., Goddard, Ray, and C. Balazs 2022; Pierce et al. 2021), security (e.g., Hoekstra, Buurman, and Ginkel 2018; W. E. Jepson et al. 2017; Octavianti and Staddon 2021), rights (e.g., Sultana and Loftus 2015), sustainability (e.g., Mehta and Movik 2014), and environmental health (e.g., Rosinger and S. L. Young 2020; Syafrudin et al. 2021; Ward et al. 2018), but not through a lens of environmental justice.

[1]This chapter was published in the journal *WIREs Water* in 2014 (DOI: 10.1002/wat2.1653) with co-authors Jessica Goddard, Alasdair Cohen, and Isha Ray. I am grateful for their contributions to this work, and for their permission to use this research as a part of my dissertation.

We evaluated studies of drinking water disparities that were explicitly framed in terms of "Environmental Justice," a discipline that arose at the intersection of advocacy and academia to make visible unjust environmental risks where people "live, work, and play" (Alston 1991). To understand how researchers captured drinking water challenges where people "live, work, and play", we reviewed papers that were at least partly based on primary data collection.

To contextualize this review, we first briefly discuss the broader literature on environmental justice and on EJ and drinking water specifically (Section 2). We then explain our sample selection and methods using transparent inclusion and exclusion criteria (Section 3). We use the following questions to organize our findings (Section 4): (1) what sampling and data collection methods are used; (2) how is (un)just access to water defined and measured; (3) what forms of environmental injustice are discussed; (4) how are affected communities resisting or coping; and (5) what, if any, mechanisms of redress are advocated? In the final segment (Section 5) we contextualize our findings within the broader literature on EJ and drinking water: what scholars mean when they reference "drinking water justice," what injustices look like, and who (if anyone) is to be held accountable. We pay specific attention to narratives on community resistance and coping in struggles to procure and maintain drinking water access, as well as to contemporary debates on the role of the state in alleviating water injustices. We show that the three main types of injustice discussed in the literature, i.e., distributive, procedural, recognitional, do not exist in isolation but often produce or reinforce one another. Ultimately, this review aims to support researchers at the intersection of EJ and drinking water in understanding critical gaps and convergences in the literature, and to highlight the value of primary-data research.

2.2 Background

Environmental Justice

Environmental justice is grounded in the principle that all people deserve access to a healthy and safe environment, and to be protected equally by environmental and public health laws and regulations. From the 1980s on, building on the lessons and strategies of the Civil Rights movement, activists and scholars have used EJ as a framework to understand how environmental hazards are distributed (e.g., Bullard 1990; Christ Commission for Racial Justice 1987; McGurty 2000; S. M. Wilson et al. 2010). EJ goals are now formally integrated into multiple levels of US government and policy (Bullard and Johnson 2000). Definitions of EJ vary across agencies and authors (Holifield 2001), but most interpretations converge on equal protection against the distribution of environmental harms, and equal access to participation in decision-making processes related to environmental hazards (Schlosberg 2004).[2]

[2]EJ can also be centered around the distribution of disproportionate benefits or privilege, but this benefit-oriented framing is less common (see Stewart, Bacon, and Burke 2014; Sunter, Castellanos, and Kammen 2019; Szasz and Meuser 1997).

Environmental injustice occurs when a community is involuntarily subjected to disproportionate environmental burdens, excluded from decision-making processes, or prevented from seeking redress. Many earlier EJ studies investigated the first phenomenon, measuring patterns of environmental burdens along race and class lines (e.g., Bullard and Wright 1993; Kuehn 2000). Research on disparities across race and class have dominated the US literature, but EJ scholars and activists are also concerned with gender, ability, indigeneity, and other axes of difference (Pellow 2017). Such selective allocation of environmental hazards (or benefits) is referred to as distributive injustice. The lack of opportunity for affected communities to participate in decision-making processes related to these hazards is referred to as procedural injustice. Examples of procedural injustices include exclusion from impact assessments of waste facility development in residential areas (Deacon and Baxter 2013) or inadequate representation in the formulation of climate adaptation policy (Holland 2017). A third variant of injustice, termed recognitional, is concerned less with the distribution of goods and services and more with the humiliation or disrespect with which minoritized communities can be treated (Honneth 2004) As we discuss below (Section 4.4), the different forms of injustice often co-occur, or one may lead to or exacerbate another.

Many EJ studies describe a collective victimization by a polluter, or a "bad actor" to be held accountable. The argument is that, given a clear connection between pollution and polluter, this actor (usually a corporation or the state) should compensate the affected communities and/or take corrective actions. Research has highlighted the health risks of living in proximity to known pollution sources (P. Brown 1995; Bullard and Wright 1993), as well as the importance of compensation from those responsible (Čapek 1993). In this way, redress by polluters, or corrective justice, has also been central to study narratives.

Beyond documenting existing injustices, EJ research also points to the structural mechanisms underlying the existence and persistence of hazard distribution. Rather than proving "discriminatory intent" (Pulido 1996), research on structural mechanisms study the "political economy of environmental discrimination" (e.g., segregation, suburbanization and white flight, historical patterns of development) (R. A. Morello-Frosch 2002). The social production of injustices has been attributed to the racialized foundations of US society (Omi and Winant 2014), and to ongoing inequalities among socioeconomic groups and institutions (Kurtz 2009; Pulido 2000; Pulido, Kohl, and Cotton 2016). Acknowledging historical inequalities were also central to the evolution of recognitional justice - i.e., recognition of diversity and group difference (Fraser 2000; Schlosberg 2004; I. M. Young 1989). One of the earliest declarative statements on EJ demands - The Principles of Environmental Justice (1991), from the First National People of Color Environmental Leadership Summit - foreshadowed these insights in its call for "respect ... of our cultures, languages and beliefs about the natural world" and "to secure our political, economic and cultural liberation that has been denied for over 500 years of colonization and oppression."

More recently, Global EJ and Critical EJ have emerged as extensions of EJ theory. Global-level EJ studies emphasize structural injustices that reflect persistent power imbal-

ances between nations (Zeitoun 2013) or between mainstream and minoritized ethnicities (Agyeman 2001). EJ scholars, in recognition of the global scale of environmental problems, have pushed the discipline beyond its original U.S. context (Walker 2009), while calling for new and culturally-contextual interpretations of EJ (Álvarez and Coolsaet 2020; Schlosberg 2004, 2013). Stakeholder understandings of justice are sensitive to time and place and may not conform to a "universal standard of justice" (Holifield 2001; Minkler et al. 2008). Critical EJ has pushed traditional EJ to take the rights of non-human species seriously, and to recognize that the liberal state may not be an effective redress mechanism for environmental injustices (Pellow 2017; Sze 2020). Both variants, however, remain true to the original EJ focus of disparate burdens, and to the blend of scholarship and advocacy that characterizes the EJ literature.

Drinking water and environmental justice

While some early EJ studies included water contamination as an indicator of injustice, it is only over the last two decades that EJ scholars have focused on inequalities in drinking water access as an environmental injustice. This has paralleled an international movement to recognize the human right to clean and safe drinking water: in 2002, the UN Committee on Economic, Social and Cultural Rights asserted "[the human right to water] is a prerequisite for the realization of other human rights" (UN Economic and Social Council 2002); and in 2015, the Sustainable Development Goals included targets for "universal and equitable access to safe and affordable drinking water for all" (UN General Assembly 2015).

Given the global scope of our review, for the sake of consistency, we employ the definition of "drinking water" provided by the Joint Monitoring Programme (JMP) for Water Supply and Sanitation (2020): "Drinking water services refers to the accessibility, availability and quality of the main source used by households for drinking, cooking, personal hygiene and other domestic uses." For instance: the physical accessibility of water can be measured by the distance to the nearest water source, or the time spent for water collection. The availability of water supply can be described by the intermittency of supply, the presence-absence of physical infrastructure, or whether the quantity is sufficient for hygiene and household needs (Lee and Schwab 2005; Martínez-Santos 2017). Quality is often measured against thresholds determined by a federal or international regulatory agency, such as the World Health Organization, or the U.S. Environmental Protection Agency (US Environmental Protection Agency n.d.; World Health Organization 2018). There are also aspects of access that do not fit into the JMP definition, such as affordability, often indicated by a ratio of water expenditures relative to income (García-Valiñas, Martínez-Espiñeira, and González-Gómez 2010; Goddard, Ray, and C. Balazs 2021). We use "drinking water access" as an umbrella term encompassing any or all of these aspects (i.e., accessibility, availability and quality), recognizing that culture and economics may determine what counts as "access" in different contexts. Drinking water injustice can involve one or more of these, and metrics to represent these aspects of water access are continually evolving.

Our review focuses on primary data collection to understand how research on EJ and drinking water has captured the lived experiences of affected communities. Several EJ and drinking water papers represent the relationship between access and social vulnerability using secondary data (e.g., C. L. Balazs, R. Morello-Frosch, et al. 2011, 2012). Using secondary data such as the U.S. Census helps to identify population-level trends and mobilize policy decisions and funding to track and rectify injustices. Another contribution of studies using secondary data is to develop narrower hypotheses for further research. Our review focuses on primary data studies because certain research questions, such as uncovering on-the-ground community experiences, are accessible only through the collection of primary data; these questions are important enough to warrant their own review. Furthermore, secondary data can potentially obscure what is happening at more granular levels: for example, it can "invisibilize" small populations such as the unhoused or undocumented (Borderon et al. 2021). By focusing specifically on studies that collected primary data, we surface on-the-ground as well as perceived injustices closest to communities. Our findings and reflections, therefore, reflect a subset of the broader environmental justice literature on drinking water.

Adjacent fields of inquiry outside of this review

There is a vast and valuable literature concerned with safe and adequate water for the underserved that is outside of this review. These works can be called EJ-adjacent, in that they situate themselves within different frameworks while sharing EJ themes.

The Human Right to Water literature argues for safe water as an inalienable right of all humans, and is less concerned with comparing communities to each other and more concerned with evaluating progress toward universal access (Heller 2022). Its concept of justice is thus absolute rather than relative, and some scholars have questioned the compatibility of rights-based versus justice-based approaches (Harris, Rodina, and Morinville 2015). Political ecology (PE) has a long tradition of research on struggles over water but PE papers may not explicitly invoke the concept of justice. One of PE's core deviations from the EJ framing is in how it "sees" the state: whereas the origins of EJ assumed a liberal state that could rectify injustices, PE comes from a Marxist lineage with a state that is to be confronted more than relied on. Ranganathan and Balazs (2015) have usefully termed PE versus EJ concepts as the "everyday" over the "regulatory" state, though, as they point out, this divide is breaking down in newer work; rights and justice movements can work with, against, and outside the state at the same time (e.g., Angel and Loftus 2019).

A third literature on water security is about people's struggles to access water and the importance of reliability, sustainability, safety, and adequacy (W. E. Jepson et al. 2017; Meehan et al. 2020). Several older themes of water access, such as welfare and long-term sustainability, have recently been covered under the "security" label (Hoekstra, Buurman, and Ginkel 2018). The fourth strand, the largest and the oldest of all of these, is Water, Sanitation and Hygiene (WASH). This globally-focused body of work is concerned with analytical studies on water quality, water systems, and their associated health outcomes,

including attention to the coping costs of not having adequate and safe water access. It is mainly a literature on ill-health or inadequate infrastructure rather than on injustice. EJ studies on drinking water build upon the contributions and critiques of these adjacent fields, but EJ's empirical core is grounded intellectually and politically in its advocacy-oriented origins, and is thus distinct from these other disciplines in theory and in action.

2.3 Methods

This study is structured as a critical review of the drinking water and EJ literature with a focus on studies reliant on primary data. To ensure comprehensive, transparent, and reproducible results, we followed study identification, screening, and selection methods used to conduct systematic reviews. Studies were considered eligible if they focused on a population's access to drinking water and made use of an environmental justice framework. We executed this search in the United States, and placed no restrictions on study setting, language, or publication date. We excluded studies if they did not explicitly focus on drinking water (e.g., those on surface or groundwater) and did not reference EJ. We pre-registered our study protocol on PROSPERO (#CRD42017069837).[3] We amended our original protocol eligibility criteria to include only studies based, fully or partially, on primary data collection (e.g., water samples, surveys, etc.). (See Table A.1 and A.2 in the supplementary materials for a completed PRISMA reporting checklist.)

We searched the following online databases for eligible studies: Web of Science, PubMed / MEDLINE, EMBASE, Cochrane Library, and ScienceDirect. We conducted the search in two phases, in 2017 and 2021 (see Figures A.1 and A.2 in the supplementary materials for disaggregated literature search flow charts). We also used Google Scholar to conduct a hand-search of additional studies using citation snowballing. We piloted search terms ahead of the full search to ensure they were sufficiently broad to identify a subset of papers that we already knew to be eligible. Table 2.1 summarizes the search terms, sets, and Boolean operators we used.

Screening and data extraction protocols

For the first screening phase, three authors (SK, AC, JG) independently screened the titles and abstracts of retrieved studies to flag those potentially meeting our inclusion criteria. We divided titles and abstracts into three random equally-sized and overlapping groups so review decisions were cross-checked by two authors. We removed all duplicate records, as well as those with missing information, at this step. For the second screening phase, the same authors assessed the full texts of potentially eligible studies. As with the first phase, two authors made each review decision, with the third author's opinion sought in

[3]PROSPERO is an online database for pre-registering (systematic) reviews, designed to minimize the possibility of duplicate research projects and to reduce reporting bias.

Table 2.1: Initial search terms and Boolean operators for review of EJ and drinking water.

Drinking Water*		Environmental Justice*
"Water"		"Environmental Justice"
"Drinking water"		"Justice"
"Potable water"		"Social Justice"
"Tap water"		"Environmental equity"
"Piped water"	AND	"Environmental equality"
"Municipal water"		"Environmental racism"
"Utility water"		"Distributive justice"
"Bottled water"		"Climate justice"
"Water system"		
"Water district"		

*Boolean operator "OR" used between search terms in each set (AND between sets)

the case of disagreement. For the final phase of data extraction, the first author used a standardized, pre-piloted spreadsheet to input and organize collected data. Extracted information included: study setting; study population and participant demographics; details of the study design; recruitment and study completion rates; and all available outcomes and times of measurement. The final set of selected papers was published between 2002 and 2020.

Data analyses

We included a plan in our PROSPERO protocol to synthesize study characteristics and outcomes, including: negative health outcomes associated with consumption of contaminated drinking water; types of drinking water access, contamination, and exposure; research methods and theoretical foundations; and the distribution of study locations.

We conducted data cleaning, analysis, and visualization using ggplot2 and rnaturalearth in RStudio using R version 4.0.5 (R Core Team 2021; RStudio Team 2020; South 2017; Wickham 2016).

2.4 Results

Following a summary of our screening process outcomes (Section 4.1) we report broad trends and patterns across the reviewed studies (Section 4.2 - 4.6) using the five organizing questions from the Introduction.

Study screening and characteristics

Our database search returned 2,423 records. After duplicate removal, we screened titles and abstracts as described above, with reviewer agreement of 90.1% (kappa = 0.33). Given this suboptimal inter-rater agreement (Landis and Koch 1977), we re-clarified our screening process and conducted a second round of review, which resulted in stronger reviewer agreement of 97.8% (kappa = 0.81). We did not find any (additional) eligible papers through citation snowballing. We deemed 33 of the 136 papers subjected to full-text review to be eligible for final inclusion (Figure 2.1).

Most eligible studies were based in low-income, non-urban, and minority communities in the US, as we would expect from an EJ-focused literature. (See Tables A.3 and A.4 in the supplemental materials for an overview of broad community characteristics.) Ten of the 33 papers are situated outside of the US. In general, author affiliations are concentrated in the global North. Figure 2.2 illustrates the global distribution of author affiliations[4] and their study sites.

What sampling and data collection methods are used?

We observed substantial heterogeneity in study designs, sampling protocols, independent variables studied, outcomes of interest, and the metrics with which these were measured. Just over half the studies in our review collected data using purposive or convenience sampling (n=17). A small number collected data through random sampling (n=5). The remaining studies (n=11) were not specific about their sampling strategy. Interviews are the most common type of data collection method (n=21), followed by surveys (n=15), water quality sampling (n=8), and focus groups (n=2). Sample sizes and sampling units vary considerably across studies (Table 2.2).

Table 2.2: A summary of data collection methods used across reviews.

Author	Study location	Used open-ended interviews (n=sample size)	Used surveys (n=sample size)	Used focus groups (n=sample size)	Collected water quality samples (n=sample size)	Sampling strategy
Williams and Florez (2002)	US	-	1183	-	-	random

[4]For authors affiliated with more than one institution, we made our best judgment on which was their primary institution at the time of publication. We included all authors for multi-authored papers.

King et al (2006)	US	12	-	-	-	non-random
Mascarenhas (2007)	Canada	27	-	-	-	unclear
Blakeney and Marshall (2009)	US	40	73	-	-	non-random
Castano-Vinyals et al (2011)	Spain	1118	-	-	-	non-random
Heaney et al (2011)	US	-	120	-	94	non-random
Otero et al (2011)	Spain	17	-	-	-	unclear
Domenech et al (2013)	Nepal	18	-	-	-	unclear
Heaney et al (2013)	US	-	27	-	20	non-random
Wedgworth and Brown (2013)	US	-	305	-	305	random
Wutich et al (2013)	Bolivia, Fiji, New Zealand, US	135	-	-	-	non-random
Balazs and Lubell (2014)	US	9	30	-	-	non-random
Balazs and Ray (2014)	US	17	-	-	-	unclear
Mehta et al (2014)	Bolivia, India	185	91	-	-	unclear

Joshi (2015)	India	-	-	-	-	unclear
Corlin et al (2016)	US	-	-	-	144	unclear
Flanagan et al (2016)	US	-	1287	-	-	random
Kozlowski and Perkins (2016)	US	20	-	-	-	non-random
Sitthikriengkrai and Porath (2017)	Thailand	-	-	-	-	unclear
Wright-Contreras et al (2017)	Vietnam	unclear	100	-	8	unclear
Eggers et al (2018)	US	30	197	-	164	non-random
Rall (2018)	US	15	-	-	-	non-random
Hargrove et al (2018)	US	unclear	50	unclear	-	unclear
Amiri and Zhao (2019)	US	-	13	23	59	unclear
Sansom et al (2019)	US	-	13	-	13	random
Krings et al (2019)	US	4	-	-	-	non-random
Hale (2019)	US	40	-	-	-	non-random
Stillo et al (2019)	US	-	76	-	-	unclear

Seamster and Purifoy (2021)	US	unclear	-	-	-	non-random
Lockhart et al (2020)	US	8	21	-	-	random
Dobbin (2020)	US	27	-	-	-	non-random

How is (un)just access to water defined and measured?

Definitions of access include physical accessibility, measured as the presence-absence of piped water infrastructure or sanitation services (Heaney, S. Wilson, et al. 2011; Lockhart, Wood, and MacDonald Gibson 2020; Seamster and Purifoy 2021; Wedgworth and J. Brown 2013)[5] and water quality, measured as microbiological or chemical contamination (Amiri and Zhao 2019; Corlin et al. 2016; Eggers et al. 2018). Intermittency and affordability, research themes in related disciplines such as water security (See Wutich, Budds, Eichelberger, et al. 2017), are not discussed in the reviewed papers.

Studies assess disparities in water access in three main ways: perceptions of drinking water access or quality; conditions of water access, i.e., piped access or measurements of water quality; and poor health outcomes. (See Table A.5 in the supplementary materials for more information on data extraction, including measures of drinking water access and ties to health outcomes.) Research on user perceptions of (in)justice illustrate how injustices are experienced and whether concerns are voiced (Dobbin 2020; Kozlowski and Perkins 2016; Rall 2018; Sansom et al. 2019; Stillo et al. 2019; B. L. Williams and Florez 2002). For example: priorities such as middle-class lifestyles may be seen as acceptable trade-offs for pollution, muting dissenting voices against water quality injustices within the community (Kozlowski and Perkins 2016); or "outsider groups" such as Mexican-Americans may be less trustful of their water supply, but also less likely to participate in civic activities concerning water (B. L. Williams and Florez 2002). While perception-based studies communicate lived experiences of injustices, they do not provide evidence on whether perceptions correspond to measured levels of water contamination - perceptions (such as of odor and taste) may indicate problems with quality, but indicator-less (e.g., smell-less, tasteless) water can also be unsafe.

The most common approach to evaluating disparities in access is the measurement of drinking water contamination in under-served communities. Studies that measure chemical or microbial contamination find elevated levels of contamination in households disconnected from public water systems, often in unincorporated communities and American Indian reser-

[5]The Seamster & Purifoy paper was included as it was published online in 2020.

vations (Eggers et al. 2018; Flanagan et al. 2016; Heaney, S. Wilson, et al. 2011). (See Table S4 in the supplementary materials for a summary of the parameters evaluated across studies with sample sizes and methods.) Three papers on microbial contamination measure E. coli in domestic well samples (Heaney, S. Wilson, et al. 2011; Heaney, Wing, et al. 2013; Wedgworth and J. Brown 2013), all in majority Black communities. Metals, organics, and disinfection byproducts have also indicated unsafe drinking water contamination in marginalized communities (Corlin et al. 2016; Eggers et al. 2018; Heaney, Wing, et al. 2013; Wright-Contreras, March, and Schramm 2017). All but one of these studies found some degree of elevated drinking water contamination.

While many studies (n=26) motivate their work citing the health risks of unsafe drinking water, only three (Amiri and Zhao 2019; Hargrove, Del Rio, and Korc 2018; Wedgworth and J. Brown 2013) report health outcome data related to drinking water contamination. Wedgworth and Brown (2013) use survey data from 264 households to show a statistical association between gastrointestinal disease and fecal contamination in water supply. In a health survey of two Texas colonias, Hargrove et al. (2018) report water insecurity-related stress, hauling-related injuries, and gastrointestinal illnesses. Amiri and Zhao (2019) distributed a general health questionnaire to their focus group participants, but found their questionnaire was not well-tailored for the literacy skills of their participants. All three studies distance their findings from any claims of causality, reflecting the difficulty of attributing health outcomes to drinking water contamination (e.g., Cumming et al. 2019).

What forms of environmental injustice are most commonly discussed?

Most studies (n=28) present drinking water disparities as distributive injustice. Distributive injustices include water quality falling short of meeting safe drinking water standards (Amiri and Zhao 2019; Corlin et al. 2016; Heaney, S. Wilson, et al. 2011; Wedgworth and J. Brown 2013), or the inability to tap into municipal services (Heaney, Wing, et al. 2013; Lockhart, Wood, and MacDonald Gibson 2020; Seamster and Purifoy 2021). Potential contamination sources range from polluting industries (Amiri and Zhao 2019; Blakeney and Marshall 2009; Sansom et al. 2019), disinfection byproducts (Castaño-Vinyals et al. 2011), or fecal contamination in private wells (Heaney, S. Wilson, et al. 2011; Stillo et al. 2019; Wedgworth and J. Brown 2013). Studies on inadequate water quantity are less common; authors attribute disparities to large-scale infrastructure projects, inadequate state planning, or deliberate exclusion from municipal services (Domènech, March, and Saurí 2013; Hale 2019; Joshi 2015; Lockhart, Wood, and MacDonald Gibson 2020; Otero et al. 2011; Seamster and Purifoy 2021).

Procedural injustice is the second-most common type of injustice discussed in our review (n=16). Cases of procedural injustices include under-representation of disadvantaged communities in regional water management planning (C. L. Balazs and Lubell 2014; Dobbin 2020); lack of voice and agency at the municipal level (Rutt and Bluwstein 2017); or

the revocation of public deliberation in the name of emergency management (Krings, Kornberg, and Lane 2019). Institutional norms and histories shape local expectations of citizen involvement and state accountability (Wutich, Brewis, et al. 2013).

While most studies ground their research designs in one of these two forms of injustice, some authors discuss two additional forms. The first of these is recognitional injustice (n=6), or the lack of "recognition of harm, and recognition of other ways of knowing" (Mascarenhas 2007). This may appear as personal attacks and discreditation of lived experiences (Rutt and Bluwstein 2017); disdain for the poor (Rall 2018); or official refusals (or inability) to recognize harm to cultural, spiritual, and ecological traditions (Mascarenhas 2007). Occupational injustice is the second form, explicitly explored by only one study that evaluated how the consequences of poor water quality spill over into individuals' abilities to carry out everyday activities (Blakeney and Marshall 2009). Activities that must be adapted because of water supply conditions – cooking routines, personal hygiene, organization of workdays – are tangible examples of how a water injustice morphs into "occupational injustice", affecting almost all aspects of daily life.

Though several distinct forms of injustice are identified in the literature, numerous studies in our review illustrate how these different types of injustice are co-produced or reinforce one another. (See Table S5 in the supplementary materials for a summary of how studies explicitly and implicitly engage with different forms of justice.) Deference to certain types of expertise can lead to the disqualification of testimonies in participatory processes (Mascarenhas 2007; Sitthikriengkrai and Porath 2017). Distributive injustice can result in psychosocial effects such as the degradation of spirit and the loss of hope (Blakeney and Marshall 2009), or manifest as occupational injustices through barriers to handwashing, laundry, or cooking (Blakeney and Marshall 2009; Lockhart, Wood, and MacDonald Gibson 2020). When recognitional justice is encouraged – such as through the availability of bilingual materials – it can lead to more fruitful and generative participatory processes (C. L. Balazs and Lubell 2014). Table 2.3 contains examples of relationships that we pulled from our reviewed works.

Table 2.3: How different forms of justice reinforce one another or are co-produced.

Author	**Examples**
Mascarenhas (2007)	(1) Relaxed monitoring and reporting requirements for industry made it difficult to have environmental harms be properly recognized. (2) Deference to specific types of expertise enabled the disqualification of testimonies in participatory processes.

Blakeney and Marshall (2009)	Distributive injustice produced psychosocial effects, affecting the ability to push for recognitional justice: "People's spirit ... has degraded ... because of the degradation of the river ... if you spend all that time being unable to combat it, sometimes you just kind of lose hope and join in and think that the river is unrecoverable."
Domenech et al (2013)	Compensation measures for environmental degradation were proposed based on direct losses related to fields, houses, or trees (i.e., to land ownership), funneling benefits to land-owning castes over landless ones.
Heaney et al (2011, 2013)	Community-based methods - specifically, the collaborative effort to collect and analyze drinking water data in data-sparce places - positioned researchers and communities to engage with multiple forms of (in)justice simultaneously.
Wutich et al (2013)	(1) Personal-level injustices short-circuited otherwise-just institutions. (2) Valuation of different justices can be hierarchical, i.e., some are more distressing than others depending on context.
Balazs and Lubell (2014)	(1) Social learning (learning across stakeholders) connected procedural justice goals to distributive justice outcomes through increased representation of traditionally marginalized groups and their water challenges. (2) Efforts at inclusivity within these meetings - such as providing bilingual options - helped attendees feel 'safe' for the first time.
Balazs and Ray (2014)	Residents from East Orosi were repeatedly turned away by water board administrators when seeking clarification on their water quality reports because of their Spanish accents.
Mehta et al (2014)	Different realities of citizenship (facilitated by recognitional and procedural injustice) has resulted in varied distributive outcomes of drinking water access.

Kozlowski and Perkins (2016)	"DuPont families" (families with kin employed by DuPont) alienated, ostracized, and even physically intimidated those critical of DuPont, making it difficult or uncomfortable to participate in social functions (such as going to church) or to voice their opinions.
Case (2017)	Grassroots activists' efforts to address procedural injustice - e.g., by writing memos, organizing public workshops, appealing to higher levels of government, or harnessing the media - paved the way for distributive justice.
Rutt and Bluwstein (2017)	Public officials suppressed civic engagement with misogynistic and paternalistic language.
Sitthikriengkrai and Porath (2017)	(1) Cultural difference and prejudice made it difficult for Karen villagers to advocate for their experience of environmental degradation - which threatened their community fabric as well as their capacity to cope - in "strict legal and/or empirical terms." (2) Engagement with civil society can exacerbate recognitional injustice if EJ communities are mischaracterized, such as when villages are reduced to "a community of beggars" and thereby stripped of their sense of dignity.
Rall (2018)	Community organizers found that fighting the stigmatization and internalized oppression of the poor was key to overcoming class divisions (for movement building), which had manifested as blaming, criticizing, criminalizing, and ignoring those unable to cope.
Amiri and Zhao (2019)	Similar to Heaney et al above, community-driven project design enabled researchers and communities to engage with multiple forms of (in)justice at once.
Krings et al (2019)	Emergency management created a special scenario in which traditional channels of participation and civic engagement were closed off, resulting in (and entrenching) distributive injustice.

Seamster and Purifoy (2021)	A predominantly white and wealthy community actively obstructed a neighboring Black community's pursuit of procedural and distributive justice by denying the acquisition of basic amenities such as water and sanitation.
Lockhart et al (2020)	Lack of access to drinking water directly affected peoples' abilities to participate in daily activities (e.g., handwashing, laundry, cooking). These effects were mitigated by personal connections or by monetary means, the latter of which is facilitated by socioeconomic status.
Dobbin (2020)	Collaborative groundwater governance increased recognitional justice, as large water agencies became (more) aware of smaller ones (and their issues) or as new relationships were established across different stakeholders.

As mentioned earlier, environmental injustices are often described in collective or community level terms. We find the papers in our review also raise individual or household-level dimensions of injustice that do not fit neatly into this distributive-procedural-recognitional framework. When the polluter is also the main regional employer, those vocal about the environmental injustice can be met with physical intimidation and social ostracization, presenting obvious challenges to accessing procedural channels (Kozlowski and Perkins 2016; Rall 2018). The pervasive influence of societal patriarchal and misogynistic norms are also apparent, when men are expected to "man up" and "sacrifice themselves for work" (Kozlowski and Perkins 2016) and women are described as "hysterical," a "crazy housewife," or asked (by a government official) whether they are "on their period" (Rutt and Bluwstein 2017). In describing how jealous, grudge-bearing vendors can deny service to would-be customers, Wutich et al (2013) contend such experiences can be more distressing than distributive or procedural injustices.

How are affected communities resisting or coping?

In the absence of drinking water access, communities may engage in active protest and resistance, as well as everyday acts of coping. In states with enforceable environmental regulations, some communities find success in partnering with nonprofits and other external actors to win monetary compensation and regulatory enforcement (Sitthikriengkrai and Porath 2017). Support from external partners may be the only way for these lived experiences to gain credibility in the eyes of the government (Krings, Kornberg, and Lane 2019). Community activists also strategically use the media (Case 2017; Sitthikriengkrai and Porath

2017) or direct action (Domènech, March, and Saurí 2013) to shame or force the government's or industry's hand into enacting change. Perhaps unsurprisingly, however, protestors and resistors have often been targeted and shut down by the state (Mascarenhas 2007; Rutt and Bluwstein 2017).

Communities also partner with community-based organizations and universities for assistance in quantifying lack of water access, fighting for political representation, or providing a platform for community voices (Rall 2018). Participatory Action Research (PAR) methods encourage more equitable community-researcher dynamics (Amiri and Zhao 2019; Blakeney and Marshall 2009; Eggers et al. 2018; Heaney, S. Wilson, et al. 2011), and shape the research design itself with local knowledge (Heaney, Wing, et al. 2013). (See Table S6 in the supplementary materials for more information on study motivations, drivers of site selection, and community partnerships.) Such collaborations are not inherently good; community partners could perpetuate injustices by prioritizing agendas that serve them more than the community itself (Sitthikriengkrai and Porath 2017). Collectively, however, PAR-based studies represent an effort to challenge extractive research dynamics by encouraging the community to harness research as an empowerment opportunity.

Outside of resistance and protest, acts of coping to procure water take place at both the individual and community level. Some examples of individual coping mechanisms are to rely on extra-household resources such as friends and family, laundromats, or hotels for cooking, washing, and bathing (Blakeney and Marshall 2009; Eggers et al. 2018; Lockhart, Wood, and MacDonald Gibson 2020). Other options are to buy bottled water or to haul water from nearby municipal water supplies, an ostensibly safer source but a significant financial burden (C. L. Balazs and Ray 2014; Blakeney and Marshall 2009; Eggers et al. 2018; Krings, Kornberg, and Lane 2019; Lockhart, Wood, and MacDonald Gibson 2020). Hargrove et al. (2018) estimate that it costs households with median annual incomes of 15,000 USD roughly 70 USD per month to haul water; at 5.6% of income, this is well above most accepted affordability thresholds (Goddard, Ray, and C. Balazs 2022). At the community level, neighborhoods pursue service extension and/or consolidation with nearby systems (Seamster and Purifoy 2021), and mutual-aid water committees step up to buy from municipal providers, or procure water through unofficial or illegal means (Mehta, Allouche, et al. 2014). Coping looks different for the rich and the poor: while the poor may rely on "hidden" pathways such as illegally siphoning water from city pipes, the rich are more easily able to augment their access by purchasing water or water purification technologies (Joshi 2015; Mehta, Allouche, et al. 2014).

What, if any, mechanisms of redress are advocated?

A core tenet of the environmental justice framework is that, when injustices are uncovered, a means of redress or mitigation is necessary. In the reviewed literature, reform proposals for corrective justice and redress are usually inferred rather than made explicit. Most calls for accountability are directed at the state (Domènech, March, and Saurí 2013; Kozlowski and

Perkins 2016; Krings, Kornberg, and Lane 2019). Some studies suggest mitigative measures to improve drinking water conditions, such as the installation of public water fountains for unhoused people (Hale 2019), or, in the case of Flint, charging city officials with involuntary manslaughter (Krings, Kornberg, and Lane 2019). Others propose policy initiatives at one or two levels above a community-specific intervention. These include calls for state policy to support more regional solutions (C. L. Balazs and Lubell 2014); increased government funding to develop technical, managerial, and financial capacity for communities (C. L. Balazs and Ray 2014); or even a broad call for "stable economic development" that would protect human health (Kozlowski and Perkins 2016). Advocacy directed at the state-run public health department can be an explicit component of community-based participatory research projects (Amiri and Zhao 2019). Non-state pathways of redress appear as supportive mechanisms towards eventual reparation; examples include advocacy for public outreach to be sensitive to socio-economic status (Flanagan et al. 2016), or for activists in White communities to challenge race and class privileges (Kozlowski and Perkins 2016).

Many under-served communities fall outside the reach of state assistance for drinking water provision, and struggle with recognition and citizenship in addition to access to water. US-based studies discuss this struggle in unincorporated and predominantly Latinx or Black communities, American Indian reservations, and US-Mexico border colonias. Some studies explicitly place the blame of drinking water disparities on the state for neglecting basic services for indigenous populations and for poor regulatory enforcement (Corlin et al. 2016; Eggers et al. 2018; Mascarenhas 2007). Others point to historical neglect of and discrimination against low-income Black communities (Lockhart, Wood, and MacDonald Gibson 2020; Rall 2018; Seamster and Purifoy 2021; Stillo et al. 2019). Internationally, this form of neglect especially pertains to peri-urban or semi-rural spaces (Mehta, Allouche, et al. 2014; Wright-Contreras, March, and Schramm 2017; Wutich, Brewis, et al. 2013). Collectively, the literature shows that the boundaries of state oversight (and thus accountability) often coincide with disparities in access to safe water.

Several studies document the paradox of redress when the state itself is complicit in the persistence of environmental injustices. For instance, histories of unequal land and water changes have been officially rewritten as heroic responses to water scarcity (Otero et al. 2011); struggles for access to water go hand-in-hand with other rights violations (Mehta, Allouche, et al. 2014); state structures have historically had discriminatory and exclusionary effects on indigenous and Black populations (Eggers et al. 2018; Rall 2018; Seamster and Purifoy 2021; Stillo et al. 2019); and where, eventually, activists have succeeded in acquiring safe water, it may have needed years of "pushing, pushing" (Case 2017). The irony of the EJ literature leaning on the state for redress when the state is also a perpetrator of injustice is a point we return to below.

2.5 Discussion

We organize our discussion around the questions and findings presented in the section above, with a focus on how they relate to the broader EJ literature.

Access

Studies on perceptions of (poor) water quality give us unique insights into the lived water access conditions of many communities, especially in places where water quality data is unavailable. While perceptions matter, they may not by themselves signal distributional injustices in water access without independent verification of water quality. Perceptions can be influenced by many factors (França Doria 2009): people can think water quality is poor when it is in fact good (Rozin et al. 2015), or believe bottled water is of high quality when it may not be (Cohen and Ray 2018; A. R. Williams et al. 2015). We contend that treating the perception of poor water quality as a distributional injustice by itself, as several of our reviewed studies have done, raises questions about construct validity. Perception-based data on a sample of respondents within areas in which distributional injustices have already been established could usefully combine the strengths of primary and secondary data.

Perceptions can, however, signal procedural injustices, or show why access inequalities persist. Perceptions of poor water quality that the water utilities pay no attention to, or dismiss out of hand, can be signs of procedural injustice. Perceptions can uncover how, even in a community with a well-documented environmental burden, there may be different views on whether that burden is an unjust one (Kozlowski and Perkins 2016). Perceptions are also central to recognitional injustice, particularly when they relate to how the community is treated by the utility or by the state (B. L. Williams and Florez 2002). Studies on community perceptions of poor water quality, or distrust of the water utility or of the state, are best conducted with primary data collection. Studies using only publicly-available data cannot reveal how under-served communities feel.

Several studies on poor water quality compare their findings against WHO or EPA benchmarks for drinking water standards, but do not refer to a comparison community (or regional / national averages) to attempt to evaluate disproportionate harms (Corlin et al. 2016; Heaney, S. Wilson, et al. 2011; Heaney, Wing, et al. 2013; Wedgworth and J. Brown 2013). Even if the study communities are minoritized because of race or ethnicity, without a comparison with a larger population, disproportionate harms cannot be specifically established. These studies, therefore, implicitly rely on an absolute framing of justice, which considers a failure to reach a minimum standard (e.g., to stay below a federal maximum contaminant level) inherently unjust (Feinberg 1974; Montague 1980). The original concept of EJ, however, was distinctive in its focus on fairness and equity for specific communities with respect to others (Taylor 2000). The convergence of injustice as a violation of universal standards with injustice as a disproportionate burden for particular communities - which we find primary studies often assume - brings justice-oriented studies closer to adjacent frame-

works such as the human right to water or water security (Cook and Bakker 2012; Gleick 1998; Sultana and Loftus 2015). Research using secondary data can more readily identify disproportionate harms using statistical analysis.

Justice

Early EJ studies have been critiqued for their characterization and analyses of distributional injustices as "snapshots" in time (Been and Gupta 1997; Pastor, Sadd, and Hipp 2001), and for ignoring the production of injustices as structurally racial projects (see Omi and Winant 2014). While this is not true of all studies in our review, many are of a descriptive, "snapshot" nature, in part because they use mainly primary data. These studies do not explain how and why water access injustices arose, nor why they persist, though understanding past and ongoing power dynamics between stakeholders is critical for understanding why otherwise unacceptable conditions of water supply come to be considered "natural" (Otero et al. 2011).

"Snapshot" studies have played a critical role in the development of EJ as a field: indeed, the discipline emerged as a form of praxis, at the convergence of environmental activism, broader social movements, and scholarly research (Sze and Jonathan K. London 2008; Taylor 2000). Even today, requests to the government for water quality testing, infrastructure assessment, and service connections or annexation are often ignored in the US (Lambrinidou 2018; Pulido 2016) and abroad (Nganyanyuka et al. 2018). Documenting distributional inequities and characterizing harms therefore remain crucial for keeping the field grounded, and for providing ongoing, evidence-based tools for environmental advocates and for communities.

Several studies explore how multiple forms of injustice - distributional, procedural and recognitional - co-occur (Mascarenhas 2007; Mehta, Allouche, et al. 2014; Sitthikriengkrai and Porath 2017; Wutich, Brewis, et al. 2013). The acknowledgement of plural injustices is in line with calls to avoid ahistorical documentation of distributional inequities, and to instead analyze how multiple forms of injustice are mutually co-constitutive (Malin and Ryder 2018; Pellow 2017; Schlosberg 2013). Throughout the reviewed papers, we see that recognitional injustice often undergirds ongoing distributional or procedural injustices (Fraser 2000; Schlosberg 2004), but also that distributive and procedural injustice can in turn engender recognitional injustice. Such recognitional injustice often results in psychosocial stressors (Blakeney and Marshall 2009), which exacerbate environmental health disparities across race and class lines (Gee and Payne-Sturges 2004; R. Morello-Frosch et al. 2011); as Sitthikriengkrai and Porath (2017) observe, "such indifference and neglect becomes part of the environmental injustice they are suffering." Our review captures ways in which distributive, procedural, and recognitional injustices are co-produced. Collectively, these studies show how community struggles to obtain drinking water access are underlain by long histories of disenfranchisement and resistance, rather than singular sources or events of degradation.

We offer the schematic in Figure 2.3 to diagram how forms of (in)justice – with the impor-

tant inclusion of occupational justice – are situated with respect to one another. The arrows emphasize the multidirectional relationships of (in)justice production and reinforcement we took note of above, with occupational justice at the center to symbolize how distributive, procedural, and recognitional injustices can morph into "daily life" effects.

Community resistance and coping

Communities and community-based organizations have responded to the lack of access to safe water with various forms of resistance (see Results Section 4.5). Protest and resistance can be effective, be ignored (Heaney, Wing, et al. 2013), or be actively shut down (Rutt and Bluwstein 2017). Communities may also have internal divisions, as early EJ literature recognized (Bullard 1990), and protest can be seen as a threat to the benefits of job security even when the harmful effects of polluting are obvious (Kozlowski and Perkins 2016). Community-based participatory research approaches have been promoted as empowering communities to own their water quality data, to demand greater access to procedural justice in the future (Heaney, S. Wilson, et al. 2011), and as part of a project of social transformation (Mertens 2010).

Organized resistance is always a risk (Scott 2012), and the complexity of tracing drinking water contamination is itself an impediment to protest (Sitthikriengkrai and Porath 2017). Establishing a causal connection between polluters, pollutants, and communities has been a major goal of traditional EJ research (Been and Gupta 1997; Mohai and Saha 2015; Pastor, Sadd, and Hipp 2001). The drinking water contamination papers in our review, however, do not always identify a polluter, and none are designed to assess a causal relationship between polluter and pollutant. For instance, studies may point to inadequate household infrastructure (such as failing on-site septic tanks, or old private wells) as potential causes of water quality degradation (Heaney, S. Wilson, et al. 2011; Heaney, Wing, et al. 2013; Wedgworth and J. Brown 2013), with no external actors held responsible. Tracing the causes of nonpoint water pollution sources that contaminate ground- and surface water, possibly over long periods, is particularly difficult (Loague, Corwin, and Ellsworth 1998). The challenges of attribution give rise to the question: if we cannot establish the entity (or entities) responsible for poor drinking water quality, how do we connect distributive injustice to corrective justice?

Where resistance and protest are not feasible, too risky, or challenging to organize due to intra-community differences (Naman and Gibson 2015), communities have to cope with inadequate water quality or quantities (Venkataramanan et al. 2020). Balazs and Ray (2014) describe a "composite drinking water burden" in which "exposure and coping are mutually constitutive, cyclical, and multilevel" - illustrating a vicious cycle in which under-resourced communities are forced to take on additional costs such as purchasing bottled water or expensive household-level filtration systems. Coping costs can also be incurred in other aspects of life, such as home maintenance costs or heightened levels of social anxiety (Blakeney and Marshall 2009; Lockhart, Wood, and MacDonald Gibson 2020), or having to share water

with neighboring households (Wutich, Budds, W. Jepson, et al. 2018). Current EJ research on drinking water disparities alludes to, but could more comprehensively assess, these financial and psychosocial coping costs in their analyses of injustice. These costs are widespread, and therefore research focused only on the direct effects of poor water quality understates the full costs of environmental injustice. It is possible to measure at least some of these costs, as studies in low-income settings have shown (Burt, Ercümen, et al. 2018; Pattanayak et al. 2005).

Less common in US studies, but quite prevalent in those based in the Global South, is the analysis of unauthorized use as a form of coping (Joshi 2015; Mehta, Allouche, et al. 2014). This can include tapping into municipal water systems without permission, refusing to pay utility bills when they are unaffordable, or illegally sinking a private or community well. These coping mechanisms are distinct from buying or boiling water; they are examples of what has been called the "quiet encroachment of the ordinary" (Bayat 2000). Or, as Scott (2012) would say, these are unobtrusive forms of lawbreaking that "do not shout". The Global EJ (GEJ) literature, theorizing from the South more broadly, has framed such acts as strategic responses to poverty or inequality (e.g., Burt and Ray 2014). The US-based literature, by contrast, discusses non-payment (and subsequent shut-offs) of water services in cities such as Flint or Baltimore as forced upon the poor (Mack and Wrase 2017; Swain, McKinney, and Susskind 2020) – but not as a form of silent insubordination under scarcity.

Overall, the drinking water as well as the broader EJ literature show that resistance is a dynamic process during which the nature of resistance, the environmental conditions, and the needs and priorities of the community itself, are mutually shaped and reshaped. The critique of "snapshot" studies of distributional disparities is even more salient with this recognition.

Redress and the role of the state

In addition to distributional, procedural and recognitional justice, redress, or corrective justice, is central to the EJ framework. Corrective justice calls for penalizing those who have broken environmental laws, and for remedying the damages inflicted as a result of lawbreaking (Kuehn 2000). Holding the polluter accountable, while a key demand of EJ activists (Bullard and Wright 1993), is a challenge for drinking water contamination because attribution of cause to nonpoint sources is difficult (as discussed above). Even when both the source and harms of pollution are evident, redress is complicated when the responsible companies are no longer in existence, as in the case of uranium mines in Navajo Nation (deLemos et al. 2009; Harmon et al. 2017). When holding a distinct polluter accountable is not feasible, the position within the traditional EJ literature is that the state should step in and either mitigate the harms to the community and/or be held accountable for its own lax enforcement (Kuehn 2000). This observation applies to most of the papers in our review, which (often only implicitly) call upon the state for redress.

Historically, the state has played a central role in corrective justice, as either the arbiter or the regulator, to remedy environmental injustices. However, Critical EJ (CEJ) scholars have argued that automatically turning to the state as part of the solution is naive, because the state has allowed injustices to continue, or has actively perpetrated injustices itself, especially in Black and Indigenous communities in the US (Kurtz 2009; Pellow 2017; Pulido, Kohl, and Cotton 2016). Many CEJ scholars have joined forces with political ecologists, arguing that the concept of justice within the EJ movement in the US is inherently liberal, and that racism and exclusionary practices have characterized such liberalism (Pulido, Kohl, and Cotton 2016; Ranganathan 2016). Reliance on the state to mete out justice, therefore, at worst reinforces the hegemonic power structure that produces injustice (Pellow 2017; Swyngedouw and Heynen 2003), and at best dilutes the transformative power of demands for justice by compromising with the liberal state (Pulido, Kohl, and Cotton 2016).

The critique that EJ falls back on the state as arbiter when the state is also an enabler, and thus the call to look beyond the regulatory state in search of transformative coalitions (Pellow 2017; Ranganathan and C. Balazs 2015), must be taken seriously. Pulido et al (2016) have argued that the EJ movement, relying on the state, has improved environments for relatively few marginalized communities. We find, however, that even the outlines of proposed transformative solutions remain vague, as does strategic guidance for the EJ movement to seek redress outside the state (see also Purucker 2021).

Extra-state redress is a particular challenge at the scale of drinking water systems. The global reality is that in countries with near-universal access to drinking water, most households have water provided or regulated by state or local water agencies (Ray and Smith 2021). In unincorporated California, London et al (2021) (2021) show how informality – taken broadly as conditions outside of formal state or market institutions – can obstruct the attainment of water justice. Some of the most deprived areas discussed in the reviewed papers were in such informal, liminal spaces, such as peri-urban areas or unincorporated communities, where the regulatory state is weak or non-existent at the necessary scale (Heaney, S. Wilson, et al. 2011; Mehta, Allouche, et al. 2014). Successful non-state community-managed drinking water systems do exist, e.g. in Paraguay, Bolivia and Peru (Loach, Melgarejo, and Lombardo 2000; Marston 2014; Whittington et al. 2009), but their potential for replication at scale is unknown.

In light of past (and ongoing) debates about redress and the primacy of the state, Sze and London (2008) have argued that the state needs to be brought "back in" (following Skocpol 1985), despite its failures, and despite its exclusions. The "state" as they see it is not a uniform entity, but a fractured one, that actively participates in the distribution of environmental protections and harms at the same time. Pushing back against the argument that EJ is too movement-oriented to be theoretically sound (see Swyngedouw and Heynen 2003), Sze and London argue that EJ has always been about understanding how power manifests itself through race and class to produce benefits and harms, and that making demands of a fractured state is a distinctive feature of EJ theory and practice. Others have

advocated for new modes of water governance that prioritize inclusivity and meaningful engagement of civil society actors (Harris, McKenzie, et al. 2017). The state can also be understood as an evolving (rather than fixed) entity whose form emerges through "the process of struggle," and that cannot be either dismissed or reified (Angel and Loftus 2019). While the drinking water justice papers in this review would (in the main) be compatible with these positions, relying on the state for redress while recognizing it as a producer of persistent inequalities remains a tension within environmental justice work.

Limitations

With our focus on primary data (as bound by our inclusion-exclusion criteria), our analysis is restricted to a subset of the broader literature on environmental justice and drinking water. A review of secondary data-based studies could have revealed additional insights, such as on population-level trends or spatial patterns of access across larger geographies.

In literature reviews, it is good practice to integrate some form of quality or bias assessment to assess the evidence (Khan et al. 2003). We found this difficult to do given the heterogeneity of research designs and methods used across the reviewed works. Standards vis-à-vis research design, data transparency, sample size, sampling units and sampling techniques can vary widely within interdisciplinary research areas (Agrawal 2003). However, even papers with similar goals in our review (e.g., studies on contamination) do not share research practices or outcome metrics. As this research area continues to grow, a comparative meta-analysis will eventually be possible, but only to the extent that researchers develop commensurable definitions and research designs. Wider adoption of commensurable metrics and methods could make disparate case studies more comparable and their results more amenable to aggregation. Measurements of chemical and biological drinking water contamination, or the affordability of supplementary water supplies, are examples of where standardization can be useful. Context-specific aspects of environmental justice, on the other hand - e.g., local and/or indigenous definitions of what is "just" - may not share common outcomes or fit neatly into a liberal juridical framework. Whatever the chosen metrics and approaches, authors should at a minimum be transparent about the strengths and limitations of their study designs.

2.6 Conclusion

We reviewed 33 primary data-based studies published between 2002 and 2020 on environmental (in)justice with respect to adequate and safe drinking water access. This review offers four key contributions to the broader literature on drinking water and EJ. First, our literature review, to our knowledge, is the first on environmental justice and drinking water specifically. By identifying the strengths of our reviewed works alongside gaps in the literature, we suggest that combining primary with secondary data can enrich EJ research (potentially beyond the drinking water domain). Second, by focusing on primary data-based

studies, we draw out insights pertaining to drinking water and EJ, such as user perceptions of water quality or activist narratives of resistance. We show, for example, how EJ struggle is a dynamic process, through which community priorities co-evolve with environmental conditions (e.g., availability of drinking water) and political responses (e.g., resistance or cooperation by government and industry) over time. We show how perceptions – how people in communities feel – give shape to cases of procedural injustice or calls for recognitional justice. These observations are necessarily excluded in papers relying mainly on secondary data but remain essential to understanding EJ struggles over drinking water access. Third, we show that primary data-based EJ studies on drinking water are predominantly predicated on redress by the liberal state, though the broader EJ literature contains active critiques of the liberal racialized state, especially in the US. Fourth, we demonstrate how multiple types of injustice are often co-produced and mutually-reinforcing.

As might be expected, most of the reviewed studies were concerned with drinking water access difficulties in marginalized communities – by race, ethnicity, and socio-economic status. While many drinking water and EJ studies refer to health impacts, relatively few are designed to measure drinking water exposures and associated health outcomes. Distributional and procedural injustices remain the dominant forms of injustice in the reviewed literature. Community protests and resistance when faced with drinking water challenges are also a common theme, but several studies describe years of "pushing" and years of neglect before state agencies offer any redress (if at all). Critical EJ theorists have therefore argued that relying on the state, which is often complicit in perpetuating injustice, is not a useful way forward for the field. However, what communities should walk towards when "walking away" (Pellow 2017) from the state is ill-defined in the EJ literature. Overall, few of the reviewed studies are designed to explicitly assess responsibility. The accountability for redress under different forms of water governance remains an important area for future research.

The original EJ framework has broadened out to GEJ, which connects local EJ struggles to larger political economy forces, and to CEJ, which argues for a more critical examination of the role of the state as well as for a more inclusive role for gender and ecology within EJ. Drinking water studies appear within the GEJ strand, but to a lesser extent in the CEJ strand. (See Table S7 in the supplementary materials for a breakdown of how our reviewed papers overlap with GEJ, CEJ, and Pellow's (2017) four pillars of CEJ.) US-based studies, in particular those influenced by CEJ, have started to analyze the structural racism that underlies the liberal US state as part of their analysis of EJ and water (Pulido 2016; Ranganathan 2016). Future EJ research can take the field towards a more nuanced understanding of how racial policies and governance structures - liberal and otherwise - have shaped the drinking water disparities of today. GEJ studies have accepted the juridical redress framework of EJ to varying degrees, but US-based studies have yet to incorporate the strengths of many global studies (such as a detailed account of coping costs, or theorizing coping as a form of resistance when legal means of access seem infeasible). What US-based EJ studies can learn from the Globally-framed EJ literature is also an important area for

future research.

References

Agrawal, Arun (2003). "Sustainable Governance of Common-Pool Resources: Context, Methods, and Politics". In: *Annual Review of Anthropology* 32.1, pp. 243–262. DOI: 10.1146/annurev.anthro.32.061002.093112. URL: https://doi.org/10.1146/annurev.anthro.32.061002.093112 (visited on 09/02/2021).

Agyeman, Julian (Jan. 1, 2001). "Ethnic Minorities in Britain: Short Change, Systematic Indifference and Sustainable Development". In: *Journal of Environmental Policy & Planning* 3.1, pp. 15–30. ISSN: 1523-908X. DOI: 10.1080/15239080108559291. URL: https://doi.org/10.1080/15239080108559291 (visited on 09/01/2021).

Agyeman, Julian et al. (2016). "Trends and Directions in Environmental Justice: From Inequity to Everyday Life, Community, and Just Sustainabilities". In: *Annual Review of Environment and Resources* 41.1, pp. 321–340. DOI: 10.1146/annurev-environ-110615-090052. URL: https://doi.org/10.1146/annurev-environ-110615-090052 (visited on 09/01/2021).

Alston, Dana (1991). "Transforming a Movement". In: *Race, Poverty & the Environment* 2.3, pp. 1–29. ISSN: 1532-2874. URL: https://www.jstor.org/stable/41555193 (visited on 03/22/2022).

Álvarez, Lina and Brendan Coolsaet (Apr. 2, 2020). "Decolonizing Environmental Justice Studies: A Latin American Perspective". In: *Capitalism Nature Socialism* 31.2, pp. 50–69. ISSN: 1045-5752. DOI: 10.1080/10455752.2018.1558272. URL: https://doi.org/10.1080/10455752.2018.1558272 (visited on 04/22/2024).

Amiri, Azita and Shuang Zhao (2019). "Working with an environmental justice community: Nurse observation, assessment, and intervention". In: *Nursing Forum* 54.2, pp. 270–279. ISSN: 1744-6198. DOI: 10.1111/nuf.12327. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/nuf.12327 (visited on 04/22/2024).

Angel, James and Alex Loftus (Jan. 1, 2019). "With-against-and-beyond the human right to water". In: *Geoforum* 98, pp. 206–213. ISSN: 0016-7185. DOI: 10.1016/j.geoforum.2017.05.002. URL: https://www.sciencedirect.com/science/article/pii/S0016718517301082 (visited on 04/22/2024).

Balazs, Carolina L. and Mark Lubell (Nov. 1, 2014). "Social learning in an environmental justice context: a case study of integrated regional water management". In: *Water Policy* 16 (S2), pp. 97–120. ISSN: 1366-7017. DOI: 10.2166/wp.2014.101. URL: https://doi.org/10.2166/wp.2014.101 (visited on 09/01/2021).

Balazs, Carolina L., Rachel Morello-Frosch, et al. (Sept. 1, 2011). "Social Disparities in Nitrate-Contaminated Drinking Water in California's San Joaquin Valley". In: *Environmental Health Perspectives* 119.9, pp. 1272–1278. DOI: 10.1289/ehp.1002878. URL: https://ehp.niehs.nih.gov/doi/full/10.1289/ehp.1002878 (visited on 09/01/2021).

Balazs, Carolina L., Rachel Morello-Frosch, et al. (Nov. 14, 2012). "Environmental justice implications of arsenic contamination in California's San Joaquin Valley: a cross-sectional, cluster-design examining exposure and compliance in community drinking water systems". In: *Environmental Health* 11.1, p. 84. ISSN: 1476-069X. DOI: 10.1186/1476-069X-11-84. URL: https://doi.org/10.1186/1476-069X-11-84 (visited on 04/25/2023).

Balazs, Carolina L. and Isha Ray (Apr. 2014). "The Drinking Water Disparities Framework: On the Origins and Persistence of Inequities in Exposure". In: *American Journal of Public Health* 104.4, pp. 603–611. ISSN: 0090-0036. DOI: 10.2105/AJPH.2013.301664. URL: https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2013.301664 (visited on 07/18/2023).

Bayat, Asef (Sept. 1, 2000). "From 'Dangerous Classes' to 'Quiet Rebels': Politics of the Urban Subaltern in the Global South". In: *International Sociology* 15.3, pp. 533–557. ISSN: 0268-5809. DOI: 10.1177/026858000015003005. URL: https://doi.org/10.1177/026858000015003005 (visited on 09/02/2021).

Been, Vicki and Francis Gupta (1997). "Coming to the Nuisance or Going to the Barrios? A Longitudinal Analysis of Environmental Justice Claims". In: *Ecology Law Quarterly* 24.1, pp. 1–56. ISSN: 0046-1121. URL: https://www.jstor.org/stable/24113440 (visited on 04/09/2024).

Blakeney, Anne B. and Amy Marshall (Jan. 1, 2009). "Water Quality, Health, and Human Occupations". In: *American Journal of Occupational Therapy* 63.1, pp. 46–57. ISSN: 0272-9490. DOI: 10.5014/ajot.63.1.46. URL: https://ajot.aota.org/article.aspx?articleid=1869980 (visited on 09/01/2021).

Borderon, Marion et al. (Dec. 6, 2021). "The risks of invisibilization of populations and places in environment-migration research". In: *Humanities and Social Sciences Communications* 8.1, pp. 1–11. ISSN: 2662-9992. DOI: 10.1057/s41599-021-00999-0. URL: https://www.nature.com/articles/s41599-021-00999-0 (visited on 02/19/2022).

Brown, P. (Apr. 1, 1995). "Race, Class, and Environmental Health: A Review and Systematization of the Literature". In: *Environmental Research* 69.1, pp. 15–30. ISSN: 0013-9351. DOI: 10.1006/enrs.1995.1021. URL: https://www.sciencedirect.com/science/article/pii/S0013935185710213 (visited on 09/01/2021).

Bullard, Robert D. (Oct. 16, 1990). *Dumping In Dixie: Race, Class, And Environmental Quality*. Avalon Publishing. 204 pp. ISBN: 978-0-8133-7954-8.

Bullard, Robert D. and Glenn S. Johnson (2000). "Environmentalism and Public Policy: Environmental Justice: Grassroots Activism and Its Impact on Public Policy Decision Making". In: *Journal of Social Issues* 56.3, pp. 555–578. ISSN: 1540-4560. DOI: 10.1111/0022-4537.00184. URL: https://spssi.onlinelibrary.wiley.com/doi/abs/10.1111/0022-4537.00184 (visited on 09/01/2021).

Bullard, Robert D. and Beverly H. Wright (Sept. 1, 1993). "Environmental Justice for all: Community Perspectives on Health and Research". In: *Toxicology and Industrial Health* 9.5, pp. 821–841. ISSN: 0748-2337. DOI: 10.1177/074823379300900508. URL: https://doi.org/10.1177/074823379300900508 (visited on 09/01/2021).

Burt, Zachary, Ayşe Ercümen, et al. (Sept. 1, 2018). "From intermittent to continuous service: Costs, benefits, equity and sustainability of water system reforms in Hubli-Dharwad, India". In: *World Development* 109, pp. 121–133. ISSN: 0305-750X. DOI: 10.1016/j.worlddev.2018.04.011. URL: https://www.sciencedirect.com/science/article/pii/S0305750X1830130X (visited on 09/02/2021).

Burt, Zachary and Isha Ray (2014). "Storage and Non-Payment: Persistent Informalities within the Formal Water Supply of Hubli-Dharwad, India". In: 7.1, p. 15.

Čapek, Stella M. (Feb. 1, 1993). "The "Environmental Justice" Frame: A Conceptual Discussion and an Application*". In: *Social Problems* 40.1, pp. 5–24. ISSN: 0037-7791. DOI: 10.2307/3097023. URL: https://doi.org/10.2307/3097023 (visited on 09/01/2021).

Case, Robert A. (Jan. 1, 2017). "Environmental oversight and the citizen activist: Lessons from an oral history of activism surrounding Elmira, Ontario's 1989 water crisis". In: *Community Development* 48.1, pp. 86–104. ISSN: 1557-5330. DOI: 10.1080/15575330.2016.1249491. URL: https://doi.org/10.1080/15575330.2016.1249491 (visited on 09/01/2021).

Castaño-Vinyals, Gemma et al. (Mar. 16, 2011). "Socioeconomic status and exposure to disinfection by-products in drinking water in Spain". In: *Environmental Health* 10.1, p. 18. ISSN: 1476-069X. DOI: 10.1186/1476-069X-10-18. URL: https://doi.org/10.1186/1476-069X-10-18 (visited on 09/01/2021).

Christ Commission for Racial Justice, United Church of (1987). *Toxic wastes and race in the United States: A national report on the racial and socio-economic characteristics of communities with hazardous waste sites*, pp. 1–79. URL: https://www.ucc.org/wp-content/uploads/2020/12/ToxicWastesRace.pdf.

Cohen, Alasdair and Isha Ray (July 2018). "The global risks of increasing reliance on bottled water". In: *Nature Sustainability* 1.7, pp. 327–329. ISSN: 2398-9629. DOI: 10.1038/s41893-018-0098-9. URL: https://www.nature.com/articles/s41893-018-0098-9 (visited on 09/02/2021).

Cook, Christina and Karen Bakker (Feb. 1, 2012). "Water security: Debating an emerging paradigm". In: *Global Environmental Change* 22.1, pp. 94–102. ISSN: 0959-3780. DOI: 10.1016/j.gloenvcha.2011.10.011. URL: https://www.sciencedirect.com/science/article/pii/S0959378011001804 (visited on 09/02/2021).

Corlin, Laura et al. (Dec. 1, 2016). "Health Effects and Environmental Justice Concerns of Exposure to Uranium in Drinking Water". In: *Current Environmental Health Reports* 3.4, pp. 434–442. ISSN: 2196-5412. DOI: 10.1007/s40572-016-0114-z. URL: https://doi.org/10.1007/s40572-016-0114-z (visited on 09/01/2021).

Cumming, Oliver et al. (Aug. 28, 2019). "The implications of three major new trials for the effect of water, sanitation and hygiene on childhood diarrhea and stunting: a consensus statement". In: *BMC Medicine* 17.1, p. 173. ISSN: 1741-7015. DOI: 10.1186/s12916-019-1410-x. URL: https://doi.org/10.1186/s12916-019-1410-x (visited on 09/02/2021).

Deacon, Leith and Jamie Baxter (June 1, 2013). "No opportunity to say no: a case study of procedural environmental injustice in Canada". In: *Journal of Environmental Planning*

and Management 56.5, pp. 607–623. ISSN: 0964-0568. DOI: 10.1080/09640568.2012.692502. URL: https://doi.org/10.1080/09640568.2012.692502 (visited on 09/01/2021).

deLemos, Jamie L. et al. (July 9, 2009). "Development of risk maps to minimize uranium exposures in the Navajo Churchrock mining district". In: *Environmental Health* 8.1. Number: 1, p. 29. ISSN: 1476-069X. DOI: 10.1186/1476-069X-8-29. URL: https://doi.org/10.1186/1476-069X-8-29 (visited on 09/02/2021).

Dobbin, Kristin B. (Dec. 1, 2020). ""Good Luck Fixing the Problem": Small Low-Income Community Participation in Collaborative Groundwater Governance and Implications for Drinking Water Source Protection". In: *Society & Natural Resources* 33.12, pp. 1468–1485. ISSN: 0894-1920. DOI: 10.1080/08941920.2020.1772925. URL: https://doi.org/10.1080/08941920.2020.1772925 (visited on 04/22/2024).

Domènech, Laia, Hug March, and David Saurí (June 1, 2013). "Contesting large-scale water supply projects at both ends of the pipe in Kathmandu and Melamchi Valleys, Nepal". In: *Geoforum* 47, pp. 22–31. ISSN: 0016-7185. DOI: 10.1016/j.geoforum.2013.02.002. URL: https://www.sciencedirect.com/science/article/pii/S0016718513000377 (visited on 09/01/2021).

Eggers, Margaret J. et al. (Jan. 2018). "Community Engaged Cumulative Risk Assessment of Exposure to Inorganic Well Water Contaminants, Crow Reservation, Montana". In: *International Journal of Environmental Research and Public Health* 15.1, p. 76. ISSN: 1660-4601. DOI: 10.3390/ijerph15010076. URL: https://www.mdpi.com/1660-4601/15/1/76 (visited on 04/22/2024).

Feinberg, Joel (1974). "Noncomparative Justice". In: *The Philosophical Review* 83.3, pp. 297–338. ISSN: 0031-8108. DOI: 10.2307/2183696. URL: https://www.jstor.org/stable/2183696 (visited on 09/02/2021).

Felton, Ryan, Lisa Gill, and Lewis Kendall (Mar. 31, 2021). "We sampled tap water across the US – and found arsenic, lead and toxic chemicals". In: *the Guardian*. URL: http://www.theguardian.com/us-news/2021/mar/31/americas-tap-water-samples-forever-chemicals (visited on 09/02/2021).

Flanagan, Sara V. et al. (Aug. 15, 2016). "Arsenic in private well water part 3 of 3: Socioeconomic vulnerability to exposure in Maine and New Jersey". In: *Science of The Total Environment* 562, pp. 1019–1030. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2016.03.217. URL: https://www.sciencedirect.com/science/article/pii/S0048969716306398 (visited on 09/02/2021).

França Doria, Miguel de (Nov. 1, 2009). "Factors influencing public perception of drinking water quality". In: *Water Policy* 12.1, pp. 1–19. ISSN: 1366-7017. DOI: 10.2166/wp.2009.051. URL: https://doi.org/10.2166/wp.2009.051 (visited on 09/02/2021).

Fraser, Nancy (May 1, 2000). "Rethinking Recognition". In: *New Left Review* 3, pp. 107–120. ISSN: 0028-6060. URL: https://www.proquest.com/docview/1301918978/citation/FC7E9F2BB9F64881PQ/1 (visited on 09/01/2021).

García-Valiñas, María De Los Ángeles, Roberto Martínez-Espiñeira, and Francisco González-Gómez (Sept. 1, 2010). "Measuring Water Affordability: A Proposal for Urban Centres in Developed Countries". In: *International Journal of Water Resources Development* 26.3,

pp. 441–458. ISSN: 0790-0627. DOI: 10.1080/07900627.2010.491971. URL: https://doi.org/10.1080/07900627.2010.491971 (visited on 09/01/2021).

Gee, Gilbert C. and Devon C. Payne-Sturges (Dec. 2004). "Environmental Health Disparities: A Framework Integrating Psychosocial and Environmental Concepts". In: *Environmental Health Perspectives* 112.17. Publisher: Environmental Health Perspectives, pp. 1645–1653. DOI: 10.1289/ehp.7074. URL: https://ehp.niehs.nih.gov/doi/full/10.1289/ehp.7074 (visited on 04/28/2024).

Gleick, Peter H. (Oct. 1, 1998). "The human right to water". In: *Water Policy* 1.5, pp. 487–503. ISSN: 1366-7017. DOI: 10.1016/S1366-7017(99)00008-2. URL: https://www.sciencedirect.com/science/article/pii/S1366701799000082 (visited on 09/02/2021).

Goddard, Jessica J., Isha Ray, and Carolina Balazs (Jan. 20, 2021). "Water affordability and human right to water implications in California". In: *PLOS ONE* 16.1, e0245237. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0245237. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245237 (visited on 09/01/2021).

— (2022). "How should water affordability be measured in the United States? A critical review". In: *WIREs Water* 9.1, e1573. ISSN: 2049-1948. DOI: 10.1002/wat2.1573. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1573 (visited on 02/19/2022).

Graham, David A. (Jan. 22, 2016). *Who Poisoned Flint?* The Atlantic. URL: https://www.theatlantic.com/politics/archive/2016/01/who-poisoned-flint/425454/ (visited on 09/01/2021).

Hale, Marcia Rosalie (Apr. 2019). "Fountains for Environmental Justice: Public Water, Homelessness, and Migration in the Face of Global Environmental Change". In: *Environmental Justice* 12.2, pp. 33–40. ISSN: 1939-4071. DOI: 10.1089/env.2018.0031. URL: https://www.liebertpub.com/doi/full/10.1089/env.2018.0031 (visited on 04/22/2024).

Hargrove, William L., Michelle Del Rio, and Marcelo Korc (Dec. 2018). "Water Matters: Water Insecurity and Inadequate Sanitation in the U.S./Mexico Border Region". In: *Environmental Justice* 11.6, pp. 222–227. ISSN: 1939-4071. DOI: 10.1089/env.2018.0022. URL: https://www.liebertpub.com/doi/full/10.1089/env.2018.0022 (visited on 04/22/2024).

Harmon, Molly E. et al. (July 2017). "Residential proximity to abandoned uranium mines and serum inflammatory potential in chronically exposed Navajo communities". In: *Journal of Exposure Science & Environmental Epidemiology* 27.4, pp. 365–371. ISSN: 1559-064X. DOI: 10.1038/jes.2016.79. URL: https://www.nature.com/articles/jes201679 (visited on 09/02/2021).

Harris, Leila M., Scott McKenzie, et al. (2017). "Water justice". In: *The Routledge Handbook of Environmental Justice*. 1st. Routledge. ISBN: 978-1-315-67898-6. URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315678986-27/water-justice-leila-harris-scott-mckenzie-lucy-rodina-sameer-shah-nicole-wilson (visited on 09/29/2021).

Harris, Leila M., Lucy Rodina, and Cynthia Morinville (Oct. 2, 2015). "Revisiting the Human Right to Water from an environmental justice lens". In: *Politics, Groups, and Identities* 3.4, pp. 660–665. ISSN: 2156-5503. DOI: 10.1080/21565503.2015.1080619. URL: https://doi.org/10.1080/21565503.2015.1080619 (visited on 09/30/2021).

Heaney, Christopher D., Sacoby Wilson, et al. (Aug. 2011). "Use of Community-Owned and -Managed Research to Assess the Vulnerability of Water and Sewer Services in Marginalized and Underserved Environmental Justice Communities". In: *Journal of Environmental Health* 74.1, pp. 8–17. ISSN: 00220892. URL: https://www.proquest.com/docview/878224975/abstract/F07775EA10ED40DBPQ/1 (visited on 09/01/2021).

Heaney, Christopher D., Steve Wing, et al. (June 2013). "Public Infrastructure Disparities and the Microbiological and Chemical Safety of Drinking and Surface Water Supplies in a Community Bordering a Landfill". In: *Journal of environmental health* 75.10, pp. 24–36. ISSN: 0022-0892. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4514614/ (visited on 09/01/2021).

Heller, Léo (May 12, 2022). *The Human Rights to Water and Sanitation*. Cambridge University Press. 453 pp. ISBN: 978-1-108-83724-8.

Hoekstra, Arjen Y., Joost Buurman, and Kees C. H. van Ginkel (May 2018). "Urban water security: A review". In: *Environmental Research Letters* 13.5, p. 053002. ISSN: 1748-9326. DOI: 10.1088/1748-9326/aaba52. URL: https://doi.org/10.1088/1748-9326/aaba52 (visited on 09/02/2021).

Holifield, Ryan (Feb. 1, 2001). "Defining Environmental Justice and Environmental Racism". In: *Urban Geography* 22.1, pp. 78–90. ISSN: 0272-3638. DOI: 10.2747/0272-3638.22.1.78. URL: https://doi.org/10.2747/0272-3638.22.1.78 (visited on 04/09/2024).

Holland, Breena (May 4, 2017). "Procedural justice in local climate adaptation: political capabilities and transformational change". In: *Environmental Politics* 26.3, pp. 391–412. ISSN: 0964-4016. DOI: 10.1080/09644016.2017.1287625. URL: https://doi.org/10.1080/09644016.2017.1287625 (visited on 09/01/2021).

Honneth, Axel (Dec. 1, 2004). "Recognition and Justice: Outline of a Plural Theory of Justice". In: *Acta Sociologica* 47.4, pp. 351–364. ISSN: 0001-6993. DOI: 10.1177/0001699304048668. URL: https://doi.org/10.1177/0001699304048668 (visited on 04/22/2024).

Jepson, Wendy E. et al. (2017). "Progress in household water insecurity metrics: a cross-disciplinary approach". In: *WIREs Water* 4.3, e1214. ISSN: 2049-1948. DOI: 10.1002/wat2.1214. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1214 (visited on 02/19/2022).

JMP (2020). *Drinking water*. Drinking water — Joint Monitoring Programme for Water Supply and Sanitation. URL: https://washdata.org/monitoring/drinking-water.

Joshi, Deepa (May 1, 2015). "Like water for justice". In: *Geoforum* 61, pp. 111–121. ISSN: 0016-7185. DOI: 10.1016/j.geoforum.2015.02.020. URL: https://www.sciencedirect.com/science/article/pii/S0016718515000640 (visited on 09/01/2021).

Khan, Khalid S et al. (Mar. 1, 2003). "Five Steps to Conducting a Systematic Review". In: *Journal of the Royal Society of Medicine* 96.3, pp. 118–121. ISSN: 0141-0768. DOI: 10.

1177/014107680309600304. URL: https://doi.org/10.1177/014107680309600304 (visited on 09/02/2021).

Kozlowski, Michelle and Harold A. Perkins (Oct. 2, 2016). "Environmental justice in Appalachia Ohio? An expanded consideration of privilege and the role it plays in defending the contaminated status quo in a white, working-class community". In: *Local Environment* 21.10, pp. 1288–1304. ISSN: 1354-9839. DOI: 10.1080/13549839.2015.1111316. URL: https://doi.org/10.1080/13549839.2015.1111316 (visited on 09/01/2021).

Krings, Amy, Dana Kornberg, and Erin Lane (July 1, 2019). "Organizing Under Austerity: How Residents' Concerns Became the Flint Water Crisis". In: *Critical Sociology* 45.4, pp. 583–597. ISSN: 0896-9205. DOI: 10.1177/0896920518757053. URL: https://doi.org/10.1177/0896920518757053 (visited on 04/22/2024).

Kuehn, Robert R (2000). "A Taxonomy of Environmental Justice". In: *Environmental Law Reporter News & Analysis* 30.9, pp. 10681–10703.

Kurtz, Hilda E. (2009). "Acknowledging the Racial State: An Agenda for Environmental Justice Research". In: *Antipode* 41.4, pp. 684–704. ISSN: 1467-8330. DOI: 10.1111/j.1467-8330.2009.00694.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8330.2009.00694.x (visited on 03/15/2024).

Lambrinidou, Yanna (2018). "When Technical Experts Set Out to "Do Good": Deficit-Based Constructions of "the Public" and the Moral Imperative for New Visions of Engagement". In: *Michigan Journal of Sustainability* 6.1. ISSN: 2332-0621. DOI: http://dx.doi.org/10.3998/mjs.12333712.0006.102. URL: http://hdl.handle.net/2027/spo.12333712.0006.102.

Landis, J. Richard and Gary G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data". In: *Biometrics* 33.1, pp. 159–174. ISSN: 0006-341X. DOI: 10.2307/2529310. URL: https://www.jstor.org/stable/2529310 (visited on 04/22/2024).

Lee, Ellen J. and Kellogg J. Schwab (June 1, 2005). "Deficiencies in drinking water distribution systems in developing countries". In: *Journal of Water and Health* 3.2, pp. 109–127. ISSN: 1477-8920. DOI: 10.2166/wh.2005.0012. URL: https://doi.org/10.2166/wh.2005.0012 (visited on 09/01/2021).

Loach, Peter W., Silvio Melgarejo, and Manuel Lombardo (2000). "The small-scale water provider in Paraguay: bringing private sector efficiency to water resource use and the provision of drinking water to the poor". In: *Natural Resources Forum* 24.4, pp. 263–271. ISSN: 1477-8947. DOI: 10.1111/j.1477-8947.2000.tb00951.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1477-8947.2000.tb00951.x (visited on 09/03/2021).

Loague, Keith, Dennis L. Corwin, and Timothy R. Ellsworth (1998). "The Challenge of Predicting Nonpoint Source Pollution". In: *Environmental Science & Technology* 32.5, p. 4.

Lockhart, Sydney, Erica Wood, and Jacqueline MacDonald Gibson (Aug. 1, 2020). "Impacts of Exclusion From Municipal Water Service on Water Availability: A Case Study". In: *NEW SOLUTIONS: A Journal of Environmental and Occupational Health Policy* 30.2,

pp. 127–137. ISSN: 1048-2911. DOI: 10.1177/1048291120932913. URL: https://doi.org/10.1177/1048291120932913 (visited on 04/22/2024).

London, J. K. et al. (2021). "Disadvantaged Unincorporated Communities and the Struggle for Water Justice in California". In: *Water Alternatives* 14.2, pp. 520–545. ISSN: 1965-0175. URL: https://escholarship.org/uc/item/984273kz (visited on 04/23/2024).

Mack, Elizabeth A. and Sarah Wrase (Jan. 11, 2017). "A Burgeoning Crisis? A Nationwide Assessment of the Geography of Water Affordability in the United States". In: *PLOS ONE* 12.1, e0169488. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0169488. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169488 (visited on 09/02/2021).

Malin, Stephanie A. and Stacia S. Ryder (Jan. 2, 2018). "Developing deeply intersectional environmental justice scholarship". In: *Environmental Sociology* 4.1, pp. 1–7. ISSN: 2325-1042. DOI: 10.1080/23251042.2018.1446711. URL: https://www.tandfonline.com/doi/full/10.1080/23251042.2018.1446711 (visited on 09/02/2021).

Marston, Andrea J (2014). "The Scale of Informality: Community-Run Water Systems in Peri-Urban Cochabamba, Bolivia". In: *Water Alternatives* 7.1, p. 17.

Martínez-Santos, Pedro (July 4, 2017). "Does 91% of the world's population really have "sustainable access to safe drinking water"?" In: *International Journal of Water Resources Development* 33.4, pp. 514–533. ISSN: 0790-0627. DOI: 10.1080/07900627.2017.1298517. URL: https://iahr.tandfonline.com/doi/full/10.1080/07900627.2017.1298517 (visited on 09/01/2021).

Mascarenhas, Michael (Dec. 1, 2007). "Where the Waters Divide: First Nations, Tainted Water and Environmental Justice in Canada". In: *Local Environment* 12.6, pp. 565–577. ISSN: 1354-9839. DOI: 10.1080/13549830701657265. URL: https://doi.org/10.1080/13549830701657265 (visited on 09/01/2021).

McGurty, Eileen Maura (June 1, 2000). "Warren County, NC, and the Emergence of the Environmental Justice Movement: Unlikely Coalitions and Shared Meanings in Local Collective Action". In: *Society & Natural Resources* 13.4, pp. 373–387. ISSN: 0894-1920. DOI: 10.1080/089419200279027. URL: https://doi.org/10.1080/089419200279027 (visited on 09/01/2021).

Meehan, Katie et al. (2020). "Exposing the myths of household water insecurity in the global north: A critical review". In: *WIREs Water* 7.6, e1486. ISSN: 2049-1948. DOI: 10.1002/wat2.1486. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1486 (visited on 02/19/2022).

Mehta, Lyla, Jeremy Allouche, et al. (July 1, 2014). "Global environmental justice and the right to water: The case of peri-urban Cochabamba and Delhi". In: *Geoforum* 54, pp. 158–166. ISSN: 0016-7185. DOI: 10.1016/j.geoforum.2013.05.014. URL: https://www.sciencedirect.com/science/article/pii/S0016718513001309 (visited on 09/01/2021).

Mehta, Lyla and Synne Movik (2014). "Liquid dynamics: challenges for sustainability in the water domain". In: *WIREs Water* 1.4, pp. 369–384. ISSN: 2049-1948. DOI: 10.1002/

wat2.1031. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1031 (visited on 02/20/2022).

Mertens, Donna M (Jan. 1, 2010). "Social Transformation and Evaluation". In: *Evaluation Journal of Australasia* 10.2, pp. 3–10. ISSN: 1035-719X. DOI: 10.1177/1035719X1001000202. URL: https://doi.org/10.1177/1035719X1001000202 (visited on 09/03/2021).

Minkler, Meredith et al. (Feb. 1, 2008). "Promoting Environmental Justice Through Community-Based Participatory Research: The Role of Community and Partnership Capacity". In: *Health Education & Behavior* 35.1, pp. 119–137. ISSN: 1090-1981. DOI: 10.1177/1090198106287692. URL: https://doi.org/10.1177/1090198106287692 (visited on 09/01/2021).

Mohai, Paul, David Pellow, and J. Timmons Roberts (2009). "Environmental Justice". In: *Annual Review of Environment and Resources* 34.1, pp. 405–430. DOI: 10.1146/annurev-environ-082508-094348. URL: https://doi.org/10.1146/annurev-environ-082508-094348 (visited on 09/01/2021).

Mohai, Paul and Robin Saha (Dec. 2015). "Which came first, people or pollution? A review of theory and evidence from longitudinal environmental justice studies". In: *Environmental Research Letters* 10.12, p. 125011. ISSN: 1748-9326. DOI: 10.1088/1748-9326/10/12/125011. URL: https://dx.doi.org/10.1088/1748-9326/10/12/125011 (visited on 04/09/2024).

Montague, Phillip (1980). "Comparative and Non-Comparative Justice". In: *The Philosophical Quarterly (1950-)* 30.119, pp. 131–140. ISSN: 0031-8094. DOI: 10.2307/2219277. URL: https://www.jstor.org/stable/2219277 (visited on 09/02/2021).

Morello-Frosch, Rachel et al. (May 2011). "Understanding The Cumulative Impacts Of Inequalities In Environmental Health: Implications For Policy". In: *Health Affairs* 30.5, pp. 879–887. ISSN: 0278-2715. DOI: 10.1377/hlthaff.2011.0153. URL: https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2011.0153 (visited on 04/22/2024).

Morello-Frosch, Rachel A (Aug. 1, 2002). "Discrimination and the Political Economy of Environmental Inequality". In: *Environment and Planning C: Government and Policy* 20.4, pp. 477–496. ISSN: 0263-774X. DOI: 10.1068/c03r. URL: https://doi.org/10.1068/c03r (visited on 04/22/2024).

Mueller, J. Tom and Stephen Gasteyer (June 22, 2021). "The widespread and unjust drinking water and clean water crisis in the United States". In: *Nature Communications* 12.1, p. 3544. ISSN: 2041-1723. DOI: 10.1038/s41467-021-23898-z. URL: https://www.nature.com/articles/s41467-021-23898-z (visited on 09/09/2021).

Naman, Julia Marie and Jacqueline MacDonald Gibson (Oct. 1, 2015). "Disparities in Water and Sewer Services in North Carolina: An Analysis of the Decision-Making Process". In: *American Journal of Public Health* 105.10, e20–e26. ISSN: 0090-0036. DOI: 10.2105/AJPH.2015.302731. URL: https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2015.302731 (visited on 09/02/2021).

Nganyanyuka, Kapongola et al. (Oct. 1, 2018). "If citizens protest, do water providers listen? Water woes in a Tanzanian town". In: *Environment and Urbanization* 30.2, pp. 613–630.

ISSN: 0956-2478. DOI: 10.1177/0956247817744942. URL: https://doi.org/10.1177/0956247817744942 (visited on 09/02/2021).

Octavianti, Thanti and Chad Staddon (2021). "A review of 80 assessment tools measuring water security". In: *WIREs Water* 8.3, e1516. ISSN: 2049-1948. DOI: 10.1002/wat2.1516. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1516 (visited on 02/20/2022).

Omi, Michael and Howard Winant (June 20, 2014). *Racial Formation in the United States*. Routledge. 345 pp. ISBN: 978-1-135-12751-0.

Otero, Iago et al. (May 15, 2011). "Water scarcity, social power and the production of an elite suburb: The political ecology of water in Matadepera, Catalonia". In: *Ecological Economics*. Special Section: Ecological Economics and Environmental History 70.7, pp. 1297–1308. ISSN: 0921-8009. DOI: 10.1016/j.ecolecon.2009.09.011. URL: https://www.sciencedirect.com/science/article/pii/S0921800909004030 (visited on 09/01/2021).

Pastor, Manuel, Jim Sadd, and John Hipp (2001). "Which Came First? Toxic Facilities, Minority Move-In, and Environmental Justice". In: *Journal of Urban Affairs* 23.1, pp. 1–21. ISSN: 1467-9906. DOI: 10.1111/0735-2166.00072. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/0735-2166.00072 (visited on 09/02/2021).

Patel, Anisha I. et al. (2020). "Drinking Water in the United States: Implications of Water Safety, Access, and Consumption". In: *Annual Review of Nutrition* 40.1, pp. 345–373. DOI: 10.1146/annurev-nutr-122319-035707. URL: https://doi.org/10.1146/annurev-nutr-122319-035707 (visited on 09/09/2021).

Pattanayak, Subhrendu K. et al. (2005). "Coping with unreliable public water supplies: Averting expenditures by households in Kathmandu, Nepal". In: *Water Resources Research* 41.2. ISSN: 1944-7973. DOI: 10.1029/2003WR002443. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003WR002443 (visited on 09/02/2021).

Pellow, David (Dec. 2017). *What is Critical Environmental Justice?* John Wiley & Sons. 200 pp. ISBN: 978-0-7456-7937-2.

Pierce, Gregory et al. (2021). "Solutions to the problem of drinking water service affordability: A review of the evidence". In: *WIREs Water* 8.4, e1522. ISSN: 2049-1948. DOI: 10.1002/wat2.1522. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1522 (visited on 02/20/2022).

Pulido, Laura (1996). "A Critical Review of the Methodology of Environmental Racism Research*". In: *Antipode* 28.2, pp. 142–159. ISSN: 1467-8330. DOI: 10.1111/j.1467-8330.1996.tb00519.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8330.1996.tb00519.x (visited on 09/01/2021).

— (Mar. 1, 2000). "Rethinking Environmental Racism: White Privilege and Urban Development in Southern California". In: *Annals of the Association of American Geographers* 90.1, pp. 12–40. ISSN: 0004-5608. DOI: 10.1111/0004-5608.00182. URL: https://doi.org/10.1111/0004-5608.00182 (visited on 09/01/2021).

Pulido, Laura (July 2, 2016). "Flint, Environmental Racism, and Racial Capitalism". In: *Capitalism Nature Socialism* 27.3, pp. 1–16. ISSN: 1045-5752. DOI: 10.1080/10455752.

2016.1213013. URL: https://doi.org/10.1080/10455752.2016.1213013 (visited on 09/02/2021).

Pulido, Laura, Ellen Kohl, and Nicole-Marie Cotton (Apr. 2, 2016). "State Regulation and Environmental Justice: The Need for Strategy Reassessment". In: *Capitalism Nature Socialism* 27.2, pp. 12–31. ISSN: 1045-5752. DOI: 10.1080/10455752.2016.1146782. URL: https://doi.org/10.1080/10455752.2016.1146782 (visited on 09/02/2021).

Purucker, David (July 3, 2021). "Critical environmental justice and the state: a critique of Pellow". In: *Environmental Sociology* 7.3, pp. 176–186. ISSN: null. DOI: 10.1080/23251042.2021.1878575. URL: https://doi.org/10.1080/23251042.2021.1878575 (visited on 03/15/2024).

R Core Team (2021). *R: A language and environment for statistical computing*. Vienna, Austria. URL: https://www.R-project.org/.

Rall, Ann (May 4, 2018). "Community Organizing and the Detroit Water Struggle: Report from the Front Lines". In: *Journal of Progressive Human Services* 29.2, pp. 103–129. ISSN: 1042-8232. DOI: 10.1080/10428232.2017.1412767. URL: https://doi.org/10.1080/10428232.2017.1412767 (visited on 04/22/2024).

Ranganathan, Malini (July 2, 2016). "Thinking with Flint: Racial Liberalism and the Roots of an American Water Tragedy". In: *Capitalism Nature Socialism* 27.3, pp. 17–33. ISSN: 1045-5752. DOI: 10.1080/10455752.2016.1206583. URL: https://doi.org/10.1080/10455752.2016.1206583 (visited on 09/02/2021).

Ranganathan, Malini and Carolina Balazs (Apr. 3, 2015). "Water marginalization at the urban fringe: environmental justice and urban political ecology across the North–South divide". In: *Urban Geography* 36.3, pp. 403–423. ISSN: 0272-3638. DOI: 10.1080/02723638.2015.1005414. URL: https://doi.org/10.1080/02723638.2015.1005414 (visited on 09/02/2021).

Ray, Isha and Kirk R Smith (Mar. 1, 2021). "Towards safe drinking water and clean cooking for all". In: *The Lancet Global Health* 9.3, e361–e365. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(20)30476-9. URL: https://www.sciencedirect.com/science/article/pii/S2214109X20304769 (visited on 09/02/2021).

Rosinger, Asher Y. and Sera L. Young (2020). "The toll of household water insecurity on health and human biology: Current understandings and future directions". In: *WIREs Water* 7.6, e1468. ISSN: 2049-1948. DOI: 10.1002/wat2.1468. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1468 (visited on 02/20/2022).

Rozin, Paul et al. (Jan. 1, 2015). "Psychological aspects of the rejection of recycled water: Contamination, purification and disgust". In: *Judgment and decision making* 10, pp. 50–63.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. Boston, MA. URL: http://www.rstudio.com/.

Rutt, Rebecca L. and Jevgeniy Bluwstein (Apr. 1, 2017). "Quests for Justice and Mechanisms of Suppression in Flint, Michigan". In: *Environmental Justice* 10.2, pp. 27–35. ISSN: 1939-4071. DOI: 10.1089/env.2016.0047. URL: https://www.liebertpub.com/doi/full/10.1089/env.2016.0047 (visited on 09/01/2021).

Sansom, Garett et al. (Jan. 2019). "Vulnerable Populations Exposed to Lead-Contaminated Drinking Water within Houston Ship Channel Communities". In: *International Journal of Environmental Research and Public Health* 16.15, p. 2745. ISSN: 1660-4601. DOI: 10.3390/ijerph16152745. URL: https://www.mdpi.com/1660-4601/16/15/2745 (visited on 04/22/2024).

Schlosberg, David (Sept. 1, 2004). "Reconceiving Environmental Justice: Global Movements And Political Theories". In: *Environmental Politics* 13.3, pp. 517–540. ISSN: 0964-4016. DOI: 10.1080/0964401042000229025. URL: https://doi.org/10.1080/09644010420000229025 (visited on 09/01/2021).

— (Feb. 1, 2013). "Theorising environmental justice: the expanding sphere of a discourse". In: *Environmental Politics* 22.1, pp. 37–55. ISSN: 0964-4016. DOI: 10.1080/09644016.2013.755387. URL: https://rsa.tandfonline.com/doi/full/10.1080/09644016.2013.755387 (visited on 09/01/2021).

Scott, James C. (Oct. 14, 2012). *Two Cheers for Anarchism*. Princeton University Press. ISBN: 978-1-4008-4462-3. URL: https://www.degruyter.com/document/doi/10.1515/9781400844623/html (visited on 09/02/2021).

Seamster, Louise and Danielle Purifoy (Apr. 3, 2021). "What is environmental racism for? Place-based harm and relational development". In: *Environmental Sociology* 7.2, pp. 110–121. ISSN: null. DOI: 10.1080/23251042.2020.1790331. URL: https://doi.org/10.1080/23251042.2020.1790331 (visited on 04/22/2024).

Sitthikriengkrai, Malee and Nathan Porath (June 1, 2017). "'Lead-polluted water changed our lives': A Thai-Karen village's quest for environmental justice". In: *South East Asia Research* 25.2, pp. 139–156. ISSN: 0967-828X. DOI: 10.1177/0967828X17706566. URL: https://doi.org/10.1177/0967828X17706566 (visited on 09/01/2021).

Skocpol, Theda (Sept. 13, 1985). "Bringing the State Back In: Strategies of Analysis in Current Research". In: *Bringing the State Back In*. Cambridge University Press. ISBN: 978-1-107-71713-8.

South, Andy (2017). *rnaturalearth: World Map Data from Natural Earth*. URL: https://CRAN.R-project.org/package=rnaturalearth.

Stewart, Iris T., Christopher M. Bacon, and William D. Burke (Dec. 1, 2014). "The uneven distribution of environmental burdens and benefits in Silicon Valley's backyard". In: *Applied Geography* 55, pp. 266–277. ISSN: 0143-6228. DOI: 10.1016/j.apgeog.2014.09.016. URL: https://www.sciencedirect.com/science/article/pii/S0143622814002264 (visited on 09/01/2021).

Stillo, Frank et al. (Oct. 10, 2019). "Well water testing in African-American communities without municipal infrastructure: Beliefs driving decisions". In: *Science of The Total Environment* 686, pp. 1220–1228. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2019.05.317. URL: https://www.sciencedirect.com/science/article/pii/S0048969719323642 (visited on 04/22/2024).

Sultana, Farhana and Alex Loftus (2015). "The Human Right to Water: Critiques and Condition of Possibility". In: *WIREs Water* 2.2, pp. 97–105. ISSN: 2049-1948. DOI: 10.1002/

wat2.1067. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1067 (visited on 09/02/2021).

Sunter, Deborah A., Sergio Castellanos, and Daniel M. Kammen (Jan. 2019). "Disparities in rooftop photovoltaics deployment in the United States by race and ethnicity". In: *Nature Sustainability* 2.1, pp. 71–76. ISSN: 2398-9629. DOI: 10.1038/s41893-018-0204-z. URL: https://www.nature.com/articles/s41893-018-0204-z/ (visited on 09/01/2021).

Swain, Marian, Emmett McKinney, and Lawrence Susskind (Feb. 21, 2020). "Water Shutoffs in Older American Cities: Causes, Extent, and Remedies". In: *Journal of Planning Education and Research*, p. 0739456X20904431. ISSN: 0739-456X. DOI: 10.1177/0739456X20904431. URL: https://doi.org/10.1177/0739456X20904431 (visited on 09/02/2021).

Swyngedouw, Erik and Nikolas C. Heynen (2003). "Urban Political Ecology, Justice and the Politics of Scale". In: *Antipode* 35.5, pp. 898–918. ISSN: 1467-8330. DOI: 10.1111/j.1467-8330.2003.00364.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8330.2003.00364.x (visited on 09/02/2021).

Syafrudin, Muhammad et al. (Jan. 2021). "Pesticides in Drinking Water—A Review". In: *International Journal of Environmental Research and Public Health* 18.2, p. 468. ISSN: 1660-4601. DOI: 10.3390/ijerph18020468. URL: https://www.mdpi.com/1660-4601/18/2/468 (visited on 02/19/2022).

Szasz, Andrew and Michael Meuser (July 1, 1997). "Environmental Inequalities: Literature Review and Proposals for New Directions in Research and Theory". In: *Current Sociology* 45.3, pp. 99–120. ISSN: 0011-3921. DOI: 10.1177/001139297045003006. URL: https://doi.org/10.1177/001139297045003006 (visited on 09/01/2021).

Sze, Julie (Jan. 7, 2020). *Environmental Justice in a Moment of Danger*. Univ of California Press. 155 pp. ISBN: 978-0-520-30074-3.

Sze, Julie and Jonathan K. London (2008). "Environmental Justice at the Crossroads". In: *Sociology Compass* 2.4, pp. 1331–1354. ISSN: 1751-9020. DOI: 10.1111/j.1751-9020.2008.00131.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-9020.2008.00131.x (visited on 04/09/2024).

Taylor, Dorceta E. (Jan. 1, 2000). "The Rise of the Environmental Justice Paradigm: Injustice Framing and the Social Construction of Environmental Discourses". In: *American Behavioral Scientist* 43.4, pp. 508–580. ISSN: 0002-7642. DOI: 10.1177/0002764200043004003. URL: https://doi.org/10.1177/0002764200043004003 (visited on 09/02/2021).

UN Economic and Social Council (2002). *General Comment No. 15: The Right to Water (Arts. 11 and 12 of the Covenant)*.

UN General Assembly (2015). *Transforming our world : the 2030 Agenda for Sustainable Development*. URL: https://www.unfpa.org/sites/default/files/resource-pdf/Resolution_A_RES_70_1_EN.pdf.

UNICEF and WHO (2019). *Progress on household drinking water, sanitation and hygiene 2000-2017*. New York.

United Church of Christ Commission for Racial Justice (1991). *The Proceedings of the First National People of Color Environmental Leadership Summit*. New York, NY.

US Environmental Protection Agency (n.d.). *40 CFR 141*. URL: https://www.epa.gov/sites/default/files/2019-03/documents/cfr-2011-title40-vol23-part141.pdf.

Venkataramanan, Vidya et al. (2020). "Coping strategies for individual and household-level water insecurity: A systematic review". In: *WIREs Water* 7.5, e1477. ISSN: 2049-1948. DOI: 10.1002/wat2.1477. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1477 (visited on 02/20/2022).

Walker, Gordon (2009). "Beyond Distribution and Proximity: Exploring the Multiple Spatialities of Environmental Justice". In: *Antipode* 41.4, pp. 614–636. ISSN: 1467-8330. DOI: 10.1111/j.1467-8330.2009.00691.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8330.2009.00691.x (visited on 09/01/2021).

Ward, Mary H. et al. (July 2018). "Drinking Water Nitrate and Human Health: An Updated Review". In: *International Journal of Environmental Research and Public Health* 15.7, p. 1557. ISSN: 1660-4601. DOI: 10.3390/ijerph15071557. URL: https://www.mdpi.com/1660-4601/15/7/1557 (visited on 02/19/2022).

Wedgworth, Jessica Cook and Joe Brown (June 1, 2013). "Limited Access to Safe Drinking Water and Sanitation in Alabama's Black Belt: A Cross-Sectional Case Study". In: *Water Quality, Exposure and Health* 5.2, pp. 69–74. ISSN: 1876-1666. DOI: 10.1007/s12403-013-0088-0. URL: https://doi.org/10.1007/s12403-013-0088-0 (visited on 09/01/2021).

Whittington, Dale et al. (Oct. 1, 2009). "How well is the demand-driven, community management model for rural water supply systems doing? Evidence from Bolivia, Peru and Ghana". In: *Water Policy* 11.6, pp. 696–718. ISSN: 1366-7017. DOI: 10.2166/wp.2009.310. URL: https://doi.org/10.2166/wp.2009.310 (visited on 09/03/2021).

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. URL: https://ggplot2.tidyverse.org/.

Williams, Ashley R. et al. (Oct. 27, 2015). "A Systematic Review and Meta-Analysis of Fecal Contamination and Inadequate Treatment of Packaged Water". In: *PLOS ONE* 10.10, e0140899. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0140899. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0140899 (visited on 09/02/2021).

Williams, Bryan L and Yvette Florez (Apr. 2002). "Do Mexican Americans perceive environmental issues differently than Caucasians: a study of cross-ethnic variation in perceptions related to water in Tucson." In: *Environmental Health Perspectives* 110 (Suppl 2), pp. 303–310. ISSN: 0091-6765. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1241177/ (visited on 09/01/2021).

Wilson, Sacoby M. et al. (Mar. 1, 2010). "Climate Change, Environmental Justice, and Vulnerability: An Exploratory Spatial Analysis". In: *Environmental Justice* 3.1. Number: 1 Publisher: Mary Ann Liebert, Inc., publishers, pp. 13–19. ISSN: 1939-4071. DOI: 10.1089/env.2009.0035. URL: https://www.liebertpub.com/doi/abs/10.1089/env.2009.0035 (visited on 09/01/2021).

World Health Organization (2018). *A global overview of national regulations and standards for drinking-water quality*. Geneva: World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/272345/9789241513760-eng.pdf.

Wright-Contreras, Lucía, Hug March, and Sophie Schramm (Mar. 1, 2017). "Fragmented landscapes of water supply in suburban Hanoi". In: *Habitat International* 61, pp. 64–74. ISSN: 0197-3975. DOI: 10.1016/j.habitatint.2017.02.002. URL: https://www.sciencedirect.com/science/article/pii/S0197397515303453 (visited on 09/01/2021).

Wutich, Amber, Alexandra Brewis, et al. (July 1, 2013). "Rules, Norms, and Injustice: A Cross-Cultural Study of Perceptions of Justice in Water Institutions". In: *Society & Natural Resources* 26.7, pp. 795–809. ISSN: 0894-1920. DOI: 10.1080/08941920.2012.723302. URL: https://doi.org/10.1080/08941920.2012.723302 (visited on 09/01/2021).

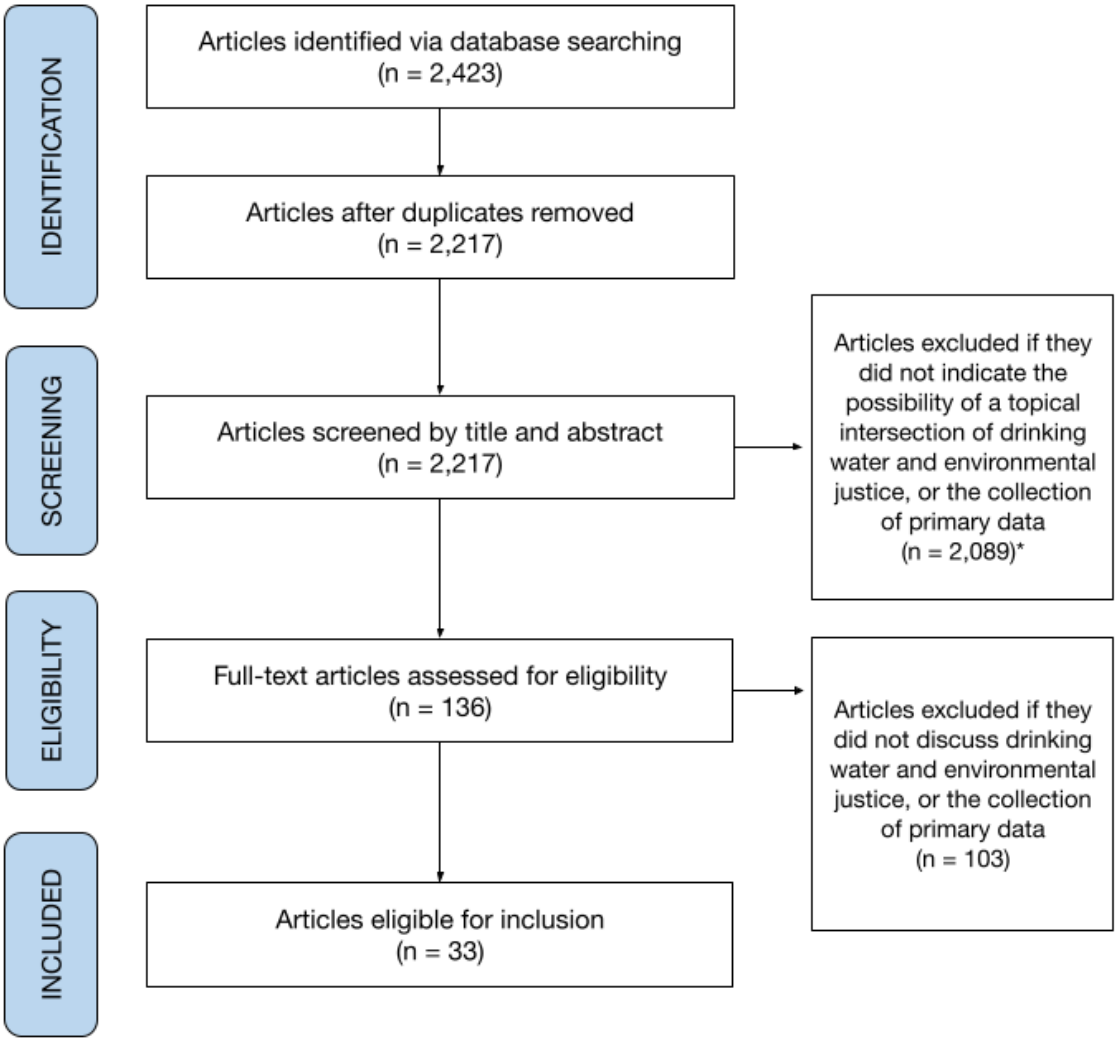
Wutich, Amber, Jessica Budds, Laura Eichelberger, et al. (Nov. 2017). "Advancing methods for research on household water insecurity: Studying entitlements and capabilities, socio-cultural dynamics, and political processes, institutions and governance". In: *Water security* 2, pp. 1–10. ISSN: 2468-3124. DOI: 10.1016/j.wasec.2017.09.001. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5844501/ (visited on 03/23/2022).

Wutich, Amber, Jessica Budds, Wendy Jepson, et al. (2018). "Household water sharing: A review of water gifts, exchanges, and transfers across cultures". In: *WIREs Water* 5.6, e1309. ISSN: 2049-1948. DOI: 10.1002/wat2.1309. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1309 (visited on 02/19/2022).

Young, Iris Marion (1989). "The Ideal of Community and the Politics of Difference". In: *Feminism/Postmodernism*. Routledge. ISBN: 978-0-203-69938-6.

Zeitoun, Mark (2013). "Global environmental justice and international transboundary waters: an initial exploration". In: *The Geographical Journal* 179.2, pp. 141–149. ISSN: 1475-4959. DOI: 10.1111/j.1475-4959.2012.00487.x. URL: https://rgs-ibg.onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-4959.2012.00487.x (visited on 09/01/2021).

Figure 2.1: Literature search flow chart (additional details in Figures S1 and S2).

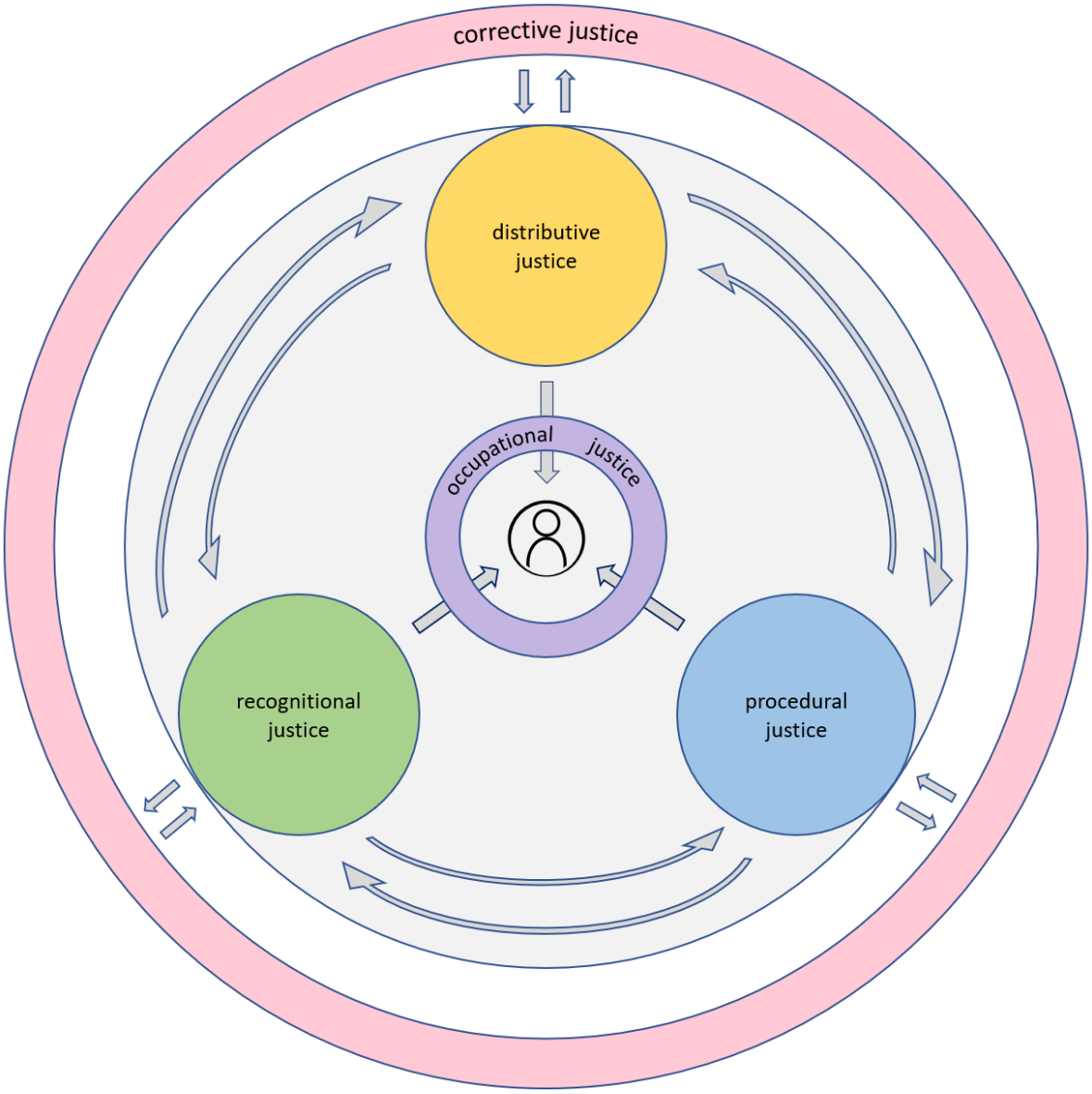


* while primary data collection was technically also a criteria of inclusion at this screening phase, we found that studies often do not elaborate on their data collection methods in their titles or abstracts; as a result, we opted to err on the side of inclusion at this step.

Figure 2.2: Distribution of drinking water studies (blue) and author affiliations (red).



Figure 2.3: Diagram of the multidirectional relationships between different types of justice.



Chapter 3

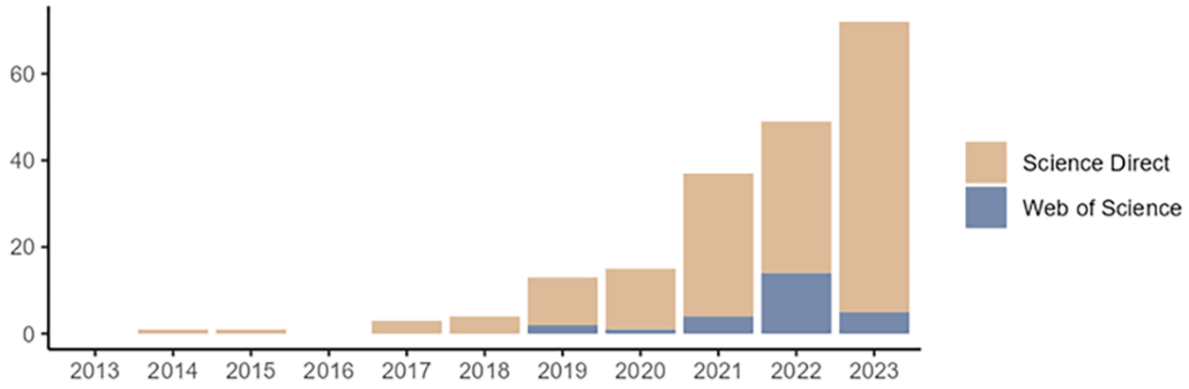
Environmental Justice is at a crossroads with machine learning[1]

Machine learning (ML) applications in environmental science (Zhong et al. 2021), monitoring, and regulatory enforcement (Hino, E. Benami, and Brooks 2018) range from estimating environmental pollution (Liu et al. 2022) to predicting future climate change (Rolnick et al. 2023) and sea level rise (Nieves, Radin, and Camps-Valls 2021). Allusions to justice implications of research findings are also common (Lobo, Laraway, and Gadgil 2022). The logic bridging ML to EJ – which is often implied and not explicitly elevated – is the premise that gaps in environmental data are themselves an injustice. Filling in such gaps through predictive modeling therefore nudges the needle towards a more just society by moving regulatory science beyond the assumption of "no data, no problem." While greater access to information can be desirable (with important caveats regarding consent, surveillance, and extractive data collection (Vera et al. 2019)), I aver that algorithms, which can inform EJ strategies in regulation and policy, leave the door open for unjust outcomes when unvetted. Thus, algorithms must be assessed to check for biases that could undermine EJ goals, and the broader connection between ML and environmentally just outcomes needs to be better elucidated both in theory and in practice.

Environmental scholarship has yet to fully assess the problem of bias and unfairness in ML. Studies in other fields (Mehrabi et al. 2021) have shown how algorithms can enforce discriminatory practices, perpetuate structural racism, and worsen disparities (Benjamin 2019). One example is how a widely-used algorithm in healthcare decision-making systematically underestimates health risks for Black patients with implications for racially-biased healthcare provision (Obermeyer et al. 2019); another is the perpetuation of negative stereotypes of people (and in particular, women) of color by search engines (Noble 2018). In reaction to such findings, some researchers have called for deliberately-antiracist positioning in approaches to research by those designing algorithms (Owens and Walker 2020). There is a

[1]This chapter is currently in peer-review.

Figure 3.1: Research articles with terms "environmental justice" and "machine learning" in two research databases.



growing consciousness among environmental scientists that ML can cause more harm than good without proper care (Hanson et al. 2023; McGovern et al. 2022), but the broader contour of recently published literature indicates this has yet to take hold.

The lack of engagement with this polemical topic within the environmental sciences is consequential because upstream modeling decisions can determine where or for whom models perform well. Take, for example, the standard practice of hyperparameter tuning, a systematic method of choosing modeling parameters that produce the "best performing" outcomes (Yang and Shami 2020). This involves running models over multiple iterations and using a specified metric to compare model performance. This process is inherently value-laden, as it requires the modeler to define the type of outcome they are optimizing for. A classification model optimized for overall accuracy will perform differently from models optimized for other metrics such as precision (minimizing false positives) or recall (minimizing false negatives) by design. The quality of input data can also affect performance, as models often perform better for groups or places that are relatively data-rich (Gao and Cui 2020). Some researchers have begun to explore distributional effects of algorithmic design (known as "allocative harm," as defined by Suresh and Guttag 2021), demonstrating how modeling decisions can alter modeling outcomes that in turn may condition environmental enforcement or funding (Elinor Benami et al. 2021; Huynh et al. 2024). This alone is reason enough to exercise more caution when using ML, but as it stands, there is no standard of practice for researchers to grapple with the EJ implications of their modeling decisions.

Until a such a standard of practice is established, I suggest the following four practices as a starting point for environmental research using machine learning methods with implications for environmental justice:

1. Embrace transparency as a quality requirement (Balasubramaniam et al. 2023), and take care to explain the reasoning behind data source selection, optimization processes, and parameter choice;

2. Check for bias and fairness in modeling outcomes (Obermeyer et al. 2019);

3. Explain the logic connecting machine learning outputs (e.g., prediction of air quality) to one or more aspects of environmental justice; and

4. Clearly communicate whether the modeling process factors justice implications into its design or parameter choices – as not all research is necessarily guided by principles of justice – to preempt misinterpretations and/or misapplications of findings.

The broader goal in suggesting these four practices together is to encourage more grounded conversations on how and why machine learning can help move us closer to an environmentally-just future. Engaging with specific aspects of justice – e.g., distributive, procedural, recognitional, or non-western indigenous frameworks – could help guide and validate research design and label choice. Drawing from the literature on community engagement in public health (e.g., community-based participatory research) to explore when or how to place the community "in the loop" is another important area of future research. The field at large would benefit from more precise definitions of justice that are tailored to local or regional contexts, rather than relying solely on broad, abstract, and universalizing descriptions of a just future.

Environmental justice is at a crossroads (Sze and London 2008) with ML, filled with promise and peril. The peril is not the use of ML itself, but when structural inequalities are not considered, the equity implications of modeling choices are obscured. As ML is used to fill data gaps, the irony is that data-scarce areas and communities are particularly at-risk of bias and unfairness, especially when there are systemic disparities in where, and for which communities, data are collected as part of regulatory monitoring and enforcement activities. Predictive algorithms do not solve the underlying, structural problems of power or voicelessness (Mah 2017). When called upon as a goal or motivation, EJ should not be restricted to being a buzzword in the conclusion section; it should have a place in the methods, too. From defining EJ (Holifield 2001; Schlosberg 2007) to characterizing the cumulative impacts of multiple environmental and social stressors (Morello-Frosch and Lopez 2006) and answering questions of 'which came first and why' (Been and Gupta 1997; Mohai and Saha 2015), EJ researchers sought to improve scientific methods to better characterize the community health and equity implications of multiple environmental hazards. As we wade into this age of machine learning, we must work to extend this framework of scientific reflexivity in how we use algorithmic approaches when there are implications for environmental justice.

References

Balasubramaniam, Nagadivya et al. (July 1, 2023). "Transparency and explainability of AI systems: From ethical guidelines to requirements". In: *Information and Software Technology* 159, p. 107197. ISSN: 0950-5849. DOI: 10.1016/j.infsof.2023.107197. URL: https://www.sciencedirect.com/science/article/pii/S0950584923000514 (visited on 04/09/2024).

Been, Vicki and Francis Gupta (1997). "Coming to the Nuisance or Going to the Barrios? A Longitudinal Analysis of Environmental Justice Claims". In: *Ecology Law Quarterly* 24.1, pp. 1–56. ISSN: 0046-1121. URL: https://www.jstor.org/stable/24113440 (visited on 04/09/2024).

Benami, Elinor et al. (Mar. 3, 2021). "The Distributive Effects of Risk Prediction in Environmental Compliance: Algorithmic Design, Environmental Justice, and Public Policy". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency. Virtual Event Canada: ACM, pp. 90–105. ISBN: 978-1-4503-8309-7. DOI: 10.1145/3442188.3445873. URL: https://dl.acm.org/doi/10.1145/3442188.3445873 (visited on 05/17/2022).

Benjamin, Ruha (Oct. 25, 2019). "Assessing risk, automating racism". In: *Science* 366.6464, pp. 421–422. DOI: 10.1126/science.aaz3873. URL: https://www.science.org/doi/full/10.1126/science.aaz3873 (visited on 03/17/2022).

Gao, Yan and Yan Cui (Oct. 12, 2020). "Deep transfer learning for reducing health care disparities arising from biomedical data inequality". In: *Nature Communications* 11.1, p. 5131. ISSN: 2041-1723. DOI: 10.1038/s41467-020-18918-3. URL: https://www.nature.com/articles/s41467-020-18918-3 (visited on 04/09/2024).

Hanson, Brooks et al. (Nov. 2023). "Garbage in, garbage out: mitigating risks and maximizing benefits of AI in research". In: *Nature* 623.7985, pp. 28–31. DOI: 10.1038/d41586-023-03316-8. URL: https://www.nature.com/articles/d41586-023-03316-8 (visited on 03/23/2024).

Hino, M., E. Benami, and N. Brooks (Oct. 2018). "Machine learning for environmental monitoring". In: *Nature Sustainability* 1.10, pp. 583–588. ISSN: 2398-9629. DOI: 10.1038/s41893-018-0142-9. URL: https://www.nature.com/articles/s41893-018-0142-9 (visited on 01/20/2022).

Holifield, Ryan (Feb. 1, 2001). "Defining Environmental Justice and Environmental Racism". In: *Urban Geography* 22.1, pp. 78–90. ISSN: 0272-3638. DOI: 10.2747/0272-3638.22.1.78. URL: https://doi.org/10.2747/0272-3638.22.1.78 (visited on 04/09/2024).

Huynh, Benjamin Q. et al. (Feb. 2024). "Mitigating allocative tradeoffs and harms in an environmental justice data tool". In: *Nature Machine Intelligence* 6.2, pp. 187–194. ISSN: 2522-5839. DOI: 10.1038/s42256-024-00793-y. URL: https://www.nature.com/articles/s42256-024-00793-y (visited on 03/21/2024).

Liu, Xian et al. (Feb. 15, 2022). "Data-Driven Machine Learning in Environmental Pollution: Gains and Problems". In: *Environmental Science & Technology* 56.4, pp. 2124–2133. ISSN:

0013-936X. DOI: 10.1021/acs.est.1c06157. URL: https://doi.org/10.1021/acs.est.1c06157 (visited on 04/09/2024).

Lobo, G. P., J. Laraway, and A. J. Gadgil (Jan. 10, 2022). "Identifying schools at high-risk for elevated lead in drinking water using only publicly available data". In: *Science of The Total Environment* 803, p. 150046. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2021.150046. URL: https://www.sciencedirect.com/science/article/pii/S0048969721051214 (visited on 01/14/2022).

Mah, Alice (Apr. 3, 2017). "Environmental justice in the age of big data: challenging toxic blind spots of voice, speed, and expertise". In: *Environmental Sociology* 3.2, pp. 122–133. ISSN: null. DOI: 10.1080/23251042.2016.1220849. URL: https://doi.org/10.1080/23251042.2016.1220849 (visited on 05/17/2022).

McGovern, Amy et al. (Jan. 2022). "Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science". In: *Environmental Data Science* 1, e6. ISSN: 2634-4602. DOI: 10.1017/eds.2022.5. URL: https://www.cambridge.org/core/journals/environmental-data-science/article/why-we-need-to-focus-on-developing-ethical-responsible-and-trustworthy-artificial-intelligence-approaches-for-environmental-science/8A32BECCE5FBB518F982855FAB6553D4 (visited on 03/23/2024).

Mehrabi, Ninareh et al. (July 13, 2021). "A Survey on Bias and Fairness in Machine Learning". In: *ACM Computing Surveys* 54.6, 115:1–115:35. ISSN: 0360-0300. DOI: 10.1145/3457607. URL: https://dl.acm.org/doi/10.1145/3457607 (visited on 04/09/2024).

Mohai, Paul and Robin Saha (Dec. 2015). "Which came first, people or pollution? A review of theory and evidence from longitudinal environmental justice studies". In: *Environmental Research Letters* 10.12, p. 125011. ISSN: 1748-9326. DOI: 10.1088/1748-9326/10/12/125011. URL: https://dx.doi.org/10.1088/1748-9326/10/12/125011 (visited on 04/09/2024).

Morello-Frosch, Rachel and Russ Lopez (Oct. 1, 2006). "The riskscape and the color line: Examining the role of segregation in environmental health disparities". In: *Environmental Research*. IG000012 102.2, pp. 181–196. ISSN: 0013-9351. DOI: 10.1016/j.envres.2006.05.007. URL: https://www.sciencedirect.com/science/article/pii/S0013935106001277 (visited on 05/03/2024).

Nieves, Veronica, Cristina Radin, and Gustau Camps-Valls (Apr. 7, 2021). "Predicting regional coastal sea level changes with machine learning". In: *Scientific Reports* 11.1, p. 7650. ISSN: 2045-2322. DOI: 10.1038/s41598-021-87460-z. URL: https://www.nature.com/articles/s41598-021-87460-z (visited on 04/09/2024).

Noble, Safiya Umoja (Feb. 20, 2018). "Algorithms of Oppression: How Search Engines Reinforce Racism". In: *Algorithms of Oppression*. New York University Press. ISBN: 978-1-4798-3364-1. URL: https://www.degruyter.com/document/doi/10.18574/nyu/9781479833641.001.0001/html (visited on 04/09/2024).

Obermeyer, Ziad et al. (Oct. 25, 2019). "Dissecting racial bias in an algorithm used to manage the health of populations". In: *Science* 366.6464, pp. 447–453. DOI: 10.1126/science.

aax2342. URL: https://www.science.org/doi/full/10.1126/science.aax2342 (visited on 03/18/2022).

Owens, Kellie and Alexis Walker (Sept. 2020). "Those designing healthcare algorithms must become actively anti-racist". In: *Nature Medicine* 26.9, pp. 1327–1328. ISSN: 1546-170X. DOI: 10.1038/s41591-020-1020-3. URL: https://www.nature.com/articles/s41591-020-1020-3 (visited on 04/22/2023).

Rolnick, David et al. (Mar. 31, 2023). "Tackling Climate Change with Machine Learning". In: *ACM Computing Surveys* 55.2, pp. 1–96. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3485128. URL: https://dl.acm.org/doi/10.1145/3485128 (visited on 05/02/2022).

Schlosberg, David (May 18, 2007). *Defining Environmental Justice: Theories, Movements, and Nature*. OUP Oxford. 253 pp. ISBN: 978-0-19-153671-7.

Suresh, Harini and John Guttag (Nov. 4, 2021). "A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle". In: *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. EAAMO '21. New York, NY, USA: Association for Computing Machinery, pp. 1–9. ISBN: 978-1-4503-8553-4. DOI: 10.1145/3465416.3483305. URL: https://dl.acm.org/doi/10.1145/3465416.3483305 (visited on 04/09/2024).

Sze, Julie and Jonathan K. London (2008). "Environmental Justice at the Crossroads". In: *Sociology Compass* 2.4, pp. 1331–1354. ISSN: 1751-9020. DOI: 10.1111/j.1751-9020.2008.00131.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-9020.2008.00131.x (visited on 04/09/2024).

Vera, Lourdes A. et al. (June 7, 2019). "When data justice and environmental justice meet: formulating a response to extractive logic through environmental data justice". In: *Information, Communication & Society* 22.7, pp. 1012–1028. ISSN: 1369-118X. DOI: 10.1080/1369118X.2019.1596293. URL: https://doi.org/10.1080/1369118X.2019.1596293 (visited on 01/18/2022).

Yang, Li and Abdallah Shami (Nov. 20, 2020). "On hyperparameter optimization of machine learning algorithms: Theory and practice". In: *Neurocomputing* 415, pp. 295–316. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2020.07.061. URL: https://www.sciencedirect.com/science/article/pii/S0925231220311693 (visited on 04/09/2024).

Zhong, Shifa et al. (Oct. 5, 2021). "Machine Learning: New Ideas and Tools in Environmental Science and Engineering". In: *Environmental Science & Technology* 55.19, pp. 12741–12754. ISSN: 0013-936X. DOI: 10.1021/acs.est.1c01339. URL: https://doi.org/10.1021/acs.est.1c01339 (visited on 04/09/2024).

Chapter 4

Using machine learning to predict drinking water quality in California[1]

4.1 Background

Applications of machine learning (ML) for research on environmental science and management range from regulatory enforcement (Hino, E. Benami, and Brooks 2018) and climate science (Huntingford et al. 2019) to planetary-scale land cover change (Gorelick et al. 2017). While contexts vary, the throughline of these studies is the use of ML to extrapolate and interpolate data to spaces or times where information is otherwise unavailable.

The potential social benefits of machine learning and artificial intelligence have been discussed extensively (Tomašev et al. 2020), but they come with causes for concern regarding embedded bias and unintended consequences. Studies in public health and elsewhere have shown algorithms to codify discrimination, exacerbate systematic inequalities, and reinforce racist outcomes (Benjamin 2019), which has prompted calls for more proactive and anti-racist approaches to research (Owens and A. Walker 2020). With the growing popularity of ML in environmental science and management, it is important for environmental researchers to learn from and build upon reflections made in other fields. This chapter builds a growing body of literature on algorithms and environmental justice (EJ) (Elinor Benami et al. 2021; Hanson et al. 2023; Huynh et al. 2024; McGovern et al. 2022), and extends the work in this area with a two-tiered approach of designing and dissecting a ML approach to EJ, paired with a discussion on the EJ implications of optimization and model design.

This chapter consists of two parts. In the first half, we present a case study of using ML for EJ-motivated research: prediction of drinking water quality. In the second half, we dissect this algorithmic approach to examine if and how our modeling decisions affected

[1]These next two chapters are currently in peer-review as one manuscript, and were co-authored by Rachel Morello-Frosch and Duncan Callaway. I am grateful for their contributions to this work, and for their permission to use this research as a part of my dissertation.

modeling outcomes – and not only how these decisions change whether the model is right or wrong, but for whom. We find that multiple decision points in the modeling process – from choosing classifiers, to labels, to input data – can lead to outcomes that differ in consequential ways. More importantly, we find that these choices can result in significant differences in demographic characteristics of false negatives compared to true positives (i.e., who the model is wrong about). We conclude by proposing a set of considerations for researchers to both follow and improve upon when using ML to address questions of environmental science, management, and justice. If justice is embedded in policy and regulatory decision-making (e.g., as signaled by US Federal Executive Order 12898), it is critical to be transparent about how justice priorities are considered in the research that is used to inform these policies. This holds particularly true as government agencies are looking towards wider adoption of ML-based approaches to address data gaps and to undertake anticipatory approaches in regulatory science (Veale and Brass 2019; Wirtz, Weyerer, and Geyer 2019).

4.2 A case study for machine learning: drinking water quality in California

In the United States (US), drinking water quality standards are set by the US Environmental Protection Agency (EPA) under the Safe Drinking Water Act (SDWA) and its amendments. Most US residents obtain their drinking water from community water systems (CWS), or public water systems that serve at least 25 people or 15 service connections year-round (Agency 2015). CWS must adhere to these federal drinking water standards and are required to regularly monitor and report their water quality to designated regulatory agencies; states, in turn, are required by the 1996 SDWA amendments to assist water systems acquire and maintain the technical, managerial, and financial capacity to comply (Soelter and Miller 1999). Water systems receive a maximum contaminant level (MCL) violation from the state if their water samples repeatedly exceed a contaminant-specific threshold set by the EPA. Water systems receive a monitoring and reporting (MR) violation if they fail to monitor and report their water quality to the state according to a regulatory schedule (Glade and Ray 2022).

In the case of a water quality violation, the next steps are relatively straightforward (though resolution can still take a very long time in practice). For example, water systems must issue boil water notices in the case of bacterial contamination, and there are state programs in place to fund interim replacement drinking water delivery or point-of-use treatment systems (California State Water Resources Control Board 2021b). In contrast, the implications of MR violations are less apparent; they can obscure public health risks as the water quality of these systems are unknown, possibly masking more serious health risks (Fedinick et al. 2017). Furthermore, an earlier report by the US Government Accountability Office found that monitoring and reporting violations are strong and significant predictors for health-based violations (Office 2011).

There is a substantial body of literature using environmental justice – the principle that "all people and communities are entitled to equal protection of environmental and public health laws and regulations" (Bullard 1993) – as a lens through which to understand the distribution of environmental hazards and burdens, including access to drinking water (and lack thereof). We consider monitoring and reporting violations as one such concern, particularly because the relationship between drinking water problems, race, and class are well-documented in California and across the US (Balazs et al. 2012; McDonald and Jones 2018; Pace et al. 2022). The ability to predict water quality in public systems could therefore help facilitate proactive measures to advance distributive environmental justice, especially for drinking water systems that have fallen out of monitoring and reporting compliance. The expediency with which ML algorithms can flag at-risk systems without the typical costs and delays associated with water sampling can also help bridge inter-system gaps in technical, managerial, and financial capacity. The specific aspects of distributive environmental justice that this chapter operationalizes are (i) access to clean, safe, and affordable drinking water is uneven; (ii) there are differences in how this uneven landscape of drinking water is monitored and reported; and (iii) these differences in monitoring and reporting have implications for nonconsensual, persistent exposures to environmental health hazards.

Researchers have applied ML as an approach to predict water quality with regards to unregulated or under-sampled contaminants or places, such as arsenic in private wells (Lombard et al. 2021) or lead in schools (Lobo, Laraway, and Gadgil 2022). To our knowledge, no one has applied ML to address lapses in monitoring and reporting or looked at how upstream decisions in the modeling process affect the *demographic* outcomes of classification exercises. In this chapter, we train a suite of classification algorithms to predict drinking water quality using historical drinking water sampling data, spatially-derived agricultural and geological information, and demographic indicators as inputs. In addition to assessing out-of-sample performance using a test set, we apply these models to predict sampling outcomes for systems following a MR violation. (We set aside data on systems that previously received MR violations from the training process to use as an additional test set.) We then "dissect" these models to evaluate the impacts of our upstream modeling decisions, such as model or tuning parameter choice, on predicted outcomes, including different demographic distributions across the four possible classification outcomes (true positive, true negative, false positive, false negative). We conclude with recommendations on how to minimize this type of bias, emphasizing the importance of both transparency and connecting intention to methods in justice-motivated research.

4.3 Methods

Data collection

We chose the state of California as a case study for its availability of water data and persistence of drinking water issues. We conducted our analysis using R version 4.3.1 (Team

2023). We extracted all the data in our analysis from publicly available sources. (See Appendix Table B.1 for the data sources of our response and feature variables.) The following packages were integral to our analysis: tidycensus (K. Walker and Herman 2024) and soilDB (Beaudette et al. 2024) for census and soil data; exactextractr (Baston 2023) and sf (Pebesma 2018) for spatial methods; and the tidymodels (Kuhn and Wickham 2023) 'meta-package' for machine learning implementation. Our process is designed to be reproducible, though results may change with future updates to available data, such as water system boundaries.

Water system data

Our spatial unit of analysis is the community water system, using data on system characteristics (California State Water Resources Control Board 2019a) and spatial boundaries (California State Water Resources Control Board 2020) provided by the State Water Resources Control Board (SWRCB) Division of Drinking Water (DDW). We used water system fee codes to select for public water systems listed as "Community" (systems that serve water to 25 year-round residents or through 15 service connections to year-round residents) or "State Small" (systems that serve water to less than 25 year-round residents through five to 14 service connections), and which aren't selling (i.e., aren't wholesalers) or purchasing their water. As a measure of quality control, we overlaid the water system boundaries with a high-resolution 100-meter population grid of California (Depsky, Cushing, and Morello-Frosch 2022) to identify those system boundaries that are likely to be inaccurate – either systems for which our derived population estimates are under 25 and thus below the population threshold for community water systems, or less than 15% of the population listed online on the Drinking Water Watch (a public drinking water database for California).

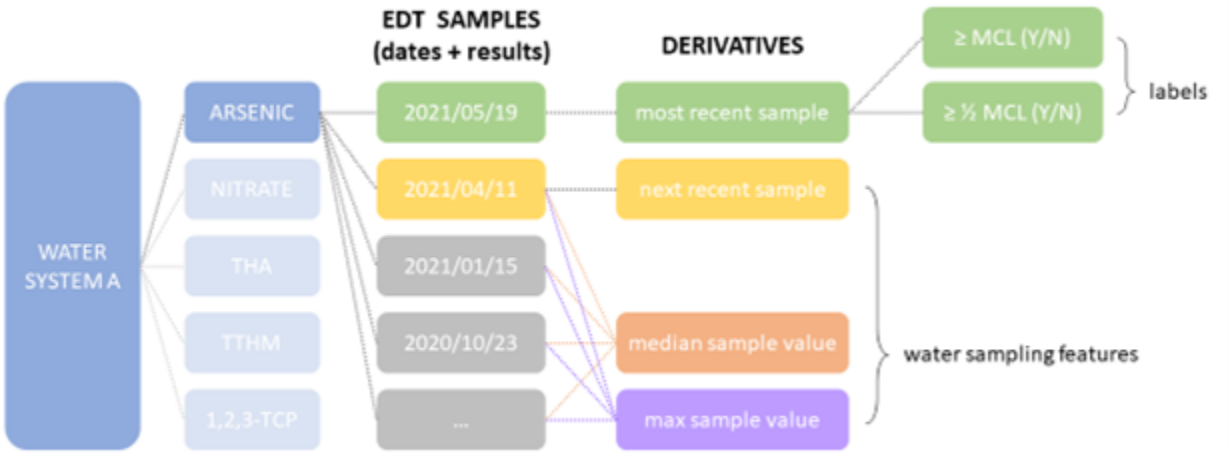
Water sampling data

We pulled historical and current drinking water violations and sampling data from the DDW's Electronic Data Transfer (EDT) library for water systems with available spatial boundaries (California State Water Resources Control Board 2024). The EDT is an online repository of water sampling data collected by water systems and analyzed by laboratories extending back to 1973. We downloaded data for five drinking water contaminants: arsenic, nitrate, total trihalomethanes (TTHM), haloacetic acids (THA), and 1,2,3-trichloropropane (1,2,3-TCP). We chose these contaminants for their public health concern and the prevalence of samples over the MCL threshold. (See Appendix Figure B.1 for histograms of sampling counts for each contaminant over our study period.) We created a binary (0/1) column to use as our response variable, on whether the most recent sample for each contaminant taken at each water system was at/above (1) or below (0) that contaminant's MCL threshold. We then calculated the following summary statistics using all samples excluding the most recent: the total number of samples; the total number of samples over the contaminant threshold (e.g., MCL); the ratio of samples over the threshold to the number of all samples taken; and the mean, median, and max sample values. Figure 4.1 illustrates how we derived label

and feature values from the EDT dataset. This analysis does not calculate or account for running averages, which are occasionally utilized in establishing MCL thresholds.

Next, we took annual compliance reporting data across all available years and identified water systems that received at least one monitoring and reporting-related violation between 2013 and 2021. We defined the "most recent sample" of these systems (the outcome we are predicting) as the sample taken immediately following their most recent monitoring and reporting violation. We excluded any sampling data collected after the "most recent," and calculated summary statistics for all preceding sampling data following the steps explained above. (See Appendix Figure S2 for an illustration of this process.)

Figure 4.1: Process for deriving label and feature values from EDT data for a given water system and contaminant.



We also supplemented EDT data with data from the United States Geological Survey's Groundwater Ambient Monitoring and Assessment (GAMA) program (California State Water Resources Control Board 2019b), a state program implemented to construct a more comprehensive understanding of groundwater quality in California. We took the ten nearest public supply wells to each water systems' boundaries by Euclidean distance, then calculated the following summary statistics for our five selected contaminants: total sample count, and the average, median, and max contaminant concentration values.

Demographic data

We estimated water system demographic characteristics as water systems do not provide data on their service area demographics and water system boundaries can transect census boundaries. If a water system fell entirely within one block group, we used that block group's

data to approximate that water system's demographics. If a water system encompassed multiple block groups, we used our derived population estimates (see above) to calculate the ratio of each block group's population to the total population in the water system service area. We then used this ratio to apportion weights to each block group's demographic profile, for sixteen demographic indicators – on race, education, income, employment, and housing – frequently associated with social vulnerability in the environmental justice literature. (See Appendix Table B.3 for a full list of these indicators.)

Land-use data

We used the water system boundaries to spatially extract land-use characteristics in and around water systems. We obtained crop-related data from the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) Cropland Data Layer (United States Department of Agriculture 2021), land-use data from the USGS National Land Cover Database (NLCD) (Service 2019), animal feeding operation information from SWRCB California Integrated Water Quality System Project (California State Water Resources Control Board 2021a), and soil data from the USDA Natural Resources Conservations Service (NRCS) Soil Survey Geographic Database (Agriculture Natural Resources Conservation Service 2023). For each of these land-use characteristics, we estimated coverage in and around water systems by counting points with a three-kilometer buffer and calculating relative proportions of coverage using pixels and area.

Data preparation

We structured our final data frame such that each row represents a water system-contaminant pair, with columns of data on water sampling history, approximated water system demographics, and nearby land-use. Each row has different values for its water sampling-related columns, but other columns may have the same values across rows within a water system. For example, water sampling data would differ for each of Water System A's rows (i.e., sampling data would differ by contaminant), but the demographic variables for Water System A would stay the same. Prior to running our analysis, we transformed our data in the following ways: dropped highly correlated features and features with low variance, coerced low frequency missing values to zero , created dummies for categorical variables, and normalized our data.

We used these input variables to train and compare the performance of a suite of models to predict the likelihood of whether the next sample taken by a community water system for a given contaminant would be predicted to exceed that contaminant's MCL.

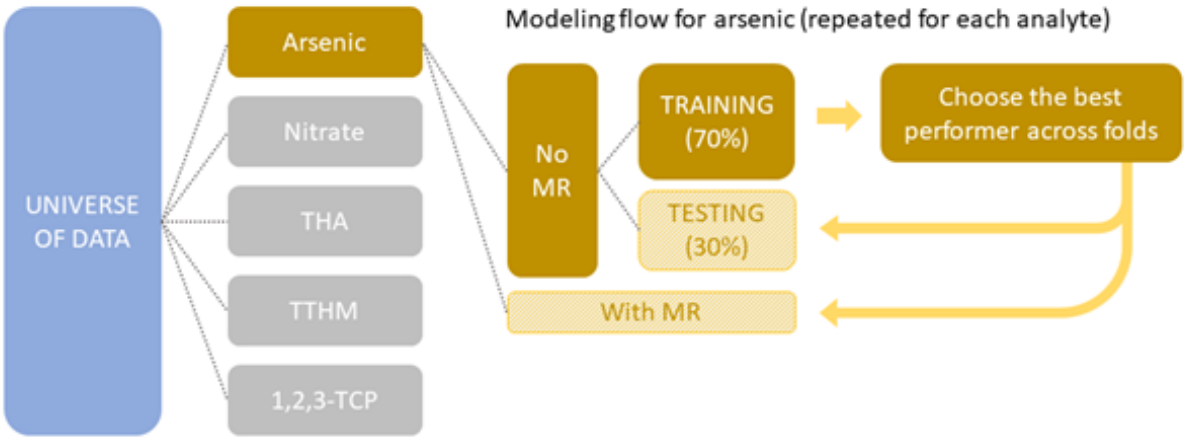
Model training and selection

As explained above, MR-violating systems are the primary type of system we are interested in predicting. As we processed our data, however, it became clear we did not have sufficient

data to train models solely using MR-violating systems (there are too few), nor were we able to meaningfully assess whether MR-violating systems materially differ from non-violating systems. We therefore decided to exclude MR-violating systems from the training process altogether and to use them instead as a secondary test set. We evaluated the goodness of fit of our models for these excluded systems by measuring their ability to classify samples taken immediately following MR violations (i.e., the next sample identified by date). We understood the risk that models trained on non-violating systems may be unable to accurately predict water quality for MR-violating systems (i.e., if MR-violating systems are fundamentally different than non-violating systems). If this were the case, we would have had to reconsider our approach, or accept and make explicit that these models should not be used to predict samples for systems out of monitoring and reporting compliance.

We divided our cleaned dataset of MR-compliant systems by contaminant, then randomly partitioned each into training (70%) and testing (30%) sets (see Figure 4.2). We stratified this train-test split using counts of water samples above the MCL. (See Appendix Table B.6 for a breakdown of these counts across train-test splits for each contaminant.) We separated our dataset by contaminant type for two reasons; first, EPA regulatory requirements related to sampling frequency and maximum contaminant levels are contaminant specific, and second, these contaminants are introduced into the water through different pathways, making it reasonable to expect that the weights of input variables might vary across contaminants.

Figure 4.2: A flowchart of our data subsetting process, using arsenic as an example.



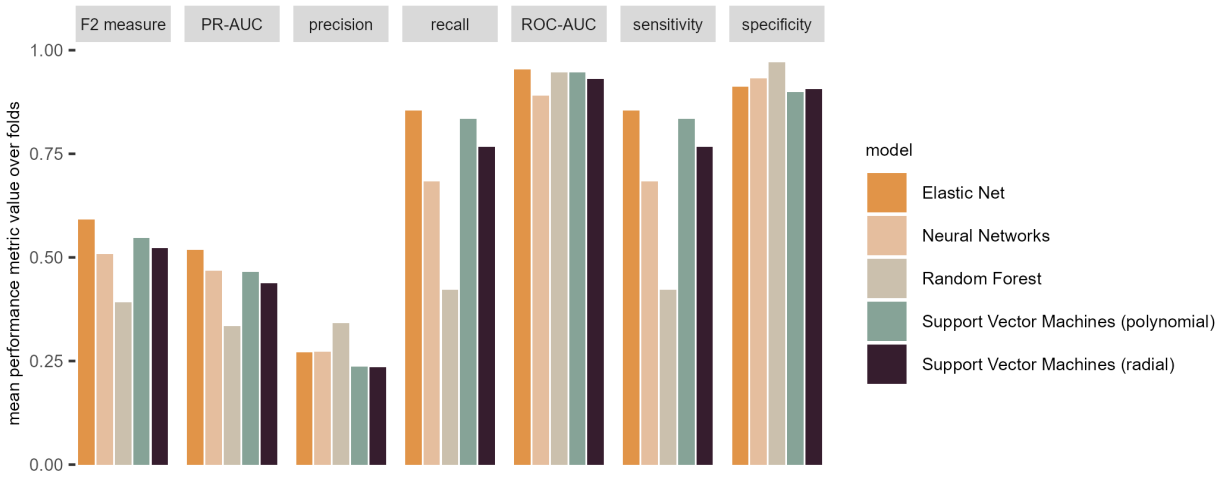
We compared the performance of five classification models with our prepared data: elastic net regularization, random forest, polynomial and radial kernel support vector machines (SVM), and neural networks. We chose these models for their abilities to handle imbalanced classification problems and their prevalence in the literature. We performed a minimal

amount of tuning for each with k-fold cross validation (k=7) before comparing outputs, using Synthetic Minority Oversampling Technique (SMOTE) to account for the imbalanceness of the response variable. Aside from the correct identification of "positive" systems, we were concerned about the environmental justice and health implications of continued exposure due to incorrect negative predictions (i.e., false negatives). We thus chose to use the F2 score (calculated using the equation for F_β shown below, with $\beta = 2$) for model evaluation and selection to reduce the rate of false negatives.

$$F_\beta score = (1 + \beta^2)\frac{precision * recall}{\beta^2 * precision + recall}$$

Relative model performance rankings differed across performance metrics. Figure 4.3 presents the mean values of each metric for each model over cross-validation folds in the training process.

Figure 4.3: Example of how metric choice can affect evaluations of model performance (contaminant: arsenic).



The ROC-AUC is one of the most used metrics for classification model tuning and selection, but it can be misleadingly optimistic (i.e., scores are high relative to other metrics) for evaluating performance on imbalanced datasets (Saito and Rehmsmeier 2015), as exemplified above. This highlights the importance of metric-selection, which we discuss later in the chapter.

Assumptions and limitations

There are several key assumptions and limitations in our approach. First, EDT data does not include information on whether sample values represent post- or pre-treatment water quality. This means that the predicted outcomes should not be equated to *exposure* to that predicted water quality. Second, our data is aggregated at the water system level, but water quality sampling data is available at the source level for recent years (2012 onward). Using sources *within* systems as the modeling unit would be an interesting follow-up to this study. Third, we assume the spatial integrity of water system boundaries for our analysis. While many contributors have put a significant amount of effort into cleaning up the available data, the boundaries are imperfect.

4.4 Results

Model performance varied by contaminant. (See Appendix Figure B.3 for a comparison of F2 scores.) No single classifier outperformed all others across all contaminants. Elastic net and both SVMs (radial and polynomial) consistently performed well, while random forest consistently performed poorest. To keep our results concise, we focus on one contaminant (arsenic) and model (elastic net) in the main body of the chapter. (See Appendix Figure S3 for results for the other contaminants for elastic net.)

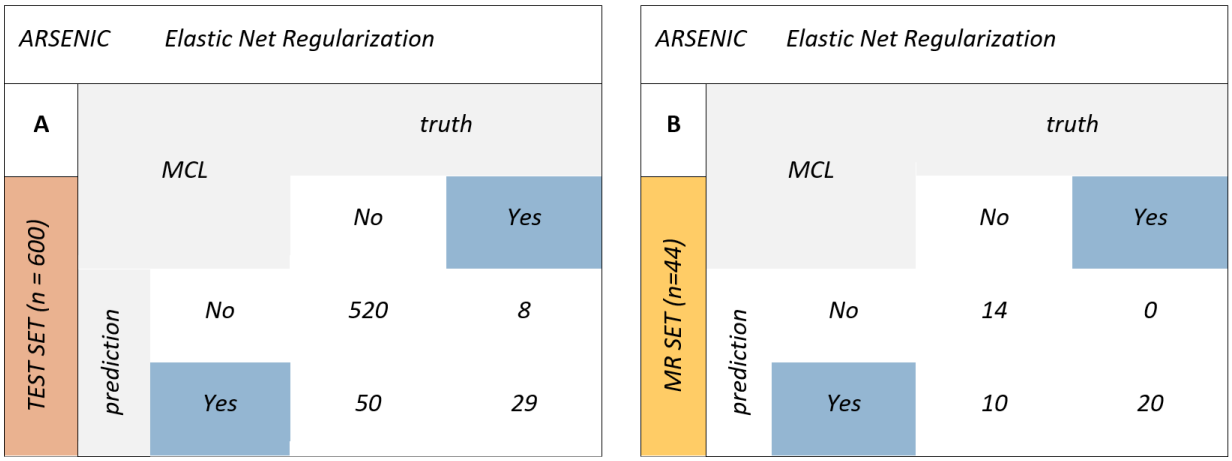
Figure 4.4 shows the result of applying a tuned elastic net regularization model for predicting arsenic on our two test sets (the test set – 30% of the data – on the left, and the set-aside MR systems on the right). We produced these confusion matrices using strict class predictions, meaning that the model translated probabilities into predictions by comparing the probability of each system belonging to the "over MCL" class to a default threshold of 0.5.

Figure 4.4A (classification outcomes for the test set) indicates 93% accuracy overall for elastic net, with a precision of 0.39 (proportion of correct guesses of "positive" over all guesses of "positive"), recall of 0.82 (proportion of correct guesses of "positive" over all occurrences of "positive"), and F2 score of 0.67. The model performed even better when predicting the next sample for MR violating systems (Figure 4.4B): it correctly identified all water systems with a most recent sample (i.e. the next sample taken after an MR violation) that exceeded the MCL, with a F2 score of 0.91.

Implications of using ML for water quality predictions

Through our exercise above, we demonstrate how a machine learning approach can accurately predict drinking water quality. Notably, models performed well not only for systems in compliance with monitoring and reporting regulations, but also for systems that "just" received a monitoring and reporting violation (as a reminder: we simulated this by excluding all sampling data taken after MR violations). In theory, these predictions could help

Figure 4.4: Example of how metric choice can affect evaluations of model performance (contaminant: arsenic).



public agencies efficiently and effectively distribute their resources in terms of monitoring and drinking water regulation enforcement activities.

References

Agency, United States Environmental Protection (Sept. 21, 2015). *Information about Public Water Systems*. URL: https://www.epa.gov/dwreginfo/information-about-public-water-systems (visited on 03/21/2024).

Agriculture Natural Resources Conservation Service, United States Department of (2023). *Soil Survey Geographic Database (SSURGO)*. URL: https://www.nrcs.usda.gov/resources/data-and-reports/soil-survey-geographic-database-ssurgo.

Balazs, Carolina L. et al. (Nov. 14, 2012). "Environmental justice implications of arsenic contamination in California's San Joaquin Valley: a cross-sectional, cluster-design examining exposure and compliance in community drinking water systems". In: *Environmental Health* 11.1, p. 84. ISSN: 1476-069X. DOI: 10.1186/1476-069X-11-84. URL: https://doi.org/10.1186/1476-069X-11-84 (visited on 04/25/2023).

Baston, Daniel (2023). *exactextractr: Fast Extraction from Raster Datasets using Polygons*. Version R package version 0.10.0. URL: https://CRAN.R-project.org/package=exactextractr.

Beaudette, Dylan et al. (2024). *soilDB: Soil Database Interface*. Version R package version 2.8.1. URL: https://CRAN.R-project.org/package=soilDB.

Benami, Elinor et al. (Mar. 3, 2021). "The Distributive Effects of Risk Prediction in Environmental Compliance: Algorithmic Design, Environmental Justice, and Public Policy". In:

Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency. Virtual Event Canada: ACM, pp. 90–105. ISBN: 978-1-4503-8309-7. DOI: 10.1145/3442188.3445873. URL: https://dl.acm.org/doi/10.1145/3442188.3445873 (visited on 05/17/2022).

Benjamin, Ruha (Oct. 25, 2019). "Assessing risk, automating racism". In: *Science* 366.6464, pp. 421–422. DOI: 10.1126/science.aaz3873. URL: https://www.science.org/doi/full/10.1126/science.aaz3873 (visited on 03/17/2022).

Bullard, Robert D. (1993). "The Legacy of American Apartheid and Environmental Racism". In: *St. John's Journal of Legal Commentary* 9, p. 445. URL: https://heinonline.org/HOL/Page?handle=hein.journals/sjjlc9&id=457&div=&collection=.

California State Water Resources Control Board (2019a). *Drinking Water - Public Water System Information*. URL: https://data.ca.gov/dataset/drinking-water-public-water-system-information (visited on 03/22/2024).

— (2019b). *Ground Water - Water Quality Results - California Open Data*. URL: https://data.ca.gov/dataset/ground-water-water-quality-results (visited on 03/22/2024).

— (2020). *California Drinking Water System Area Boundaries*. URL: https://gis.data.ca.gov/datasets/fbba842bf134497c9d611ad506ec48cc.

— (2021a). *California Integrated Water Quality System Project - Interactive Regulated Facilities Report*. URL: https://www.waterboards.ca.gov/water_issues/programs/ciwqs/publicreports.html.

— (2021b). *SAFER Drinking Water*. URL: https://www.waterboards.ca.gov/safer/ (visited on 03/23/2024).

— (2024). *EDT Library and Water Quality Analyses Data and Download Page*. URL: https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/EDTlibrary.html.

Depsky, Nicholas J., Lara Cushing, and Rachel Morello-Frosch (July 14, 2022). "High-resolution gridded estimates of population sociodemographics from the 2020 census in California". In: *PLOS ONE* 17.7, e0270746. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0270746. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270746 (visited on 07/20/2022).

Fedinick, Kristi Pullen et al. (2017). *Threats on tap: widespread violations highlight need for investment in water infrastructure and protections*. National Resources Defense Council, pp. 1–26. URL: https://www.nrdc.org/sites/default/files/threats-on-tap-water-infrastructure-protections-report.pdf.

Glade, Sara and Isha Ray (Mar. 2022). "Safe drinking water for small low-income communities: the long road from violation to remediation". In: *Environmental Research Letters* 17.4, p. 044008. ISSN: 1748-9326. DOI: 10.1088/1748-9326/ac58aa. URL: https://dx.doi.org/10.1088/1748-9326/ac58aa (visited on 03/21/2024).

Gorelick, Noel et al. (Dec. 1, 2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone". In: *Remote Sensing of Environment*. Big Remotely Sensed Data: tools,

applications and experiences 202, pp. 18–27. ISSN: 0034-4257. DOI: 10.1016/j.rse.2017.06.031. URL: https://www.sciencedirect.com/science/article/pii/S0034425717302900 (visited on 03/24/2024).

Hanson, Brooks et al. (Nov. 2023). "Garbage in, garbage out: mitigating risks and maximizing benefits of AI in research". In: *Nature* 623.7985, pp. 28–31. DOI: 10.1038/d41586-023-03316-8. URL: https://www.nature.com/articles/d41586-023-03316-8 (visited on 03/23/2024).

Hino, M., E. Benami, and N. Brooks (Oct. 2018). "Machine learning for environmental monitoring". In: *Nature Sustainability* 1.10, pp. 583–588. ISSN: 2398-9629. DOI: 10.1038/s41893-018-0142-9. URL: https://www.nature.com/articles/s41893-018-0142-9 (visited on 01/20/2022).

Huntingford, Chris et al. (Nov. 2019). "Machine learning and artificial intelligence to aid climate change research and preparedness". In: *Environmental Research Letters* 14.12, p. 124007. ISSN: 1748-9326. DOI: 10.1088/1748-9326/ab4e55. URL: https://dx.doi.org/10.1088/1748-9326/ab4e55 (visited on 03/24/2024).

Huynh, Benjamin Q. et al. (Feb. 2024). "Mitigating allocative tradeoffs and harms in an environmental justice data tool". In: *Nature Machine Intelligence* 6.2, pp. 187–194. ISSN: 2522-5839. DOI: 10.1038/s42256-024-00793-y. URL: https://www.nature.com/articles/s42256-024-00793-y (visited on 03/21/2024).

Kuhn, Max and Hadley Wickham (2023). *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles*. URL: https://www.tidymodels.org.

Lobo, G. P., J. Laraway, and A. J. Gadgil (Jan. 10, 2022). "Identifying schools at high-risk for elevated lead in drinking water using only publicly available data". In: *Science of The Total Environment* 803, p. 150046. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2021.150046. URL: https://www.sciencedirect.com/science/article/pii/S0048969721051214 (visited on 01/14/2022).

Lombard, Melissa A. et al. (Apr. 20, 2021). "Machine Learning Models of Arsenic in Private Wells Throughout the Conterminous United States As a Tool for Exposure Assessment in Human Health Studies". In: *Environmental Science & Technology* 55.8, pp. 5012–5023. ISSN: 0013-936X. DOI: 10.1021/acs.est.0c05239. URL: https://doi.org/10.1021/acs.est.0c05239 (visited on 04/27/2023).

McDonald, Yolanda J. and Nicole E. Jones (Oct. 2018). "Drinking Water Violations and Environmental Justice in the United States, 2011–2015". In: *American Journal of Public Health* 108.10, pp. 1401–1407. ISSN: 0090-0036. DOI: 10.2105/AJPH.2018.304621. URL: https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2018.304621 (visited on 04/25/2023).

McGovern, Amy et al. (Jan. 2022). "Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science". In: *Environmental Data Science* 1, e6. ISSN: 2634-4602. DOI: 10.1017/eds.2022.5. URL: https://www.cambridge.org/core/journals/environmental-data-science/article/why-we-need-to-focus-on-developing-ethical-responsible-

and-trustworthy-artificial-intelligence-approaches-for-environmental-science/8A32BECCE5FBB518F982855FAB6553D4 (visited on 03/23/2024).

Office, United States Government Accountability (2011). *Unreliable State Data Limit EPA's Ability to Target Enforcement Priorities and Communicate Water Systems' Performance*. Report to Congressional Requesters. United States Government Accountability Office, pp. 1–74. URL: https://www.gao.gov/assets/gao-11-381.pdf.

Owens, Kellie and Alexis Walker (Sept. 2020). "Those designing healthcare algorithms must become actively anti-racist". In: *Nature Medicine* 26.9, pp. 1327–1328. ISSN: 1546-170X. DOI: 10.1038/s41591-020-1020-3. URL: https://www.nature.com/articles/s41591-020-1020-3 (visited on 04/22/2023).

Pace, Clare et al. (Jan. 2022). "Inequities in Drinking Water Quality Among Domestic Well Communities and Community Water Systems, California, 2011–2019". In: *American Journal of Public Health* 112.1, pp. 88–97. ISSN: 0090-0036. DOI: 10.2105/AJPH.2021.306561. URL: https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2021.306561 (visited on 03/21/2024).

Pebesma, Edzer (2018). "Simple Features for R: Standardized Support for Spatial Vector Data". In: *The R Journal* 10.1, pp. 439–446. DOI: 10.32614/RJ-2018-009. URL: https://doi.org/10.32614/RJ-2018-009.

Saito, Takaya and Marc Rehmsmeier (Mar. 4, 2015). "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets". In: *PLOS ONE* 10.3, e0118432. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0118432. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118432 (visited on 05/11/2022).

Service, United States Geological (2019). *National Land Cover Data Set*. URL: https://www.usgs.gov/centers/eros/science/national-land-cover-database.

Soelter, Alan D. and Ellen G. Miller (1999). "Capacity development: the small system perspective". In: *Journal AWWA* 91.4, pp. 110–122. ISSN: 1551-8833. DOI: 10.1002/j.1551-8833.1999.tb08617.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1551-8833.1999.tb08617.x (visited on 03/21/2024).

Team, R Core (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. URL: https://www.R-project.org/.

Tomašev, Nenad et al. (May 18, 2020). "AI for social good: unlocking the opportunity for positive impact". In: *Nature Communications* 11.1, p. 2468. ISSN: 2041-1723. DOI: 10.1038/s41467-020-15871-z. URL: https://www.nature.com/articles/s41467-020-15871-z (visited on 09/29/2023).

United States Department of Agriculture, National Agricultural Statistics Service (2021). *Cropland Data Layer*. URL: https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php.

Veale, Michael and Irina Brass (2019). *Administration by Algorithm? Public Management Meets Public Sector Machine Learning*. Rochester, NY. URL: https://papers.ssrn.com/abstract=3375391 (visited on 05/02/2023).

Walker, Kyle and Matt Herman (2024). *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*. Version R package version 1.6.3. URL: https://walker-data.com/tidycensus/.

Wirtz, Bernd W., Jan C. Weyerer, and Carolin Geyer (May 19, 2019). "Artificial Intelligence and the Public Sector—Applications and Challenges". In: *International Journal of Public Administration* 42.7, pp. 596–615. ISSN: 0190-0692. DOI: 10.1080/01900692.2018.1498103. URL: https://doi.org/10.1080/01900692.2018.1498103 (visited on 05/02/2023).

Chapter 5

Dissecting an algorithmic approach to environmental justice

5.1 Dissecting the model to explore justice implications of modeling decisions

In this chapter, we explore relationships between our modeling decisions, predictions, and water system demographics. Our aim is to show how modeling decisions affect outcomes beyond "performance." Specifically, we investigate what types of systems or populations would receive more attention (true positives) or receive less (false negatives) if a regulatory agency were to adopt this ML approach to guide the targeting of regulatory enforcement activities. In the following section, we examine: the effect of changing the label on model performance; how feature choice affects modeling performance; how demographic characteristics of modeling outcomes differ based on upstream decisions; and whether sampling history may bias results. This list of modeling decisions is not meant to be comprehensive. We conclude with an exploration on reducing classification bias.

Does changing the label affect predictions or performance?

As the EPA sets MCLs as a practical and enforceable regulatory threshold, they are higher than health-based standards such as MCL goals (MCLGs) or Public Health Goals (PHGs) (Agency 2015). Past research has shown how examining relationships between water system characteristics and water quality at levels below the MCL can reveal statistical relationships that would have otherwise been masked or muted at higher levels (Mulhern et al. 2022).

Model performance improved (apart from THA) when we changed the label threshold from the MCL to half the MCL. (See Appendix Figure 4.3.) This improvement is to be expected, as we are effectively making the dataset less imbalanced by loosening the requirement threshold of the minority class. However, we find that this F2 score improvement obscures

other changes that would have important public health implications in an applied setting. When using the $\frac{1}{2}$ MCL as the label threshold, the model was 91% accurate overall for the test set (down from 93%), with a precision of 0.71 (up from 0.39), recall of 0.73 (down from 0.82), and F2 score of 0.73 (up from 0.67). For MR violating systems, the model had both a precision and recall of 0.94 (a gain in precision but a loss in recall), and a F2 score of 0.94 (up from 0.91). (See Appendix Figure B.5 for these results as confusion matrices.)

While neither of these models and their respective labels are 'better' or 'worse' than the other in absolute terms, their differences highlight tradeoffs inherent in label selection. Here, the tradeoff is between how *accurate* the positive predictions are (precision) versus how *complete* (i.e., the proportion of total positive predictions they predict correctly, or recall). These tradeoffs would have real-world implications in a situation where policy makers are required to weigh the benefits and drawbacks between the two in an applied public health context. While there is no universally applicable formula to weighing precision versus recall and what balance is "best" is context-dependent, minimizing false negatives is a defensible priority for public health.

One option for a regulator would be to use the probability scores output by a less restrictive label (e.g., the $\frac{1}{2}$ MCL) to inform a ranked prioritization system for resource allocation in terms of regulatory enforcement and monitoring. In addition to having better precision, it is reasonable to expect (though not guaranteed) that samples predicted to exceed the $\frac{1}{2}$ MCL are also likely to be predicted to exceed the MCL. Regulators can then take these probabilities and pass it through additional considerations, such as past funding history or community demographics. (See Appendix Figure B.6 for an example of how classification probabilities might be used in a prioritization system.)

How does withholding input features affect modeling performance?

To assess the importance of different data types on modeling performance, we organized the feature variables into three categories: drinking water data, demographic data, and physical (geological, agricultural) data. (See Appendix Table B.8 for more information on how feature variables were categorized.) We then repeated the model training process on the same split of training-testing data for arsenic and its MCL (just one of many contaminant-label combinations we could have chosen), but this time using different combinations of these data group categories (e.g., only water data, water and demographic data, water and physical data, only demographic data, etc.). When trained using these data categories in isolation, models with only drinking water data were much better at predicting water quality than models with only physical data or only demographic data. When two or more of the three data categories were paired together, model performance either stayed constant or slightly improved.

Even without a causal component to our analysis, we can infer from these results that

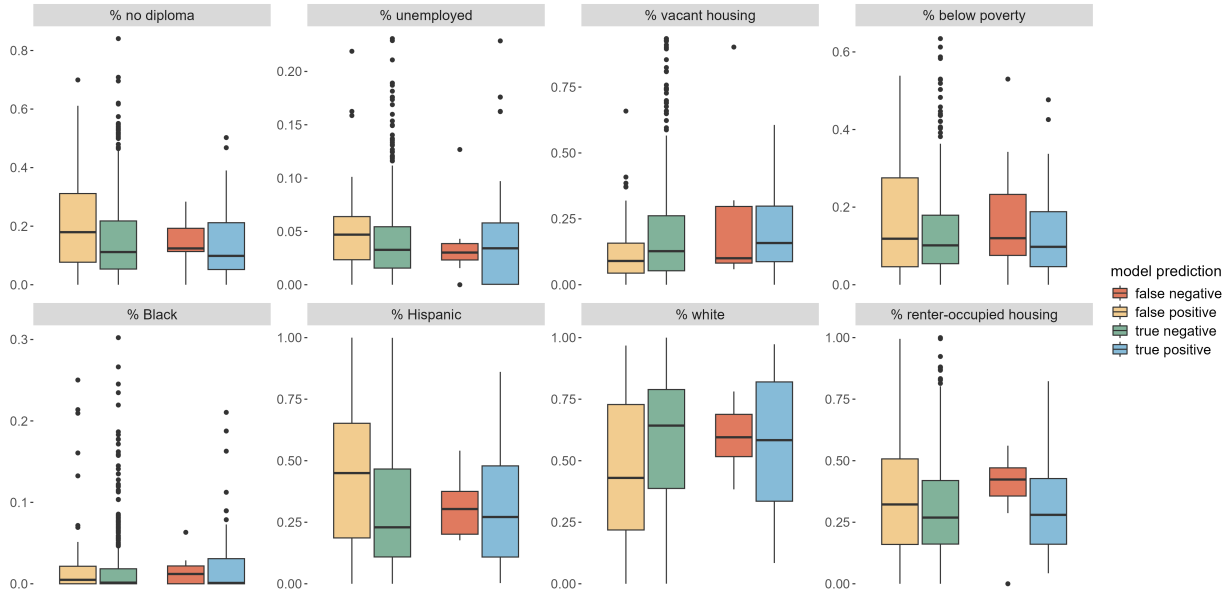
systems' sampling history likely play a large role in determining prediction accuracy. This suggests that predictions made for systems with sparser sampling histories should be subject to closer scrutiny, particularly those with samples that are predicted to be under the threshold value.

How are modeling decisions, predictions, and demographics connected?

We can add a dimension to the confusion matrices shown above by describing each of the quadrants (true positive, true negative, false positive, false negative) using demographic indicators. Knowing the demographic distribution of each of these four classification outcomes is of clear interest in a real-world application. We can also compare if or how these indicators shift within and across categories with different modeling decisions.

Figure 5.1 presents an example distribution of different demographic variables for each of the four model prediction categories. This is the output for arsenic MCL exceedances using elastic net . The figure shows that percentages of certain demographic indicators (% below poverty, % Black, % renter-occupied housing) appear higher for false negatives (red box) in comparison to true positives (blue box).

Figure 5.1: Distribution of demographic indicators across the four confusion matrix quadrants, predicting for arsenic (MCL) using elastic net.



When we ran the Mann-Whitney U test (a non-parametric alternative to independent

samples t-tests) across each of these selected demographic variables and paired model prediction groups, % Black and % renter-occupied housing came back as significantly different when comparing false negatives to true positives (p = 0.048 and 0.018, respectively). These statistically-significant differences between false negatives and true positives – alongside the other visually-evident differences – change or disappear when we change the label threshold from the MCL to $\frac{1}{2}$ MCL, or the model itself to a different classifier (Table 3). (See Appendix Figure B.7 for box plots of other contaminants.) Our point is not to show that there are certain models or thresholds that result in biased or unfair outcomes, but rather to emphasize that the potential for such unfair outcomes should not be ignored.

It is challenging to quantify these relationships in a statistically meaningful way given our sample size: the classification counts for false negatives and true positives are small due to the imbalanced nature of the data. In the example above, there are only n=5 systems in the false negative group for the MCL models (compared to n=26 for the $\frac{1}{2}$ MCL). The demographic data is also not normally distributed.

Does sampling history (or the ability to sample) bias results?

We can use the approach above to examine the relationship between sample counts and different classification outcomes. If there are statistically significant relationships between the number of samples taken and whether predictions are false negatives or true positives, this raises a potential concern of embedding bias: if some systems are sampling less because they lack capacity, policy initiatives informed by a model biased by sample counts could further exacerbate gaps in sampling ability.

While there is no statistically significant relationship between classification outcomes and sampling history for models using the MCL label, there are for those using the $\frac{1}{2}$ MCL label. (See Appendix Table B.9 for Mann-Whitney U test results for sample counts and model outcomes for arsenic.) One interpretation of these results might be that, as the label threshold is lowered and the data becomes less imbalanced, sampling history (and in this case, sampling counts) becomes a much more important predictor. This suggests the possibility of bias towards systems with larger sampling histories, or richer data.

What can we do to reduce bias in classification outcomes?

We presented above that relaxing the model label threshold from MCL to $\frac{1}{2}$ MCL is one option to lessen bias in our outcomes. Depending on the context, however, this may not be advisable or appropriate. The relationships between MCL thresholds and detection limits vary by contaminant, as well as by where the samples are sent to be tested.

As an alternative, we explored the effects of adjusting the *probability threshold* for which the model classifies a prediction as positive versus negative. The default threshold for classification models is typically 0.5, but this is not a hard-and-fast rule; moving this threshold is an established approach for imbalanced classification (Zou et al. 2016). Below, we took our

trained elastic net model for arsenic (label MCL) and re-classified the predicted outcomes using an array of thresholds for positive outcomes. We repeated this for all threshold values between 0.2 and 0.7, at increments of 0.1. For each threshold, we compared the average demographic profiles of "false negative" systems and "true positive" systems for the four selected indicators below (% Black, % below poverty, % unemployed, and % renter).

Table 4 illustrates how the statistical relationships between classification categories and demographic indicators can shift when probability thresholds for classification are changed. In this case, only the threshold of 0.2 had p-values above 0.05 for each of these four demographic indicators. However, due to its relative relaxedness, this threshold had the lowest F2 score. This approach is context-specific (dependent on input data, chosen model, chosen parameters, etc.), and cannot be generalized.

5.2 Discussion

In this analysis, we examined the potential environmental justice implications of using ML for environmental monitoring. In the first half, we demonstrate the potential of ML as a resource- and time-efficient tool to strengthen environmental justice initiatives under resource-constrained regulatory settings through a case study on drinking water quality prediction. In the second half, we underscore its pitfalls by dissecting our own modeling process, showing how different choices can lead to different outcomes. In this discussion, we focus on one point in particular: the relationship between justice and the pursuit of model optimization. We conclude with a set of recommendations for future work.

Optimize to what end?

Our analysis shows how modeling choices can have significant implications with regards to demographic differences in predictive outcomes. This is by design, as modelers must define the types of outcomes they are optimizing for, whether explicitly or implicitly. We call attention to the issue that there is no standard of practice for researchers undergoing justice-driven projects to be deliberate about *which modeling choices and approaches* are most appropriate to advance *which specific dimension of EJ* in research frameworks and for regulatory decision-making. In our example, we define our environmental injustice of concern to be continued, inadvertent exposure due to incorrect predictions (a distributive injustice), and thus select the F2 score to minimize false negatives through model selection. The label choice can also change the structural composition of the data (e.g., imbalanced versus balanced), which can affect precision and recall. Labels, reflecting broader social context, can also embed bias. Furthermore, our results underscore how maximizing performance may not maximize environmentally just outcomes. Chasing model performance can improve results for some communities (e.g., those with more or better-quality data) but not others.

How the practice of algorithmic optimization can exacerbate and entrench disparities

has been observed in the form of "algorithmic redlining" (Nost 2019; Safransky 2020). Encouragingly, research indicates that label-driven biases are addressable through label choice (Obermeyer et al. 2019). Building upon this work, our research underscores the importance of assessing the implications of modeling decisions to ensure they don't exacerbate or entrench existing disparities. We offer our case study as an alternative way of "thinking algorithmically" to resist the hegemonizing of algorithmic knowledge (Machen and Nost 2021) in the realm of environmental justice policy and regulatory decision-making. Government agencies are obliged to evaluate their regulations, policies, and practices to ensure that they do not undermine equal protections and opportunities based on race or otherwise, and the use of machine learning is no exception. While this approach does not necessarily break the cycle of extractive logic, or "the logic of pulling relations out from bodies and lands into data directed towards other ends," (Vera et al. 2019) there is value in its direct relevance for policies being discussed and adopted now; our hope is that this way of algorithmic thinking pushes the needle towards a more just future.

5.3 Future recommendations

The predictive capabilities of ML algorithms hold promise for strengthening our collective pursuit towards an environmentally-just future. Under the right circumstances, a carefully designed and implemented model can both help address data gaps as well as inform proactive policy making and investment. However, we also find that the modeling process itself – as typically presented in environmental research – is fraught with the possibility of embedding inequality and bias. We present the following set of recommendations as a starting place for researchers looking to apply ML to inform environmentally-just outcomes in decision-making.

1. Ensure that the connection between the context-specific definition of justice and the decisions made across the modeling process are clear, logical, and well-documented;
2. Apply a range of label definitions and thresholds as a sensitivity test;
3. Clearly delineate any significant demographic patterns across modeling outcomes (correct and incorrect), if and when predictions can be validated; and
4. Report and discuss the sociodemographic characteristics of communities predicted to be at risk, as well as the implications for where regulatory resources would ostensibly be targeted.

With this, we build upon existing guidance for environmental justice model selection (Gardner-Frolick, Boyd, and Giang 2022) by prioritizing justice-based logic for parameter choices and recommending the evaluation of demographic variables across modeling outcomes.

It is exciting to consider the implications of machine learning for environmental science, management, and justice. Environmental data and its derivatives – e.g., algorithmic predictions and forecasts – are only going to increase in their availability, accessibility, and pervasiveness. Our research underscores why it is important for researchers to be specific and transparent about which environmental justice considerations they considered in their machine learning workflows. Our set of recommendations is neither complete nor in its final form; this confluence of machine learning, environmental justice, and responsible research design is a promising and important area of future research. Our hope is that this work encourages others to join this conversation and improve and build upon the approaches we have suggested here.

References

Agency, United States Environmental Protection (Sept. 1, 2015). *How EPA Regulates Drinking Water Contaminants*. URL: https://www.epa.gov/sdwa/how-epa-regulates-drinking-water-contaminants (visited on 03/23/2024).

Gardner-Frolick, Rivkah, David Boyd, and Amanda Giang (Mar. 1, 2022). "Selecting Data Analytic and Modeling Methods to Support Air Pollution and Environmental Justice Investigations: A Critical Review and Guidance Framework". In: *Environmental Science & Technology* 56.5, pp. 2843–2860. ISSN: 0013-936X. DOI: 10.1021/acs.est.1c01739. URL: https://doi.org/10.1021/acs.est.1c01739 (visited on 09/30/2023).

Machen, Ruth and Eric Nost (2021). "Thinking algorithmically: The making of hegemonic knowledge in climate governance". In: *Transactions of the Institute of British Geographers* 46.3, pp. 555–569. ISSN: 1475-5661. DOI: 10.1111/tran.12441. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/tran.12441 (visited on 09/30/2023).

Mulhern, Riley et al. (Mar. 1, 2022). "A new approach to a legacy concern: Evaluating machine-learned Bayesian networks to predict childhood lead exposure risk from community water systems". In: *Environmental Research* 204, p. 112146. ISSN: 0013-9351. DOI: 10.1016/j.envres.2021.112146. URL: https://www.sciencedirect.com/science/article/pii/S001393512101447X (visited on 01/17/2022).

Nost, Eric (Nov. 1, 2019). "Climate services for whom? The political economics of contextualizing climate data in Louisiana's coastal Master Plan". In: *Climatic Change* 157.1, pp. 27–42. ISSN: 1573-1480. DOI: 10.1007/s10584-019-02383-z. URL: https://doi.org/10.1007/s10584-019-02383-z (visited on 09/30/2023).

Obermeyer, Ziad et al. (Oct. 25, 2019). "Dissecting racial bias in an algorithm used to manage the health of populations". In: *Science* 366.6464, pp. 447–453. DOI: 10.1126/science.aax2342. URL: https://www.science.org/doi/full/10.1126/science.aax2342 (visited on 03/18/2022).

Safransky, Sara (2020). "Geographies of Algorithmic Violence: Redlining the Smart City". In: *International Journal of Urban and Regional Research* 44.2, pp. 200–218. ISSN: 1468-

2427. DOI: 10.1111/1468-2427.12833. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-2427.12833 (visited on 09/30/2023).

Vera, Lourdes A. et al. (June 7, 2019). "When data justice and environmental justice meet: formulating a response to extractive logic through environmental data justice". In: *Information, Communication & Society* 22.7, pp. 1012–1028. ISSN: 1369-118X. DOI: 10.1080/1369118X.2019.1596293. URL: https://doi.org/10.1080/1369118X.2019.1596293 (visited on 01/18/2022).

Zou, Quan et al. (Sept. 1, 2016). "Finding the Best Classification Threshold in Imbalanced Classification". In: *Big Data Research*. Big data analytics and applications 5, pp. 2–8. ISSN: 2214-5796. DOI: 10.1016/j.bdr.2015.12.001. URL: https://www.sciencedirect.com/science/article/pii/S2214579615000611 (visited on 03/21/2024).

Chapter 6

On strengthening the rigor, relevance, reach, and reflexivity of science[1]

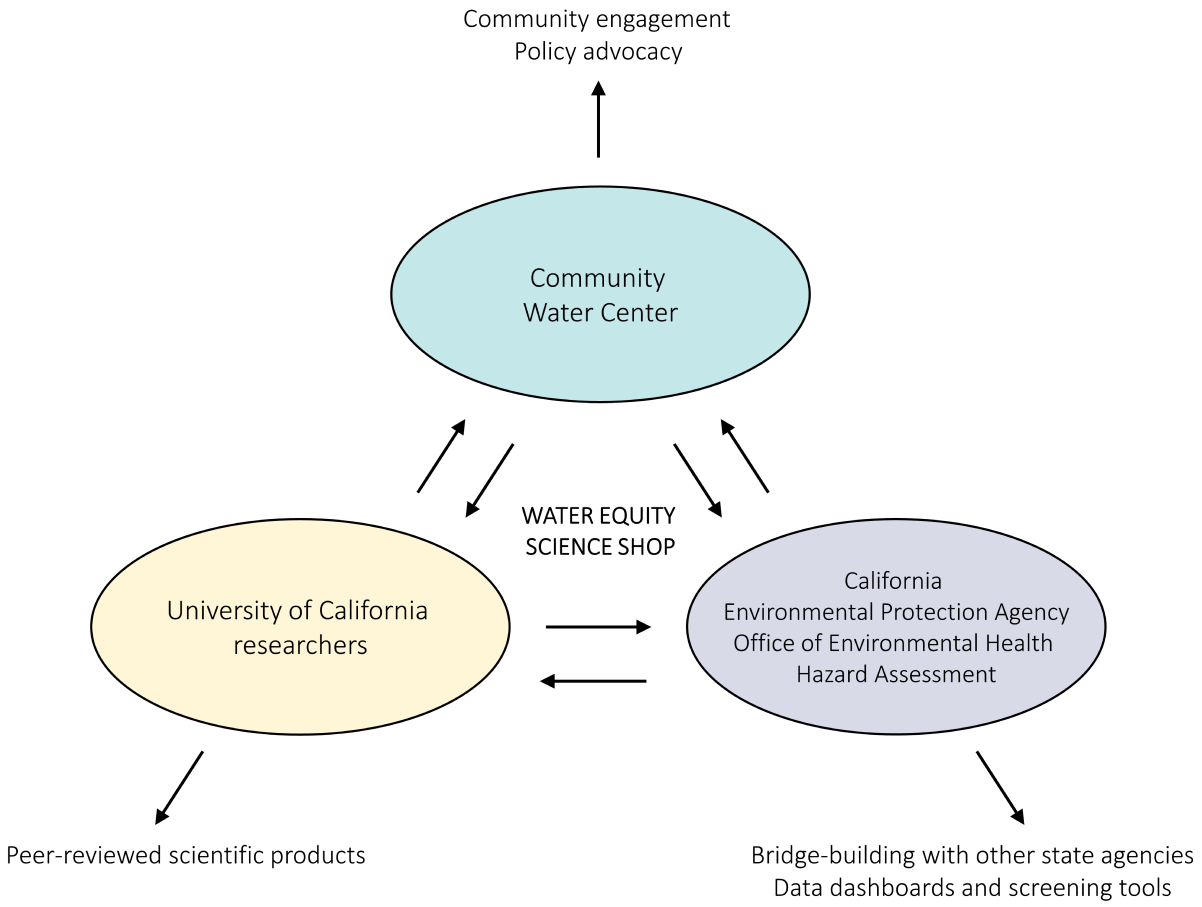
6.1 Background

Community-academic partnerships have undertaken participatory research on environmental health and justice that centers communities' priorities (Salimi et al. 2012) and strengthens science (Balazs and Morello-Frosch 2013). Our research collaborative, the Water Equity Science Shop (WESS; Figure 1), synthesizes the European Science Shop model (Leydesdorff and Ward 2005)] and principles of community-engaged research practices (Mikesell, Bromley, and Khodyakov 2013). One fruit of this collaboration is the Drinking Water Tool (DWT), a free online mapping platform owned and managed by Community Water Center (CWC) (Pace et al. 2022).

Launched in 2020, the DWT centralizes information on (i) water quality data and possible contamination sources; (ii) projected groundwater supply impacts from future droughts; and (iii) who makes key water decisions and how to get involved in water governance. The bilingual English/Spanish DWT is designed for diverse end-users, including community-based organizations (CBOs), government agencies, and the public. The DWT addressed the need for more accessible data on drinking water, which at the time was not easily available nor interpretable. (California has since developed platforms such as the SAFER dashboards (2024) to improve drinking water-related data accessibility.) We updated the DWT in 2023 after rounds of community-engaged feedback (Figure 2), adding refined data on domestic well areas and contaminant sources alongside improvements in mapping water quality, identifying monitoring gaps, and communicating potential water quality threats. In this practice brief, we present lessons-learned in community engagement, EJ research, and knowledge-sharing

[1]This chapter is currently in peer-review, and was co-authored by Arianna Libenson, Tien Tran, Komal Bangia, Lara Cushing, Jenny Rempel, Laura August, Lauren Baehner, Rachel Morello-Frosch, and Clare Pace. I am grateful for their contributions to this work, and for their permission to use this research as a part of my dissertation.

Figure 6.1: Water Equity Science Shop. Collaboration outputs are listed next to their primary agents.

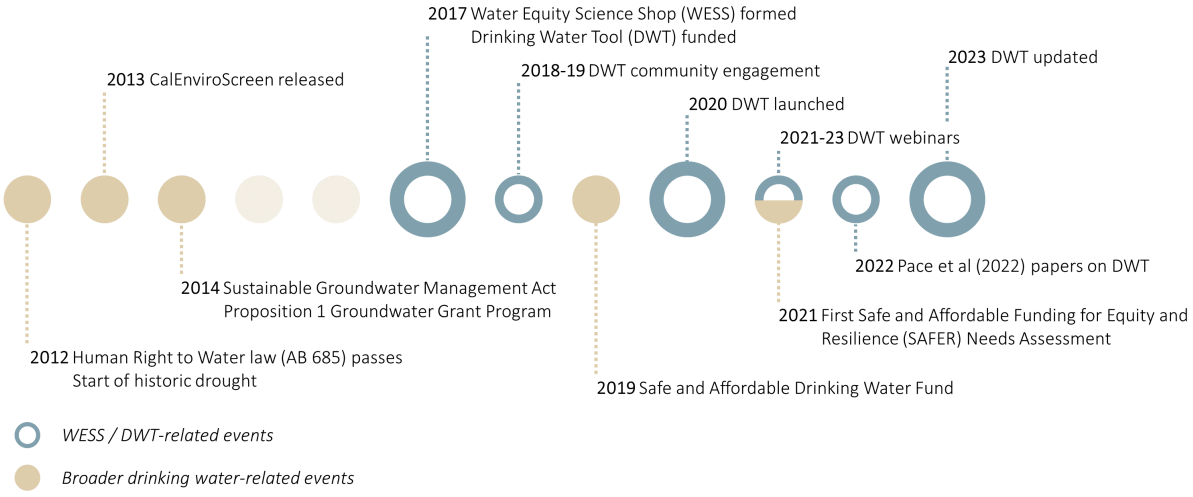


through WESS and the DWT.

6.2 Findings

We apply a "4 Rs" framework to assess WESS's outputs. Balazs and Morello-Frosch introduced the "3 Rs" framework to show how community-engaged research can strengthen the rigor, relevance, and reach of science (Balazs and Morello-Frosch 2013). They define relevance as "whether science is asking the right questions"; rigor as the "practice and promotion of good science"; and reach as "[how] knowledge is disseminated to diverse audiences and translated into useful tools." Recent scholarship has suggested reflexivity – self-examination of positionality, intention, and other aspects of research – as a fourth "R" (Cordner et al. 2015; Lockie 2018; Raphael 2019).

Figure 6.2: Timeline of the Water Equity Science Shop's (WESS) development, the evolution of its Drinking Water Tool (DWT), and the broader landscape of California drinking water policy.



Relevance through responsiveness

In 2013, the Office of Environmental Health Hazard Assessment (OEHHA) launched CalEnviroScreen, the first state-managed tool for mapping environmental justice (EJ) and cumulative health impacts (Alexeeff et al. 2012; Cushing et al. 2015). Similar tools have since been developed in other states (Min et al. 2019; Williams et al. 2022) and at the federal level (Council on Environmental Quality 2024). The first iteration of the DWT incorporated CalEnviroScreen's drinking water quality indicator (Bangia et al. 2020). Strengths of such government tools include their rigor, accessibility, and legitimacy in the eyes of policymakers. However, state agencies must navigate layers of red tape for such tools to be approved and updated. In contrast, responding quickly to community needs and changing conditions is of high priority for CWC (as a CBO). In California, climate change and EJ concerns have brought the importance of such responsiveness into sharp relief, as communities recover from thousands of domestic wells running dry during the recent historic drought (Jasechko and Perrone 2020) and environmental health science on contaminants such as PFAS advance at a rapid pace (Johnston and Cushing 2020).

Tools managed by non-government actors can also function as a proof of concept to pilot methods prior to state adoption. While planning the latest DWT update, community partners and CWC raised the addition of PFAS data as a high priority, compelled by emerging research on human exposure through drinking water (Domingo and Nadal 2019). At this point, the State Water Resources Control Board (SWRCB) had made its PFAS sampling data available through GeoTracker, an environmental regulatory database (Beckley et al.

2022). Though a powerful resource as-is, the GeoTracker platform did not include EJ concerns as an analytical lens. To address this gap, we integrated SWRCB's data into the DWT to overlay PFAS data with demographic characteristics and domestic well area boundaries, and tailor visualizations based on new regulatory standards. WESS is using these spatial layers together to identify partner communities for a drinking water sampling project to test for PFAS in unregulated domestic well areas. This EJ and health-motivated approach to mapping PFAS data is now being considered for the next CalEnviroScreen update.

Rigor through collaborative methods development

Our facilitation of diverse partnerships and conversations about drinking water justice represents our effort to practice and promote good science. For example, prior to building the DWT, we organized a meeting in 2018 to encourage dialogue across its anticipated userbase. Our goal was to collectively identify data gaps and brainstorm how to synthesize existing resources to address pressing drinking water issues, such as providing interim water access to domestic well users.

One challenge that came up was how to handle imperfect data. Given the state's incomplete domestic well drilling records, state and local agencies were struggling with how (and whether) to include domestic well communities in their sustainable groundwater planning decisions. WESS recognized that estimating domestic well reliance using incomplete data was preferable to excluding domestic well communities from planning altogether. To enhance the accuracy of our domestic well use estimates, we integrated residential parcel data with water systems' service area boundaries to better define the spatial extent of domestic well communities (Rempel et al. 2023). This enabled us to estimate domestic well populations across multiple spatial scales, and to develop the first statewide estimates of populations relying on domestic wells versus community water systems. WESS is using these data refinements to strategize outreach to communities at risk of elevated contaminant exposure and cumulative health effects.

Reflexivity through multidirectional learning

WESS has encouraged multidirectional learning between community partners, CWC, OEHHA, and university researchers. Additionally, over years of collaboration, many WESS collaborators have worn different "hats" of advocacy, academia, and government, resulting in overlapping experiences, shared understandings, and trust. This has strengthened WESS' collective ability to reflect on intended and potential unanticipated impacts of projects and tools.

Some EJ researchers have challenged the EJ movement's potential over-reliance on the state for solutions or reparations (Kurtz 2009). In response, others have characterized this critique as too reductionist of the complex relationships between EJ movements and government actors (Harrison 2023; Purucker 2021). Our experience suggests that such relationships,

while sometimes challenging, enable EJ communities to effectively shape environmental policy and regulation as well as the science that informs them.

6.3 Practice Recommendations

We conclude with three practice recommendations that have been instrumental in our work.

Recommendation 1: leverage non-governmental tools to respond to evolving crises

One of the primary strengths of the DWT is the nimbleness with which it can respond to evolving EJ priorities and emerging research. Academics have adopted the use of *pre-prints* – research manuscripts made available prior to peer review – when quick dissemination of results is warranted (e.g., during pandemics) (Fraser et al. 2021). In this vein, we recommend the development and use of non-governmental tools for their responsiveness, in ways that support their future integration into state-managed regulatory instruments.

Recommendation 2: work with the state to enhance the reach and sustainability of advocacy and research

The challenge of ensuring the longevity of such tools outside of state contexts is in securing sustainable resources. WESS has successfully tapped into state, federal, and foundation funding sources that encourage such community partnerships, which has been critical for WESS' long-term sustainability. Collaborating directly with state agencies has also helped translate WESS's research and CWC's advocacy into policy decisions concerning drinking water access and resource distribution to impacted communities.

Recommendation 3: invest in sustainable relationships that enable collaboratives to weather ebbs and flows in funding

WESS partners co-developed federal, state and foundation funding proposals to support community- and data-driven research that advances water justice goals in policymaking. Our core strength stems from our prior successful and iterative work together that has sustained our collaborative over the long-term, both when funding has been abundant and when it has temporarily run dry. Together, WESS partners have ensured a nimbleness in research aims and approaches, which integrate primary and secondary data analysis to enable the collaborative to keep policy relevant work active, even during temporary shortfalls in funding streams.

6.4 Funding

This project was supported by:

- the National Institute of Environmental Health Sciences (award #P42ES004705); and
- the Water Quality, Supply, and Infrastructure Improvement Act of 2014 and through an agreement with the State Department of Water Resources (award #4600012684)

References

Alexeeff, George V. et al. (Feb. 2012). "A Screening Method for Assessing Cumulative Impacts". In: *International Journal of Environmental Research and Public Health* 9.2, pp. 648–659. ISSN: 1660-4601. DOI: 10.3390/ijerph9020648. URL: https://www.mdpi.com/1660-4601/9/2/648 (visited on 02/11/2022).

Balazs, Carolina L. and Rachel Morello-Frosch (Feb. 2013). "The Three Rs: How Community-Based Participatory Research Strengthens the Rigor, Relevance, and Reach of Science". In: *Environmental Justice* 6.1, pp. 9–16. ISSN: 1939-4071. DOI: 10.1089/env.2012.0017. URL: https://www.liebertpub.com/doi/full/10.1089/env.2012.0017 (visited on 05/03/2022).

Bangia, Komal et al. (2020). "Assessment of contaminants in California drinking water by region and system size". In: *AWWA Water Science* 2.5, e1194. ISSN: 2577-8161. DOI: 10.1002/aws2.1194. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/aws2.1194 (visited on 03/15/2024).

Beckley, Lila et al. (2022). "The California GeoTracker Database: A Unique Public Resource for Understanding Contaminated Sites". In: *Groundwater Monitoring & Remediation* 42.3, pp. 105–115. ISSN: 1745-6592. DOI: 10.1111/gwmr.12520. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/gwmr.12520 (visited on 03/15/2024).

California State Water Resources Control Board (Mar. 1, 2024). *SAFER Dashboard*. SAFER Dashboard. URL: https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/saferdashboard.html (visited on 03/01/2024).

Cordner, Alissa et al. (2015). "Reflexive Research Ethics for Environmental Health and Justice: Academics and Movement Building". In: *Research Ethics and Social Movements*. Routledge. ISBN: 978-1-315-74240-3.

Council on Environmental Quality (Mar. 1, 2024). *Explore the map - Climate & Economic Justice Screening Tool*. URL: https://screeningtool.geoplatform.gov/en/ (visited on 03/01/2024).

Cushing, Lara et al. (Nov. 2015). "Racial/Ethnic Disparities in Cumulative Environmental Health Impacts in California: Evidence From a Statewide Environmental Justice Screening Tool (CalEnviroScreen 1.1)". In: *American Journal of Public Health* 105.11, pp. 2341–2348. ISSN: 0090-0036. DOI: 10.2105/AJPH.2015.302643. URL: https://

ajph.aphapublications.org/doi/full/10.2105/AJPH.2015.302643 (visited on 03/15/2024).

Domingo, José L. and Martí Nadal (Oct. 1, 2019). "Human exposure to per- and polyfluoroalkyl substances (PFAS) through drinking water: A review of the recent scientific literature". In: *Environmental Research* 177, p. 108648. ISSN: 0013-9351. DOI: 10.1016/j.envres.2019.108648. URL: https://www.sciencedirect.com/science/article/pii/S0013935119304451 (visited on 03/15/2024).

Fraser, Nicholas et al. (Apr. 2, 2021). "The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape". In: *PLOS Biology* 19.4, e3000959. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.3000959. URL: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000959 (visited on 03/15/2024).

Harrison, Jill Lindsey (Dec. 1, 2023). "Environmental justice and the state". In: *Environment and Planning E: Nature and Space* 6.4, pp. 2740–2760. ISSN: 2514-8486. DOI: 10.1177/25148486221138736. URL: https://doi.org/10.1177/25148486221138736 (visited on 03/15/2024).

Jasechko, Scott and Debra Perrone (2020). "California's Central Valley Groundwater Wells Run Dry During Recent Drought". In: *Earth's Future* 8.4, e2019EF001339. ISSN: 2328-4277. DOI: 10.1029/2019EF001339. URL: https://onlinelibrary.wiley.com/doi/abs/10.1029/2019EF001339 (visited on 03/15/2024).

Johnston, Jill and Lara Cushing (Mar. 1, 2020). "Chemical Exposures, Health, and Environmental Justice in Communities Living on the Fenceline of Industry". In: *Current Environmental Health Reports* 7.1, pp. 48–57. ISSN: 2196-5412. DOI: 10.1007/s40572-020-00263-8. URL: https://doi.org/10.1007/s40572-020-00263-8 (visited on 03/15/2024).

Kurtz, Hilda E. (2009). "Acknowledging the Racial State: An Agenda for Environmental Justice Research". In: *Antipode* 41.4, pp. 684–704. ISSN: 1467-8330. DOI: 10.1111/j.1467-8330.2009.00694.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8330.2009.00694.x (visited on 03/15/2024).

Leydesdorff, Loet and Janelle Ward (Oct. 1, 2005). "Science shops: a kaleidoscope of science–society collaborations in Europe". In: *Public Understanding of Science* 14.4, pp. 353–372. ISSN: 0963-6625. DOI: 10.1177/0963662505056612. URL: https://doi.org/10.1177/0963662505056612 (visited on 02/11/2022).

Lockie, Stewart (Apr. 3, 2018). "Privilege and responsibility in environmental justice research". In: *Environmental Sociology* 4.2, pp. 175–180. ISSN: null. DOI: 10.1080/23251042.2018.1460936. URL: https://doi.org/10.1080/23251042.2018.1460936 (visited on 03/15/2024).

Mikesell, Lisa, Elizabeth Bromley, and Dmitry Khodyakov (Dec. 2013). "Ethical Community-Engaged Research: A Literature Review". In: *American Journal of Public Health* 103.12, e7–e14. ISSN: 0090-0036. DOI: 10.2105/AJPH.2013.301605. URL: https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2013.301605 (visited on 03/15/2024).

Min, Esther et al. (Jan. 2019). "The Washington State Environmental Health Disparities Map: Development of a Community-Responsive Cumulative Impacts Assessment Tool". In: *International Journal of Environmental Research and Public Health* 16.22, p. 4470. ISSN: 1660-4601. DOI: 10.3390/ijerph16224470. URL: https://www.mdpi.com/1660-4601/16/22/4470 (visited on 03/15/2024).

Pace, Clare et al. (Jan. 2022). "The Drinking Water Tool: A Community-Driven Data Visualization Tool for Policy Implementation". In: *International Journal of Environmental Research and Public Health* 19.3, p. 1419. ISSN: 1660-4601. DOI: 10.3390/ijerph19031419. URL: https://www.mdpi.com/1660-4601/19/3/1419 (visited on 03/15/2024).

Purucker, David (July 3, 2021). "Critical environmental justice and the state: a critique of Pellow". In: *Environmental Sociology* 7.3, pp. 176–186. ISSN: null. DOI: 10.1080/23251042.2021.1878575. URL: https://doi.org/10.1080/23251042.2021.1878575 (visited on 03/15/2024).

Raphael, Chad (Nov. 17, 2019). "Engaged Communication Scholarship for Environmental Justice: A Research Agenda". In: *Environmental Communication* 13.8, pp. 1087–1107. ISSN: 1752-4032. DOI: 10.1080/17524032.2019.1591478. URL: https://doi.org/10.1080/17524032.2019.1591478 (visited on 03/15/2024).

Rempel, Jenny et al. (2023). *Domestic well areas version 2.0, Update for the Drinking Water Tool*. URL: https://drinkingwatertool.communitywatercenter.org/wp-content/uploads/2023/09/DWA_v2_plss_020824_Metadata.pdf.

Salimi, Yahya et al. (June 2012). "Is Community-based Participatory Research (CBPR) Useful? A Systematic Review on Papers in a Decade". In: *International Journal of Preventive Medicine* 3.6, pp. 386–393. ISSN: 2008-7802. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3389435/ (visited on 03/15/2024).

Williams, Elizabeth et al. (Dec. 2022). "MD EJSCREEN v2.0: Visualizing Overburdening of Environmental Justice Issues Using the Updated Maryland Environmental Justice Screening Tool". In: *Environmental Justice* 15.6, pp. 385–401. ISSN: 1939-4071. DOI: 10.1089/env.2020.0055. URL: https://www.liebertpub.com/doi/full/10.1089/env.2020.0055 (visited on 03/15/2024).

Chapter 7

Conclusion

The primary motivation of my research is the existence and persistence of environmental injustices, which research has shown are spatially distributed along systemic and societal inequities such as race and class. In the latter half of my years in ERG, my interests coalesced on the plane where two facets of popular environmental justice epistemology meet: specifically, on the challenges and opportunities of using big data and machine learning for prediction in conjunction with participatory research methods and community engagement.

This was not by design, but my dissertation structure roughly follows the evolution of my academic interests and intellectual growth. I entered the program motivated to study environmental justice and drinking water disparities in California. Conducting a literature review (Chapter 2) on this topic was a logical and practical step to familiarize myself with the current state of knowledge. (The fact that it was the first of its kind was an additional bonus.) The decision to focus on primary data studies emerged out of a desire to better understand both on-the-ground and perceived injustices closest to communities. Though I had not anticipated it at the time, this also became the lens through which I came to better understand the strengths and shortcomings of secondary data analysis, which machine learning and big data are built upon.

For myself and for many others, the COVID pandemic (roughly 2020 – 2023) altered the opportunities for fieldwork, and my attention settled on exploring and practicing the use of secondary data analysis to address drinking water injustices. My interest was initially limited to the potential application of emerging predictive algorithms to questions of injustice (Chapter 4), but I soon found myself even more excited by the prospect of how algorithmic approaches may be embedding bias and unfairness into their modeling outcomes in ways that were detrimental for advancing environmental justice (Chapter 5). I felt strongly enough about this issue to summarize my opinions into a perspective piece (Chapter 3) as I was finishing up the paper behind chapters 4 and 5. It was important to me that, in contrast with the other manuscripts I've included in this dissertation, I was the single author for this piece. It was important because a significant part of my PhD journey (with implications reaching beyond my academic work) was building up the confidence and desire to stick my

neck out. This was only made possible through many conversations with supportive and engaged mentors and peers, but it was nonetheless meaningful to me.

Lastly, I feel incredibly fortunate to have had the opportunity to propose and lead a piece on collaborative research for environmental justice involving research teams, community organizations, and state agencies (Chapter 6). The piece itself is short and concise, but our reflections were distilled from years of collaboration co-developing the Drinking Water Tool. More than any other chapter, it most closely represents and aligns with the title of my dissertation. In many ways, this chapter embodies the type of collaborative work I'd like to dedicate myself to past graduation.

There is also a considerable amount of personal and professional experience I gained through ERG that is not represented in this dissertation. It does not cover the year I spent as a Fulbright researcher in China, struggling with the implications of leadership changes (and therefore research priority changes) in my proposed partner organization, the Chinese Center for Disease Control and Prevention; the various projects I was a part of as a graduate student researcher, which formed the basis of my experience with community engagement; or the work that I hope to continue on bridging participatory methods and big data approaches.

In sum, my PhD journey was not a linear one, and I am grateful for all the lessons, insights, and different ways-of-thinking I was exposed to as a direct result. It was only because I was given the room to explore that I was able to build out my spatial and statistical toolbox while deepening my relationship with environmental justice theory and praxis. We are at the doorstep of a new era of environmental research (and by association, environmental justice research) characterized by increasingly complex statistical methods and "black box" machine learning. My hope is that my research on community-engaged research can ground and inform future conversations on potential tensions and synergies between algorithmic prediction and community-centered research design.

Appendix A

First Appendix: for chapter 2

Table A.1: PRISMA 2020 main checklist

Topic	No.	Item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Page 1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 1
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 1
METHODS			

Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 4
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 4
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 5
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 5

Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 5
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 5

	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Page 5
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	N/A
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	N/A

Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item 5)).	N/A
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	N/A
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Page 6, 8
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	N/A

	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Page 6

	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A
Study characteristics	17	Cite each included study and present its characteristics.	Page 6-13
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	N/A
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	N/A
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	N/A

	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			

Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 13-17
	23b	Discuss any limitations of the evidence included in the review.	Page 14
	23c	Discuss any limitations of the review processes used.	Page 18
	23d	Discuss implications of the results for practice, policy, and future research.	Page 18-19
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Page 4
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 4
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Page 4

Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	End
Competing interests	26	Declare any competing interests of review authors.	End
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	End

Table A.2: PRISMA 2020 abstract checklist

Topic	No.	Item	Reported?
TITLE			
Title	1	Identify the report as a systematic review.	Yes
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Yes
METHODS			

Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	No
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	No
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	No
Synthesis of results	6	Specify the methods used to present and synthesize results.	Yes
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Yes
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Yes

DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	No
Interpretation	10	Provide a general interpretation of the results and important implications.	Yes
OTHER			
Funding	11	Specify the primary source of funding for the review.	No
Registration	12	Provide the register name and registration number.	No

Information on PRISMA protocol were taken from: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. MetaArXiv. 2020, September 14. DOI: 10.31222/osf.io/v7gm2. For more information, visit: www.prisma-statement.org

Table A.3: Studies' community characteristics (general demographic).

Author	Study location	racial or ethnic identity	class
Williams and Florez (2002)	US	Mexican-American	mixed
King et al (2006)	US	unspecified	unspecified
Mascarenhas (2007)	Canada	First Nations	unspecified
Blakeney and Marshall (2009)	US	unspecified	unspecified

Castano-Vinyals et al (2011)	Spain	unspecified	mixed
Heaney et al (2011)	US	Black American	low-income
Otero et al (2011)	Spain	Spanish	high-income
Domenech et al (2013)	Nepal	mixed	mixed
Heaney et al (2013)	US	Black American	low-income
Wedgworth and Brown (2013)	US	Black American	low-income
Wutich et al (2013)	Bolivia, Fiji, US, New Zealand	mixed	mixed
Balazs and Lubell (2014)	US	Latino/a/x	low-income
Balazs and Ray (2014)	US	mixed	mixed
Mehta et al (2014)	Bolivia, India	unspecified	low-income
Joshi (2015)	India	Gorkha	unspecified
Corlin et al (2016)	US	Navajo	unspecified
Flanagan et al (2016)	US	unspecified	unspecified
Kozlowski and Perkins (2016)	US	White American	working class
Case (2017)	Canada	unspecified	unspecified
Rutt and Bluwstein (2017)	US	Black American	low-income
Sitthikriengkrai and Porath (2017)	Thailand	Karen (Thai)	low-income
Wright-Contreras et al (2017)	Vietnam	unspecified	mixed
Eggers et al (2018)	US	Apsaálooke (Crow)	low-income
Rall (2018)	US	Black American	low-income
Hargrove et al (2018)	US	Hispanic	low-income

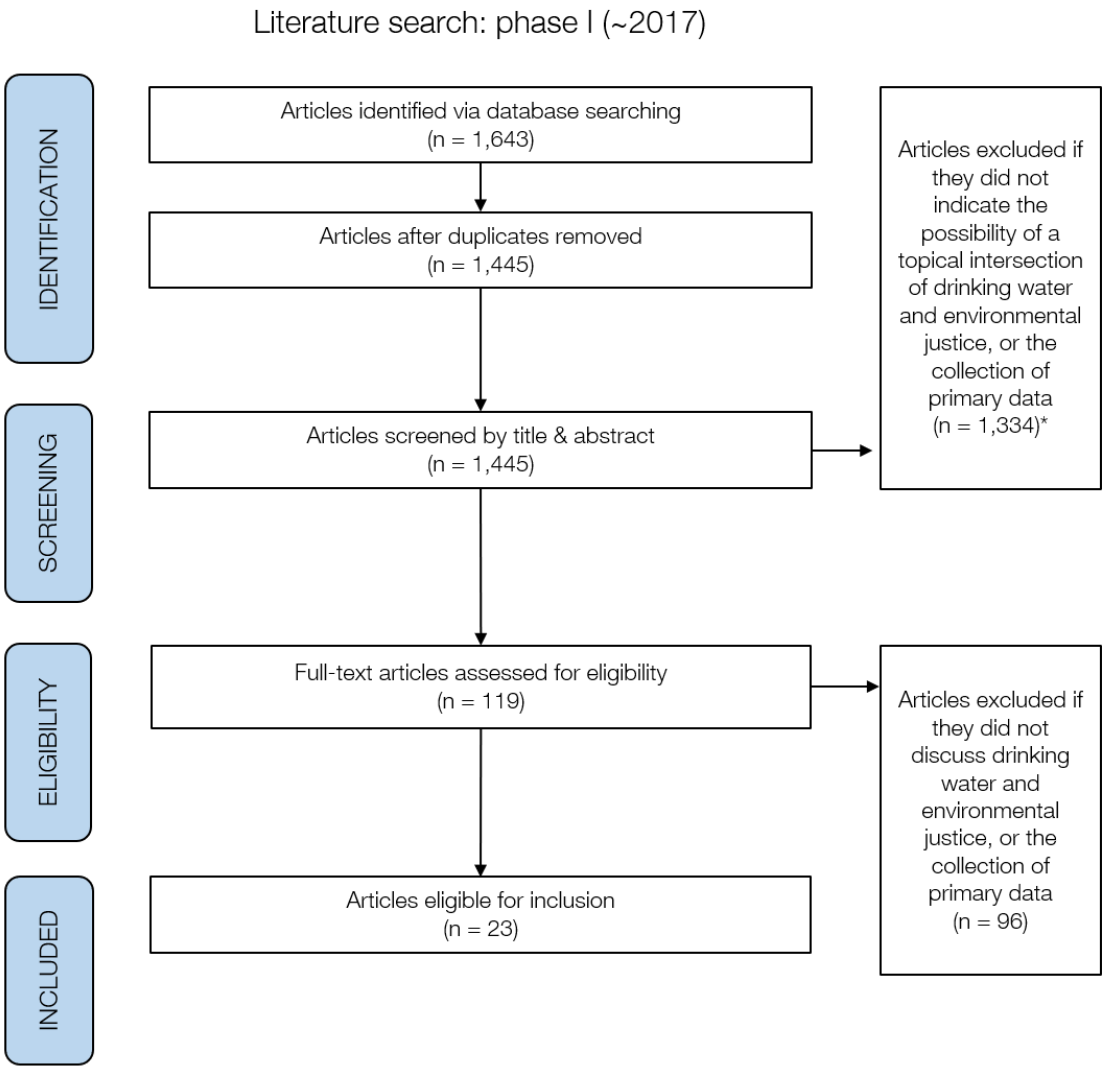
Amiri and Zhao (2019)	US	Black American	low-income
Sansom et al (2019)	US	unspecified	low-income
Krings et al (2019)	US	Black American	low-income
Hale (2019)	US	unspecified	unhoused
Stillo et al (2019)	US	Black American	low-income
Seamster and Purifoy (2021)	US	Black American	low-income
Lockhart et al (2020)	US	Black American	low-income
Dobbin (2020)	US	unspecified	unspecified

Table A.4: Studies' community characteristics (water system-related).

Author	infrastructure	incorporation	locale type
Williams and Florez (2002)	piped	incorporated	urban
King et al (2006)	mixed	unspecified	unspecified
Mascarenhas (2007)	unspecified	unspecified	unspecified
Blakeney and Marshall (2009)	mixed	unspecified	rural
Castano-Vinyals et al (2011)	mixed	unspecified	mixed
Heaney et al (2011)	mixed	unspecified	rural-urban boundary
Otero et al (2011)	mixed	unspecified	urban
Domenech et al (2013)	mixed	unspecified	mixed
Heaney et al (2013)	domestic wells / septic	unincorporated	unspecified
Wedgworth and Brown (2013)	mixed	unspecified	rural
Wutich et al (2013)	mixed	unspecified	mixed

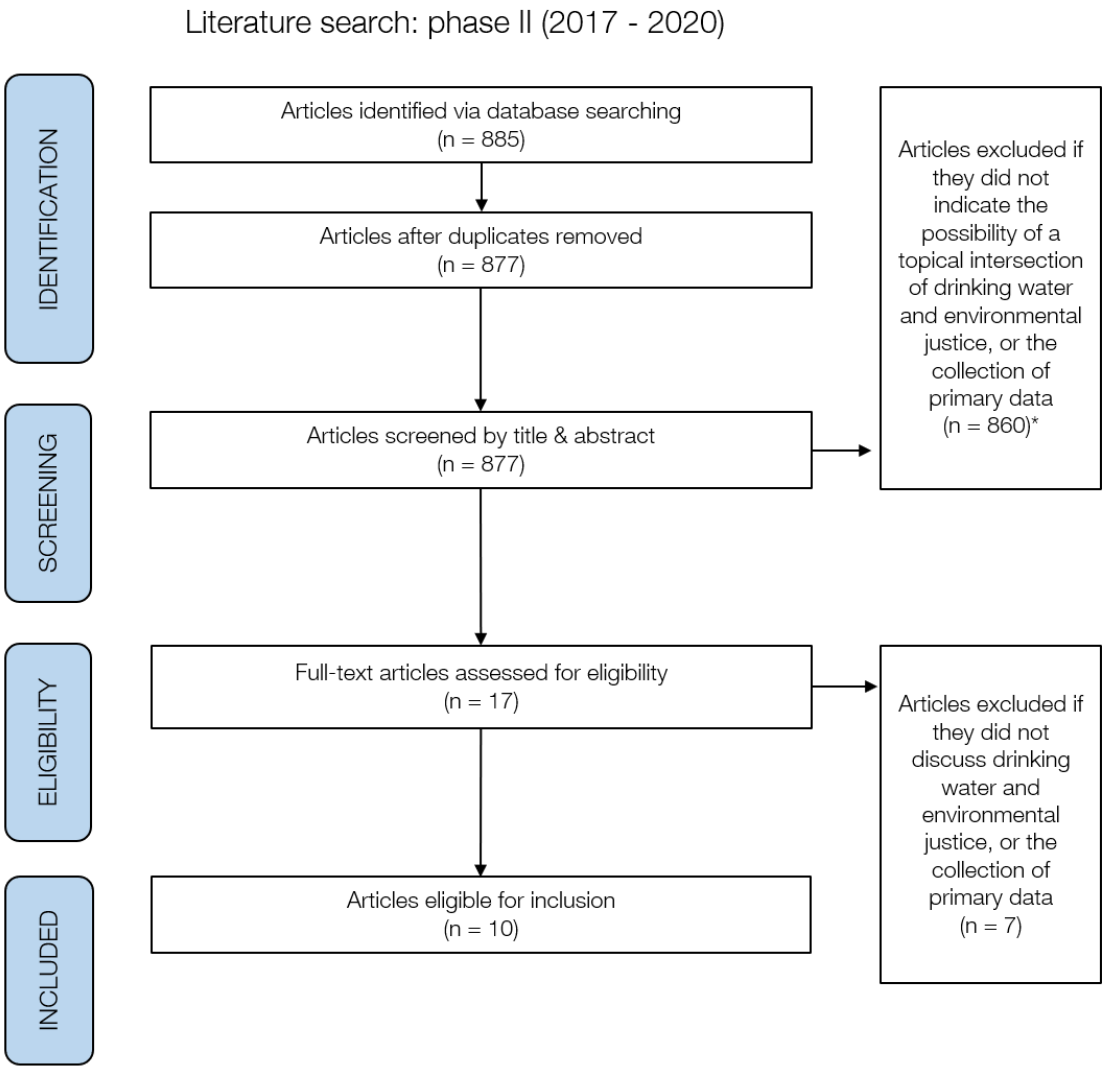
Balazs and Lubell (2014)	mixed	mixed	mixed
Balazs and Ray (2014)	mixed	mixed	mixed
Mehta et al (2014)	lack of water mains / sewer lines	unspecified	peri-urban
Joshi (2015)	mixed	unspecified	urban
Corlin et al (2016)	domestic wells	unspecified	unspecified
Flanagan et al (2016)	domestic wells	unspecified	unspecified
Kozlowski and Perkins (2016)	mixed	unspecified	rural
Case (2017)	piped	unspecified	rural
Rutt and Bluwstein (2017)	piped	incorporated	urban
Sitthikriengkrai and Porath (2017)	mixed	unspecified	rural
Wright-Contreras et al (2017)	mixed	unspecified	peri-urban
Eggers et al (2018)	domestic wells	unspecified	rural
Rall (2018)	piped	incorporated	urban
Hargrove et al (2018)	mixed	unincorporated	rural
Amiri and Zhao (2019)	mixed	unspecified	rural
Sansom et al (2019)	piped	unspecified	urban
Krings et al (2019)	mixed	unspecified	rural
Hale (2019)	water fountain	unspecified	urban
Stillo et al (2019)	domestic wells	unincorporated	peri-urban
Seamster and Purifoy (2021)	piped	unincorporated	unspecified
Lockhart et al (2020)	domestic wells / septic	unincorporated	peri-urban
Dobbin (2020)	unspecified	unspecified	rural

Figure A.1: Screening & selection flow chart, phase I (2017)



* while primary data collection was technically also a criteria of inclusion at this screening phase, we found that studies often do not elaborate on their data collection methods in their titles or abstracts; as a result, we opted to err on the side of inclusion at this step.

Figure A.2: Screening & selection flow chart, phase II (2020)



* while primary data collection was technically also a criteria of inclusion at this screening phase, we found that studies often do not elaborate on their data collection methods in their titles or abstracts; as a result, we opted to err on the side of inclusion at this step.

Note: in the original publication, Tables A.3 and A.4 were one large table. I split them up for ease of formatting in LaTeX.

Table A.5: Additional information on data extraction.

Author	Primary measure of DW access	Reports when data was collected	Conceptual ties between drinking water quality and health	Collected potential health outcome data
Williams and Florez (2002)	both	no	yes	no
King et al (2006)	unspecified	no	yes	no
Mascarenhas (2007)	quality	yes	yes	no
Blakeney and Marshall (2009)	both	yes	yes	no
Castano-Vinyals et al (2011)	quality	yes	yes	no
Heaney et al (2011)	quality	no	yes	no
Otero et al (2011)	quantity	no	no	no
Domenech et al (2013)	quantity	yes	no	no
Heaney et al (2013)	quality	yes	yes	no
Wedgworth and Brown (2013)	quality	yes	yes	yes
Wutich et al (2013)	both	no	yes	no
Balazs and Lubell (2014)	unspecified	no	no	no
Balazs and Ray (2014)	quality	yes	yes	no

Mehta et al (2014)	quality	yes	yes	no
Joshi (2015)	quantity	no	no	no
Corlin et al (2016)	quality	yes	yes	no
Flanagan et al (2016)	quality	yes	yes	no
Kozlowski and Perkins (2016)	quality	no	yes	no
Case (2017)	quality	no	yes	no
Rutt and Bluwstein (2017)	quality	yes	yes	no
Sitthikriengkrai and Porath (2017)	quality	no	yes	no
Wright-Contreras et al (2017)	both	yes	yes	no
Eggers et al (2018)	quality	no	yes	no
Rall (2018)	quantity	yes	no	no
Hargrove et al (2018)	both	no	yes	yes
Amiri and Zhao (2019)	quality	yes	yes	yes
Sansom et al (2019)	quality	yes	yes	no
Krings et al (2019)	quality	no	yes	no
Hale (2019)	quantity	no	yes	no
Stillo et al (2019)	quality	no	yes	no
Seamster and Purifoy (2021)	unspecified	yes	no	no

Lockhart et al (2020)	quantity	no	yes	no
Dobbin (2020)	both	yes	no	no

Appendix B

Second Appendix: for chapters 4-5

Table B.1: List of response and feature variables' data sources.

Response variables	**Source**
Most recent water system WQ sampling data	State Water Resources Control Board (SWRCB) Electronic Data Transfer (EDT) library

Feature variables	**Source**
Data pertaining to water system and drinking water	
Water system boundaries	SWRCB
Population demographics (block-group level)	United States Census (2017-2021 American Community Survey)
Basic water system characteristics	Safe Drinking Water Information System (SDWIS)
Historical water system WQ sampling data	SWRCB EDT library
Historical sampling data from public supply wells (ten nearest neighbors)	Groundwater Ambient Monitoring and Assessment Program (GAMA)
Data pertaining to other relevant characteristics	
$\tilde{1}31$ crop- and landcover-specific classes	National Agricultural Statistics Service (NASS)
$\tilde{1}6$ broader land cover categories	National Land Cover Database (NLCD)
Attributes of nearby confined animal feeding operations	SWRCB

Table B.3: Census variables and their codes (2017-2021 5-year American Community Survey). We did not use some of these variables in the analysis if they were highly collinear or redundant.

Category	Variable	Code
Race	White	B03002_003
	Black	B03002_004
	Asian	B03002_006
	Hispanic	B03002_012
Education	No schooling completed	B15003_002
	Nursery school	B15003_003
	Kindergarten	B15003_004
	1^{st} grade	B15003_005
	2^{nd} grade	B15003_006
	3^{rd} grade	B15003_007
	4^{th} grade	B15003_008
	5^{th} grade	B15003_009
	6^{th} grade	B15003_010
	7^{th} grade	B15003_011
	8^{th} grade	B15003_012
	9^{th} grade	B15003_013
	10^{th} grade	B15003_014
	11^{th} grade	B15003_015
	12^{th} grade, No Diploma	B15003_016
Income	Below poverty line	B17101_002
	At or above poverty line	B17101_003
Employment	In labor force	B23025_002
	In civilian labor force	B23025_003
	In civilian labor force (employed)	B23025_004
	In civilian labor force (unemployed)	B23025_005
Housing and tenure	Occupied	B25002_002
	Vacant	B25002_003
	Tenure, owner occupied	B25003_002
	Tenure, renter occupied	B25003_003

Table B.5: Land-use variables

Source	Variable
NASS	Single variables: alfalfa, almonds, barren, developed (open space, low intensity, medium intensity, high intensity), fallow / idle, forest (deciduous, evergreen, mixed), grassland / pasture, open water, shrubland, wetlands (herbaceous, woody), winter wheat / corn (double crop), cucumbers, oats / corn (double crop), grapes, olives, other crops, other tree, sod / grass / seed, cotton, citrus, triticale / corn (double crop), perennial ice / snow, avocadoes
	Grouped variables: "other nuts" (pistachios, walnuts, pecans), "wheats" (winter wheat, durum wheat, spring wheat), "legumes / nightshades" (tomatoes, potatoes, peppers, dry beans, chick peas, peas, vetch), "corns" (corn, sweet corn), "other grains" (barley, oats, triticale, rye, sorghum, rice, double crop winter wheat / sorghum, millet), "fruits" (plums, cherries, pomegranates, oranges, peaches, apples, pears, cantaloupes, nectarines, watermelons, strawberries, blueberries, miscellaneous vegetables and fruits, honeydew melons), "flowers" (clover / wildflowers, sunflower, safflower), "roots" (onions, sugarbeets, garlic, carrots, turnips), "brassica" (lettuce, greens, broccoli, cauliflower, cabbage, mustard), "herbs" (herbs, mint), gourds (squash, pumpkins)
NLCD	Developed (open space), developed (low-intensity), developed (medium-intensity), developed (high-intensity), evergreen forest, mixed forest, herbaceous, open water, barren land, shrub scrub, woody wetlands, emergent herbaceous wetlands, deciduous forest, cultivated crops, hay pasture, unclassified
SSURGO	drainage class – dominant condition, septic tank absorption fields – dominant condition, slope gradient – weighted average, manure and food processing waste – weighted average

Table B.6: Distribution of MCL and half-MCL occurrences in training and testing sets.

		training		testing	
		at or over	under	at or over	under
arsenic	MCL	52	1363	37	570
	$\frac{1}{2}$ MCL	198	1217	103	504
nitrate	MCL	26	1307	9	563
	$\frac{1}{2}$ MCL	159	1174	79	499
THA	MCL	12	913	7	390
	$\frac{1}{2}$ MCL	63	862	31	366
TTHM	MCL	22	1355	4	587
	$\frac{1}{2}$ MCL	121	1256	66	525
1,2,3-TCP	MCL	40	1255	15	540
	$\frac{1}{2}$ MCL	45	1250	15	540

Table B.8: Model input features organized into broad categories.

Drinking water	Demographic	Physical
EDT water sampling history; Nearby GAMA well samples (nearest neighbors)	Commonly used social indicators in environmental justice research for measuring vulnerability	Proximate crops; Proximate land uses; Presence of animal feeding operations

Table B.9: Mann-Whitney U test results for sample counts and model outcomes (false negatives and true positives) for arsenic

	elastic net	neural net	svm (radial)	svm (polynomial)	random forest
sample count (MCL)	0.98	0.77	0.99	0.99	0.99
sample count (1/2 MCL)	0.0016	0.0024	0.003	0.003	0.18

Figure B.1: Time series plot of EDT sample counts over time. (A) depicts the proportion of MCL to non-MCL samples, and (B) depicts the same for the $\frac{1}{2}$ MCL.

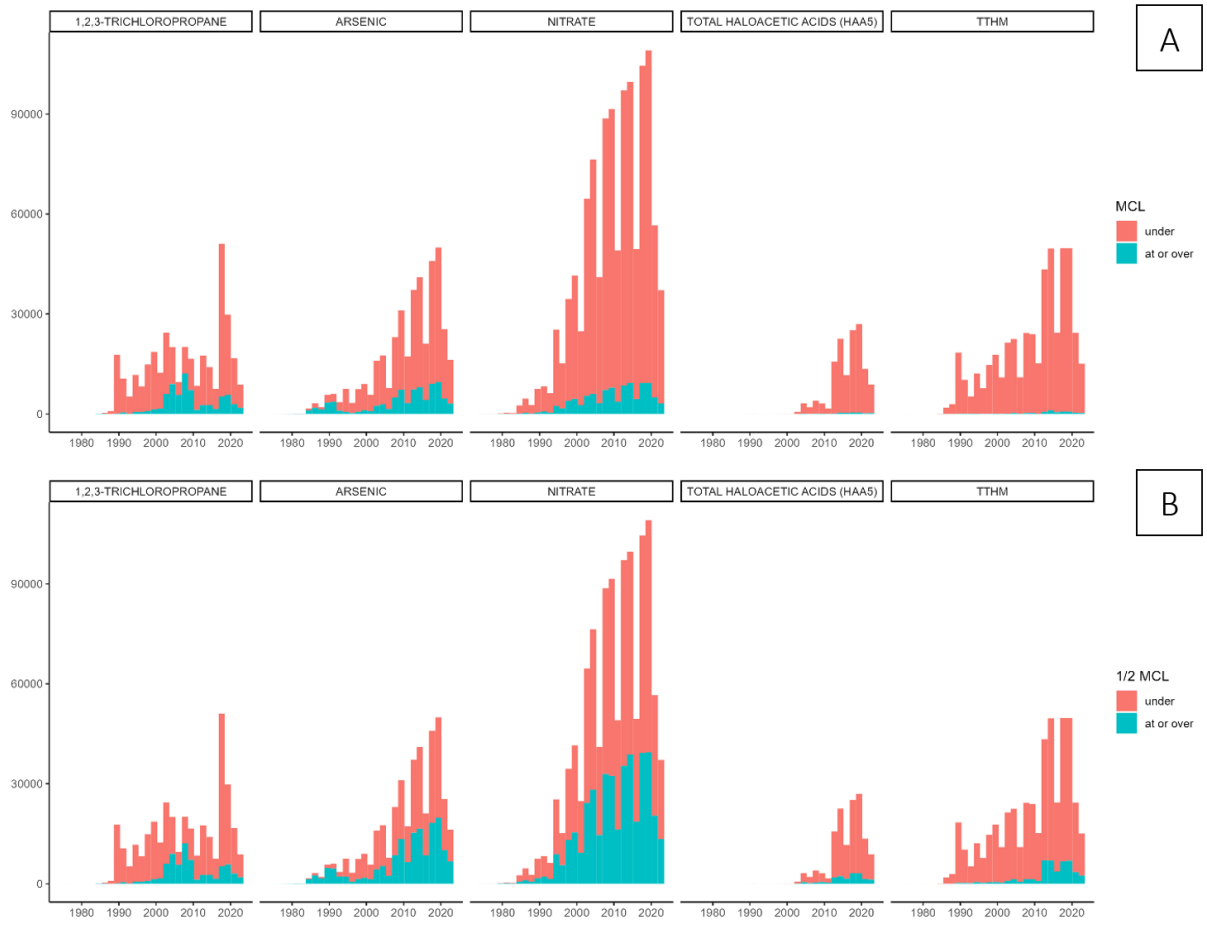


Figure B.2: Process of determining the "most recent sample" for systems with MR violations.

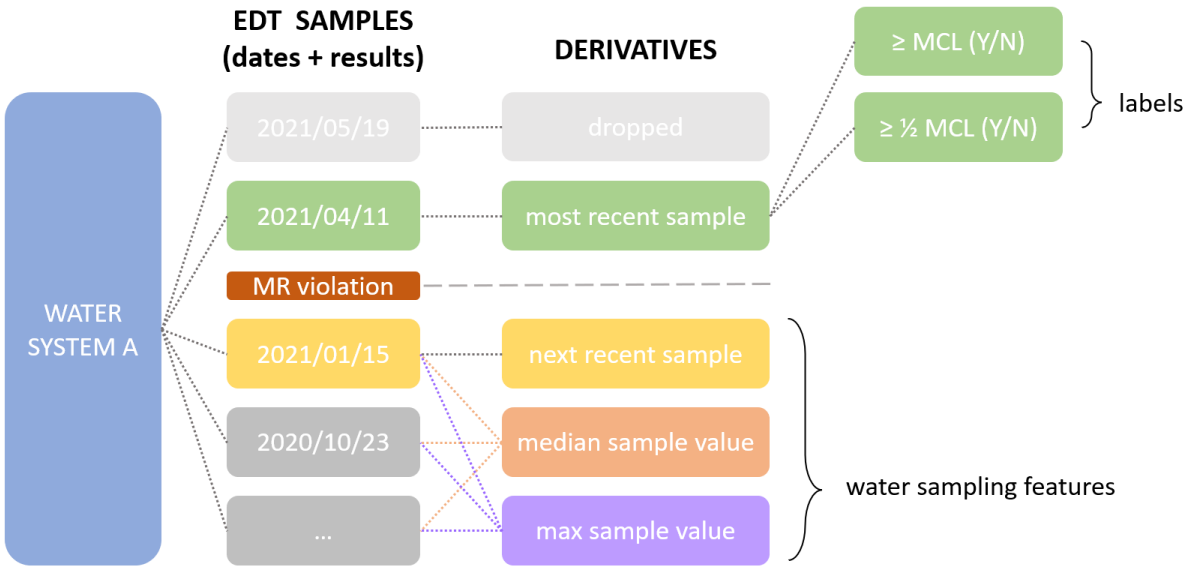


Figure B.3: Model performance for all models and contaminants, MCL and half MCL (F2 score).

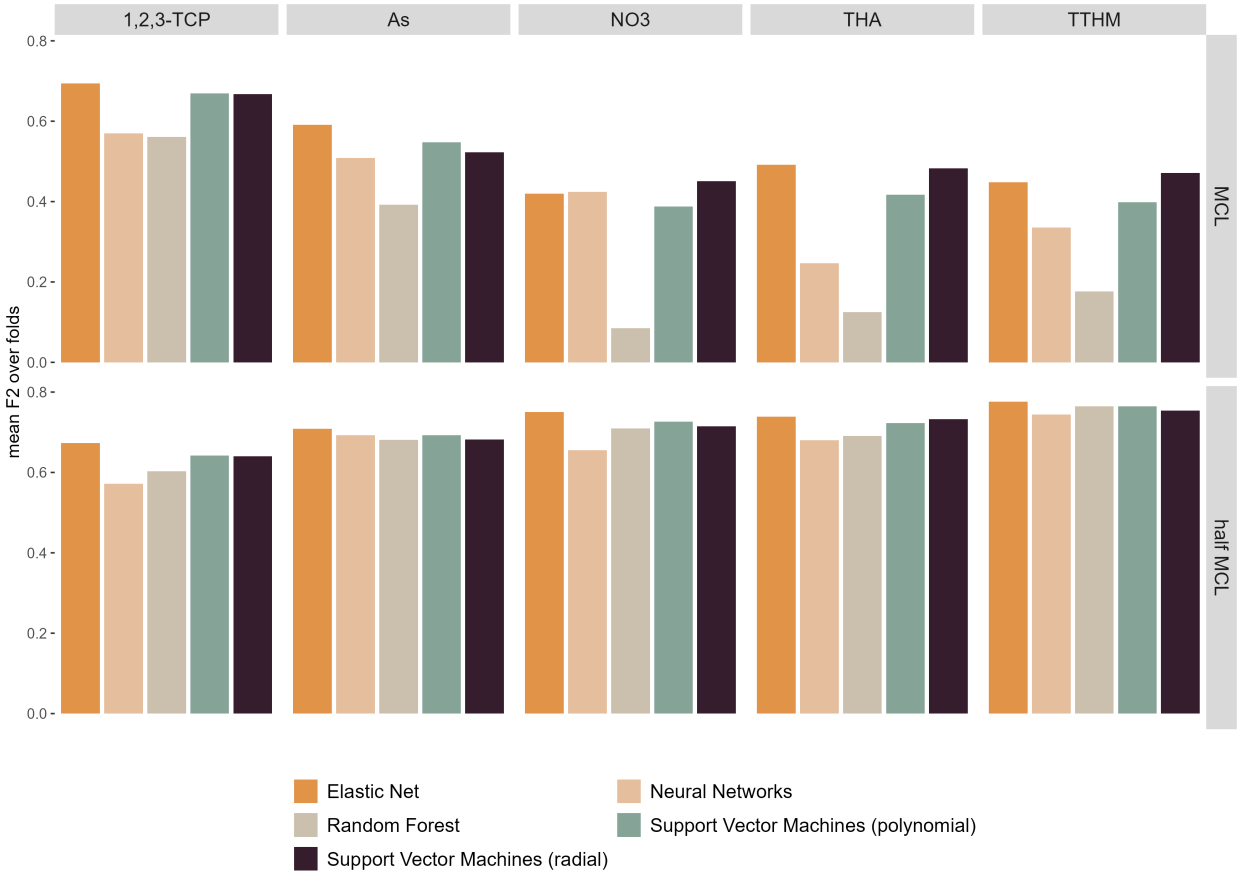


Figure B.4: Confusion matrices for test sets and MR sets using Elastic Net (label = MCL), all contaminants. Results for all other model and contaminant combinations available upon request.

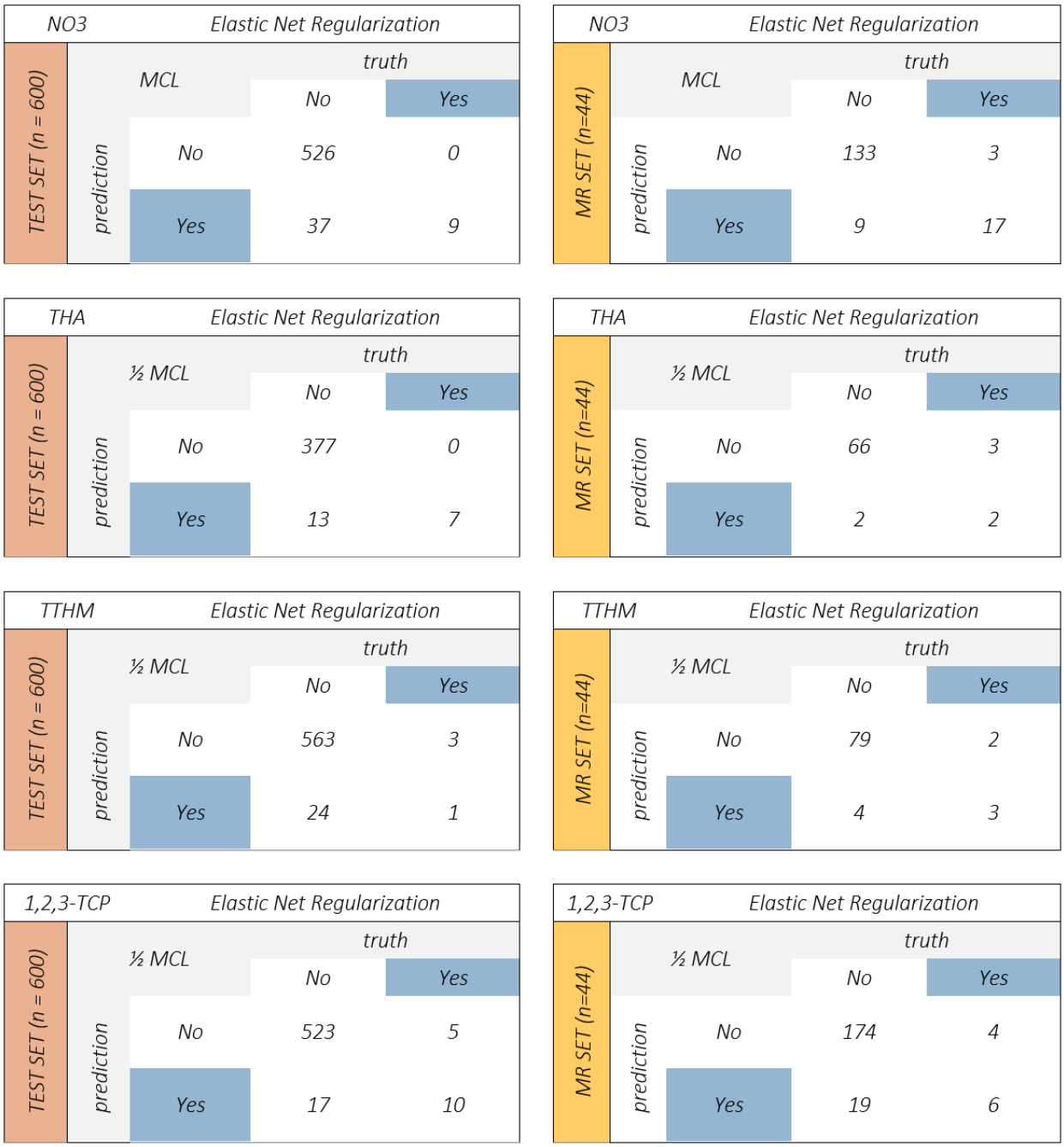


Figure B.5: Confusion matrices for test sets and MR sets using Elastic Net (label = $\frac{1}{2}$ MCL), arsenic.

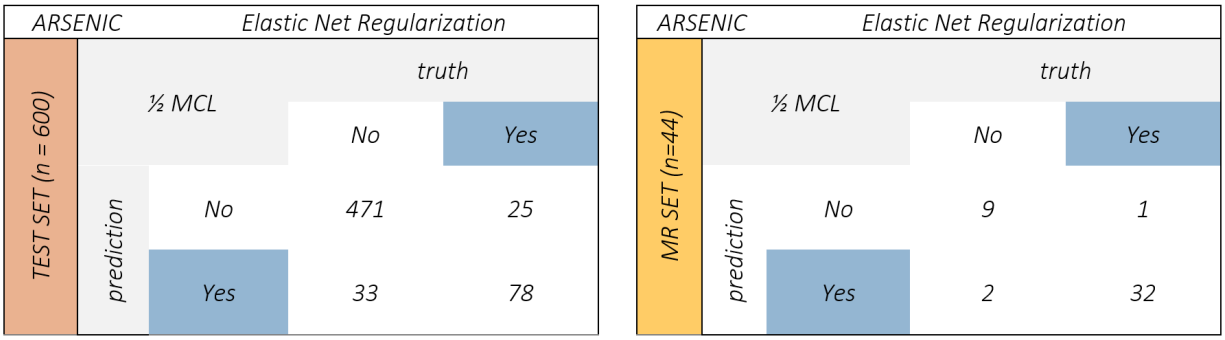


Figure B.6: Example of a resource prioritization system using classification probabilities.

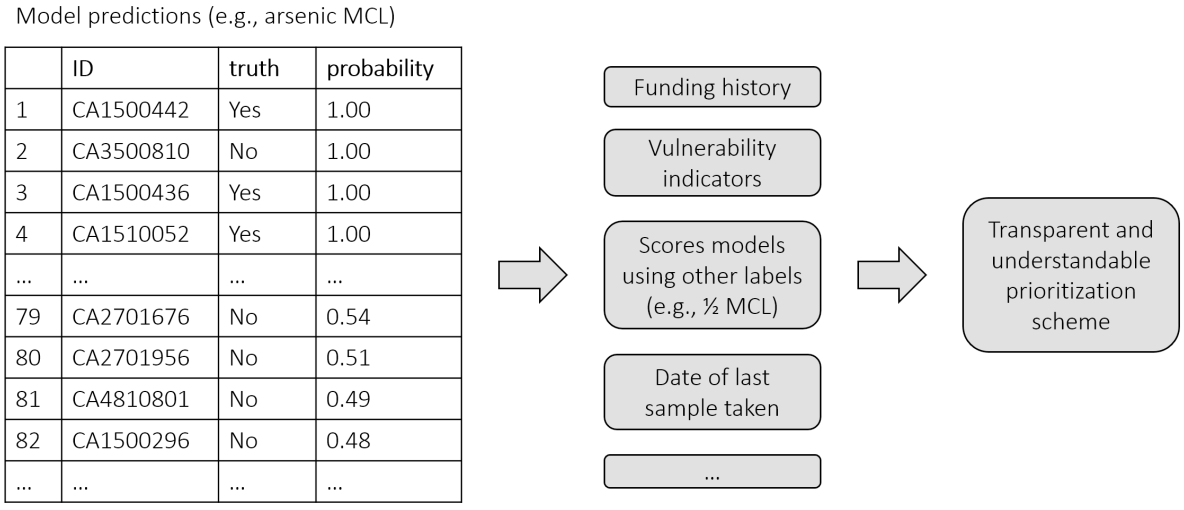
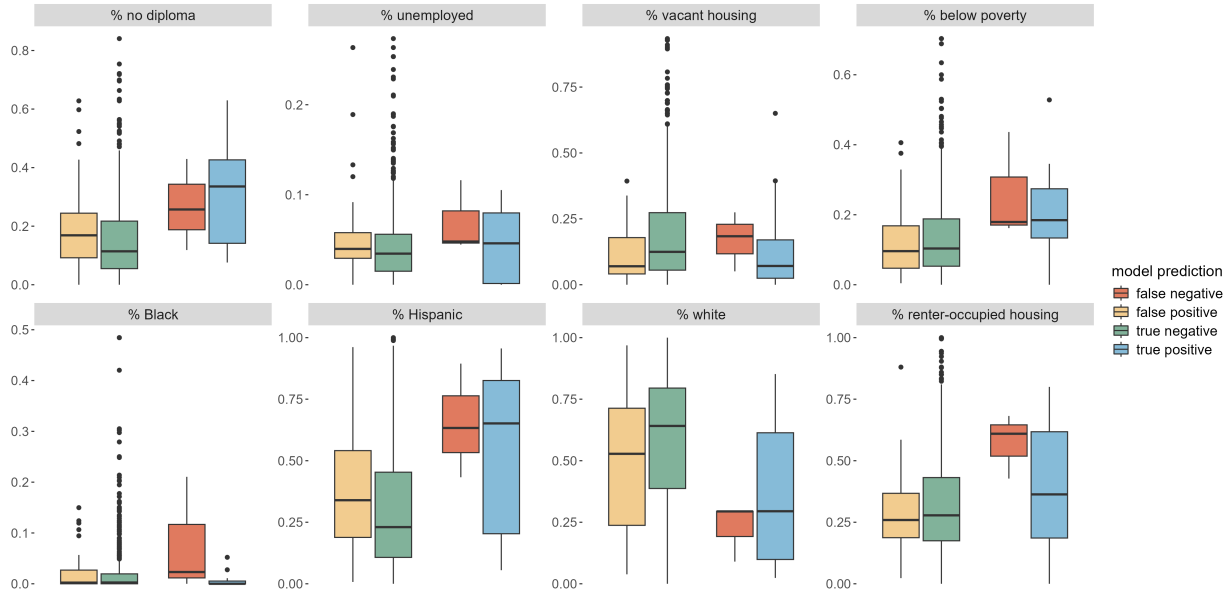


Figure B.7: Distribution of demographics indicators for the four other contaminants across classification outcomes using Elastic Net.

(a) NO_3



(b) 1,2,3-TCP

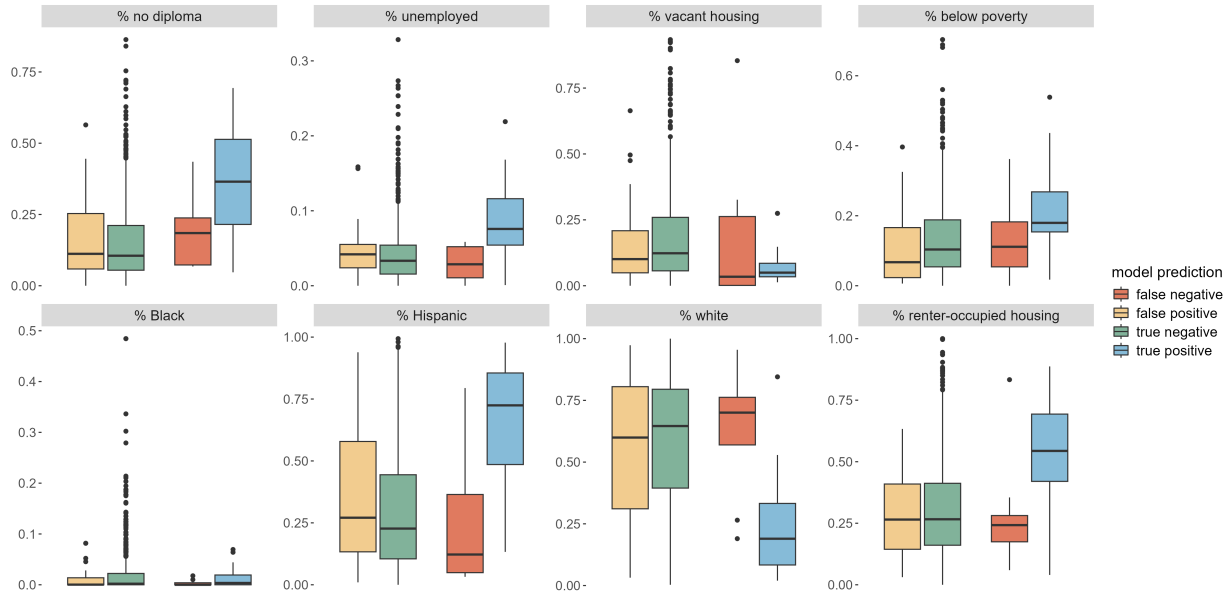
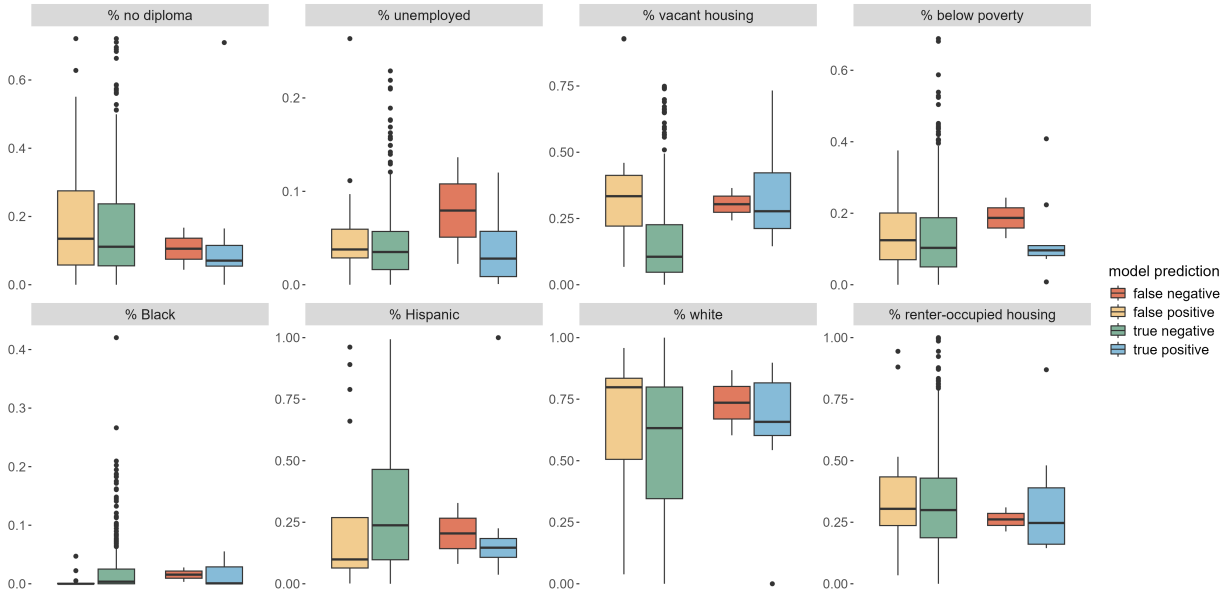


Figure B.7: Distribution of demographics indicators for the four other contaminants across classification outcomes using Elastic Net.

(c) THA



(d) TTHM

