Essays in Monitoring and Enforcement of Environmental and Agricultural Policies

by

Peiley Lau

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael L. Anderson, Chair
Professor Solomon M. Hsiang
Associate Professor M. James Sallee

Spring 2021

Essays in Monitoring and Enforcement of Environmental and Agricultural Policies

Abstract

Essays in Monitoring and Enforcement of Environmental and Agricultural Policies

by

Peiley Lau

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Michael L. Anderson, Chair

Chapter One examines how firms respond to changes in environmental regulations that rely on self-reporting. Routine natural gas flaring at oil wells in the U.S. has soared in the last decade as advances in hydraulic fracturing enabled oil production in previously unprofitable regions, leading to concerns over climate and health damages. What is the effect of environmental policy aimed at reducing gas flaring? How do responses in self-reported versus remotely detected flaring data compare? I employ a difference-in-differences approach comparing North Dakota to Montana wells before and after the policy implementation to estimate the effect of the policy on reported and remotely-detected flaring. I construct a novel monthly well level remotely detected flared volume dataset, that is comparable to administrative well level flared volume data between 2012-2019 in the Bakken Shale. I document both an unprecedented uptick in natural gas flaring and an increasing gap between self-reported and remotely detected flaring. During the first two years of the policy, wells reduced flaring in compliance with the policy and reported truthfully; however, in the last three years, compliance fell, and half the gains in reported flaring reduction are from misreporting. This paper fits into a growing body of literature utilizing newly available satellite data to evaluate the effectiveness of environmental policy and measure the activity of the oil and gas industry. This paper fits into a growing body of literature utilizing newly available satellite data to evaluate the effectiveness of environmental policy and measure the activity of the oil and gas industry.

Chapter Two presents a novel hydrological approach to identifying and addressing nonpoint source pollution from agricultural runoff in national-scale river networks.

Freshwater and coastal ecosystems worldwide increasingly suffer from eutrophication caused by nonpoint source agricultural runoff, resulting in detrimental environmental, social, and economic outcomes.

The main challenge for policymakers in regulating and reducing nonpoint source water pollution is the difficulty in identifying the source of nutrient runoff and measuring whether policy changes effectively reduce pollution.

Our paper addresses this problem through modeling pollutant flux in national-scale river networks and identifying which catchments exhibit higher nutrient loss at a spatially and temporally granular level.

Our spatial approach combines publicly available water data and hydrographic methods to localize annual nitrogen and phosphorus based pollutant loads at the sub-watershed level in the United States Mississippi River Basin (USMRB) and the country of New Zealand between 1981 – 2018.

We found that the distribution of nitrogen and phosphorus pollutant loads are spatially heterogeneous even within the same region in the two watersheds that we evaluated. We found that average total nitrogen and total phosphorus runoff was consistently highest in the Southland and Gisborne Regional Councils in New Zealand. In the USMRB, average total nitrogen and total phosphorus runoff was highest in the central region of the watershed.

Our results are useful for evaluating historical water pollution and establishing baseline pollutant loads, measuring changes in agricultural runoff over time, and monitoring the effects of changes in water quality regulations and land use activities at annual and sub-annual time steps. The use of our spatial model can be used to shed light on the effect of water pollution on human health, how water pollution differentially impacts disadvantaged populations and communities of color, and the distribution of water pollution under varying climate change scenarios.

Chapter Three investigates whether there is evidence of multinational oil producers engaging in base erosion and profit shifting practices. I do this through estimating the responsiveness of reported profits of oil companies with respect to corporate tax rate differences. The elasticity of reported profits with respect to changes in corporate tax rates helps to elucidate whether companies employ artificial profit shifting techniques. My identification strategy exploits exogenous changes in corporate tax rate differentials for multinational corporations whose foreign subsidiaries pay taxes in the country where they are located. Identifying a significant and substantive corporate tax differential elasticity to reported profits will provide evidence that firms extralegally choose where to report their profits based on the locations of their subsidiaries, thereby artificially shifting their profits from high to low tax jurisdictions.

Contents

List of Figures

List of Tables

Acknowledgements

This journey would not have been possible without the unwavering support and encouragement from so many wonderful people. Thank you from the bottom of my heart.

Sol, thank you for taking me on as a research assistant during my first year and giving me the opportunity to join the lab. Little did I know what a large and overwhelmingly positive impact that would have on my research career. Joining GPL has fundamentally shaped how I think about and approach research, and also what my role and responsibility as an economist should be. You have taught me how to take risks and to not let my fear of the unknown or failure get in the way of working on hard and important problems.

Jim, thank you for your never ending fountain of wisdom and advice. I am so grateful for your mentorship, and have benefited beyond words from your insightful comments, careful feedback, and unwavering support. Thank you for encouraging me to pursue research in the space of oil and gas industry - something I would have never considered without your suggestion; going down this avenue has proven to be a very fruitful and rewarding line of research that I am excited to continue on.

Michael, thank you for taking the time to dispense sage econometric advice and being available to talk through any questions that I had related to my job market paper.

Sofia and Max, thank you for your help and encouragement during the job market search in the middle of a pandemic.

Carmen and Diana, you are the bedrock of ARE. Thank you for the Sliver pizzas that fueled us through studying for midterms, delicious pan dulce that always brings a smile to my face, and the kindness and generosity that you show to the ARE student community. You truly make ARE feel like a family.

Thank you to all the ARE students who over the years have advocated fearlessly for equity, inclusivity, and diversity both within the ARE community as well as economics discipline at large.

To Tamma, JP, Jon, Megan, Terin, Andy, Ian, Rachel, Hannah, Gabe, Luna, Esther, Seb, Jaecheol, Daniel, Jack, and Tiffany - thank you all for making the Global Policy Lab such an exciting and intellectually rewarding community. I have learned such much from the interesting, challenging, and important interdisciplinary environmental and climate research that you all do. I am humbled by your dedication to making this world a better place.

To Jeanette, Julia, Sandy, Emma, and Kenny - wow, getting to work with such a talented, passionate, hardworking, funny, kind, and supportive team has been a dream come true. I had no idea when I joined the New Zealand team just how truly lucky I was. I have learned so much from each of you, and you have helped me to become a better researcher, mentor,

and collaborator. More importantly, getting to see your smiles each morning made the bike ride uphill to the stadium worth it. I am already counting down the days until we are able to go tramping in New Zealand again.

To my ARE 2015 cohort: Kate, Scott, Gabe, Karl, Eleanor, Megan, Hal, Andy, Ben, Claire, Wenfeng, Elisa, and Derek - you are all such amazing individuals and I am deeply grateful to have shared this roller coaster of a journey that is graduate school with you all. You have kept me sane, comforted me through disappointing real analysis and micro midterms, cheered me on during seminar presentations and research breakthroughs, and brought me so much laughter and joy over these past six years. I am so excited to celebrate your continuing academic and professional accomplishments in this next chapter of our careers, and to also call you friends for life. I look forward to many many more river floats, camping, mini-golfing, broom balling, and general merry making with you all.

Mom, Dad, and Gordon, thank you for being there for me since day one. You have contributed to my achievements more than you can ever know – from fostering my love of reading from an early age to encouraging me to be a lifelong learner to supporting my dreams at every step of the way.

Dan, as the great e.e. cummings once said, "you are my sun, my moon, and all my stars." Thank you for sharing this one wondrous life with me.

Chapter 1

The Extraordinary Rise of Unreported Gas Flaring Among Fracking Firms[1]

1.1 Introduction

Policies designed to correct market externalities are more effective when regulators are able to monitor and enforce them. Otherwise, without the ability to effectively ensure compliance, such policies are largely non-binding guidance (Gray and Shimshack 2011; Shimshack 2014). Detecting instances of strategic misreporting and evasion by firms and individuals has spanned multiple areas of economic inquiry (Zitzewitz 2012), ranging from multinational corporate tax evasion (Fisman and Wei 2004; Zucman 2015; Alstadsæter, Johannesen, and Zucman 2019; Guyton et al. 2020), to trade (Fisman, Moustakerski, and Wei 2008; Mishra, Subramanian, and Topalova 2008) to finance (Kacperczyk, Sialm, and Zheng 2008; Ljungqvist, Malloy, and Marston 2009), to natural resource extraction and environmental pollution. Given that such policies are designed to correct for market failures that generate social externalities, profit-maximizing firms or utility-maximizing agents subject to such regulations are privately motivated to circumvent these policies by obscuring their true behavior, and thus not have to internalize the negative social externalities of their actions. While the specific economic outcomes of interest differ, forensic economic inquiry rests on the ability to discern reported or observed behavior from true, but often deliberately obfuscated, behavior.

In many economic and policy applications, it can be difficult to directly measure hidden or illegal behavior (Zitzewitz 2012). This is particularly challenging when the outcome of interest involves the transaction of intangible assets, such as the sale of intellectual property. In contrast, environmental policies that regulate natural resources, such as surface water or

[1]I thank Solomon Hsiang, James Sallee, and Michael Anderson for their invaluable advice and support. I also thank Karl Dunkle-Werner, Gabriel Englander, Scott Kaplan, Kate Pennington, Megan Lang, Andy Hultgren, Ian Bolliger, Emma Krasovich, Julia Longmate, Jeanette Tseng, Sandy Sum, Kendon Bell, Alan Krupnick, Brian Prest, William Wheeler, Tamma Carleton, Susanna Berkouwer, Nick Hagerty, and Louis Preonas for thoughtful comments and feedback. I also thank seminar participants at University of California Berkeley, Stanford University, Resources for the Future, Environmental Protection Agency, and The Workshop in Environmental Economics and Data Science.

fossil fuels, regulate tangible assets. When individual or firm behavior affects these amenities, their impact can be directly measured, such as through hand collected water samples or physical audits of oil wells. However, given that some of these firms operate in more remote or inaccessible regions, a difficulty that persists for policymakers is how to monitor in these areas. In practice, enforcement of environmental policy enforcement often depends on using self-reported data to determine compliance.

In this paper, I use the context of the burning of natural gas, known as flaring, to examine the effectiveness of an environmental policy that relies on self-reported data. Flaring, or burning, of natural gas is a common method employed by the oil industry worldwide to dispose of natural gas produced from oil wells when gas capture is uneconomical. This is because for oil producers, their primary commodity is crude oil, rather than the co-produced natural gas. Furthermore, unlike oil, which can be transported by truck from the well location, gas can only be shipped through gas gathering pipelines from the well to processing facilities. Given that the construction of gas capture infrastructure is costly, oil producers may choose to burn the gas if capturing the gas is uneconomical. Regulators seek to limit flaring because it emits greenhouse gases and pollutants, and also because routine flaring burns off methane that could be captured and used for energy consumption.

I provide evidence of significant industry wide misreporting, and show how policymakers can leverage new satellite data to overcome monitoring challenges. Specifically, I use a differences-in-differences specification to examine the causal impact of a series of regulatory changes in North Dakota gas flaring policy on gas flaring behavior and environmental misreporting, with Montana wells serving as my control group. To my knowledge, this is the first paper to combine administrative and satellite data on flaring that span close to a decade to evaluate how the oil companies' response to policy evolves over longer time frame, as both the policy undergoes several iterations as well as significant changes in market and political conditions during my sample time period. I also examine the role of and to what extent three main mechanisms, (1) oil and gas prices, (2) well connection to gas gathering pipelines, and (3) congestion at regional gas processing facilities, explain flaring behavior.

In order to measure the extent of misreporting, I construct a novel monthly well level remotely detected measure of flared natural gas volumes, which is directly comparable to the self-reported well level flared volume monthly data released by North Dakota and Montana oil and gas operators. I find evidence that during the first two and a half years of the policy implementation, the policy was largely successful in increasing gas capture both at the firm and well level, and the self-reported and satellite measured flaring behavior closely track each other. However, I demonstrate that a wedge between the two datasets emerges in 2017, of which only half of the flared volume is explained by rising oil prices and gas capture infrastructure availability. This suggests that the other half of the flared gas volume isn't attributable to market dynamics or infrastructure constraints, *per se*, but rather deliberate misreporting.

My paper builds on prior work by Lade and Rudik (2020) and Lee (2020) that examine the impact of the North Dakota flaring policy on natural gas flaring and misreporting. Lade and Rudik (2020) find that the flaring regulation reduced flaring rates at wells drilled after Octover 2014 by 10-19.5 percentage points. Lee (2020) finds evidence of misreporting in the 18 months following the October 2014 implementation, and that misreporting accounts

for one-third of the 16.8% decline in flaring rates. My paper validates the conclusions from both Lade and Rudik (2020) and Lee (2020), but I utilize a richer dataset and extend my sample for an additional three to four years, which allows me to capture significant changes in flaring behavior in response to changes in market conditions.

My research contributes to three main strands of literature. The first is research that examines the economics and impacts of the unconventional shale oil and gas boom. This includes work evaluating the welfare and market impacts of newly accessible petroleum resources, which find positive wage rate and housing value impacts for local communities (Bartik et al. 2019; Feyrer, Mansur, and Sacerdote 2017; Jacobsen 2019) as well as positive welfare impacts for natural gas consumers (Hausman and Kellogg 2015). However, there is evidence that there are negative impacts on housing value for houses reliant on groundwater (Gopalakrishnan and Klaiber 2014; Muehlenbachs, Spiller, and Timmins 2015). In addition to the economics of fracking, much recent work has also examined the public health and environmental impacts of fracking, and found evidence that it has lead to increased water pollution (Olmstead et al. 2013; Hill and Ma 2017) and negative health outcomes for infants who had *in utero* exposure (Hill 2018; Currie, Greenstone, and Meckel 2017).

In addition to the local impacts of unconventional shale boom, there is also work examining firm learning and decision making within the context of the oil and gas industry (Covert 2015; Fitzgerald 2015; Steck 2018; Fetter et al. 2018; Lange and Redlinger 2019; Agerton 2020), and how firms respond to changes in environmental regulations (Muehlenbachs 2015; Fetter et al. 2018; Boomhower 2019; Lee 2020).

The second strand is a growing body of economic and policy work utilizing remotely sensed data on extractive resources and/or environmental outcomes to detect illegal behavior or ground truth limited monitoring data. A lot of this work utilizes satellite measurements of air quality concentrations to overcome limited air pollution monitors (De Marchi and Hamilton 2006; Grainger, Schreiber, and Chang 2016; Karplus, Zhang, and Almond 2018; Fowlie, Rubin, and Walker 2019; Grainger, Schreiber, and Chang 2016). In particular, Karplus, Zhang, and Almond (2018) document environmental misreporting by firms. They find evidence that the decline in SO2 emissions reported by Chinese coal fired power plants after the implementation of a new national air emissions standard are twice as large as the decline detected in satellite air quality concentration measurements, and that this effect is magnified in regions with stricter limits. Within the context of extractive resources, Saavedra and Romero (2017) use satellite imagery to measure the extent of illegal mining in Colombia to evaluate the effect of a tax reform. With respect to satellite imagery in the oil and gas industry, Do et al. (2018) utilize the NASA-NOAA VIIRS satellite product to estimate oil production in ISIS-controlled regions where such data are difficult to come by. Using the same dataset, Lee (2020) also examines the effect of the North Dakota flaring regulation on reported and remotely detected flaring, and finds evidence of misreporting.

Lastly, and more broadly, my research contributes to work examining the effectiveness of environmental regulations and monitoring. Gray and Shimshack (2011) and Shimshack (2014) review empirical evidence related to environmental regulatory institutions and behavior and the effect they have on deterring polluting behavior. By and large, they find that regulatory monitoring and enforcement is effective at both reducing firm specific polluting behavior and general pollution within the industry, and that regulations do reduce

emissions. However, Shimshack (2014) points out that regulations are effective when they are enforceable. Recent work ranging from oil discharges in the shipping industry (Vollaard 2017), to automobile emissions (Oliva 2015) and air pollution (Zou 2018) find evidence of strategic noncompliance to bypass environmental regulatory standards.

The rest of the paper is structured as follows: Section 1.2 provides context about natural gas flaring and the oil and gas industry in North Dakota; Section 1.3 explains data used and constructed for the empirical analysis; Section 1.4 provides details on the causally identified estimation specification; Section 1.5 presents the results of policy impact on flaring and misreporting in North Dakota; Section 1.6 documents additional analyses to tease out the mechanisms driving the misreporting; and Section 1.7 concludes.

1.2 North Dakota Flaring Context

1.2.1 Unconventional oil and gas development and hydraulic fracturing in the Bakken Formation

In North America, the Bakken Formation located in the Williston Basin and underlying parts of North Dakota and Montana, in the US, and Manitoba and Saskatchewan, in Canada, has become one of the top oil producing regions in recent years. Within North Dakota, the majority of the oil boom is located in the Three Forks Formation within the Bakken Formation.

Given that the presence of the oil and gas industry in the Bakken is relatively nascent, capital investment in and availability of costly infrastructure such as oil and gas pipelines or processing facilities, has been relatively sparse. For the associated gas produced at the oil wells, capturing and transporting this gas is a larger challenge. Unlike crude oil, the method of transporting natural gas from the wellhead to a gas processing facility is via a gas gathering pipeline. However, connecting a well to a gas gathering pipeline is costly and requires time to construct. For well operators, whose main interest is in commencing oil production, and for which gas capture is an afterthought, they can resort to other means for dealing with the associated gas instead.

One option is to use the natural gas to power equipment at the well site. Once this need has been met, any excess natural gas that isn't captured, can be disposed of through venting, which is direct release of the natural gas into the atmosphere, or flaring, which is burning off of the natural gas.

1.2.1.1 Historical Context

The combination of technological advances and favorable crude oil prices increased oil production in the Bakken, with oil prices reaching over $100 per barrel in early 2014 (average West Texas Intermediate crude oil price was just over $100 in March 2014) and statewide oil production peaking in December 2014 at 1.23 million barrels per day. However, as global oil prices dramatically fell to under $50 per barrel by January 2015 and stayed low over the following two years, Bakken oil production also stagnated, dipping below 900,000 barrels per

day in late 2016. During this time period, pipeline and gas plant infrastructure construction and investment also dropped. However, a lot of oil companies invested in R&D to improve fracking technology in order to decrease the break-even price of extracting a barrel of oil.

In 2017, the global oil prices began to rise again, as a result of both OPEC curtailing their oil production and disruption to Venezuela's oil production. As oil prices climbed back up to $60 per barrel by the end of the year, this led to a renewal of oil production, both in the number of rigs drilling new wells and the amount of oil produced in North Dakota. Higher oil prices meant that producers could drill in less profitable areas that have higher break even costs. Monthly North Dakota statewide oil and gas production continuously broke prior records, resulting from (i) increased drilling leading to record number of producing wells, (ii) increased productivity at each well (due to improved technology), and (iii) drilling in the productive Bakken core, which has a higher gas to oil ratio.

Additionally, the election of Donald Trump as U.S. president in November 2016, and his appointing of Scott Pruitt as director of the EPA resulted in favorable conditions for oil and gas companies. The Trump administration rolled back many environmental regulations affecting oil and gas production passed by the Obama administration, and this new regulatory paradigm increased certainty and decreased costs for oil companies. In February 2017, Donald Trump approved the North Dakota Access Pipeline (NDAPL) after many months of protests and uncertainty over the future of the pipeline. The pipeline was a game changer for North Dakota crude oil since it had a 500,000 barrel per day capacity (or about half of the state's oil production) and significantly lowered the shipping cost by $3-$5 per barrel. This further increased confidence in the Bakken oil landscape and helped the state to secure corporate investment, as the savings in transportation cost was an important factor for oil companies in deciding whether to increase oil production in North Dakota.

In December 2017, North Dakota Lieutenant Governor Brent Sanford said North Dakota crude oil was more competitive thanks to NDAPL, and that the pipeline led to an increase in oil production. But he also expressed concern about availability of pipeline and gas processing facilities to handle increasing gas production. Furthermore, he said that oil production increased more than anticipated, in particular, what gas companies had anticipated. As such, he predicted that flaring percentage will increase because of bottlenecks in the ability to capture the gas. During this time period, the drilling of new oil wells and commensurate gas production associated with these wells did indeed outpace the construction of gas transport and processing infrastructure.

1.2.2 Flaring

The flaring of natural gas by the oil industry occurs on all continents (see Figure 1.1) (with the exception of Antarctica), with Russia, Iraq, Iran, Nigeria, Venezuela, Algeria, the United States, Libya, Mexico, and Kazakhstan consistently ranked among the top ten nations that flare natural gas. Annually, oil flaring from around the globe releases 300 million tons of carbon dioxide per year, generating air pollution that contributes to climate change and releases carcinogens and other toxins linked to increased respiratory diseases in local communities.

1.2.3 Externalities associated with flaring

While flaring has increased in the U.S. as the number of fracked wells has grown, flaring has and continues to be a challenge globally in oil and gas production. Flaring in many regions around the world do not result from fracking, but rather oil production in general. According to the World Bank (World Bank 2020), in 2019, 150 billion cubic meters of natural gas were flared at the well site around the world. This amount of methane, which was just burned off, was enough to supply the entirety of the continent of Africa's electricity consumption needs, or the equivalent of generating over 800 billion kWh of electricity. This flaring released over 400 million tons of carbon dioxide equivalent emissions, equal to $20 billion, when using a social cost of carbon priced at $50 per ton. The World Bank Zero Routine Flaring by 2030 on flaring reduction calls for governments, oil companies, and other institutions to achieve zero routine flaring by 2030 (World Bank 2015), where routine flaring is defined as flaring of natural gas during oil production when there is insufficient gas capture infrastructure, rather than combusting of natural gas for safety or non-routine reasons. Elvidge et al. (2018) estimated that globally, eliminating routine oil flares would help achieve two percent of the greenhouse gas emission reductions countries committed to in the Nationally Determined Contributions (NDC) outlined in the Paris Climate Agreement, and that five countries (Yemen, Algeria, Iraq, Iran, Gabon, and Ecuador) could entirely meet their NDC reduction targets.

Locally, the effects of flaring have been causally linked to negative health outcomes in North Dakota and Texas (Blundell and Kokoza 2018; Cushing et al. 2020). When flaring combustion is complete, then the output of the flaring process is mainly releases CO_2 and water The EPA models flare combustion with a 98% flare efficiency for methane and ethane, but this is highly variable to wind speed, gas flow rate, and aeration, and clean combustion is rare (Egwurugwu and Nwafor 2013; Gvakharia et al. 2017). Gvakharia et al. (2017) analyzed the plume sample from 37 unique flares in the Bakken, and found much higher rates of incomplete combustion than what was modelled by the EPA. This is problematic because they found that incomplete combustion emits black carbon, methane, ethane, and other pollutants in to the atmosphere, and that flare plumes with poor combustion efficiences have a disproportionate impact on methane emissions. Under their estimates, methane emissions from flaring would actually double, from around 1.1 - 1.9 million tons of methane/year to closer to 2.8-4.6 million tons methane/year. When incomplete combusion occurs, then can release pollutants such as benzene, VOCs, polycyclic aromatic hydrocarbons, nitrogen monoxide, carbon, and carbon monoxide that have been linked to cancerous or negative respiratory health outcomes.

1.2.4 North Dakota Industrial Commission Policy Regulating Routine Flaring

After increasing public outcry and lawsuits over the quantity of natural gas flared in the Bakken, in March 2014, the North Dakota Industrial Commission (NDIC), the state regulatory agency that oversees oil and gas production in North Dakota, introduced new rules aimed at reducing the number of wells that routinely flare, the volume of gas that is flared,

and the duration of flaring. The policy regulated the monthly gas capture percentage, defined as the percentage of natural gas produced that is sold, used at the well site, or in another NDIC approved manner. A higher gas capture percentage indicates that more of the natural gas produced is being used, and thus not flared.

This policy, NDIC Order 24665, applies to all existing and new wells in the Bakken. The policy was implemented on October 1, 2014, with gas capture goals becoming increasingly stringent over the following six years. The policy schedule was announced with the introduction of the regulation and the goals ("North Dakota Industrial CommissionOrder 24665 Policy/Guidance Version 041718" 2014) are as such:

- 74% October 1, 2014 - December 31, 2014
- 77% January 1, 2015 - March 31, 2016
- 80% April 1, 2016 - October 31, 2016
- 85% November 1, 2016 - October 31, 2018
- 88% November 1, 2018 - October 31, 2020
- 91% November 1, 2020 - onward

According to the NDIC ("North Dakota Industrial CommissionOrder 24665 Policy/Guidance Version 041718" 2014), the "the Commission will evaluate compliance with the gas capture goals statewide, by county, by field, then by well for each operator." Newly drilled wells are exempt from meeting gas capture targets during the first 60 days of production. The operator is allowed to apply for exemptions to meeting the gas capture target from the Commission under certain circumstances.

If the operator is unable to meet the NDIC's gas capture goals, then the Commission can restrict the oil production at the non-complying well(s) to 200 barrels of oil per day if at least 60% of the monthly volume of associated gas produced at that well is captured. However, if less than 60% of the monthly volume of gas is captured, then it can further restrict oil production to no more than 100 barrels of oil per day at the well. Furthermore, if the operator is unable to meet state gas capture goals and does not restrict oil production at the non-complying wells, and fails to apply for a hearing with the Commission, then the Commission could result in a penalty of up to $12,500 per month, beginning at $1,000 the first month, and doubling each additional month the operator is in violation of the Order. Furthermore, if the operator has received notice by the Commission that one or more of their wells failed to meet the gas capture goal, and does not restrict oil production at the non-complying well, then starting in the third month of violation of the production restrictions, the Commission can impose a penalty of up to $12,500 per well per day that the well has been in violation.

1.3 Data

My analysis utilize monthly well level oil and gas production and flaring data from North Dakota and Montana wells between March 2012 – November 2019. The self reported administrative data are publicly available from the North Dakota and Montana oil and gas regulatory agencies. I construct a novel remotely detected flared volume dataset, and these data, are to the best of my knowledge, the first to estimate flared volumes at the monthly well level. Additionally, I incorporate data on oil and gas firm operators and network size, panel data on well connectivity to gas gathering lines, and panel data on gas processing plant processing capacity and congestion.

The wells included in my main analysis are wells drilled between 2010 – 2019 and lie within the Bakken Shale boundary. My outcome variable of interest is the gas capture percentage (GCP). The GCP is computed either from data on the volume of gas flared (in MCF) from the self-reported administrative data, or estimated using satellite data. I use the GCP as my outcome of interest since that is the relevant outcome targeted by the NDIC flaring policy. GCP is defined as the fraction of gas used in a government approved manner (e.g. gas sold, gas used on site, etc). The support of GCP is between 0 and 1, where a GCP of zero means that 0% of the gas produced in that month at that well was captured and a GCP of one means that 100% of the gas produced was captured.

A higher GCP is desired by regulators, because mechanically that means flaring of natural gas is lower. Given that the government doesn't collect data on all of the ways wells and firms use gas in a government approved manner, I compute GCP as:

$$\text{GCP}_{it} = 1 - \frac{\text{flared}_{it}}{\text{gas produced}_{it}}$$

for well i in month-year t. This is the upper bound on the amount of gas that is captured, since this method of computing GCP does not account for potential venting or leakage, which would further decrease the GCP.

1.3.1 Self-Reported Administrative Data

The monthly well level oil and gas production dataset are available through state regulatory agencies that oversee the oil and gas industry within their respective states. In North Dakota, the data are downloaded from the North Dakota Industrial Commission (NDIC), and in Montana, the data are downloaded from the Montana Board of Oil and Gas Conservation (MTBOGC). The two agencies set and enforce flaring and venting emissions by oil and gas operators within the state, and require that well operators report monthly oil and gas production, as well as the quantity of gas flared, by each well. In addition to the monthly production and flared volume data, the datasets also contain well level covariate data including information on the well's location, spud-date (the day when the well was drilled), name of well operator, and volume of gas sold each month.

For the wells in my sample, I include all the wells that are included in the monthly production data files between March 2012 - November 2019. At the beginning of my sample, there are 113 active wells in Montana, operated by 19 firms, and there are 2,212 active wells

in ND, operated by 52 firms. By the end of my sample, there are 466 active wells in Montana, operated by 22 firms, and there are 12,040 wells in North Dakota, operated by 60 firms.

In terms of oil and gas production across this subset of Bakken wells, at the beginning of our sample, Montana is producing a total of 633,140 barrels of oil (an average of 5603 bbl/well) and 561,827 MCF of natural gas (an average of 4972 MCF/well) in March 2012. North Dakota produced 13,064,821 barrels of oil (an average of 5,906 bbl/well) and 1,3744,686 MCF of natural gas (an average of 6,214 MCF/well) in the same month. By the end of our sample, Montana is producing 1,007,328 barrels of oil (an average of 2,162 barrels/well) and 1,775,850 MCF (an average of 3,811 MCF/well) of natural gas. North Dakota is producing 39,922,616 barrels of oil (an average of 3,315 barrels/well) and 87,045,047 MCF of natural gas (an average of 7,230 MCF/well).

1.3.2 Satellite Data

I construct the remotely detected monthly well level flared volume data from raw daily flares data recorded by the NOAA-NASA Visible Infrared Imaging Radiometer Suite (VIIRS) Class Suomi National Polar-orbiting Partnership (NPP) 2.0 and Gravitate 3.0 JOI and NPP satellite products. The VIIRS satellite was launched on October 28, 2011 and daily observations are available starting in March 2012. The VIIRS sensor aboard the satellite collects imagery and measurements using multiple instruments aboard. For the purposes of detecting flare observations, these observations are recorded by the sensors in the infrared bands of the electromagnetic spectrum. The sensor has a swath width of 3,060 km when it is at its average altitude (of 829 km); this width allows it to provide complete coverage of the globe throughout the day. The spatial resolution of the infrared sensor used for detecting combustion sources is 750m pixel resolution (Cao et al. 2013).

There are five spectral bands on the VIIRS sensor that collect low light imaging in the near-, medium- and short-wave infrared channels which allows a more precise detection of combustion sources. The Earth Observation Group, a research team at the Colorado School of Mines that specializes in constructing intermediate data products from the VIIRS satellite observations, developed an algorithm, VIIRS NightFire (VNF), to filter the raw data captured by VIIRS to remove cloudy, sunlit, moonlit, and other noisy observations. This VNF processed dataset contains daily pixel-level observations of detected combustion sources (Elvidge et al. 2015; Elvidge et al. 2018; Elvidge et al. 2019).

This processed daily flare observation dataset record the location of the flare, datetime of the flare, the radiant heat (effectively the temperature) of the flare, as well as the source area (effectively the footprint) of the flare measurement. I filter out all observations for which radiant heat and source area values are not available. I also filter out all flare observations that are recorded beyond the geographical boundary of Montana and North Dakota.

Using the conversion methodology developed by the Earth Observation Group (Elvidge et al. 2015; Elvidge et al. 2019) which maps radiant heat and source area to a quantity of natural gas that has been combusted, I convert the daily flare measurement to a volume of flared gas that was observed at a given location on a given datetime.

Next, in order to generate well level flare volume estimates, I assign a fraction (between zero and one) of the flared volume to wells that are spatially proximal to the recorded

location of the recorded flare. The reason why I am unable to precisely attribute the exact well that generated that volume of flared gas is because the spatial resolution of the sensor is 750 meters, whereas multiple wells can often be located within 750 meters of each other. Instead, I assign a fraction of the the observed flared volume to nearby wells taking into account (1) Euclidean distance between the recorded flare location and well coordinates, (2) age of the well, and (3) whether the well reported producing any oil or gas that month.

More specifically:

1. *Distance*: I use a Gaussian probabilistic assignment based off of distance, where I assume that the closer the well coordinate is to the recorded flare coordinate, the higher the likelihood that the well had emitted that flare. I restrict the candidate of potential wells to those that are within six σ of the recorded flare location, where σ is calculated through the image processing of raw satellite imagery of the flares.

2. *Age of well*: Oil and gas production at a given well follows a well documented exponential decay functional form (Arps et al. 1945). In particular, production at an unconventional fracked well faces a much steeper decline in production as the well ages. This means that at any given fracked well, the majority of its oil and gas production occurs within its first few years. As such, I assign a higher likelihood of the flared volume coming from a newer well rather than an older well. I calculate the well's maturity using its spud-date. The exponential decay parameter that I use to characterize the likelihood function is estimated using data from monthly reported flared volumes from North Dakota wells through regressing reported monthly well level flare volume on a continuous variable for age of the well.

3. *Well activity*: Lastly, I assume that the observed flare is emitted by a well that is actively producing oil and/or gas (rather a well that is not producing any oil or gas) in the month that the flare was detected. Therefore, I filter out candidate wells that report producing zero MCF of natural gas and zero barrels of oil.

For each month by well observation, I sum up the total estimated flared volume that has been assigned to that well through the aforementioned process.

1.3.3 Additional Data

1.3.3.1 EIA Tight Oil and Shale Gas Plays Lower 48 States Shapefile

The EIA Tight Oil Shale Gas Plays Shapefile dataset contains geolocated boundaries of the tight oil and shale gas plays in the continental US. I subset this shapefile to the Bakken shale polygon to obtain the boundary of the play, so that I can determine which wells in Montana and North Dakota are located in the Bakken Shale.

1.3.3.2 EIA Oil and Gas Prices

The EIA has daily Henry Hub (HH) natural gas spot price and West Texas Intermediate (WTI) oil FOB spot price. I compute the monthly average HH and WTI prices, as well as the ratio of the monthly average prices using this dataset.

1.3.3.3 Rextag Pipeline Shapefile

Rextag pipeline shapefile data contain geolocated gathering and transmission line pipeline data in the United States. For each gathering and transmission pipeline segment, the dataset also includes the information on the unique identifier for that segment, the length (in miles) of the segment, and also the operator of the segment.

1.3.3.4 Gas Processing Facility Data

For gas processing plant capacity and congestion data, I combine North Dakota Pipeline Authority's annual gas plant capacity data and North Dakota Industrial Commission's geolocated gas plant data to obtain monthly plant level data on the quantity of congestion at the processing facilities.

During my sample time period, there are 17 gas processing plants that are actively processing natural gas in the Bakken. A map of these plants overlaid with the pipeline infrastructure can be seen in Figure 1.2.

1.4 Estimation Specification

In order to evaluate the causal impact of the NDIC flaring policy on gas capture rates in North Dakota, I implement a difference-in-differences estimation specification. The first difference is before and after the NDIC flaring order was announced and implemented, and the second difference is wells that are subject to the regulation (in North Dakota) and those that serve as the control group (in Montana).

1.4.1 Parallel Trends

I verify that my treatment and control groups satisfy the parallel trends condition in the pre-treatment period (March 2012 - February 2014). My treatment and control groups are the subset of Montana and North Dakota wells drilled in the Bakken Shale between 2010 - 2019. The Montana wells in the Bakken serve as a suitable control group for the North Dakota wells in the Bakken given that they are both accessing similar underlying petroleum resources and also drilled using the same technology. To test for the parallel trends, I compute the moving average in the pre-treatment period by state for the outcome of interest, which is the well level residualized gas capture rate, $\widetilde{gcp}_{icsmt}$.

The firm and month-of-year fixed effects are estimated in state specific regressions using the event study regression specification:

$$gcp_{icsmt} = \beta_0 + \beta_s * \text{policy} + \eta_{sc} + \psi_{sm} + \epsilon_{icsmt}$$

for well i, in county c, in state s for $s \in \{\text{ND, MT}\}$, in month-of-year m and month-year t. (Where month-of-year is January, ..., December and controls for seasonal effects. Month-year is unique month-year, such as January 2013, February 2013, ..., December 2013.) This specification is estimated separately for North Dakota and Montana wells, and also separately for self-reported GCP and remotely detected GCP.

In Figure 1.7, I display the moving average well level residualized remotely detected gas capture rate comparing Montana and North Dakota wells. In the figure, the policy announcement (in March 2014) and the subsequent five iterations of the NDIC flaring policy are indicated by the bolded vertical lines. In the pre-policy period, the outcome of interest in Montana and North Dakota follow similar patterns, where the residualized GCP trend upwards between March 2012 and January 2014.

In Figure 1.11, I display the moving average well level residualized remotely detected and self-reported gas capture rate for North Dakota wells drilled between 2010-2019. The average residualized GCP are similar for the two flaring measures in the pre-policy period. They both trend upwards between beginning of 2012 through late 2013, and fall in the few months preceding the policy announcement.

In North Dakota, the average remotely detected well level GCP was 64.5 percent, whereas the average self-reported well level GCP was 64.3 percent during the pre-policy period. In Montana, the remotely detected well level GCP was 85.5 percent and the average self-reported well level GCP was 99.9 percent. Note that Montana well operators report virtually no flaring between 2012 - 2019. Across the 695,181 unique month-well observations in Montana, only 0.2% of all month-well observations reported flaring a quantity greater than zero MCF, whereas 6.7% of all MT month-well observations had remotely detected flare volumes greater than 0 MCF. The number of remotely detected flare observations (n = 47,104) is 27 times larger than the self-reported number (n = 1,687). In contrast, in North Dakota, out of the 634,566 unique month-well observations, 60.3% of the month-well observations reported a flare volume > 0 MCF, compared to 70.5% of the remotely detected month-well observations with a flare volume > 0. The number of remotely detected well-month observations with flare volumes > 0 (n = 741,785) is only 1.2 times greater than the number of reported well-month observations with flare volumes > 0 (n = 634,566).

I will estimate the causal effect of the NDIC flaring policy using both self-reported and remotely detected Montana flared volumes as control, but my main results rely on using the remotely detected Montana flare estimates as my control group.

1.5 Results

I first estimate the difference-in-differences using North Dakota and Montana remotely sensed GCP. The results from this specification reflect the impact of the policy on the actual flaring behavior of North Dakota wells. Subsequently, I estimate the difference-in-differences using the self-reported North Dakota and remotely sensed Montana GCP. These results reflect the impact of the policy on the reported flaring behavior of North Dakota wells. The difference between the two sets of result indicate discrepancy between actual and reported flaring behavior. The identifying assumption in the difference-in-differences specification is

conditional on the full set of controls, absent the announcement and implementation of the NDIC flaring policy, the flaring behavior in the Bakken North Dakota wells would follow the same trend as the Bakken Montana wells.

My estimation specification is:

$$\widetilde{gcp}_{icsmt} = \beta_0 + \beta_1 \mathbb{1}(\text{ND})_s + \sum_{p=1}^{6} \mathbb{1}\eta_p \text{policy}_p + \sum_{p=1}^{6} \mu_p \mathbb{1}(\text{ND} \times \text{policy}_p) + X\Gamma' + \text{FE} + \epsilon_{icsmt}$$

where ND is an indicator variable that is equal to 1 if the well is located in North Dakota, and 0 otherwise; and $policy_1, \ldots, policy_6$ are mutually exclusive indicator variables that are equal to 1 during the time period when that policy version is enforced, and 0 otherwise. Specifically, the six policy variables are:

1. $\text{policy}_1 = 1 \forall t \in$ (Mar. 2014 – Sept. 2014) when the policy was announced and the months following the announcement, and 0 otherwise;
2. $\text{policy}_2 = 1 \forall t \in$ (Oct. 2014 – Dec. 2014) when the GCP threshold is at 74%, and 0 otherwise;
3. $\text{policy}_3 = 1 \forall t \in$ (Jan. 2015 – Mar. 2016) when the GCP threshold is at 77%, and 0 otherwise;
4. $\text{policy}_4 = 1 \forall t \in$ (Apr. 2016 – Oct. 2016) when the GCP threshold is at 80%, and 0 otherwise;
5. $\text{policy}_5 = 1 \forall t \in$ (Nov. 2016 – Oct. 2018) when the GCP threshold is at 85%, and 0 otherwise;
6. $\text{policy}_6 = 1 \forall t \in$ (Nov. 2018 – Oct. 2020) when the GCP threshold is at 88%, and 0 otherwise;

The coefficients of interest are μ_p for $p \in 1, \ldots, 6$. The interpretation of μ_p is the average treatment effect of the policy p on the gas capture percentage in North Dakota relative to the pre-policy (March 2012 – February 2014) gas capture percentage, conditional on fixed effects.

1.5.1 Difference-in-differences results

The results from the difference-in-differences estimation are graphically illustrated in Figure 1.9, and presented in Tables 1.1 and 1.2. In Figure 1.9 I plot the coefficient estimates of the six different policy effects, differentiating between self-reported and remotely detected North Dakota flaring data. The interpretation of the coefficient estimates is the average effect of the binary policy variable turning on on the well level GCP relative to the pre-policy well level GCP.

As shown in Figure 1.9, the policy announcement had a negligible effect on the self-reported GCP, and had a small and negative effect on the remotely detected GCP. The first

implementation of the policy (GCP threshold = 74%) increased the self-reported GCP by 5 percentage points, and had a small and close to zero effect onth e remotely detected GCP. In the second iteration of the policy (GCP threshold = 77%), the point estimates of the policy on both the self-reported and satellite GCP are similar, at increasing the GCP by 7 and 6 percentage points, respectively. In the third iteration of the policy (GCP threshold = 80%), the point estimates are again similar between the two outcomes, with the policy increasing the GCP by 5 to 6 percentage points, respectively. However, in the fourth iteration of the policy (GCP = 85%), the policy estimates diverge, with the coefficient on the self-reported data increasing to 11 percentage points, whereas the coefficient on the satellite data fell to 4 percentage points. The divergence continues in the last version of the policy (GCP threshold = 88%), where the coefficient estimate on the self-reported data increase to 19 percentage points, whereas the coefficient estimate on the satellite data only increase to 13 percentage points.

Differences in the results between Table 1.1 and Table 1.2 illustrate similar policy effects across the two dependent variable datasets. These results demonstrate that in addition to similar well level average GCP in the pre-policy data between the self-reported and remotely detected data, that the policy had similar effects on the two outcomes during the early part of the sample. The policy was effective in increasing the well level GCP, and firms were reporting truthfully. However, in the latter half of the sample, the reported versus actual flaring outcomes differed, with well operators reporting larger fraction of gas captured relative to the fraction of gas they actually flared.

Another way to visualize the effect of the policy and the growing divergence between the two datasets is illustrated in Figures 1.10 and 1.11. In both figures, I plot the average well level GCP for the self-reported (dashed line) versus satellite data (solid line) in North Dakota Bakken wells, drilled between 2010 - 2019. In Figure 1.10, I compute the average from the raw data, and in Figure 1.11, I control for time invariant month-of-year and firm fixed effects.

In Figure 1.10, we see that the average well level GCP in both the satellite and self-reported data increase in the first half of the sample, are roughly similar magnitudes, and are above the state mandated gas capture threshold. However, starting around late 2016-early 2017, the two time series diverge. The reported average GCP continues to comply with the flaring policy, hovering above a gas capture threshold of 85%. However, the remotely detected average GCP drops below the 85% threshold in mid-2017, and continues to fall. By the last iteration of the flaring policy (GCP threshold = 88%), the self-reported average GCP is also no longer in compliance, and has fallen. The remotely detected average GCP falls at a steeper rate, with the gap between the two time series widening.

In Figure 1.11, the average well level residualized GCP are moving in sync over time between the beginning of the sample through late 2016, when both time series peaks. The self-reported residualized time series plateaus and declines gradually. In contrast, the remotely detected time series exhibits a sharp drop starting in late 2016, and the sharp decline continues through the end of the sample.

The regressions results and descriptive figures demonstrate that during the first part of the time series, well operators were both improving their gas capture ability in response to the newly implemented flaring policies, and also reporting truthfully. However, during the

last three years of the sample (2017 – 2019), both self-reported and remotely detected gas capture fell, and a wedge between the two datasets emerged. Understanding the mechanisms that both contributed to the decline of the GCP, and also the growing divergence between the two time series is the focus of the next section.

1.5.1.1 Difference-in-differences, using North Dakota remotely detected × Montana remotely detected GCP

The results from the difference-in-differences specification using remotely detected Montana and North Dakota GCP is in Table 1.1. In all four columns, I control for well productivity and age of the well. In column (1) I do not include any fixed effects, in column (2) I control for firm and month-year fixed effects. In column (3), I add state fixed effects, whereas in column (4) I add county level fixed effects instead. In each of these four specifications, I cluster the standard errors at the 1°x1° pixel and compute robust standard errors.

As stated earlier, my coefficients of interest are μ_p, or the estimates associated with the variables X%GCP in ND. The interpretation of the coefficients on the X%GCP in ND variables is the average treatment effect of the X% gas capture policy on remotely detected gas capture percentage in North Dakota relative to the pre-treatment period. A positive coefficient means that the policy increased the capture of natural gas, and thus reduced flaring. Furthermore, if, for example, the coefficient estimated is 0.01, then the policy on average increased the gas capture rate by 1 percentage point during the time interval that that version of the policy was active relative to the pre-policy period.

Across all four specifications, the announcement of the policy had a modest but not statistically significant effect on reducing the GCP between 0.8 to 4.1 percentage points. The first version of the policy, which stipulated a 74% GCP, also had a not statistically significant effect on the GCP. The second version of the policy, which stipulated a 77% GCP, was statistically significant and increased the gas capture percentage between 5.8 – 8.0 percentage points. The third version of the policy, which stipulated a 80% GCP, was statistically significant and increased the gas capture percentage 5.7 – 7.6 percentage points. However, under the fourth version of the policy, which stipulated a 80% GCP, only saw an average increase of 4.2–5.9 percentage points increase relative to the pre-policy period, and in the fifth iteration of the policy, which stipulated a 88% GCP, the GCP increased 6.6–13.1 percentage points. This provides evidence that the increase in gas capture in North Dakota tapered off and started falling during the implementation of the fifth version of the policy (between November 2016 – October 2018).

1.5.1.2 Difference-in-differences, using North Dakota self-reported × Montana remotely detected GCP

In this section, I re-run the same specification as from the previous section, but use the self-reported GCP for the ND wells as the outcome variable. The results are in Table 1.2, with columns (1) – (4) following the same specification as in Table 1.1.

The announcement of the policy did not have a statistically significant effect on the GCP; the point estimates are small at 0.3 - 1.4 percentage points, and not statistically different than

0. The implementation of the first version of the policy increased the self-reported GCP by 6.6-7.9 percentage points relative to the pre-policy self-reported GCP. The second iteration of the policy had a positive but not statistically significant effect of increasing GCP between 5.4 – 6.4 percentage points. The third iteration of the policy increased the GCP by 4.8 – 6.4 percentage points. The last two iterations of the policy were statistically significant and large, where the fourth iteration of the policy increased the GCP by 11.3 – 12.4 percentage points, and the last iteration of the policy increased it by 19.3 – 19.7 percentage points.

To put these results into context, the average pre-policy GCP in North Dakota was comparable between the remotely detected and self-reported flared volumes, at 64.5 percent and 64.3 percent, respectively. By the end of our sample, the flaring policy stipulated a gas capture of 88 percent. If we were to use the self-reported ND data in estimating the causal effect, then the last iteration of the policy would have increased gas capture percentage on average close to 20 percentage points, or increasing the average GCP from 64 to 84 percentage points, or only 4 percentage points below the threshold. However, using the remotely detected flaring data however, the GCP only increased between 6.6–13.1 percentage points. Thus the satellite data suggest that the average GCP rose from 64 percent to between 70–77 percent, significantly below the GCP reported in the administrative data.

1.6 Mechanisms driving decline in GCP and increase in misreporting in North Dakota

In order to examine why the NDIC flaring policy was effective during the onset of the policy at reducing both self-reported and remotely detected flaring, but failed to increase gas capture in recent years, I explore market drivers and infrastructure constraints that could affect well operator behavior. I test three primary mechanisms that could explain (1), changes in self-reported and remotely detected flared volumes (MCF), (2) the decline in both self-reported and remotely detected GCP, and, (3), the growing wedge between the self-reported and remotely detected GCP that emerged at the end of 2016/beginning of 2017. The three main mechanisms I examine are: (i) oil and gas prices, (ii) availability of gas gathering pipelines, and (iii) congestion at regional natural gas processing plants.

I will first give an overview and descriptive statistics about these three primary drivers in the context of oil and gas production in North Dakota. Then, I will in turn examine the impact of these drivers on the three outcomes of interest: flared volumes, GCP, and the gap in GCP.

1.6.1 Context about mechanisms

1.6.1.1 Crude oil and natural gas prices

First, I identify oil and gas prices as an important factor because of their volatility during my sample time period. Oil prices are particularly relevant, since, as I mentioned earlier, crude oil is the petroleum commodity of interest in the Bakken Shale. At the beginning of

my sample, a barrel of crude oil hovered between $90- $100 between spring 2012 through summer 2014. However, after peaking at $105 per barrel in June 2014, oil prices plummeted to below $40 per barrel by January 2015. Prices bottomed out in early 2016, reaching a low of $30 per barrel in February 2016. Oil prices begin to climb over the course of 2016, and break $50 per barrel by end of 2016/early 2017.

Given that oil prices are strong predictors of oil production decisions on the extensive margin (Anderson, Kellogg, and Salant 2018), increases in oil prices is positively correlated with new wells being drilled (see Figure 1.13). The number of new wells drilled in this time period peaked in September 2014, with 242 new wells drilled. The number of new wells drilled in a given month plummets, falling to 28 new wells drilled in February 2016. The number of new wells starts to climb over the course of 2016 through mid 2017, reaching 111 new wells drilled in June 2017. As new wells come online, the quantity of associated gas produced mechanically increases as well.

Over this time period, natural gas prices also follow a similar trajectory. In Figure 1.12, we see that Henry Hub gas prices peak at $6.00 in February 2014, and then sharply drop in subsequent months. Natural gas prices bottom out at $1.73 in March 2016, and start to increase later on in that year. Between late 2016 and late 2018, Henry Hub gas prices fluctuate between $3 to $4 per thousand cubic feet (MCF). Even though Bakken well operators are not specifically targeting the sale of natural gas as their main revenue source, the increase in natural gas prices increases the opportunity cost of flaring the gas rather than bringing the gas to market.

1.6.1.2 Connection to gas gathering pipelines

As I mentioned in the Data Section, I construct a panel dataset when the individual the gas gathering pipeline segments were constructed. Using this panel dataset, I can examine the roll out of gas gathering infrastructure and the fraction of wells that were connected to a gas gathering pipeline over time.

At the beginning of my sample, there were roughly 4,050 miles of gas pipelines. By the end of the sample, the total number of miles rose to 7,110 miles of pipelines (see Figure 1.14). The number of miles of new pipeline construction was largest in the first two years of the sample (between 2012–2014), and plateaued to around 10–15 miles of new pipeline constructed each month from 2016 onward. Again, the caveat for this dataset is that in order for me to identify the month when a pipeline was "constructed," I need for that pipeline segment to be connected to a well that is also selling gas. Otherwise, if the pipeline segment is constructed at some moment in time, but is not connected to a well that sells gas, then I am not able to use the month when the well reported in the production data when it first started selling gas as a proxy for when the pipeline first became active. As such, I expect the cumulative number of miles of gas gathering pipeline and the number of miles of pipelines constructed each month to be a lower bound.

At the beginning of the sample, 66% of the North Dakota wells in my sample were connected to a gas gathering line. By the end of the sample, this increased to 88% of North Dakota wells (see Figure 1.14).

1.6.1.3 Congestion at regional gas gathering plants

Anecdotally, North Dakota regulators and industry spokespeople point to lack of gas processing infrastructure as a large contributor to the flaring issue in the Bakken. During the mid- to late-2010s, petroleum companies invested in either increasing the capacity at existing plants or constructing new plants. However, there was often a short run temporary or spatial mismatch between where the existing gas plants were located and where new oil wells co-producing associated gas were located. This mismatch in space and time meant that there was gas produced at connected wells that were not able to be processed by gas processing facilities.

In order to examine whether flaring and the misreporting gap are caused by congestion issues (which I define as lack of gas plant processing capacity), I incorporate panel North Dakota gas processing plant capacity data into my flaring analysis.

In Figure 1.15, I plot the total quantity of natural gas produced in North Dakota (across all active wells, not just wells drilled between 2010-2019) in March 2012 – November 2019 (solid black line). I also plot the total self-reported flared volume (solid pink line) and remotely detected flared volume (solid red line). Additionally, I plot the quantity of natural gas produced (dashed black), flared in the self reported (dashed pink line), and remotely detected (dashed red) for the connected wells. Of the gas produced in North Dakota, at least 80-90% of the gas produced in any given month is coming from a connected well.

In Figure 1.16, I include the cumulative gas processing capacity over time. This is plotted in the step function dark blue line, and aside from a few months in 2012, and briefly in late 2017 and early 2018, lies below the total quantity of gas produced at connected wells (dashed black line). The difference between these two lines, or the cumulative statewide congestion, is the dashed blue line also plotted in the figure. Lastly, I also plot the total quantity of self-reported flared gas at connected wells in the dashed pink line, and the total quantity of remotely detected flared gas in the dashed red line. The cumulative congestion (dashed navy blue) is roughly on the same order of magnitude as the quantity of gas reported and detected as being flared. However, the cumulative congestion is only the difference between the total quantity of gas produced and gas processing capacity - but it does not take into account the network of connectivity between wells and nearby gas plants, and regional congestion.

When I refer to regional congestion, what I mean is that the relevant metric of congestion is whether the local gas processing plant that a well is connected to (which I assume to be the nearest active processing plant) has processing capacity. If the local gas processing plant is congested but there is a plant with capacity located far away and not within the well's network, then that is irrelevant. So instead, I sum up congestion using this definition, where congestion is defined as anytime the difference between the total quantity of gas produced at connected wells and the processing capacity at the nearest gas plant is positive. Using this method of computing congestion (as seen in Figure 1.17), the congestion quantity is greater than the total self-reported and remotely detected flared volumes. In particular, between 2012–2015, and from 2017–2019, the trends between the quantity of regional congestion and the remotely detected flared volumes is highly correlated.

As can be seen in Figure 1.17, the quantity of congestion (when accounting for regional connectivity) grows over time. The amount of congestion at the start of our sample is close

to 0 MMCF, whereas by November 2020, it has grown close to 37,000 MMCF. On average, this is about 4,500 MMCF of congestion per plant (Figure 1.18), although as I will show in Figure 1.21, there is substantial heterogeneity across plants.

There is substantial heterogeneity in regional congestion and self-reported and remotely detected flared volumes by gas processing plant (see Figure 1.21). In this figure, I plot the congestion (defined as the difference between total gas produced at wells connected to the plant, and the processing plants' current capacity in that month-year) in the dashed line. The quantity of detected flared volume at wells connected to that gas plant at plotted in the solid line, and the self-reported flared volume is plotted in the dotted line.

Across these 17 processing plants located in the North Dakota side of the Bakken Shale, 11 processing plants (Belfield, County Line, Dewitt, Lignite, LM4, McKenzie Grasslands, Norse Gas Plant, Roosevelt, Targa Badlands, Tioga, and Watford Gas Plant) have mostly little to no congestion during my sample of interest. The self-reported and remotely detected flared volumes are also relatively low, with the majority of monthly flared volumes under 2,000 MMCF.

However, there are six gas plants (see Figure 1.22) that both have larger congestion issues, and also larger volumes of remotely detected and self-reported flared volumes. These six plants are 1804 Springbrook, Bear Creek, Garden Greek, Ray, Robinson Lake, and Wild Basin.

Furthermore, I compare monthly plant level congestion against self-reported flared volumes (in Figure 1.19) and against remotely detected flared volumes (in Figure 1.20). In the scatterplots, the x-axis the the congestion quantity and the y-axis is the flared volume. The diagonal line is the 45° line. Points falling below the 45° line indicate that flaring is less than the amount of congestion at that plant in that month. In Figure 1.19, for months where plants have congestion issues, the quantity of gas flared at the local connected wells follows a positive linear relationship, with slopes less than 1 (roughly between 0.5 - 0.75). In Figure 1.20, for the same months where plants have congestion issues, the remotely detected flared volumes are positive and follow a positive linear relationship as well. However, the slope is closer to 1, indicating a 1:1 relationship between congestion and flared volumes.

The difference between the two slopes indicates a discrepancy between the reported flared data and the remotely detected flared data. In particular, the satellite flared dataset indicates that there is a tighter relationship between congestion and flaring.

1.6.2 Impact of primary mechanisms on flared volumes

While the NDIC flaring policy explicitly regulates the percentage of gas captured, one aim of the policy is to reduce the overall volume of gas flared in the state. In Figures 1.23 and 1.24 I plot the cumulative volume of gas flared across all North Dakota Bakken wells that were drilled between 2010-2019, comparing the volume of flared gas that was reported (solid line) to remotely detected (dashed line).

Akin to the trend in self-reported versus remotely detected average well level GCP (Figure 1.11), we see that the self-reported and remotely detected flared volumes were of comparable magnitudes between 2012 - 2016. The total volume of gas flared increased between 2012 - 2014, and decreased between 2014-2016. Across both datasets, the volume of gas flared

increased in 2017 onward. However, the wedge between the two time series emerged in January 2017, and continued to grow.

Between 2012-2016, the total amount of self-reported flared gas was 474,082 MMCF, or an average of 118,520 MMCF/year. During this four year time interval, the total amount of remotely detected flared gas was 486,302 MMCF, or an average of 121,575 MMCF/year. The cumulative difference in flared volume between the two time series during these four years was 12,220 MMCF. The remotely detected volume of flared gas was about 2.5% larger than the self-reported volume of flared gas, which is not a sizeable difference. However, Between 2017 - 2019, the total volume of self-reported flared gas, was 391,616 MMCF, whereas the volume of remotely detected flared volume was 679,547 MMCF. The difference between the two measurements are 287,931 MMCF. The remotely detected volume of flared gas was 74% larger than the self-reported volume of flared gas during this time interval.

Next, I examine the impact of the three primary mechanisms on self-reported flared volumes, and on remotely detected flared volumes (see Tables 1.3 and 1.4). I compare the coefficients from column (3) in Table 1.3 to the coefficients from column (3) in Table 1.4 in Figure 1.25. In these two regressions, I regress whether a well is connected to a gas gathering line or not, the level of congestion at the well's local gas gathering plant, oil and gas prices, on the volume of flared gas and control for the age of the well, productivity of the well, month-of-year fixed effects, year fixed effects, and time-invariant firm fixed effects. I rescale the coefficient estimates by two standard deviations so that the coefficients on the continuous and binary variables are directly comparable (Gelman 2008).

Looking at Figure 1.25, we see that connecting a well has a statistically significant effect on reducing the quantity of gas flared in both the self-reported and remotely detected measurements. Connecting a well decreases the volume of flared gas by 4,155 MCF in the administrative data, and by 2,725 MCF in the satellite data. A one standard deviation increase in congestion at the regional gas processing plant then it increases the quantity of gas flared at the well site, by 111.4 MCF in the administrative data, and by 467.1 MCF in the satellite data. As for changes in oil and gas prices: a one standard deviation increase in the price of WTI crude oil leads to an incease in flared gas by 20 MCF in the administrative data and 272 MCF in the satellite data. Lastly, a one standard deviation increase in the HH natural gas price increases the volume of flared gas by 143.9 MCF, but decreases the volume by 159.4 MCF in the satellite data. Overall, aside from the coefficients on gas price, the sign of the coefficients on the primary mechanisms are the same for the self-reported versus remotely detected data. We also see that the regression coefficient on the reduction in flared volume is greater but that the increase in flared volume is smaller for the administrative dataset.

1.6.3 Impact of primary mechanisms on GCP

Next, in this section, I evaluate the impact of the three primary mechanisms, oil and gas prices, well connection to a gas gathering pipeline, and congestion at a regional gas processing plant, on well level GCP. Here, I estimate the effects jointly, in order to discern the relative magnitudes of each of the three main mechanisms. In the appendix, I include robustness checks where I estimate the effects separately and use different estimation specifications. My

main specification is the same as the regression specification used in the previous section, except the outcome of interest in this section is well level gas capture percentage, rather than volume of gas flared. The results comparing the self-reported GCP and remotely detected GCP are in Tables 1.5 and 1.6, respectively. The rescaled coefficients from column (3) from each of the two tables are displayed in Figure 1.26. Overall the signs of the coefficient estimates between the self-reported and the remotely detected outcomes are in the same direction, with the exception of the sign on the gas price.

In both outcomes, connecting a well increases the well's GCP, which intuitively makes sense, since a gas gathering line allow for the well to increase its ability to capture gas. However, the effect on well GCP is much larger in the administrative outcome, where the connecting of a well increases the GCP by 71 percentage points, whereas in the satellite outcome, it only increases the well's GCP by 25 percentage points. Secondly, congestion at the well's regional gas plant decreases gas capture in both the administrative and satellite outcomes. The decrease in GCP in the self-reported data is smaller, where a one standard deviation increase in congestion at the plant decreases the GCP by 2 percentage points. However, in the satellite data, a one standard deviation increase in congestion decreases the GCP at the well by 15 percentage points. Lastly, for oil and gas prices, a one standard deviation increase in WTI oil price decreases both administrative and satellite GCP (by 1 percentage point and 8 percentage points, respectively). A one standard deviation increase in HH gas price decreases GCP by 2 percentage points in the self-reported data, but increases GCP by 2 percentage points in the satellite data.

1.6.4 Impact of primary mechanisms on GCP gap

Another way to visualize the differential impact of the primary mechanisms on the gas capture percentage is to regress the mechanisms on the difference of the well level GCP. Here, the difference is defined as the well level self-reported GCP less the remotely detected GCP (see Table 1.7). The rescaled coefficient estimates for column (3) is presented in Figure 1.29. Mechanically, the rescaled coefficients are the differences between the rescaled coefficient estimate from the administrative data and the rescaled coefficient estimate from the satellite data in Figure 1.26 for each policy.

The interpretation of these coefficients is a one standard deviation increase for the continuous variables or the value of indicator variable switching from zero to one, results in a X percentage point change in the difference between the two GCP estimates. If there were no misreporting, then the change should not be statistically significantly different than zero. However, if the change is positive and statistically significantly different than zero, then there is evidence of well operators reporting capturing more gas than what they in reality captured. If the change is negative and statistically significant, then the opposite would be the case.

With respect to the well connectivity, a connected well on average has a 44 percentage point GCP larger than a non-connected well. Intuitively, this makes sense because a non-connected well has less options for how it handles is associated gas co-product. A well not connected to a gas gathering line can either use the gas to power well site operations, re-inject it, or flare the gas. A connected well has an additional alternative for how it handles

its natural gas, and that is to transport the gas to a processing site. The operator of a connected well can either truthfully transport the gas and report the quantity captured, or it could not transport the gas, but still report the quantity captured.

Secondly, a standard deviation increase in the congestion at a local gas plant facility increases the misreporting gap by 13 percentage points. This means that if the regional gas plant is congested because its processing capacity is lower than the total quantity of gas produced by wells connected to the plant, then those wells, on average, report that the fraction of gas captured is 13 percentage points higher than what was actually captured.

Lastly, a one standard deviation increase in the price of oil increases the misreporting gap by 7 percentage points, whereas a one standard deviation increase in the price of natural gas reduces the misreporting gap by 4 percentage points. Intuitively, this also makes sense. When the price of oil increases, the incentive to drill additional wells to increase oil production also increases, and the flaring policy constraint becomes more binding. However, if a well operator misreports the quantity of gas flared, then they can relax the policy constraint. Conversely, as the price of natural gas rises, then the incentive to capture the gas increases. As such, there is less incentive to flare the gas, and thus, the incentive to misreport decreases.

1.6.5 Contribution of the main mechanisms to the quantity of gas flared

In this section, I use the estimated marginal effects to compute the contribution of the three main mechanisms to the volume of flared gas over time. For decomposing the relative contributions of the three main mechanisms, I evaluate the effect of the mechanisms on flared volumes at the monthly-firm, rather than the monthly-well level. This allows me to identify off of the changes that are occurring within a firm over time, rather than at an individual well. In other words, this allows me to control for the drilling of new wells over time.

Given that the analysis is at the well level, I aggregate up the RHS variables as such: (1) **percent connected** is defined as the percentage of the firm's wells that are connected in that month-year; (2) **congestion at gas plant** is defined as amount of congestion that a given firm's wells is exposed to in a given month. If the firm's wells are all connected to the same gas processing facility, then the volume of congestion is only the quantity of congestion at the facility in that month. However, if the firm has wells that are connected to two or more facilities, then the congestion is the sum of the congestion at each of those plants. (3) **number of new wells** is defined as the number of new wells drilled by that firm in that month-year. The outcome of interest the monthly volume of gas flared by the firm.

Next, I take the estimated effects from Table 1.8 to compute the contributions of well connectivity, gas plant congestion, changes in oil price, and control for new wells, on the overall flared volume in North Dakota's Bakken wells. In the self-reported data (see Figure 1.27), these four four factors explain, on average, 68-100% of the quantity of gas flared between the beginning of the sample to end of 2017. In the last two years, they explain, on average, 53% of the flared volumes in 2018, and 43% of the flared volumes in 2019. In decomposing the quantity of flared gas explained by these four factors, while the time series of the total volume is relatively stable, the largest two factors are well connectivity and

congestion. Over time, as more wells in the Bakken are connected, the amount of gas flared decreases. However, congestion at gas plants increases over time, driving up the quantity of gas flared.

Similarly, I repeat this exercise using the estimated effects from the satellite data (see Figure 1.28). Again, on average, the contribution of these four factors accounts for all of the flaring between 2012-2017. By 2018, these four mechanisms only explain 55% of the flared volume, and by 2019, 48% of the flared volume. The amount of gas flared explained by these four factors is trending upwards in 2018-2019, and this is driven by the increased flaring due to congestion. However, the upward trend in congestion-induced-flaring is not as steep as the sharp increase in flaring during this time period.

This provides evidence that in the first seven years of my sample, the amount of gas flared could be explained by the infrastructure constraints and changes in oil prices. However, in 2018 and 2019, these mechanisms only explain half of the amount of gas flared. The other half of the gas flared is unexplained, and is indicative of well operators flaring and misreporting, for reasons beyond just infrastructure constraints and changing oil prices.

1.6.6 Quantifying flaring damages

1.6.6.1 Back of the envelope calculation of fines

I perform a back-of-the-envelope calculation of the total amount of fines the North Dakota Industrial Commission (NDIC) could levy given the penalties they have set forth in the North Dakota flaring policy. The NDIC can penalize wells that are in violation of the GCP threshold and are also in excess of the oil production restrictions. Specifically, if a well is either (i) capturing at least 60% of the monthly volume of associated gas (but less than the GCP threshold) and restricts production to 200 barrels of oil per day, or (ii) capturing less than 60% and restricts production to 100 barrels of oil per day, then it will not be subject to penalties. However, if a well is not in compliance, than the NDIC regulators can fine the well of up to $12,500 for each day that the well is in violation.

Using this penalty provision, I compute the amount of fines NDIC could collect under the self-reported flaring and the remotely detected flaring data (see 1.30). If the regulators were to fine all wells out of compliance, then using the self-reported data, they would be able to assess $13.16 billion between 2012-2019 (an average of $1.65 billion dollars per year). However, if they used the remotely detected dataset, they would assess $17.50 billion (an average of $2.19 billion per year), or an increase of $4.34 billion over this eight-year period.

To put these numbers into context, in 2017, the oil and gas industry paid a total of $1.63 billion in taxes (between production and extraction taxes), and this accounted for 45% of North Dakota's tax revenue. Additionally, between 2008-2018, North Dakota collected almost $18 billion in oil and gas extraction and production taxes, or 44% of the state's tax revenue across the eleven-year period.

Given that North Dakota collects, on average, close to $4 billion in taxes every year, if the state were to use the remotely detected data to assess compliance and implement the penalties laid out in their flaring regulation, then they could increase the revenue by more than 50%. Currently, the oil and gas production and extraction taxes are allocated towards

funds including the Conservation Grant Fund, Renewable Energy Development Fund, Common Schools Trust Fund, Abandoned Well and Reclamation Fund, Outdoor Heritage Fund, additional school funding, and General Fund. While it is unclear how the proceeds from the NDIC flaring penalties are allocated, if they were allocated towards the same funds that the other revenues generated from the oil and gas industry were, these penalties would great increase the environmental and educational outcomes of the state.

1.6.6.2 Social cost of flaring

Next, I compute the social cost of flaring by these North Dakota Bakken wells (see Figure 1.31). Using the approach put forth by Agerton, Gilbert, and Upton Jr (2020), where they use the EIA emission conversion (54.75 kg CO_2/MCF of gas flared) and a \$40 social cost of carbon, then I can price one MCF of natural gas flared as generating \$2.19 in climate damages. It is worth noting that this value is only pricing climate change damages, and not the health damages from exposure to the air pollutants generated by flaring. Also, this pricing assumes the EIA emission conversion, however studies suggest this might be a lower bound since in practice flares have been found to be less efficient and can release a larger quantity of their methane into the atmosphere (Lyon et al. 2020; Zhang et al. 2020; Anchondo 2019; Malik 2019). Even under this lower bound estimation of the social cost of flaring, I find that the costs are quite substantial. In 2012, the social cost of flaring is around \$200 million using either the satellite or administrative records of flared volumes. However, by 2019, the social cost of flaring under the administrative estimates are close to \$400 million, whereas using the remotely detected estimates, the costs are around \$700 million.

1.7 Conclusion

In conclusion, I used a new satellite method to document both a large increase in flared volumes and growing divergence between self-reported and actual flaring behavior. I do so by overcoming traditional challenges of relying on self-reported administrative data, and make use of a novel remotely detected spatially and temporally resolved dataset that I construct. I find evidence that the North Dakota state policy is effective and increases gas capture when the opportunity cost of reducing crude oil production is low. I show that the NDIC flaring policy on average increased gas capture by 5-6 percentage points in the administrative and satellite data in the first two years of the policy (Oct. 2014 - Oct. 2016), and firms report truthfully.

However, I show that when the opportunity cost of not producing oil increases (via the global price of crude oil), the incentive to misreport flaring increases starting in late 2016. This is illustrated by not only the declining gas capture rate in the self-reported data, but also the growing gap between the self-reported and remotely detected flare data. Between Nov. 2016 - Nov. 2019, I find that on average the policy increases gas capture by 14 percentage points with respect to the pre-policy period in administrative data, but by only 7 percentage points in the satellite.

Overall, I find that the average well level gas capture percentage is in compliance for both the self-reported and remotely detected measurements of flaring through 2017. However, between 2017-2019, well operators continue to report gas capture values that are in compliance with the flaring policy, however the satellite data measure a sharp drop-off in the well level gas capture. The wedge between the well level GCP as measured by the two datasets continue to grow through the end of my sample. I find evidence that the policy reduced well level flaring by 6.7 (5.2 - 7.1) points in the remotely detected data, and by 6.1 (5.3 - 6.6) points in the self-reported data between October 2014 - October 2016, relative to flaring rates in the pre-policy period. However, I find evidence of misreporting November 2016 onward; I estimate that the policy on average reduced well level flaring by 6.7 (5.9 - 8.9) points in the remotely detected data, but that it reduced flaring by 15.8 (15.7 - 16.2) points in the self-reported data, relative to the pre-policy flaring rates. Furthermore, while the total volume of gas flared in the self-reported and remotely detected dataset are comparable between 2012-2016 (474,082 MMCF and 486,302 MMCF respectively), the difference between the two emerge in 2017. Between 2017-2019, the total amount of reported gas flared was 391,616 MMCF, whereas the volume of remotely detected flared gas was 679,547 MMCF, or 74% larger than the self-reported volume.

I explore the mechanisms that explain both the changes in flared volume, and this growing gap between the two measurements, and find evidence that (i) the ratio of crude oil to natural gas prices, (ii) well connection to gas gathering lines, and (iii) congestion at gas processing plants, have a positive impact on increasing the well level misreporting gap. Lastly, I decompose the contribution of the mechanisms to the overall volume of flared gas, and find that between 2012-2017, the main mechanisms explain between 70-100% of the flared volume, but only between 45-55% of the flared volume in 2018 and 2019. This shows that economic factors explain close to 100% of the flaring between 2012-2016, but only half 2017-2019, and there is a large residual left unexplained.

My paper builds off of prior work examining the impact of the regulation on gas capture and misreporting, and documents the evolution of firm response to the policy. Previous work demonstrated that the policies were effective during the first 18 months of its implementation (Lade and Rudik 2020), although up to a third of the reduction in flaring during this time period were attributable to misreporting (Lee 2020). However, I extend the period of analysis by an additional 3-4 years to cover a time period of rapidly changing market dynamics. I also increase the resolution of the dataset (through constructing a remotely detected monthly well level flaring dataset), thus allowing this paper to examine how the flaring behavior respond to changes in to market dynamics and oil and gas infrastructure.

My paper shows the value of utilizing newly available, and temporally and spatially resolved satellite data to overcome the challenge of imperfect monitoring. My back of the envelope calculation of the penalties that could be assessed under the flaring policy using the remotely detected data are an average of $2.19 billion per year between 2012-2019, more than doubling North Dakota's annual tax revenue. Given that enforcing policy compliance is a large and continuing challenge, the ability to utilize this technique to monitor oil and gas activity is promising for oil and gas regulators, both in the Bakken Shale and worldwide.

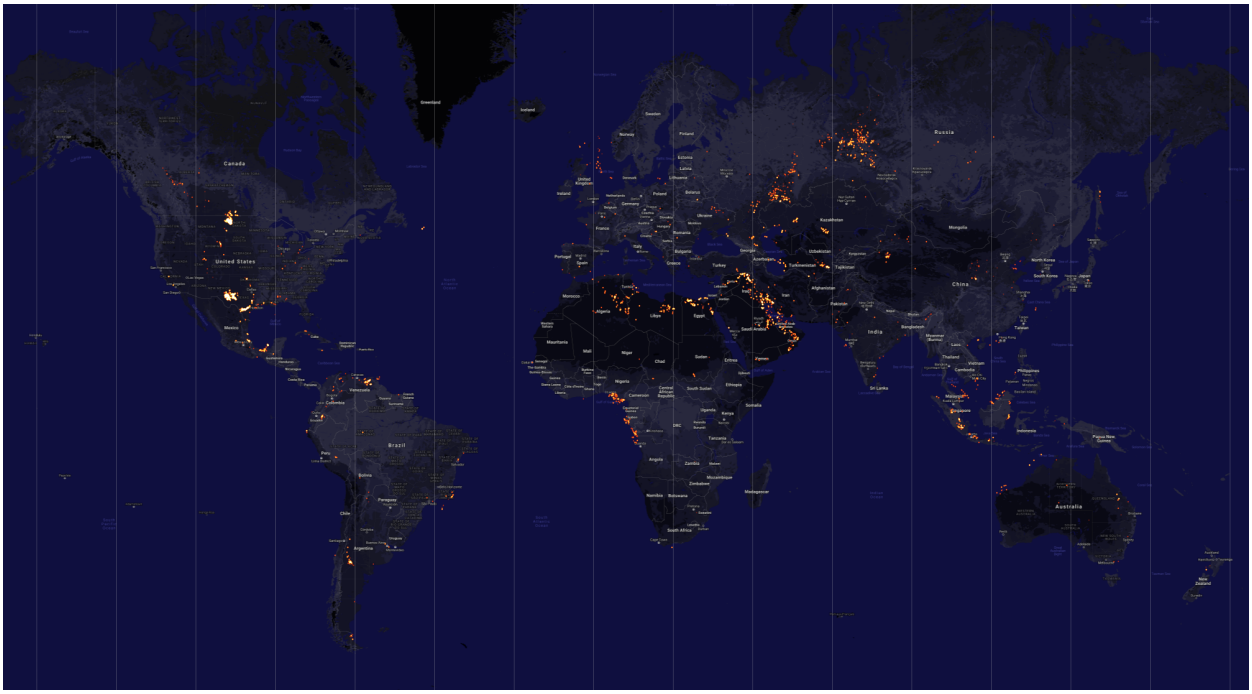


Figure 1.1: Daily flares events recorded by NASA-NOAA VIIRS satellite in August 2019: This image is a snapshot of the remotely detected oil flares worldwide captured by the NASA-NOAA VIIRS satellite instrument in August 2019.

Figures

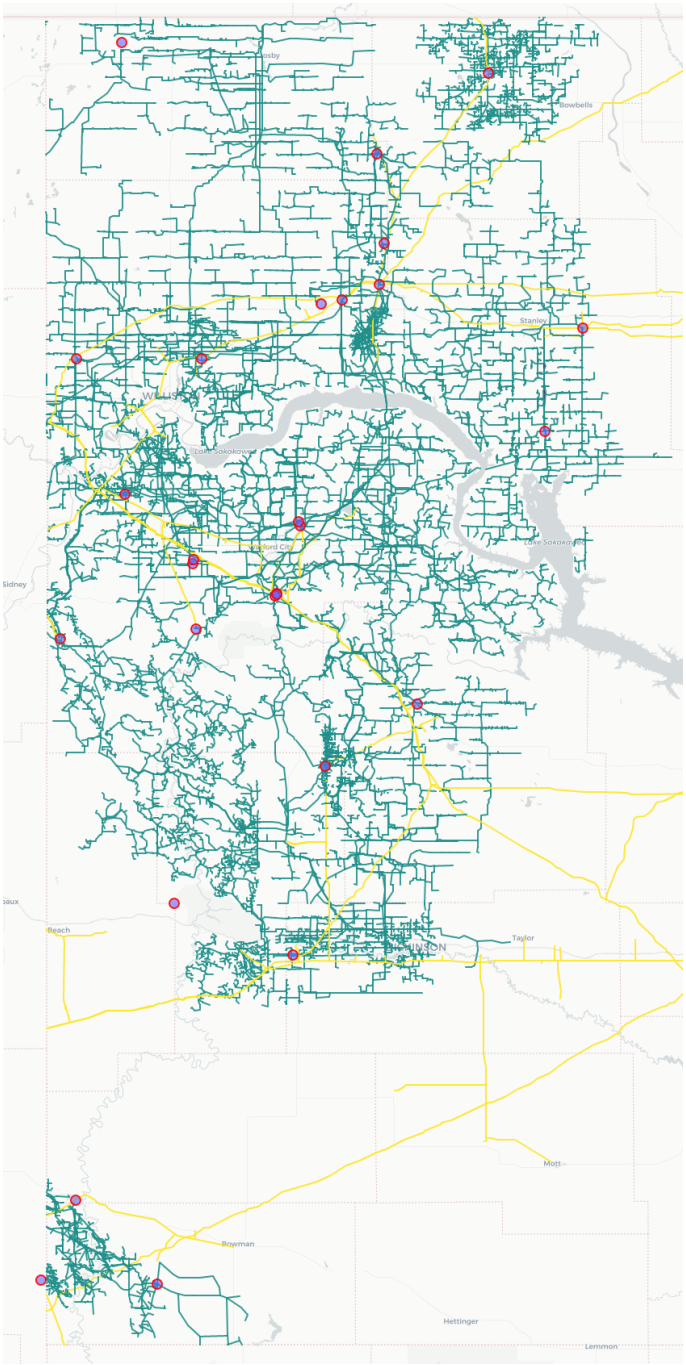


Figure 1.2: Gas capture infrastructure in North Dakota: This map shows the gas processing facilities (red circles) and the gas pipelines in the Bakken, as of June 2020. The green pipelines are the smaller gas gathering lines that connect wells to gas processing facilities. The yellow pipelines are the larger gas transmission lines that transport gas from the gas processing facilities to distances further away.

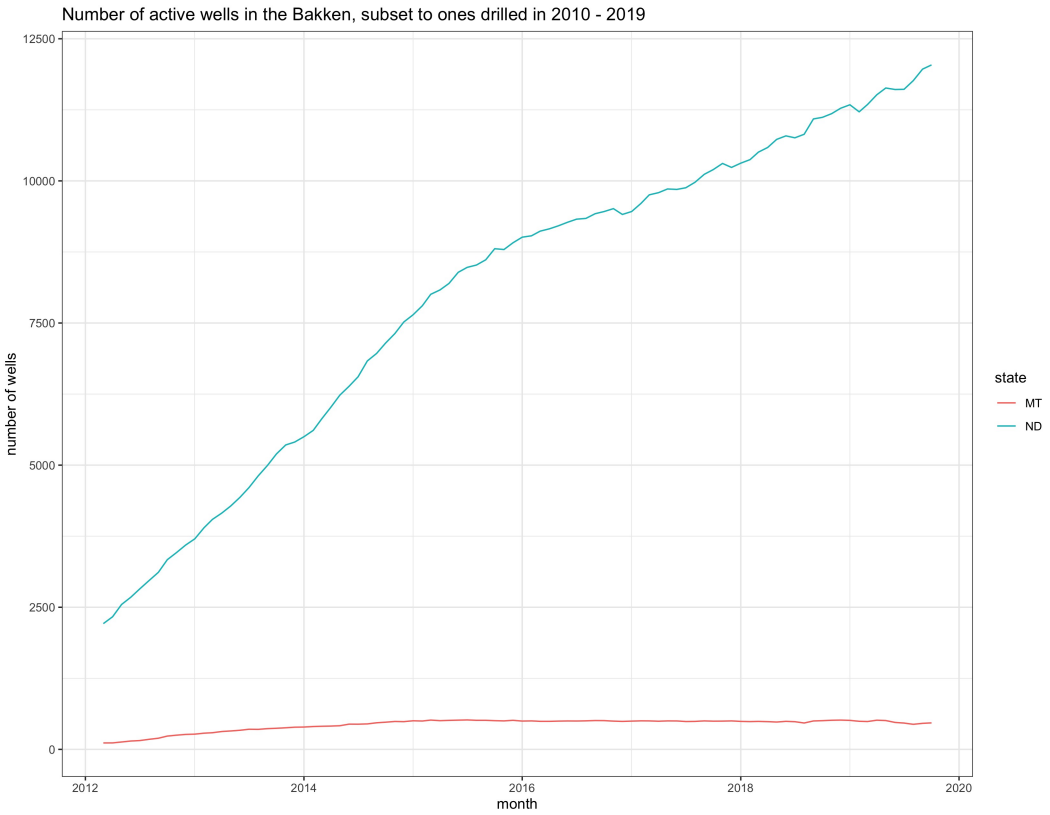


Figure 1.3: Number of active wells in the Bakken by state, subset to wells drilled between 2010-2019

Figure 1.4: Number of firms in the Bakken by state, subset to ones operating wells drilled between 2010-2019

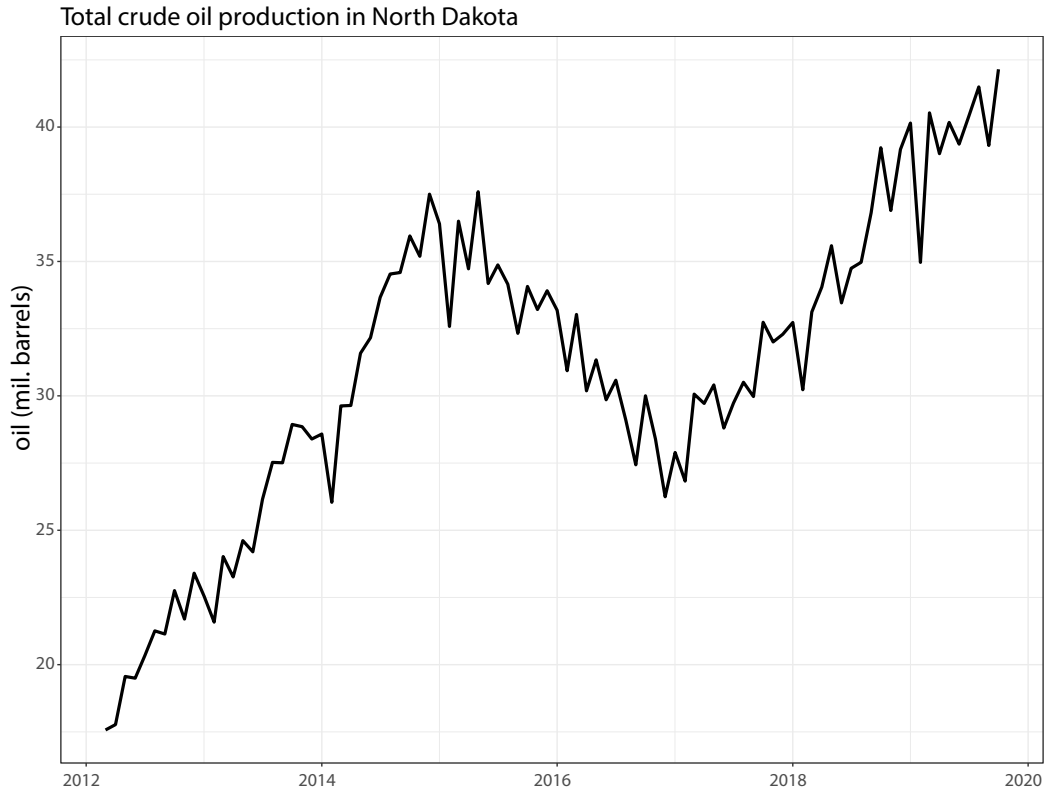


Figure 1.5: Oil production in North Dakota, subset to production from wells drilled between 2010-2019

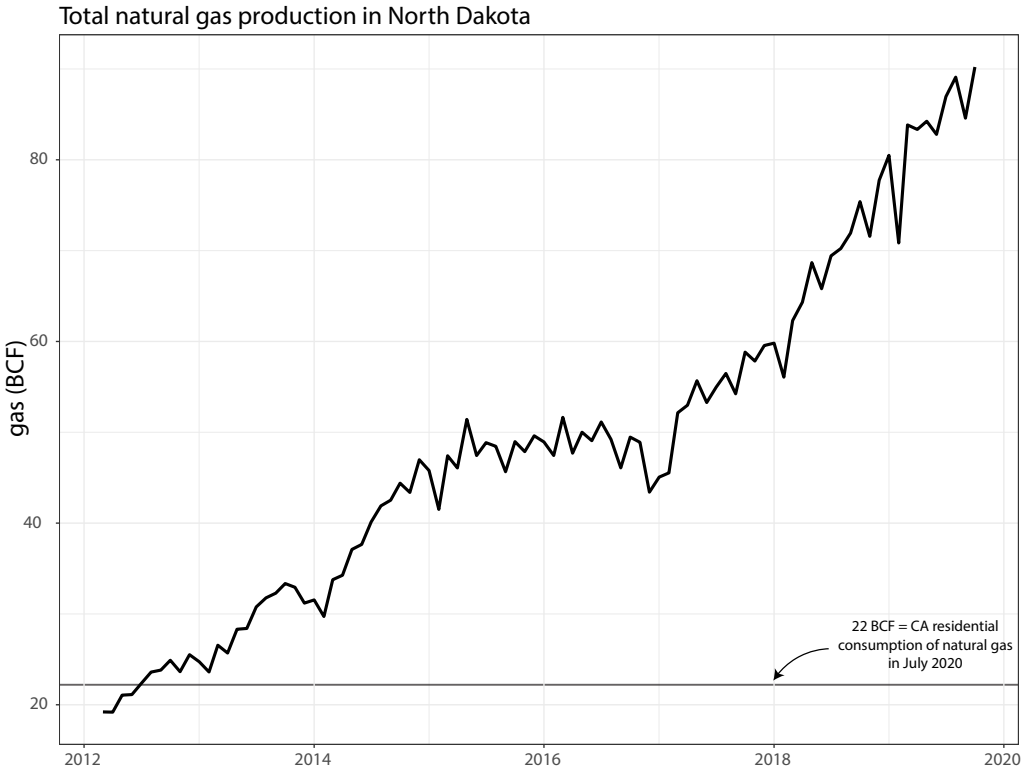


Figure 1.6: Gas production in North Dakota, subset to production from wells drilled between 2010-2019

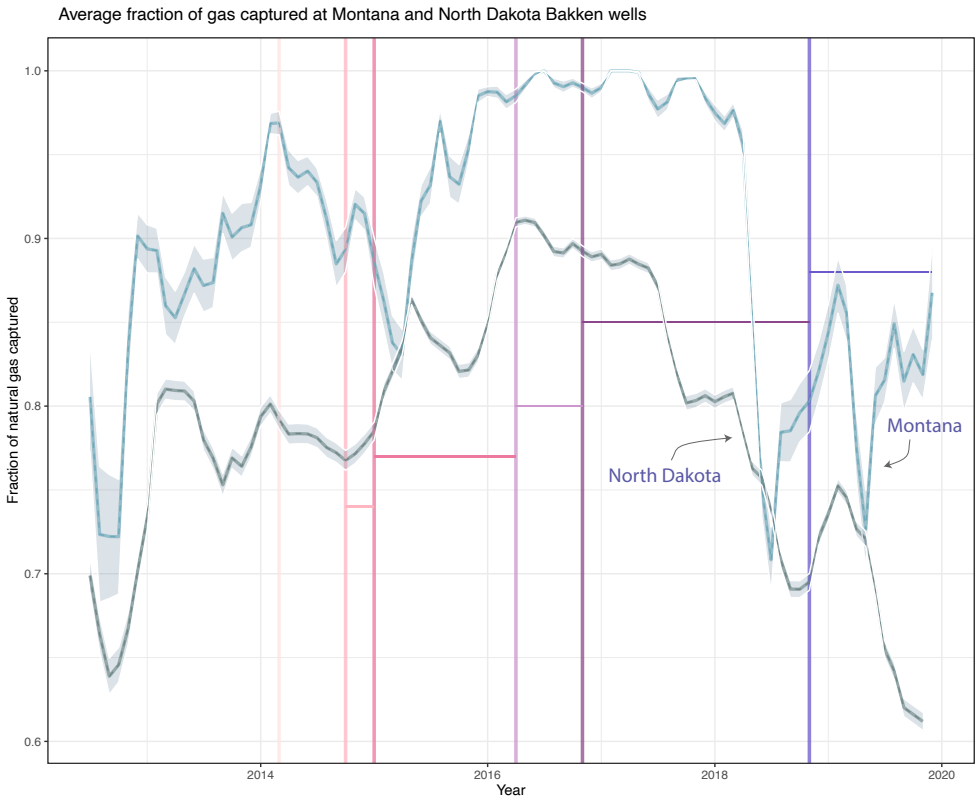


Figure 1.7: Average well level gas capture percentage at North Dakota and Montana Bakken wells, using satellite data, between March 2012-November 2019. The timing of the North Dakota flaring policies are demarcated by the vertical lines, and the policy thresholds are indicated in the horizontal step function.

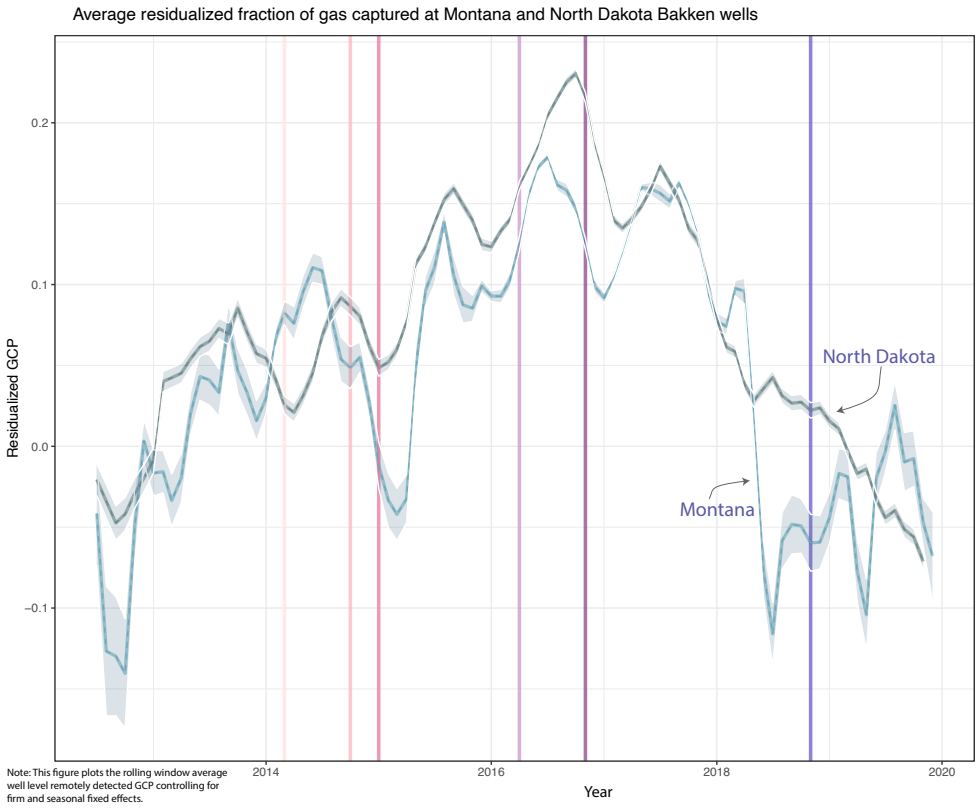


Figure 1.8: Average residualized well level gas capture percentage at North Dakota and Montana Bakken wells, using self-reported data, between March 2012-November 2019. The timing of the North Dakota flaring policies are demarcated by the vertical lines. The residualized well level GCP is computed by subtracting out time invariant month-of-year and firm fixed effects, where the fixed effects are estimated separately by state.

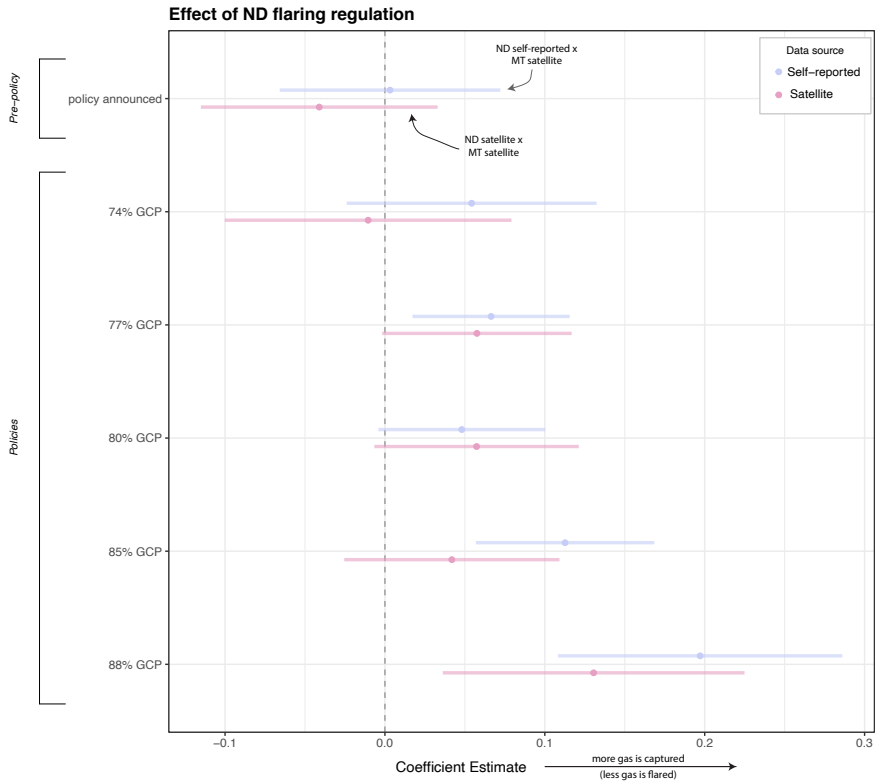


Figure 1.9: Policy effects estimated using a difference-in-differences specification, comparing between results when estimated using ND self-reported x MT satellite data versus ND satellite x MT satellite data. An increase in the coefficient estimate means that the policy increased gas capture, thereby reducing flaring.

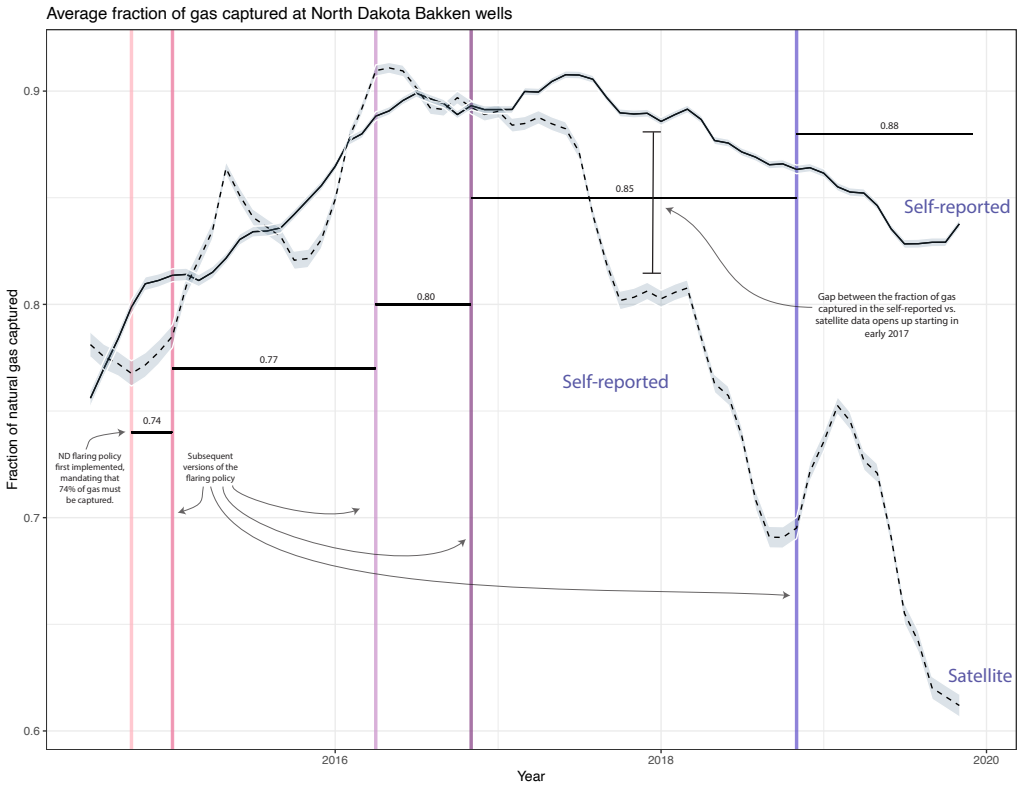


Figure 1.10: Average well level gas capture percentage at North Dakota, comparing estimates using self-reported versus satellite data, between March 2012-November 2019. The timing of the North Dakota flaring policies are demarcated by the vertical lines, and the policy thresholds are indicated in the horizontal step function.

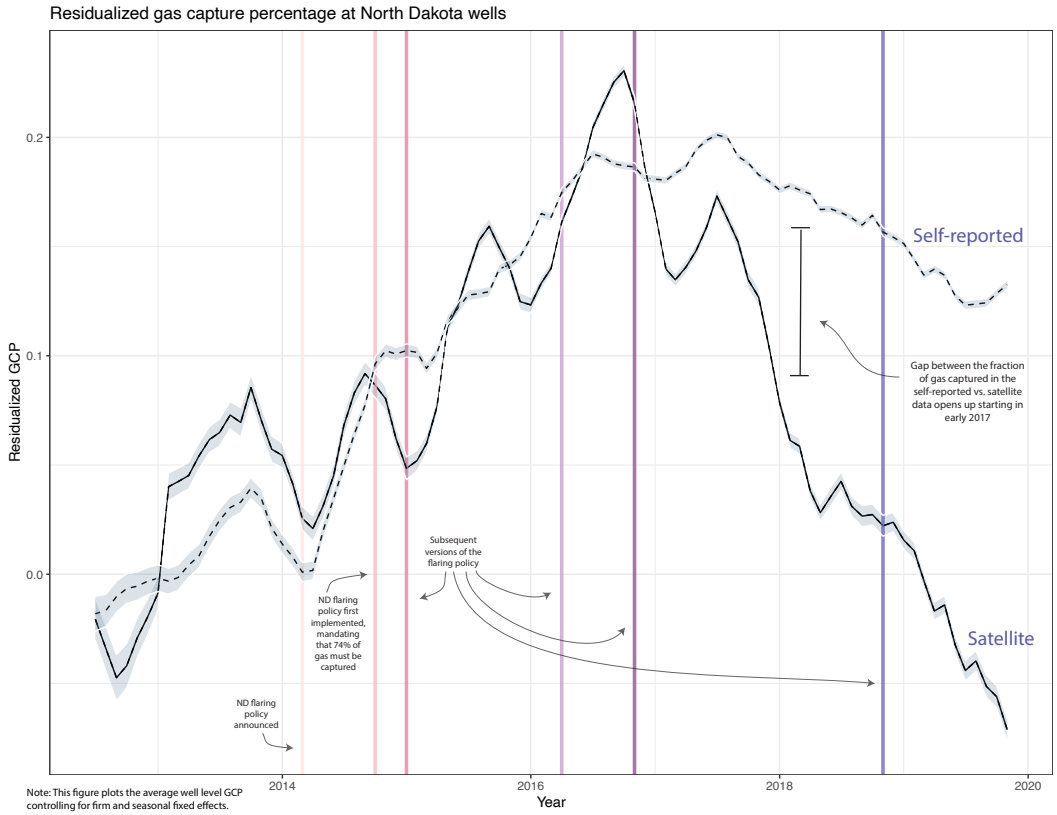


Figure 1.11: Average residualized well level gas capture percentage at North Dakota, comparing estimates using self-reported versus satellite data, between March 2012-November 2019. The timing of the North Dakota flaring policies are demarcated by the vertical lines. The residualized well level GCP is computed by subtracing out time invariant month-of-year and firm fixed effects, where the fixed effects are estimated separately for the self-reported and satellite datasets.

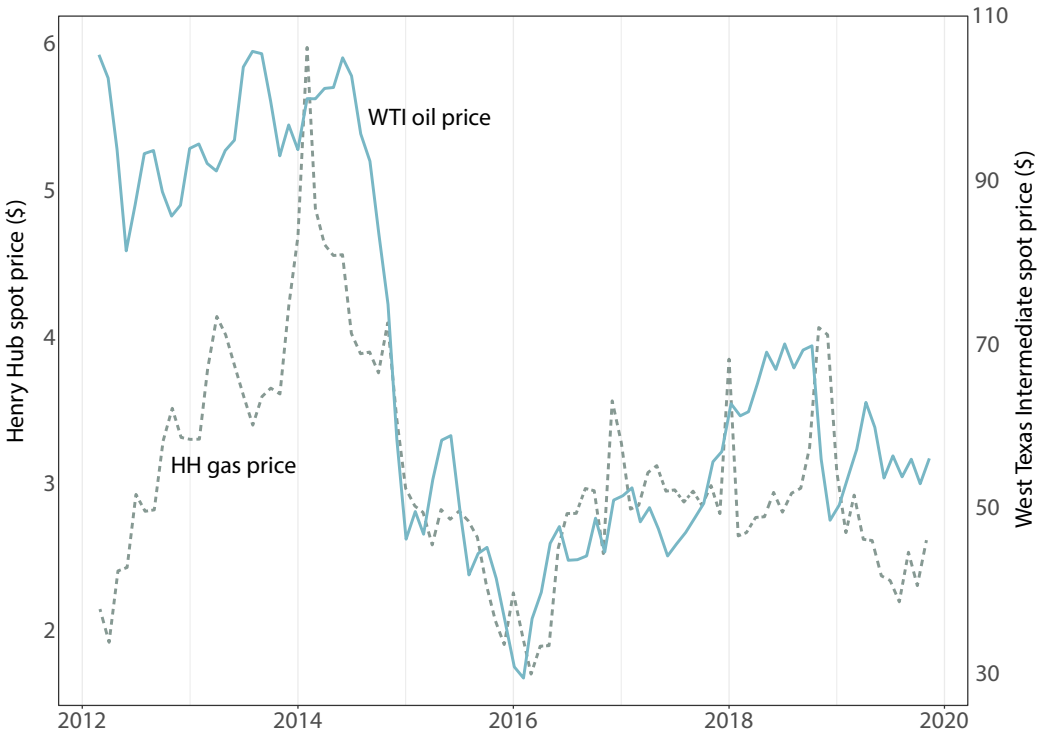


Figure 1.12: Monthly average West Texas Intermediate crude oil spot prices and Henry Hub natural gas prices, between March 2012 and November 2019.

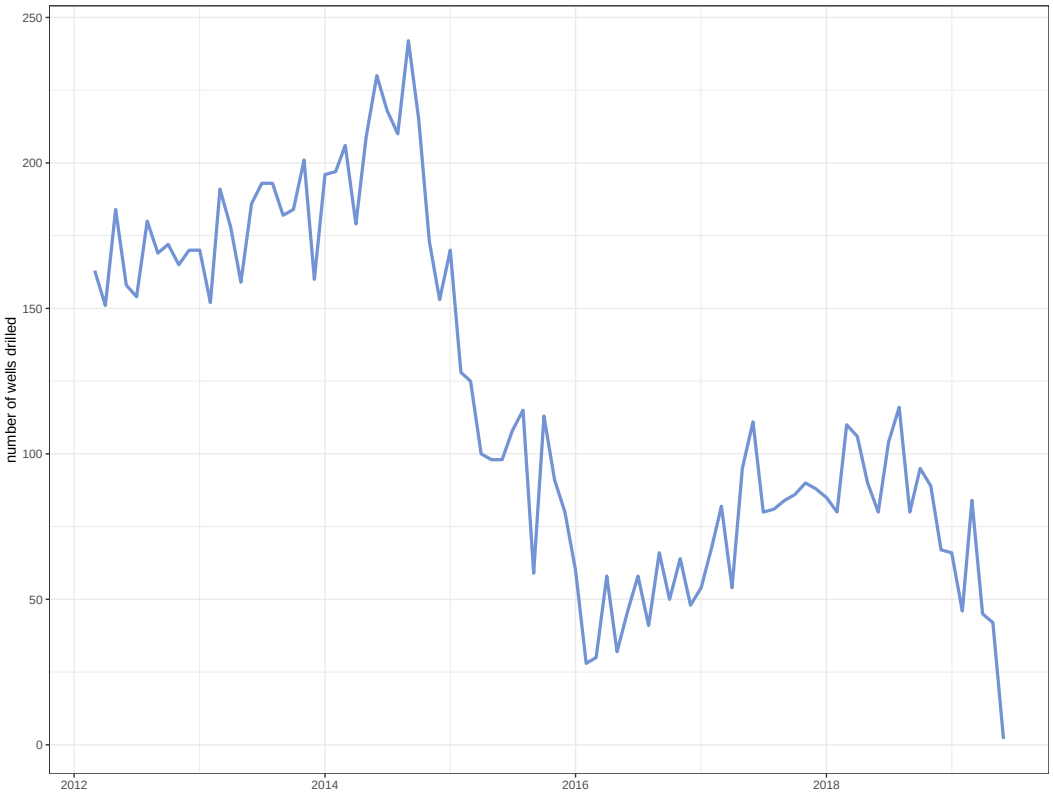


Figure 1.13: Number of wells drilled in North Dakota Bakken Shale between March 2012 - November 2019.

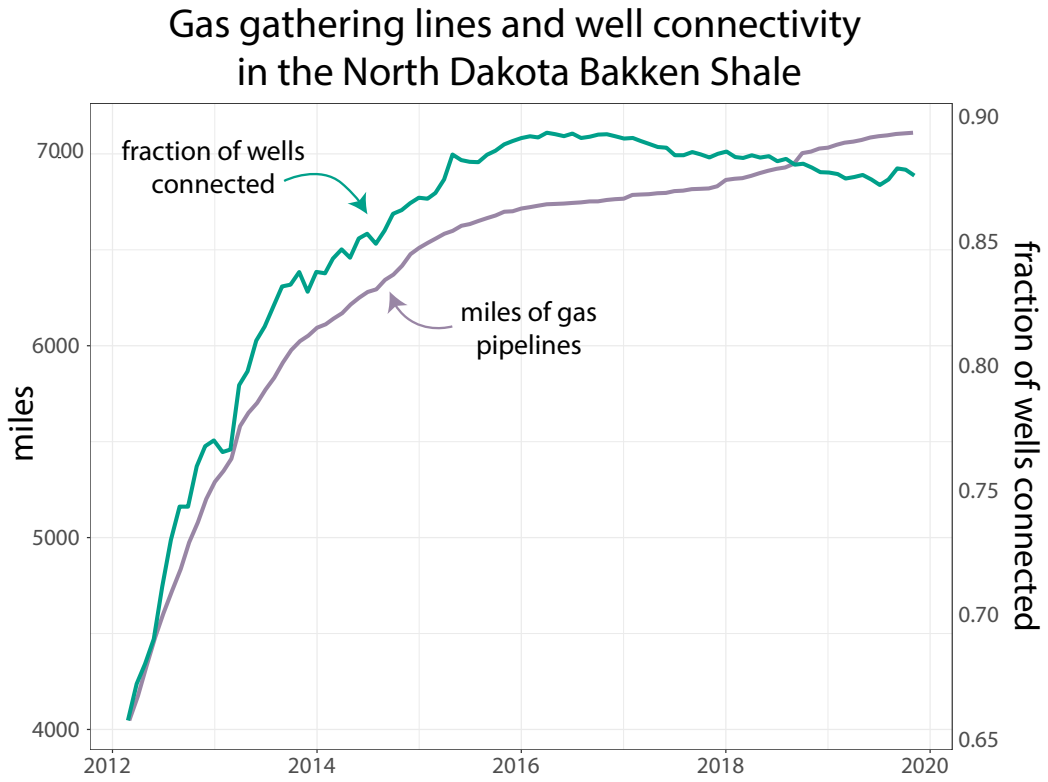


Figure 1.14: Miles of gas gathering lines constructed, and fraction of wells that were drilled between 2010 - 2019 that were connected to a gas gathering pipeline in North Dakota Bakken between 2012 - 2019.

Figure 1.15: Statewide congestion and total flaring vs. flaring at connected wells

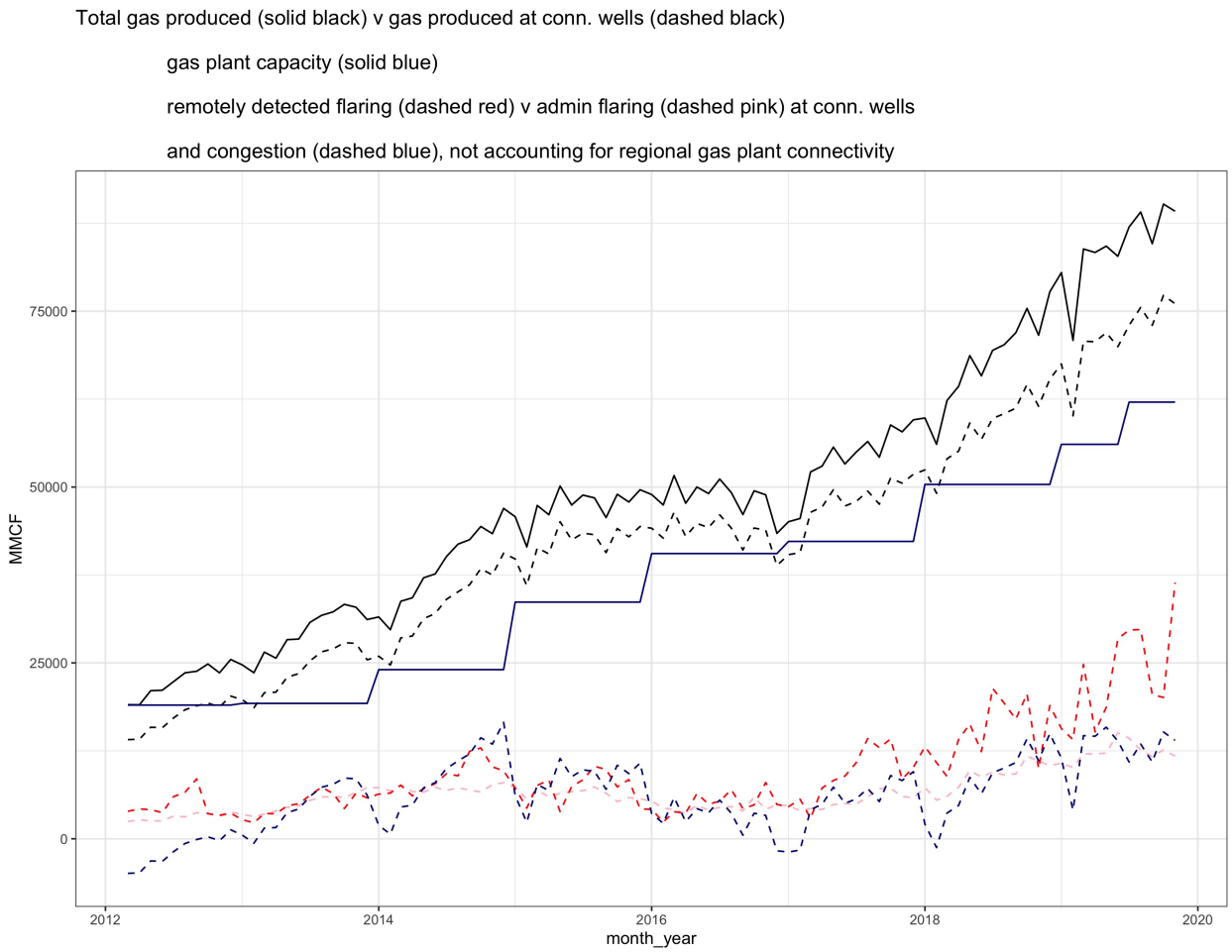


Figure 1.16: Gas production, congestion, and flaring in the Bakken, not accounting for well connectivity to plants

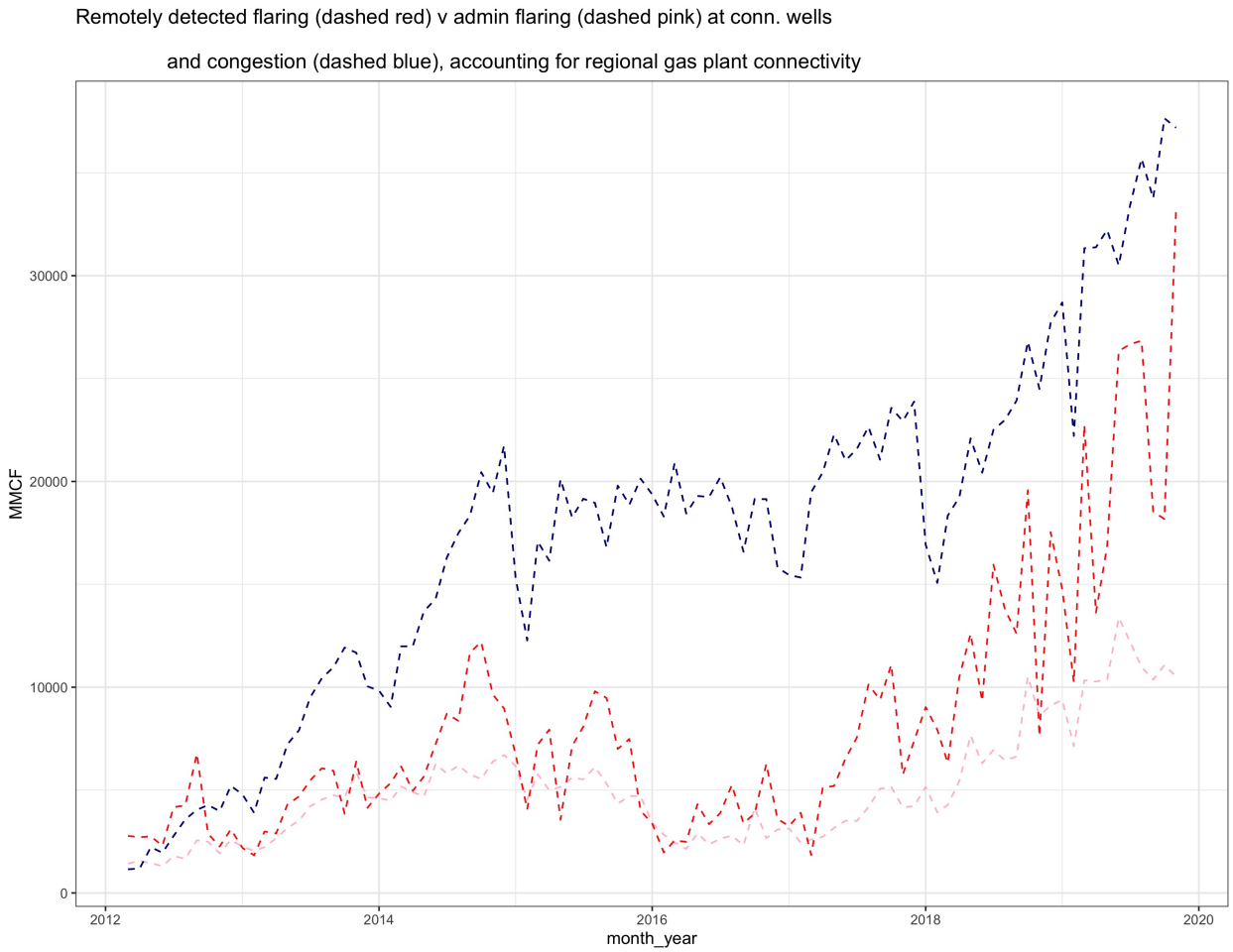


Figure 1.17: Congestion and flaring in the Bakken, accounting for well connectivity to plants

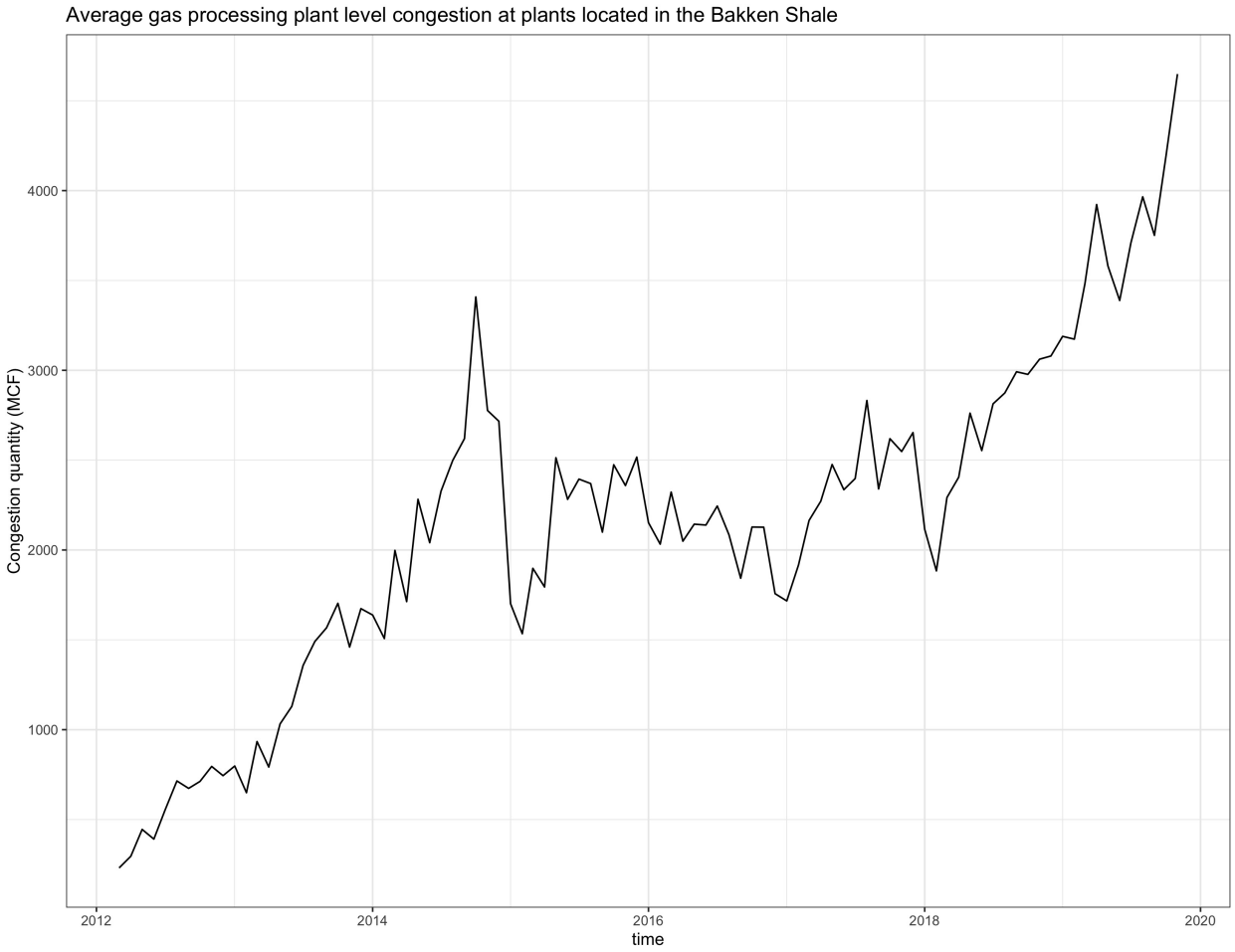


Figure 1.18: Average plant level congestion at gas processing plants in the Bakken

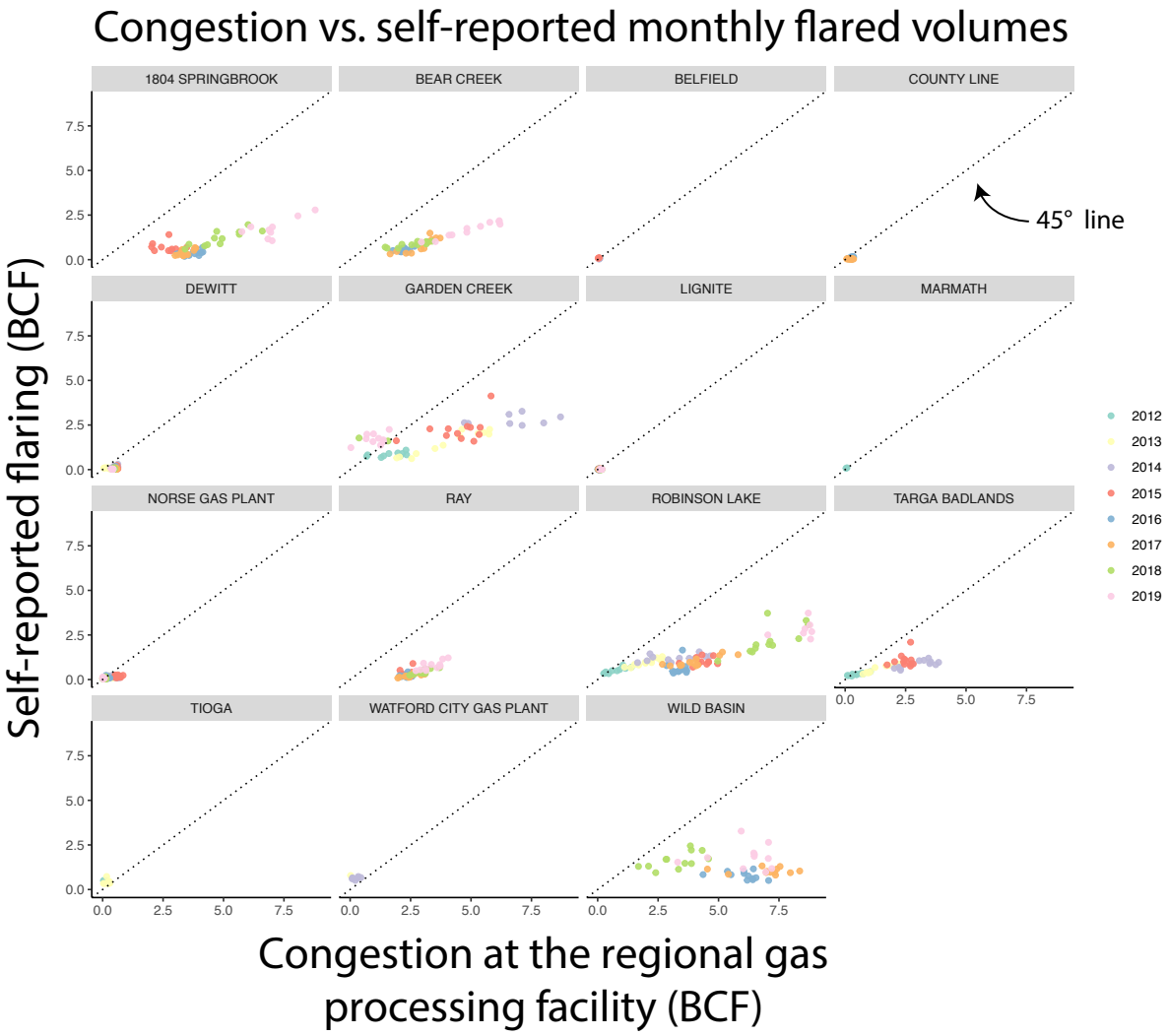


Figure 1.19: Monthly plant level congestion vs. self-reported flared volumes at connected wells

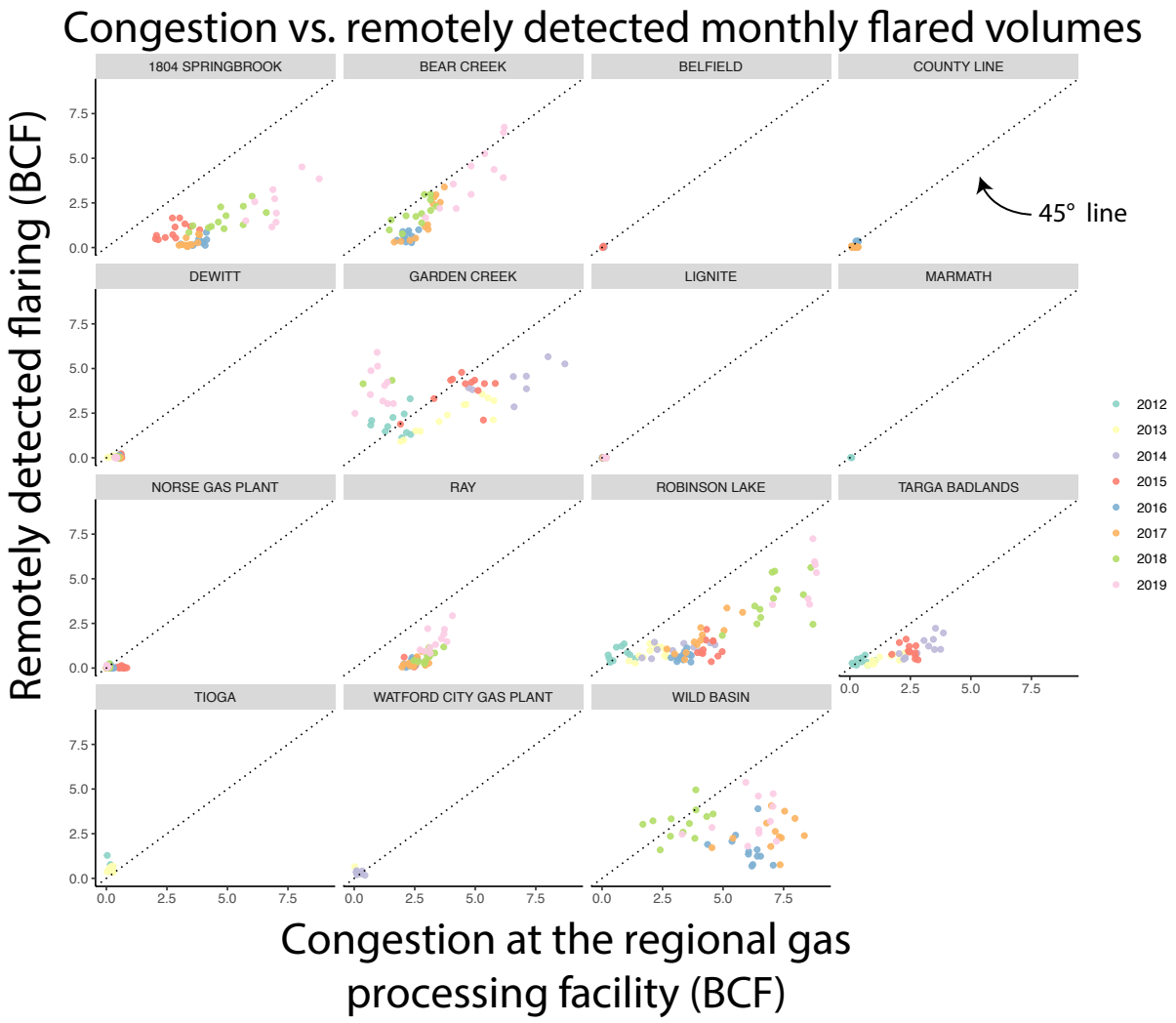


Figure 1.20: Monthly plant level congestion vs. remotely detected flared volumes at connected wells

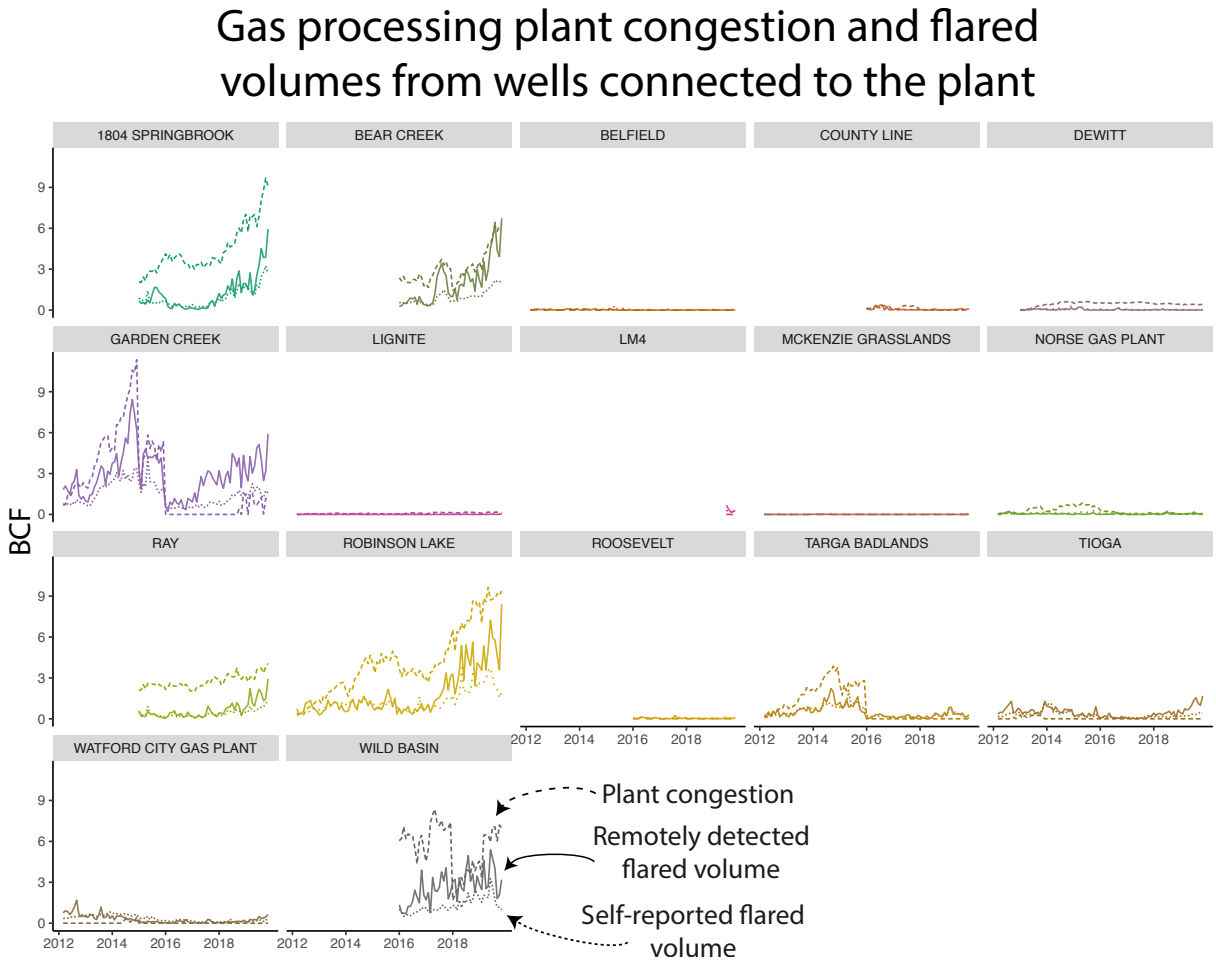


Figure 1.21: Congestion and flaring by gas plant

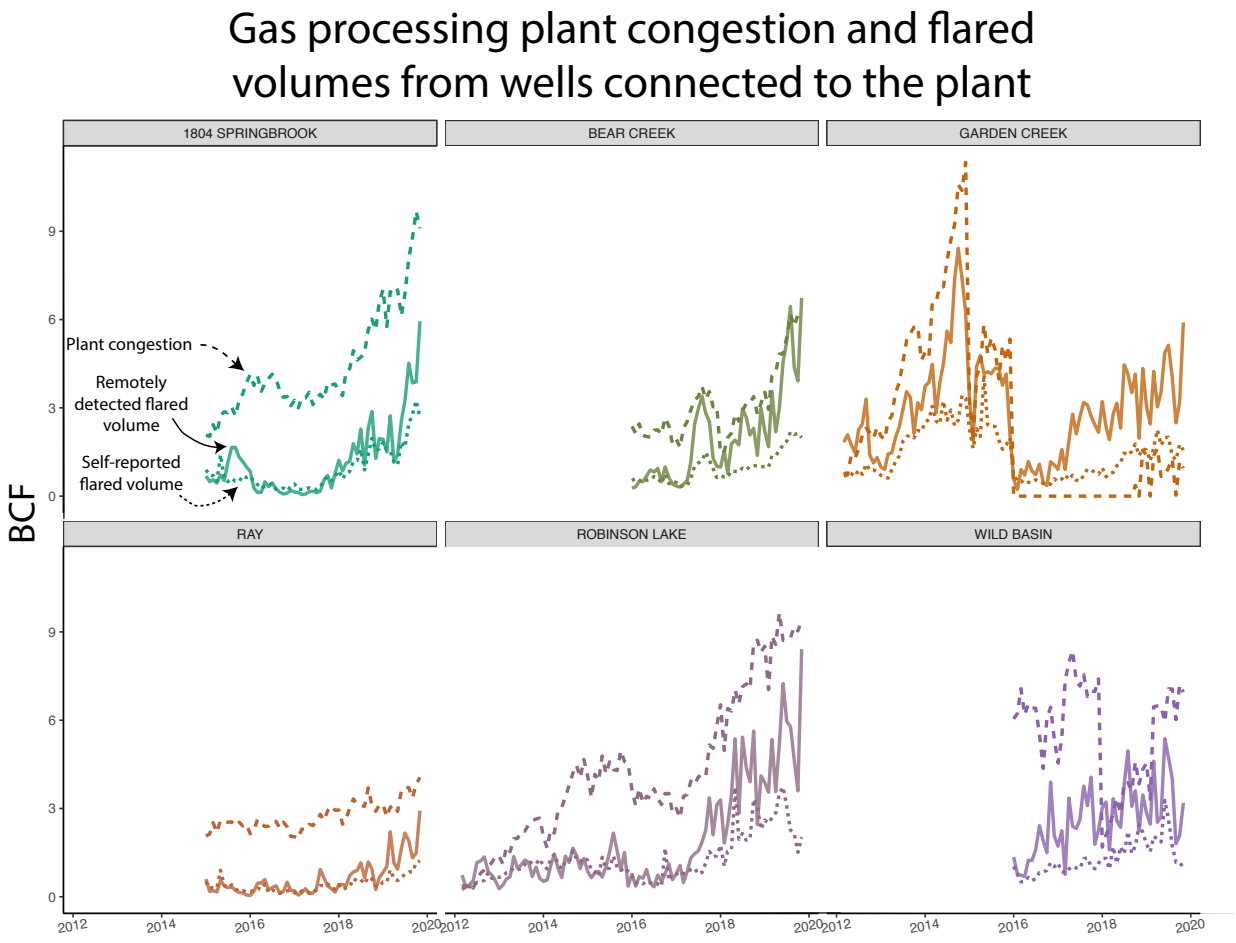


Figure 1.22: Congestion and flaring by gas plant, at the six plants with the largest flares

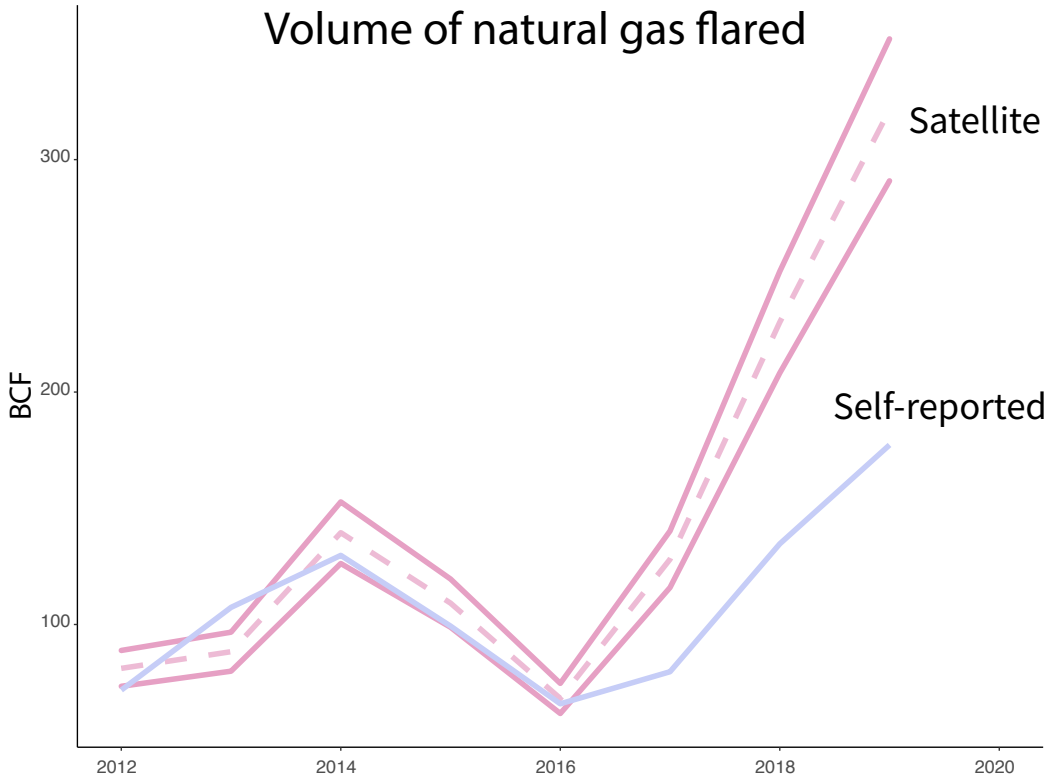


Figure 1.23: Total quantity of annualized flared gas by ND Bakken wells (that were drilled between 2010-2019), comparing volumes in the self-reported data to volumes in the remotely detected data.

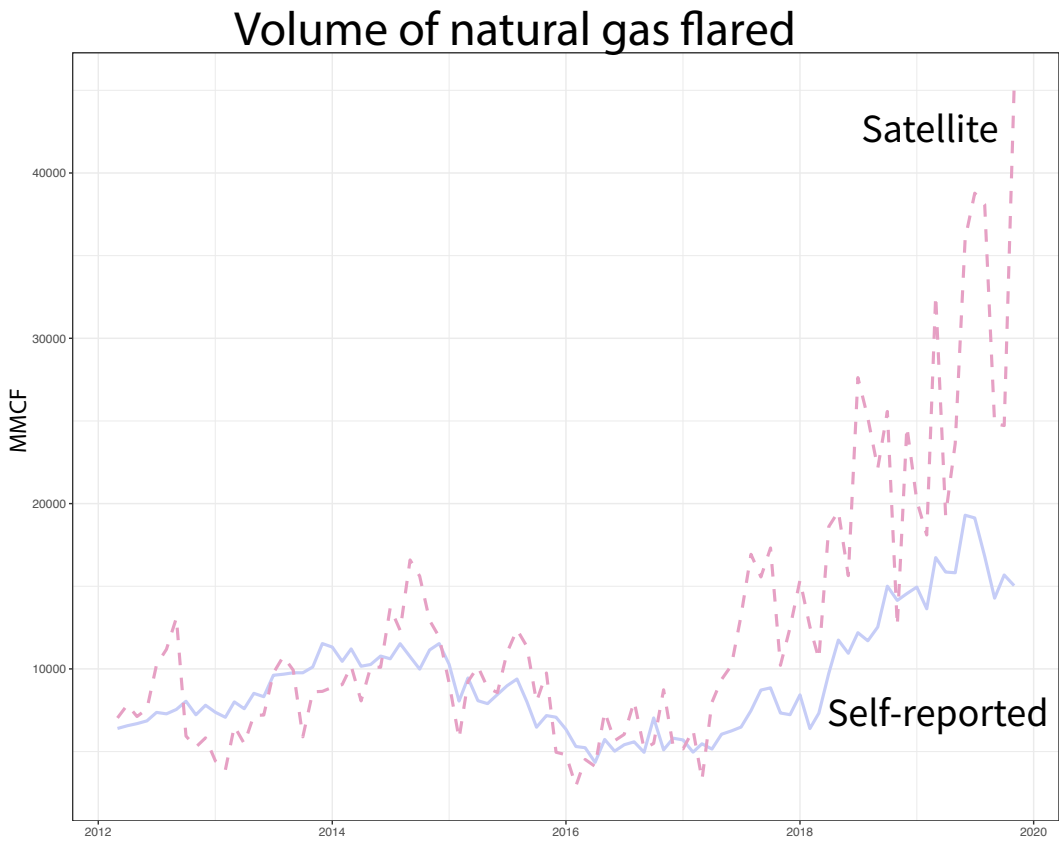


Figure 1.24: Total quantity of monthly flared gas by ND Bakken wells (that were drilled between 2010-2019), comparing volumes in the self-reported data to volumes in the remotely detected data.

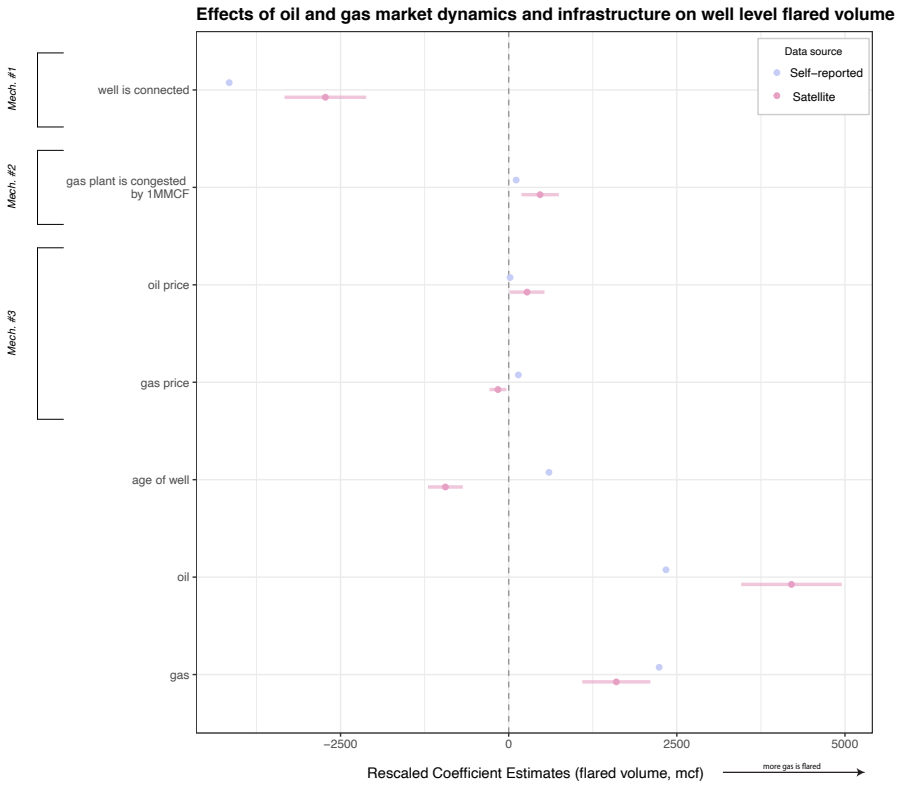


Figure 1.25: Effect of market mechanisms on well level flared gas volume

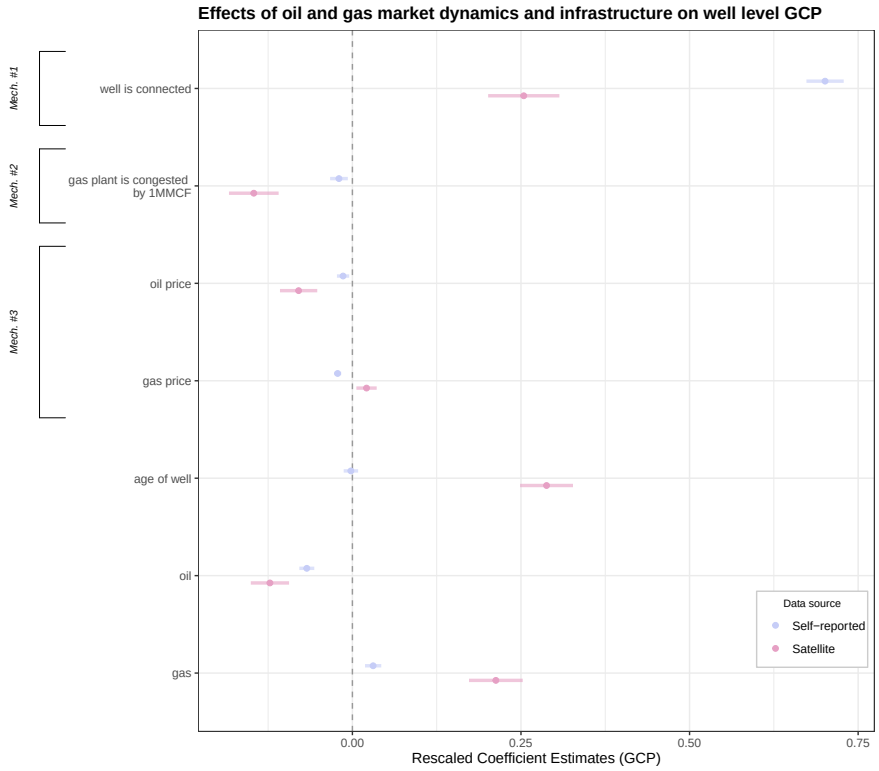


Figure 1.26: Effect of market mechanisms on well level gas capture percentage

‘

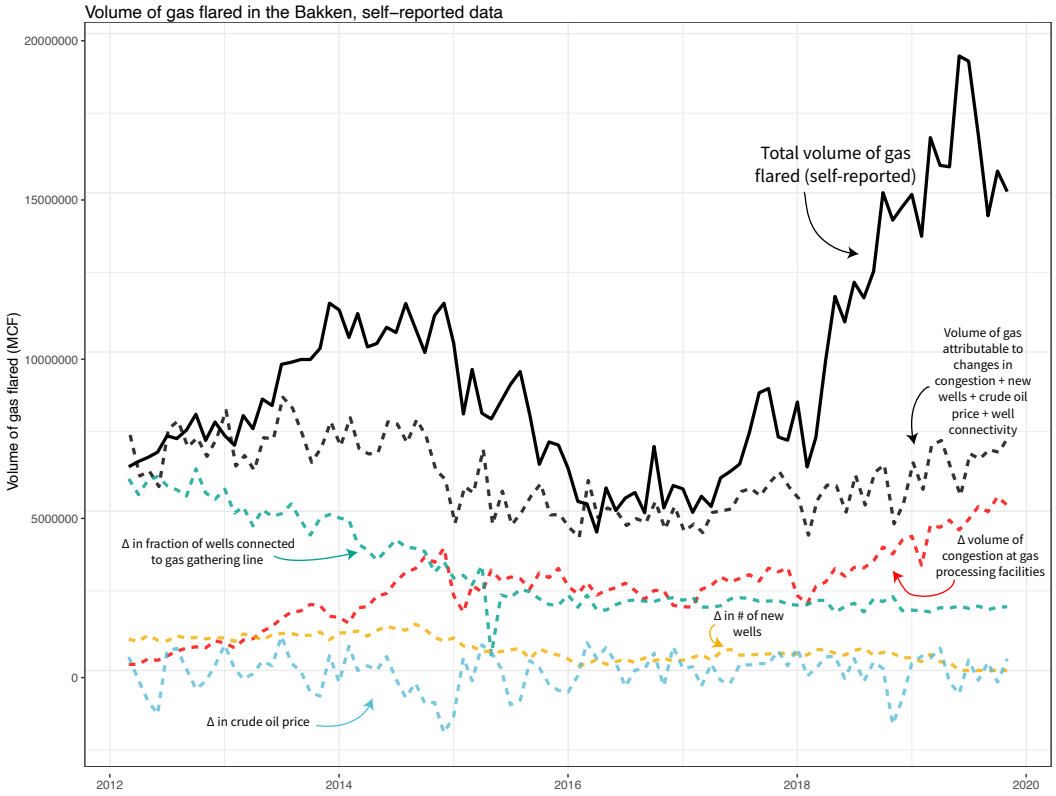


Figure 1.27: Volume of self-reported flared gas explained by well connectivity, changes in oil price, congestion at gas processing plants, and new wells

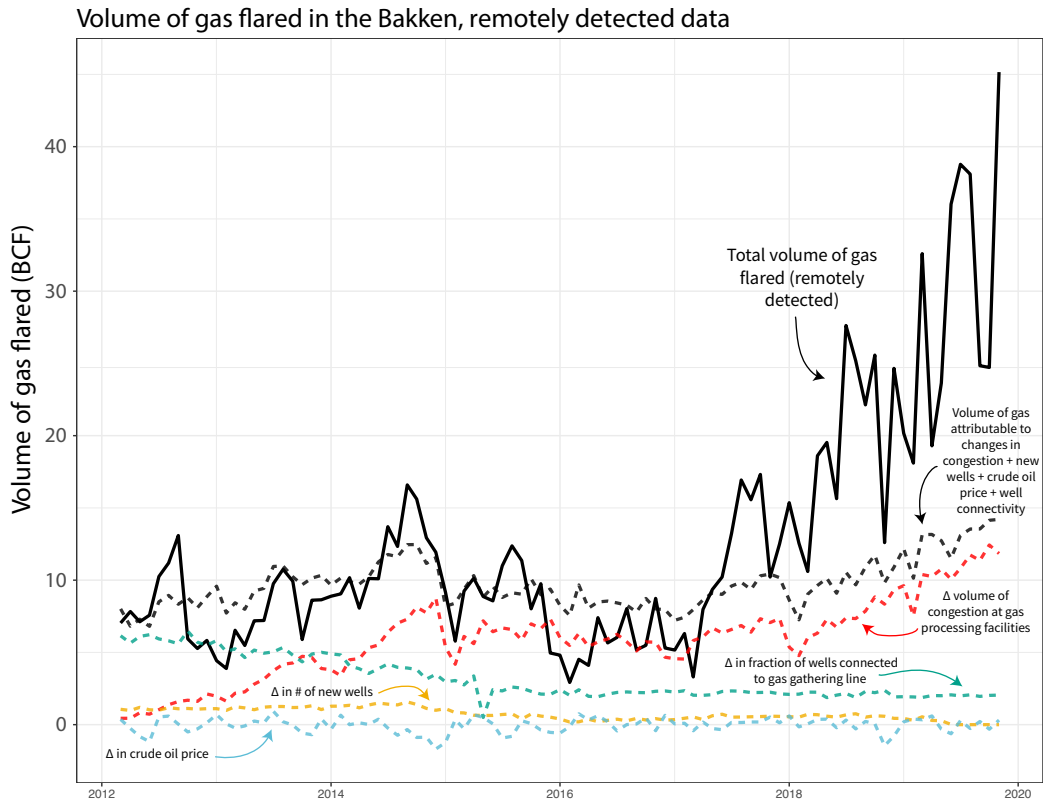


Figure 1.28: Volume of remotely detected flared gas explained by well connectivity, changes in oil price, congestion at gas processing plants, and new wells

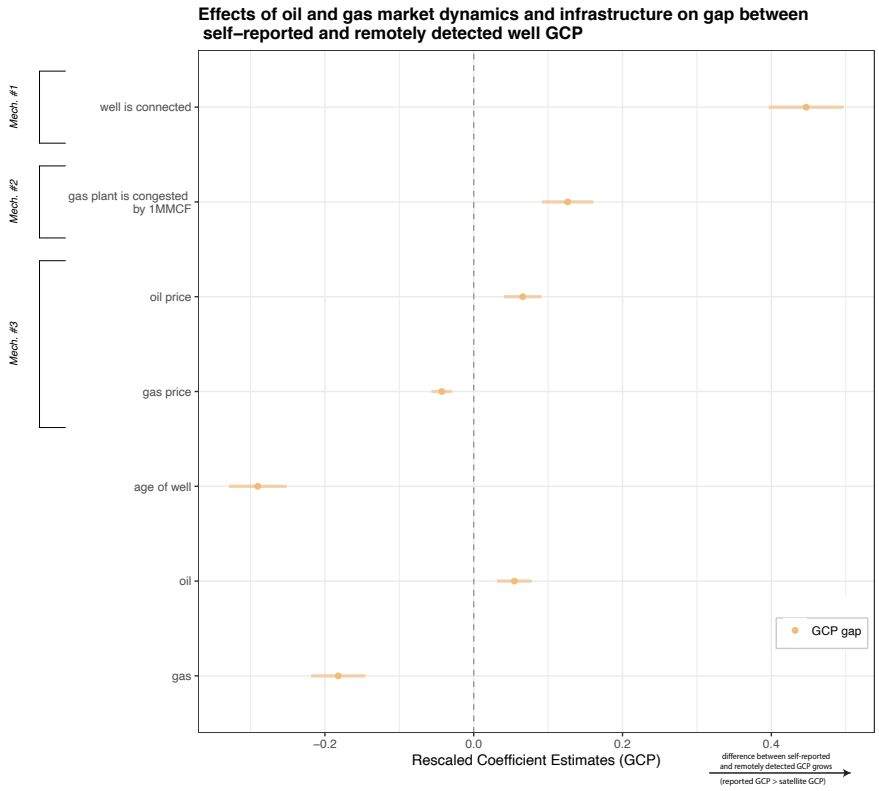


Figure 1.29: Effect of market mechanisms on the difference between well level remotely detected and self-reported gas capture percentage

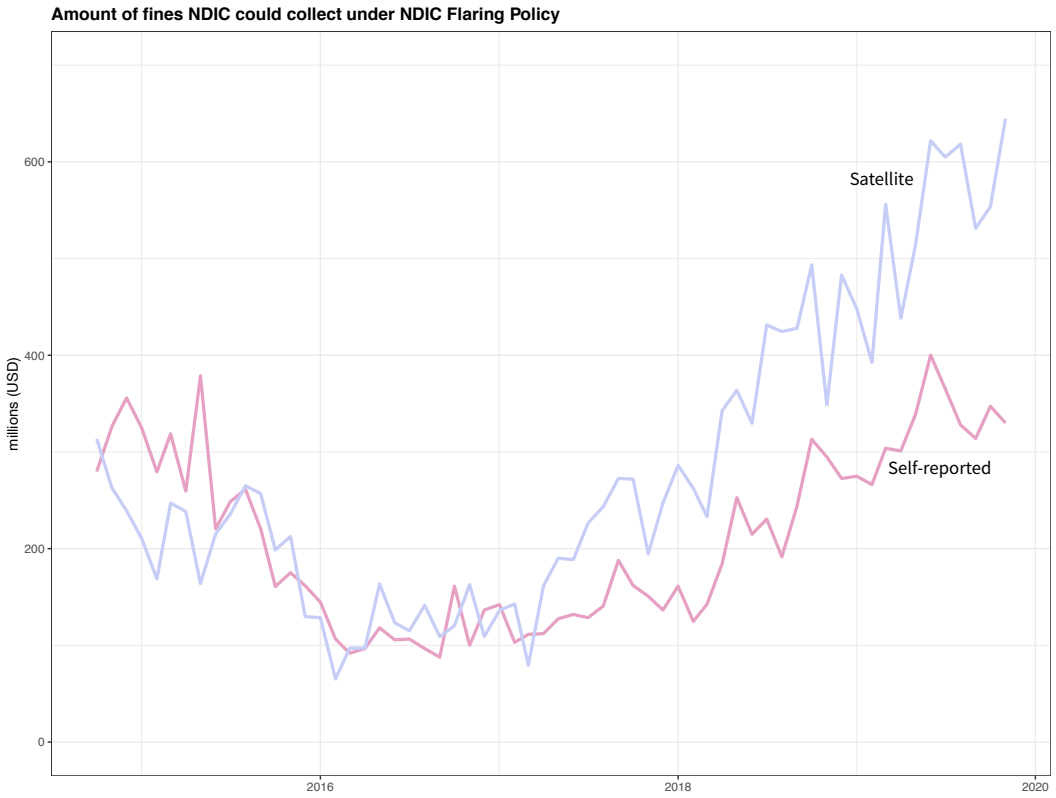


Figure 1.30: Time series of monthly fines the NDIC could assess using the self-reported versus the remotely detected GCP data.

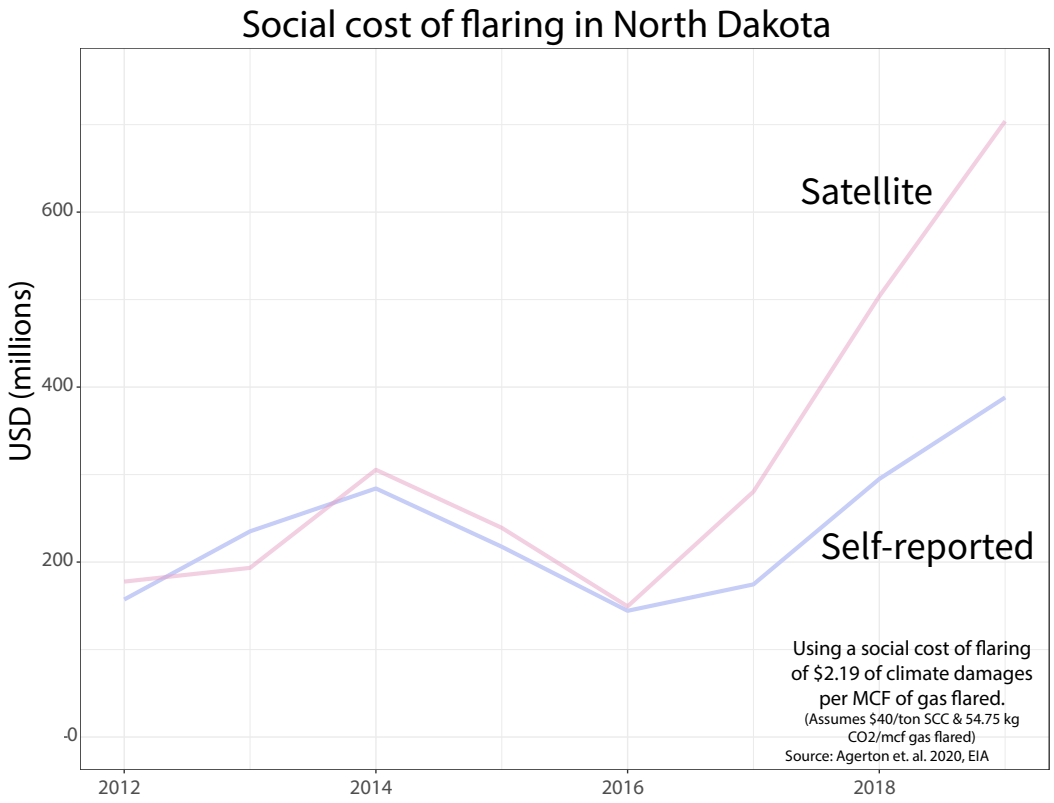


Figure 1.31: Social cost of flaring, accounting for only climate damages ($40 social cost of carbon)

Tables

Table 1.1: Difference in differences using MT remotely detected and ND remotely detected well-level GCP

	Dependent variable:			
	remotely detected GCP			
	(1)	(2)	(3)	(4)
policy announcement in ND	−0.008	−0.023	−0.023	−0.041
	(0.035)	(0.036)	(0.036)	(0.038)
74% GCP in ND	0.023	0.008	0.008	−0.010
	(0.042)	(0.043)	(0.043)	(0.046)
77% GCP in ND	0.080***	0.074***	0.074***	0.058*
	(0.027)	(0.028)	(0.028)	(0.030)
80% GCP in ND	0.073**	0.076**	0.076**	0.057*
	(0.029)	(0.030)	(0.030)	(0.033)
85% GCP in ND	0.051*	0.059*	0.059*	0.042
	(0.031)	(0.032)	(0.032)	(0.034)
88% GCP in ND	0.066*	0.074*	0.074*	0.131***
	(0.040)	(0.040)	(0.040)	(0.048)
ND	−0.197***	−0.186***		
	(0.031)	(0.034)	(0.000)	(0.000)
asinh(oil)	−0.015	−0.017**	−0.017**	−0.027***
	(0.009)	(0.008)	(0.008)	(0.007)
asinh(gas)	0.054***	0.088***	0.088***	0.116***
	(0.011)	(0.010)	(0.010)	(0.011)
age of well	0.0002***	0.0002***	0.0002***	0.0003***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Constant	0.444***			
	(0.066)			
Firm FE		✓	✓	✓
Month-year FE		✓	✓	✓
Well FE			✓	
County FE				✓
Observations	828,669	827,860	827,860	820,958
R^2	0.050	0.111	0.111	0.125
Adjusted R^2	0.050	0.111	0.111	0.125
Residual Std. Error	0.587	0.569	0.569	0.566
DF	828652	827671	827671	820752

Note: All regressions clustered at the lat-lon pixel. *p<0.1; **p<0.05; ***p<0.01

Table 1.2: Difference in differences using MT remotely detected and ND self-reported well-level GCP

	Dependent variable:			
	GCP (self-reported)			
	(1)	(2)	(3)	(4)
policy announcement in ND	0.014	0.009	0.009	0.003
	(0.033)	(0.033)	(0.033)	(0.035)
74% GCP in ND	0.064*	0.060	0.060	0.054
	(0.037)	(0.038)	(0.038)	(0.040)
77% GCP in ND	0.079***	0.073***	0.073***	0.066***
	(0.024)	(0.024)	(0.024)	(0.025)
80% GCP in ND	0.064***	0.058**	0.058**	0.048*
	(0.024)	(0.025)	(0.025)	(0.027)
85% GCP in ND	0.124***	0.120***	0.120***	0.113***
	(0.026)	(0.027)	(0.027)	(0.028)
88% GCP in ND	0.197***	0.193***	0.193***	0.197***
	(0.038)	(0.037)	(0.037)	(0.045)
well is in ND	−0.172***	−0.141***		
	(0.024)	(0.025)	(0.000)	(0.000)
asinh(oil)	−0.031***	−0.024***	−0.024***	−0.036***
	(0.004)	(0.004)	(0.004)	(0.003)
asinh(gas)	0.035***	0.028***	0.028***	0.039***
	(0.004)	(0.004)	(0.004)	(0.004)
age of well	0.00004***	0.00003***	0.00003***	0.00003***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Constant	0.819***			
	(0.033)			
Firm FE		✓	✓	✓
Month-year FE		✓	✓	✓
Well FE			✓	
County FE				✓
Observations	828,669	827,860	827,860	820,958
R^2	0.050	0.084	0.084	0.099
Adjusted R^2	0.050	0.083	0.083	0.098
Residual Std. Error	0.308	0.302	0.302	0.300
DF	828652	827671	827671	820752
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 1.3: Mechanisms driving self-reported flared volumes

	Dependent variable:			
	volume of gas flared (MCF)			
	(1)	(2)	(3)	(4)
well is connected	−4,472.771***	−4,156.160***	−4,155.554***	−2,840.242***
	(20.596)	(16.824)	(16.822)	(20.133)
congestion at gas plant (MMCF)	140.768***	20.125***	20.615***	906.730***
	(1.653)	(1.363)	(1.363)	(7.723)
WTI crude oil price	6.796***	2.777***	0.532	0.614
	(0.640)	(0.522)	(0.554)	(0.549)
HH gas price			103.439***	97.855***
			(8.545)	(8.468)
age of well		0.394***	0.394***	0.400***
		(0.006)	(0.006)	(0.006)
oil (bbl)		0.247***	0.247***	0.238***
		(0.001)	(0.001)	(0.001)
gas (MCF)		0.129***	0.129***	0.131***
		(0.001)	(0.001)	(0.001)
well is connected × congestion at gas plant (MMCF)				−901.508***
				(7.736)
Month FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Observations	746,182	746,182	746,182	746,182
R^2	0.095	0.399	0.399	0.409
Adjusted R^2	0.095	0.398	0.399	0.409
Residual Std. Error	3,380.160	2,755.600	2,755.331	2,730.594
DF	746093	746090	746089	746088
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 1.4: Mechanisms driving satellite flared volumes

	Dependent variable:			
	volume of gas flared (MCF)			
	(1)	(2)	(3)	(4)
well is connected	−3,379.345***	−2,725.077***	−2,725.749***	−1,016.791***
	(32.724)	(27.802)	(27.801)	(33.387)
congestion at gas plant (MMCF)	252.145***	86.976***	86.434***	1,237.744***
	(2.626)	(2.252)	(2.253)	(12.807)
WTI crude oil price	9.011***	4.566***	7.053***	7.159***
	(1.017)	(0.862)	(0.915)	(0.910)
HH gas price			−114.573***	−121.829***
			(14.121)	(14.043)
age of well		−0.619***	−0.619***	−0.610***
		(0.011)	(0.011)	(0.011)
oil prod. (bbl)		0.444***	0.444***	0.433***
		(0.002)	(0.002)	(0.002)
gas prod. (MCF)		0.092***	0.092***	0.095***
		(0.001)	(0.001)	(0.001)
well is connected × congestion at gas plant (MMCF)				−1,171.311***
				(12.829)
Month FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Observations	746,182	746,182	746,182	746,182
R^2	0.070	0.331	0.331	0.339
Adjusted R^2	0.070	0.331	0.331	0.339
Residual Std. Error	5,370.652	4,553.745	4,553.547	4,528.323
DF	746093	746090	746089	746088
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 1.5: Impact of primary mechanisms on self-reported well GCP

	Dependent variable:			
	GCP (self-reported)			
	(1)	(2)	(3)	(4)
well is connected		0.701***	0.701***	0.724***
		(0.014)	(0.014)	(0.012)
congestion at gas plant (MMCF)		−0.004***	−0.004***	0.012**
		(0.001)	(0.001)	(0.005)
WTI crude oil price	−0.003***	−0.001***	−0.0004***	−0.0004***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)
HH gas price			−0.016***	−0.016***
			(0.002)	(0.002)
age of well		−0.00000	−0.00000	−0.00000
		(0.00000)	(0.00000)	(0.00000)
oil prod. (bbl)		−0.00001***	−0.00001***	−0.00001***
		(0.00000)	(0.00000)	(0.00000)
gas prod. (MCF)		0.00000***	0.00000***	0.00000***
		(0.00000)	(0.00000)	(0.00000)
well is connected × congestion at gas plant (MMCF)				−0.016***
				(0.005)
Constant	1.008***			
	(0.009)			
Month FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Observations	747,169	746,182	746,182	746,182
R^2	0.034	0.314	0.314	0.315
Adjusted R^2	0.034	0.314	0.314	0.315
Residual Std. Error	0.277	0.234	0.234	0.234
DF	747167	746090	746089	746088
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 1.6: Impact of primary mechanisms on remotely detected well GCP

	Dependent variable:			
	GCP (remotely detected)			
	(1)	(2)	(3)	(4)
textwell is connected	0.287***	0.254***	0.254***	0.160***
	(0.031)	(0.027)	(0.027)	(0.025)
congestion at gas plant (MMCF)	−0.027***	−0.027***	−0.027***	−0.091***
	(0.003)	(0.003)	(0.003)	(0.012)
WTI crude oil price	−0.002***	−0.002***	−0.002***	−0.002***
	(0.0003)	(0.0004)	(0.0004)	(0.0004)
HH gas price			0.015***	0.015***
			(0.006)	(0.006)
age of well		0.0002***	0.0002***	0.0002***
		(0.00001)	(0.00001)	(0.00001)
oil prod. (bbl)		−0.00001***	−0.00001***	−0.00001***
		(0.00000)	(0.00000)	(0.00000)
gas prod. (MCF)		0.00001***	0.00001***	0.00001***
		(0.00000)	(0.00000)	(0.00000)
well is connected × congestion at gas plant (MMCF)				0.065***
				(0.012)
Month FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Observations	746,182	746,182	746,182	746,182
R^2	0.083	0.116	0.116	0.118
Adjusted R^2	0.082	0.115	0.116	0.117
Residual Std. Error	0.585	0.574	0.574	0.574
DF	746093	746090	746089	746088
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 1.7: Impact of mechanisms on the GCP gap

	Dependent variable:			
	GCP gap			
	(1)	(2)	(3)	(4)
well is connected	0.419***	0.447***	0.447***	0.564***
	(0.028)	(0.026)	(0.026)	(0.025)
congestion at gas plant (MMCF)	0.023***	0.024***	0.023***	0.103***
	(0.003)	(0.003)	(0.003)	(0.011)
WTI crude oil price	0.001***	0.001***	0.002***	0.002***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
HH gas price			−0.031***	−0.031***
			(0.005)	(0.005)
age of well		−0.0002***	−0.0002***	−0.0002***
		(0.00001)	(0.00001)	(0.00001)
oil prod. (bbl)		0.00001***	0.00001***	0.00001***
		(0.00000)	(0.00000)	(0.00000)
gas prod. (MCF)		−0.00001***	−0.00001***	−0.00001***
		(0.00000)	(0.00000)	(0.00000)
well is connected × congestion at gas plant (MMCF)				−0.081***
				(0.011)
Month FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
Observations	746,182	746,182	746,182	746,182
R^2	0.083	0.111	0.111	0.114
Adjusted R^2	0.083	0.111	0.111	0.114
Residual Std. Error	0.593	0.583	0.583	0.582
DF	746093	746090	746089	746088
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 1.8: Impact of increasing connectivity, gas plant congestion, and oil prices, on firm level flaring

	Dependent variable: Flared volume (MCF)	
	self-reported	satellite
	(1)	(2)
percent wells connected	−2,794.000***	−2,856.000**
	(979.000)	(1,367.000)
congestion at gas plant (MMCF)	6,766.000	15,432.000*
	(4,524.000)	(8,503.000)
WTI oil price	2,049.000***	1,762.000
	(659.000)	(1,148.000)
number of new wells	6,017.000	6,507.000
	(3,675.000)	(5,177.000)
mean age of well	−216.000**	−576.000***
	(88.000)	(170.000)
Month FE	✓	✓
Year FE	✓	✓
Firm FE	✓	✓
Robust SE	✓	✓
Firm cluster	✓	✓
Observations	1,825	1,825
R^2	0.650	0.670
Adjusted R^2	0.640	0.660
Residual Std. Error (df = 1750)	218,926.000	361,007.000
Note:		*p<0.1; **p<0.05; ***p<0.01

Chapter 2

Who is responsible for damaging the commons? Identifying nonpoint source polluters in national-scale river networks[1]

2.1 Introduction

Worldwide, freshwater and coastal ecosystems increasingly suffer from the effects of eutrophication (Diaz and Rosenberg 2008). Excess nutrients in water systems cause detrimental ecological, social, and economic outcomes (Breitburg et al. 2018; Yu et al. 2019). Nonpoint source agricultural runoff resulting from excess fertilizer application and manure is a primary driver of water pollution (Wuepper et al. 2020) because the runoff source and quantity are difficult to monitor. In contrast, the origin and trajectory of point source pollution are observable, enabling legislation to target and reduce point source pollution (Lapworth et al. 2012; Keiser and Shapiro 2019; He, Wang, and Zhang 2020). The main obstacle policymakers face when regulating nonpoint source water pollution is localizing the land based pollutant source, since environmental policies need the ability to target the polluting area and evaluate whether the policy was effective in reducing pollution (Stuart, Schewe, and McDermott 2014; Reimer, Denny, and Stuart 2018; Kanter et al. 2020).

[1]This chapter is coauthored with Julia Longmate, Emma Krasovich, Jeanette Tseng, Kendon Bell, Sandy Sum, and Solomon Hsiang. We thank Gabriel Englander, Andy Hultgren, Tamma Carleton, Ellen Bruno, Nick Hagerty, Joseph Shapiro, Nick Depsky, seminar participants at University of California Berkeley, AERE Summer Conference, Global Policy Lab Work-in-Progress Workshop, Columbia Interdisciplinary PhD Workshop in Sustainable Development, and Stanford University for their helpful comments. We also thank Jon Marks at the Greater Wellington Regional Council, Debbie Eastwood and Bevan Jenkins at the Waikato Regional Council, the Horizons Regional Council, Andy Hicks and Jeff Cooke at the Hawke's Bay Regional Council, Darren Gerretzen and Lisa Bevan at the Bay of Plenty Regional Council, Steffi Henkel and Val Hadsworth at the Marlborough Regional Council, Stefan Beaumont and Jonny Horrox at the West Coast Regional Council, Marc Ettema, Pete Stevenson, and Andrew Egan at the Otago Regional Council, Gail Townsend and Jason Donaghy at the Northland Regional Council, Fiona Jansma and Fiza Hafiz at the Taranaki Regional Council, Matt McLarin and Trevor James at the Tasman Regional Council, the Environment Data Team at the Auckland Regional Council, Joe Alipin at Gisborne Regional Council, and LAWA for their help in acquiring water data.

Existing water quality and nonpoint source hydrological research combines empirical- and process-based engineering models (Alexander et al. 2008; Schwarz et al. 2006; Elliott et al. 2005; Ruddy, Lorenz, and Mueller 2006; Garcia, Hoos, and Terziotti 2011; Smith et al. 2012; Garcıa et al. 2016; Al-Hamdan et al. 2017; Fu et al. 2019; Yuan, Sinshaw, and Forshay 2020). Their main objectives are modeling biophysical processes (e.g. how pollutants move across land or in water) and predicting water quality outcomes (e.g. how nutrient loading changes under different climatic conditions or management practices). These mechanistic approaches rely on parameters that are either directly measured, estimated from the literature, or calibrated, in order to model three main processes: (i) how excess nutrients are generated on land; (ii) how these nutrients are transported from land-to-stream; and (iii) and how nutrients are transported in-stream.

These hydrological models are applied to catchments of varying sizes. Models applied to large catchments with few water quality monitoring stations trade off between generating spatially and temporally granular predictions with the number of assumptions made when selecting model parameters. Moreover, mechanistic models often assume that explanatory variables, such as land use type or farm management practices, uniformly impact water quality outcomes.

In contrast, given that our goal is to identify the land-based origin of pollutants, our methodology relies on fitting much fewer model parameters. Our paper takes an on-the-ground data driven approach to localizing nutrient runoff, and relies on measuring pollutant loads at adjacent water monitoring stations. Our result is an annual sub-watershed-level dataset on nitrogen- and phosphorus-based pollutant loads. This dataset highlights which sub-watershed regions exhibit higher nutrient loss within the United States Mississippi River Basin (USMRB) and the country of New Zealand.

2.2 Context

2.2.1 Agricultural land use

The USMRB and New Zealand have thriving farming and livestock industries. As a direct result, their water ecosystems suffer from agricultural runoff-related water pollution.

2.2.1.1 US Mississippi River Basin

As the largest river basin in the US and the fourth largest in the world, the USMRB covers 1,245,000 square miles and has 31 states draining into it. It has four major rivers: Missouri, Arkansas, Mississippi, and Ohio Rivers. Agriculture has long been a major land use activity within the USMRB, and much of the United State's farmland remains within this watershed. The USMRB accounts for over 90% of the country's agricultural exports (National Park Service 2021) and contains 80% of the U.S maize and soybean acreage (Broussard and Turner 2007). It also contains the majority of America's most intensively farmed cropland where the bulk of synthetic nitrogen and phosphorus based fertilizers are applied. Additionally, the watershed also raises most of the country's livestock and pigs.

Detrimental impacts of high levels of nutrient pollution are exemplified by the Mississippi River Basin, which drains into the Gulf of Mexico and has been linked to algal blooms and eutrophication, ultimately depleting oxygen from the ecosystem and causing massive fish kills. The oxygen-depleted area, known as the hypoxic or the "dead" zone, has covered areas ranging from 6,000 to 7,000 square miles. This has potential to massively disrupt the US seafood industry since the Gulf supplies 72% of U.S. harvested shrimp, 66% of harvested oysters, and 16% of commercial fish (Bruckner 2019; Potash and Institute 1999).

Excess fertilizer from this watershed flows into the Mississippi River and eventually into the Gulf of Mexico; such eutrophication results in hypoxic conditions in the Gulf known as the "dead zone," which had a five-year average size of 5400 square miles in 2020 (Environmental Protection Agency 2020).

2.2.1.2 New Zealand

The livestock industry plays both an important role in the country's GDP and is also a primary component of the land use across the North and South Islands. In particular, the livestock industry is dominated by the pastoral farming of dairy cattle, beef cattle, and sheep. There have been significant changes to the livestock industry between 1990 - 2019 (see Figure 2.1). In particular, the number of sheep have dropped by over 50%, from close to 58 million sheep to 27 million sheep. During this time period, the number of beef cattle have decreased by 15%, from 4.6 million to 3.9 million cattle However, the number of dairy cattle has increased by over 80%, from 3.4 million to 6.3 million cattle. At the same time, livestock farming detrimentally impacts the local environment and 95% of rivers that flow through pastoral land are classified as polluted (*Our Freshwater Report 2020* 2020).

During the early 1990s, the farms with the highest density of sheep were located in Southland and Otago (see left panel in Figure 2.4). Some farms had over 800 sheep per square kilometer. Other areas with high sheep density were also located in the South Island, along the coastline in Canterbury. In the North Island, farms in the regional councils of Wellington, Manawatu-Wanganui, and Hawke's Bay also had fairly high sheep density (ranging from 200 – 800 sheep per square kilometer). However, by 2017 (see right panel in Figure 2.4), the density of sheep on farms had declined substantially. While sheep density was still higher in the same regions as in the early 1990s, the sheep density had dropped to around 200 – 400 sheep per square kilometer.

During the early 1990s, the farms with the highest density of beef cattle were located in the North Island, especially around the Northland, Auckland, Waikato Manawatu-Wanganui, Wellington regional councils, and part of Gisborne (see left panel in Figure 2.2). The density of beef cattle on farms in these regions hovered between 50-100 beef cattle per square kilomter. Similar to sheep density, beef density also fell by 2017 (see right panel in Figure 2.2). The beef cattle density remained highest in the North Island, especially in Northland, western Waikato, and eastern portions of Gisborne and Hawke's Bay, as well as across Manawatu-Wanganui. For these farms, the beef cattle density in these regions was between 25-75 cattle per square kilometer.

Lastly, the dairy cattle industry really took off in the thirty years between 1994 - 2017 (see Figure 2.3). In 1994, dairy cattle farming was localized in Waikato and Taranki regional

councils. Dairy cattle density on farms in those areas was above 100 dairy cattle per square kilometer. In 2017, dairy cattle density remained at similarly high levels in those two regional councils. However, dairy cattle farming also grew in Bay of Plenty and Manawatu-Wanganui in the North Island, as well as in Canterbury and Southland in the South Island.

2.2.2 Regulations targeting water quality

Despite the mounting environmental, health, and social costs caused by eutrophication, legislation aimed at reducing nonpoint source pollution from agricultural activity largely does not exist in both countries. The United State's Clean Water Act, which passed in 1972 and legislates national water quality, explicitly exempts regulating agricultural runoff. There are regional voluntary compliance initiatives, such as the USGS' thirteen-state Mississippi River Basin Healthy Watersheds Initiative (United States Department of Agriculture 2021). This initiative pays landowners to implement on-farm conservation practices to reduce nutrient runoff, and the USGS uses an edge-of-field (EOF) monitoring technique (United States Geological Service 2016) to measure the impact of these activities on nutrient losses. While EOF monitoring is effective at precisely measuring the impact of farming practices on nutrient runoff, it requires specialized equipment, manpower and year round monitoring, and is costly. As such, it can be difficult implement this approach at the larger scale needed to support a nationwide water monitoring campaign. Instead, EOF monitoring needs to be complemented with a low cost solution that can be more widely implemented.

In New Zealand, a series of national freshwater management regulations have been passed in recent years (*National Policy Statement for Fresh Water Management 2014* 2017), with the latest set of policies, collectively named the Essential Freshwater package, implemented in September 2020 (Ministry for Primary Industries Manatu Ahu Matua 2021). These policies govern polluting farming practices, such as restrictions on animal grazing and fertilizer application, however, they do not directly measure how changes to these farming practices reduce nutrient runoff (Ministry for the Environment 2021).

2.2.3 Nitrogen and phosphorous water pollution

In our paper, we focus on nutrient runoff that is associated with agricultural activity. The dominant chemicals found in synthetic fertilizers and animal waste, and also the cause of eutrophication in waterways, are phosphorus (P)- and nitrogen (N)-based chemical compounds. As such, we estimate the P- and N-based pollutant loads generated at basins within New Zealand and United States MRB. The P- and N-based pollutants we are able to model depend on the water quality data that is collected by government water quality monitoring sites in these two countries. The exact nutrient compounds differ between the USMRB and New Zealand given the different sources of pollution.

For the purposes of the USMRB, given USGS and EPA's focuses and data availability, our main results will focus on pollutant loads from total nitrogen, total Kjedahl nitrogen, and total phosphorus. For the context of New Zealand, given the regional councils' focus and data availability, our main results will focus on pollutant loads from total nitrogen, ammonia (as nitrogen), total phosphorus, and dissolved reactive phosphorus. For all the N- and P-

compounds in our analysis, we are giving the concentration associated with only the N or P component of that molecule. For example, ammonia is NH_3, or three hydrogen molecules and one nitrogen molecule. For our analysis, we only refer to the one nitrogen molecule that is part of NH_3, which is why specify "ammonia (as nitrogen)." In our results, when we discuss the kilograms of nitrogen or phosphorus pollution, the is the quantity associated with only N- or P- pollution (rather than also including the weights associated with the other molecules that may be part of the chemical compound).

Next, we provide background on the N- and P-chemical compounds that we focus on:

2.2.3.1 Total nitrogen

Nitrogen, an essential chemical needed for plant and animal growth, is a common chemical component of synthetic fertilizers and animal waste, and also contributes to eutrophication of waterways. The three main nitrogen forms found in water systems are nitrate, nitrite, and ammonia. Total nitrogen is the sum of these three forms of nitrogen.

2.2.3.2 Total phosphorus

Similarly, phosphorus is also a chemical vital for plant and animal life. This chemical is also an integral component of synthetic fertilizers and animal waste, and contributor of eutrophication. Phosphorus tends to bind to soil particles, and runoff into the waterway (both ground and surface water) via soil particles. Total phosphorus is the sum of orthophosphate (also known as soluble reactive phosphorus or dissolved reactive phosphorus) and the suspended phosphorus found in plant and animal fragments in water. TP loads tend to be more stable throughout the year and are a better measure of water quality.

2.2.3.3 Dissolved reactive phosphorus

In contrast, dissolved reactive phosphorus, as mentioned above, is the chemically active form of phosphorus that is dissolved in the water solution and is directly available for uptake by plants. DRP levels in a given body of water can change often since this phosphorus form can be quickly absorbed by plants.

2.2.3.4 Total Kjedahl nitrogen

TKN is a water quality measurement method for measuring the sum of organic nitrogen and the inorganic nitrogen compounds of ammonia (NH_3) and ammonium (NH_4^+). Other forms in inorganic nitrogen, such as nitrate, are not included in this measurement method. TKN is a common method used to measure the amount of nitrogen in water samples, such as at wastewater treatment plants. (Note, inorganic and organic nitrogen are both sources of nitrogen. Organic nitrogen is the form of nitrogen found in organic compounds, i.e. compounds that contain carbon. Inorganic nitrogen is the chemical form of carbon found in compounds that don't contain carbon. However, plants can only uptake inorganic nitrogen. Organic nitrogen compounds can be converted into inorganic nitrogen compounds.)

2.2.3.5 Ammonia

Ammonia is a type of inorganic nitrogen compound, and is a common input into synthetic fertilizers. This inorganic nitrogen compound can also be found in animal manure. High concentrations of ammonia in water ecosystems can directly harm existing aquatic organisms through causing the accumulation of toxins in the bloodstream or tissues.

2.2.4 Water quality monitoring and data quality in the USMRB

In the United States, hundreds of water monitoring organizations, ranging from local agencies to state, regional, and federal bodies have collected water quality information on the nation's 3.5 million miles of waterways since the early 1800s (Sprague, Oelsner, and Argue 2017; Myers 2015; Read et al. 2017). In 2012, the Water Quality Portal (WQP) was established by the government as a cooperative database that serves as the largest, centralized access point for publicly available water quality data. The WQP collates data from a number of sources, including the USGS National Water Information System (NWIS), the EPA STOrage and RETrieval (STORET) Data Warehouse, and the USDA ARS Sustaining The Earth's Watersheds - Agricultural Research Database System (STEWARDS) (WQP, read2017water, sprague2017challenges). At the time of this writing, the WQP contains water quality data dating back to the 1892 from over 400 organizations, reporting over 342 million records at more than 900,000 water sampling sites across all 50 states (read2017water, data.gov, congress-report2017). While the WQP has the potential to be an invaluable resource in assessing water quality issues across the country, the database suffers from a severe lack of standardization between contributing organizations' methodologies for monitoring, reporting, and storing water quality data (Sprague, Oelsner, and Argue 2017; Read et al. 2017).

Specifically, the lack of standardization in water quality metadata is pervasive across the WQP. Water quality metadata are particularly important to accurately interpreting water quality information, and include critical details such as the units of measurement (e.g. mg/l or ppm), chemical form of a water quality parameter (e.g. nitrate or nitrogen), or a measured concentration value (e.g. 10), among others (Sprague, Oelsner, and Argue 2017; Gray et al. 2005). Without these details, secondary-users, such as researchers and policymakers, may not be able to accurately interpret data from the WQP (Sprague, Oelsner, and Argue 2017). As a result, water quality metadata must be harmonized in order to identify and compare water quality trends across a region in which water quality data are collected by multiple organizations.

We retrieve and harmonize WQP water quality data from 7,036 monitoring sites located within the US Mississippi River Basin (USMRB) that have sampled water quality parameters associated with agricultural activities, such as chemicals containing nitrogen and phosphorus, between 1980 and 2018. Our data collation and cleaning process use best practices detailed by key water quality monitoring organizations, including the EPA, USGS, and USDA, to remove and remedy inconsistencies between organizations. As a result, we are able to construct a standardized water quality dataset that can analyze non-point source pollution over the course of a 40-year span in the MRB.

It has been widely documented that the MRB suffers from the effects of NPS nutrient pollution due to agricultural activities (Howard 2014; (EPA) 2015). Our dataset provides water quality observations that are comparable across sites and time, which is key to assessing pollution patterns in the USMRB. To our knowledge, this is the first time that a dataset of standardized nitrogen (N) and phosphorus (P) water quality concentrations from almost four decades of observations over the vast expanse of the Mississippi River Basin will be made publicly available.

2.2.4.1 Data queried from Water Quality Portal

We selected nitrogen- and phosphorus-based nutrient compounds associated with agricultural activity and non-point source pollution. Data were retrieved for 19 differently named nutrients using an R-package specific to the WQP, dataRetrieval. These nutrient compounds include: ammonia-nitrogen, ammonia-nitrogen as N, kjeldahl nitrogen, nitrate, nitrate as N, nitrogen, nitrogen (mixed forms: NH3, NH4, NO2, NO3), organic nitrogen, total kjeldahl nitrogen (organic N NH3), total nitrogen (mixed forms), organic phosphorus, orthophosphate, orthophosphate as P, orthophosphate as PO4, phosphate-phosphorus, phosphate-phosphorus as P, phosphate-phosphorus as PO4, phosphorus, total phosphorus (mixed forms).

2.2.4.2 Cleaned water quality data

After cleaning the queried water quality data (see Krasovich et al. (2021) for steps on data cleaning), we kept 1.4 million unique water quality observations collected at over ten thousand water quality monitoring sites located within the USMRB.

2.3 Model approach and contribution

The fundamental challenge of targeting and reducing nonpoint source water pollution is identifying its source. As such, our spatial methodology addresses this problem through recovering the land-based sources of annual nitrogen- and phosphorous-based pollutant loads. We collate and harmonize extensive publicly available water quality and flow data from 128 regional environmental agencies spanning 31 states in the USMRB and 14 regional councils in New Zealand. We combine this with a hydrographically processed topographical dataset to model flow direction in order to build off of existing hydrological models that rely on the fundamental assumption that water flows downhill (see Methods). Our model contribution is the ability to partition large-scale watersheds into smaller tractable regions so that we can distinguish between when pollutants appear in the water system form where they first originated.

2.3.0.1 Summary of data contribution:

The output from our model has 7,159 unique sub-watersheds (which we call basins in this paper) for 16 nutrients in the USMRB and 1,619 unique basins for 14 nutrients in New Zealand between 1981 - 2018. The area and boundary of a basin depends on the location

of the water quality monitoring site, as well as the local topography (see Methods for more detail on how we construct a basin unit). On average in the USMRB, we have 1,275 basins measuring total nitrogen, 2,692 basins measuring total Kjedahl nitrogen, and 2,546 basins measuring total phosphorus. The average size of a basin is 800 km^2. Colorado has the largest number of basins (577 basins with an average size of 600 km^2) and Michigan and South Carolina have the fewest number of basins (each having only 1 basin, with sizes of numeric(0) km^2 and 520 km^2, respectively).

On average, in New Zealand, we have 783 basins measuring ammonia, 771 basins measuring total nitrogen, 784 basins measuring total phosphorus, and 819 basins that measure dissolved reactive phosphorus. The average size of a basin is 100 km^2. Canterbury Regional Council has the largest number of basins (245 basins with an average size of 200 km^2), and Nelson Regional Council has the fewest number of basins (20 basins with an average size of 0 km^2).

2.3.0.2 Water quality and flow datasets

The USMRB instantaneous and daily flow values are collected by the United States Geological Survey. The water quality data are retrieved from the Water Quality Portal, which is an amalgamation of data from 400 separate water agencies. To the best of our knowledge, we are the first to clean and harmonize the largest set of water quality observations from the USMRB (see Krasovich et al. (2021) for our data cleaning process; cleaned data are open access). We have 914,080 unique cleaned water quality observations from 128 water agencies and 73,600,337 unique cleaned flow observations. We also have the largest set of collated water quality and flow data for the country of New Zealand. We obtained the data from 14 out of the 16 regional councils in New Zealand through contacting local water agencies. We have 690,913 unique water quality observations and 9,943,010 unique flow observations.

2.3.0.3 High level overview of modeling approach

If we observe that the pollutant load measured at two adjacent water monitoring sites differs, we attribute the differences in nutrient loading to changes that happened at some location between the two sites (e.g. changes in farm management practices, changes in soil structure leading to soil erosion, etc). Our model estimates pollutant load using the connectivity of water monitoring stations along a river as well as measurements of both water quality and flow at the same location on the same date.

We implement temporal and spatial interpolation techniques to impute for missing daily flow values. We apply hydographic processing methods to digital elevation models to to correct for flow direction across our regions of interest and to determine the network connectivity across water monitoring stations (see Fig. 2.11, Fig. 2.12, Fig. 2.13, Fig. 2.14, Fig. 2.15).

Our unit of analysis is at the basin level, which we define as a collection of land pixels whose runoff would first flow past the same downstream monitoring station. We compute the difference of pollutant loads at adjacent water quality monitoring sites; we attribute this quantity of nutrient runoff to the basin flanked by the two monitoring sites. We use external

precipitation data to corroborate the transport of water under our modeling assumptions (see Figure 2.17 and Figure 2.18) to validate our spatial approach.

2.4 Results

Our model estimates annual cumulative (see Table 2.1 and marginal (see Table 2.2) pollutant loads at the basin level for 14 nutrient compounds in New Zealand, and 16 nutrient compounds in the USMRB. Using this dataset, we are able to estimate both the average marginal pollutant loads and the changes in pollutant loads over time. For the country of New Zealand, we are focusing on the pollutant loads associated with four main pollutant compounds: ammonia (as nitrogen), total nitrogen, dissolved reactive phosphorus, and total phosphorus. We focus on these four compounds because these excess nutrients are associated with animal waste runoff and we have the most observations for these four compounds. In the USMRB, we are focusing on pollutant loads associated with three main pollutant compounds: total Kjedahl nitrogen, total nitrogen, and total phosphorus. We focus on these three compounds because these excess nutrients are associated with chemical fertilizer runoff and we have the most observations for these three compounds.

2.4.1 Average area normalized marginal pollutant loads in New Zealand

Average area normalized marginal pollutant loads are computed by estimating the mean of the basin-level area normalized marginal pollutant load over time. This statistic allows us to compare the spatial distribution of pollution. In other words, we can compare the average pollution load across neighbors to identify which regions exhibit higher or lower nutrient loss.

2.4.1.1 Total nitrogen

The distribution of basin-level area normalized total nitrogen pollution is plotted in the choropleth in Figure 2.20. The average area normalized pollution ranges from $\leq$ -11 to $\geq$ 6,206 kg/km2/year. The 50th percentile of pollution ranges from 1 to 6 kg/km2/year. To put this value into context, high leaching potential in the regional council of Waikato, a predominant dairy producing region, is 14,000 kg of nitrogen per square kilometer (Waikato Regional Council 2021). For total nitrogen, many of the regions that were highest polluting for ammonia as N also exhibit high total nitrogen runoff. In the North Island, areas with the highest polluting basins are located in Waikato, Hawke's Bay, Northland, and Gisborne. In the South Island, the highest polluting basins are located in Southland and Canterbury.

Basins that have negative total nitrogen loads are scattered throughout the country, and are located in Hawke's Bay, Manawatu-Wanganui, Canterbury, and Southland. Basins that generate minimal or close-to-zero total nitrogen pollution are located in Taranaki, the Waitomo region of Waikato, Auckland, and Otago.

While there are clusters of basins that all generate higher or lower pollution levels, there is also substantial heterogeneity at the local level. In particular, Southland, Hawke's Bay, and Manawatu-Wanganui contain basins with positive pollutant loads flanking basins with negative pollutant loads.

2.4.1.2 Ammonia as nitrogen

The distribution of basin-level area normalized ammonia (as nitrogen) pollution is plotted in the choropleth in Figure 2.19. The average area normalized pollution ranges from $\leq$ -2 kg/km2/year to $\geq$ 161 kg/km2/year. The 50th percentile of pollution ranges from 0.02 to 0.5 kg/km2/year. In the North Island, the highest polluting basins are located in the South Waikato region in Waikato, in Gisborne, in the Wairoa region of Hawke's Bay, and in Taranaki. Within the South Island, the highest polluting basins are located in the Hurunui region of Canterbury, and in Southland.

Under twenty percent of the basins have negative average marginal loads, which means that they on average absorb pollution (e.g. if the pollution enters into the groundwater, or is diverted from the surface water system, or the nitrogen is aerosolized). These basins with negative loads are scattered throughout the country, with some located in Waikato, Manawatu-Wanganui, Hawke's Bay, and Southland.

Thirty percent of the basins also exhibit very low (close to zero) average area normalized ammonia as N pollution. This suggests that these basins produce very minimal ammonia pollution that enters into the freshwater system.

2.4.1.3 Total phosphorus

The distribution of basin-level area normalized total phosphorus pollution is plotted in the choropleth in Figure 2.21. The average area normalized pollution ranges from $\leq$ -1 kg/km2/year to $\geq$ 229 kg/km2/year, with the 50th percentile of pollution ranging from 0.03 to 0.5 kg/km2/year. About ten percent of the basins absorb total phosphorus, another 20 percent generate minimal or close-to-zero total phosphorus pollution, and the remainder generate positive total phosphorus pollution.

We observe that much of the same nitrogen polluting regions (see in Figures 2.20 and 2.19) also exhibit high TP losses. The highest polluting basins in the North Island can be found in Waikato, parts of Gisborne, parts of Hawke's Bay, and a handful in Northland. Within the South Island, the highest polluting basins are located in Southland, and in Canterbury.

There are fewer basins that have negative average TP loads. These basins are located primarily in the North Island in the Hawke's Bay and Manawatu-Wanganui regional councils. There is also a cluster of basins with negative TP loads located in the Southland regional council in the South Island.

Basins that generate minimal or close-to-zero TP runoff are located in Taranaki regional council, parts of Manawatu-Wanganui, parts of Wakiato, Northland, and in the Southland regional councils.

2.4.1.4 Dissolved reactive phosphorus

Lastly, the distribution of basin-level area normalized dissolved reactive phosphorus (DRP) pollution is plotted in the choropleth in Figure 2.22. The average area normalized pollution ranges from $\leq$ -0.4 kg/km2/year to $\geq$ 106 kg/km2/year, with the 50th percentile of pollution ranging from -0.02 kg/km2/year to 0.4 kg/km2/year. There is more heterogeneity in the spatial distribution of DRP runoff compared to TP runoff. Within the North Island, the highest polluting basins are located in Waikato, Gisborne, Hawke's Bay, and Taranaki. In the South Island, the highest polluting basins are located in Canterbury and Southland.

There are fewer basins with negative loads. In the North Island, these basins are located primarily in Manawatu-Wanganui and Hawke's Bay. In the South Island, these basins are predominantly clustered in the South Island.

Lastly, the basins that generate minimal or close-to-zero DRP runoff are located primarily in the North Island, especially within Auckland, Northland, and the Hastings region of Hawke's Bay. In the South Island, there are some basins with minimal DRP pollution located in Tasman, Marlborough, and Southland.

Across the four nutrient compounds, we observe that the highest polluting basins for a given nutrient are also more likely to generate high quantities of the other nutrient loads. This suggests that there are location specific characteristics, such as farm management practices or soil type, that contribute to overall likelihood of runoff.

Additionally, we observe that in certain regional councils, the area normalized pollutant loads estimated at neighboring basins are heterogeneous. In other words, there are basins that generate very high levels of pollution runoff that neighbor basins with very low or even negative amounts of pollution runoff. This provides evidence that nutrient runoff at nearby farms with similar land use type can differ substantially within the New Zealand context.

2.4.2 Average marginal loads in USMRB

2.4.2.1 Total nitrogen

The distribution of basin-level area normalized total nitrogen (TN) pollution is plotted in the choropleth in Figure 2.24. The range in the average area normalized TN pollution load attributed to a given basin ranges from $\leq$ -2.7 million kg/km2/year to $\geq$ 5.7 million kg/km2/year. The 50th decile spans -137 kg/km2/year to 6,762 kg/km2/year.

Over 50 percent of basins have a positive average value of TN runoff. In otherwords, these basins are responsible for generating TN runoff that enters into nearby waterways. Basins that have the highest average marginal load are located at the confluence of the Missouri and Mississippi River (in central Missouri and southern Illinois), along the Ohio River (in southern Ohio and western West Virginia), and in portions of Iowa, Indiana, Nebraska, and Kansas.

Basins that exhibit moderate increase or close-to-zero increase in pollution tend to be located in the western portion of the USMRB. In particular, they are located in western Kansas, eastern Colorado, western North Dakota, and parts of Montana, South Dakota, and Wyoming.

Basins with negative TN runoff are areas where on net, nitrogen is leaving the system. These basins are located along the major rivers, epseically near the lower half of the USMRB.

2.4.2.2 Total Kjedahl nitrogen

The distribution of basin-level area normalized total Kjedahl nitrogen pollution is mapped in Figure 2.23. The values of the average area normalized TKN pollution range from $\leq$ -1.2 million kg/km2/year to $\geq$ 1.4 million kg/km2/year. The 50th decile ranges from -23 kg/km2/year to 3,072 kg/km2/year.

Basins that have the highest average TKN runoff are dispersed throughout the USMRB, and not necessarily located next to the four major rivers. There are more high polluting basins in the central and eastern portion of the USMRB. It is worth nothing that many of the high polluting basins are located next to basins that generate negative loads.

Of the basins that generate negative average TKN loads, most are located along the four major rivers. In particular, they are clustered in the lower half of the USMRB. Basins with the largest negative loads are located in the northeastern corner of Kansas, at the confluence of the Mississippi River and Ohio River at the southern tip of Illinois, and along the eastern portion of Arkansas. However, many of the other basins with negative average TKN marginal loads are located across the other states.

Basins that generate low levels of TKN pollution are clustered in the western portion of the USMRB. In particular, they are located in western Kansas, Nebraska, eastern Colorado, eastern South Dakota. There are also basins with low levels of pollution located in southern Minnesota, and swathes of Iowa.

2.4.2.3 Total phosphorus

The distribution of basin-level area normalized total phosphorus (TP) pollution is plotted in the choropleth in Figure 2.25. The average values in area normalized TP runoff range from $\leq$ -360,384 kg/km2/year to $\geq$ 381,018 kg/km2/year. The 50th decile spans 1 kg/km2/year to 576 kg/km2/year.

Basins with positive or negative average TP runoff don't appear to be clustered spatially. There are more basins with negative TP runoff flanking the major rivers. In particular, these basins are located in the stretch of the Mississippi River south of the confluence with the Ohio River and north of the confluence with the Arkansas River. The central portion of the Missouri River (along the eastern portion of Nebraska) also has basins with negative loads.

As for basins with the highest average TP pollution, they are located in central Missouri, northwestern and central Illinois, central Arkansas, southern Ohio, eastern Kentucky, scattered throughout Indiana, and eastern Kansas.

2.4.3 Linear trends in New Zealand

In addition to estimating average basin loads which characterize the spatial distribution of runoff, we are also interested in estimating how the quantity of nutrient runoff evolves over time. Pollution trends are essential for evaluating whether runoff problem has steadily

worsened or improved over time. We estimate pollution trends separately for each basin. The changes in pollutant loads over time, or the linear trend, is computed by regressing basin level marginal pollutant loads on time, given by this equation:

$$\text{load}_{ivt} = \beta_0 + \beta_1 t + \epsilon_{ivt}$$

for basin i, nutrient v, in year t. β_1 is the linear trend in pollution for that nutrient compound in that given basin over time. $\beta_1 < 0$ indicates that pollution for that given nutrient is decreasing, whereas $\beta_1 > 0$ indicates that pollution for that given nutrient is increasing.

2.4.3.1 Total nitrogen

For changes in total nitrogen runoff over time in New Zealand see Figure 2.27. The range in the distribution of the changes in total nitrogen runoff trends span $\leq$ 212 kg/km2/year to $\geq$ 311 kg/km2/year. Again, the interpretation of these values, is that for the highest polluting basins, the runoff that they are generated are increasing annually by 311 kg/km2 on average.

It is worth noting that compared to the spatial distribution of average TN pollution loads, the distribution of linear trends is clustered over space. Within the North Island, there is a large pocket of basins whose TN runoff is increasing over time. These basins are located primarily in the Ruapehu and Rangitikei regions of Manawatu-Wanganui. The basins located in Wellington demonstrate either little to no change, or minimal increase in TN runoff, over time. Similarly, the basins in Auckland also demonstrate little to no change in TN runoff over time. The rest of the North Island is actually exhibiting an improvement in the quantity of TN runoff that enters into the freshwater system.

Within the South Island, basins that are worsening over time are primarily located in Canterbury. There are also a few basins located in Southland that are also worsening over time. The basins in Otago either exhibit modest increases in pollution runoff or close-to-zero change in their pollution runoff over time. Lastly, basins that are improving over time are located primarily in Southland, although there are a small number of basins in Canterbury that are also decreasing in TN runoff as well.

2.4.3.2 Ammonia (as nitrogen)

While the spatial distribution in the changes in ammonia as N pollution follow a similar distribution as the TN linear trends, a larger percentage of basins within the country display a decline in the quantity of ammonia as N pollution over time.

The range of ammonia as N linear trends is $\leq$ -10 kg/km2/year to $\geq$ 5 kg/km2/year. The interpretation of this is that for the basins exhibiting the highest average ammonia runoff increase, the amount of ammonia as N pollution is increasing annually at a rate of 5 or more kg/km2.

Within the North Island, the basins with worsening ammonia as N pollution are clustered in Manawatu-Wanganui. It is also worth noting that there are also basins located in Waikato, Hawke's Bay, and Taranaki that demonstrate worsening ammonia pollution over time. Basins with little change in ammonia pollution are located in Wellington and Auckland. Lastly,

basins that are reducing the amount of ammonia pollution are located in Taranaki, Waikato, Gisborne, Hawke's Bay, and parts of Manawatu-Wanganui.

In the South Island, the basins with the highest average increases in ammonia runoff are located in Southland. There are also a handful of basins in Canterbury with worsening ammonia pollution. Basins in Otago continue to exhibit little improvement or modest increase in ammonia pollution over time. There are also basins in Southland, West Coast, and Canterbury, as well as across Marlborough that are demonstrating a reduction in ammonia runoff.

2.4.3.3 Total phosphorus

The spatial pattern in the total phosphorus trends (see Figure 2.28) are similar to the patterns exhibited for changes in nitrogen-based compounds (see Figures 2.27 and 2.26). The range in changes in TP runoff over time is from $\leq$ -8 kg/km2/year to $\geq$ 11 kg/km2/year. The 50th percentile of the linear trend in TP runoff is close to zero kg/km2/year. About 50 percent of the basins have decreasing TP runoff over time, about 30 percent have little change to a modest increase in TP runoff, and the remaining basins demonstrate an increase in TP runoff over time.

In the North Island, the basins with the highest average increases in TP pollution are scattered throughout, with basins located in Waikato, Manawatu-Wanganui, Taranki, Hawke's Bay, and Gisborne. The majority of basins in central Manawatu-Wanganui are increasing the amount of TP runoff over time. In Wellington, the basins either display no change on average or minimal increase in TP. For the remainder of the basins in North Island, most are showing improvement, and are actually decreasing the quantity of TP runoff over time.

In the South Island, the basins with the largest increases in TP are in Southland, with a few located in Canterbury. Across Otago, all basins either demonstrate no improvement or minimal increases in TP runoff over time. Finally, basins have on average reduced their TP runoff over time are located in Southland and Canterbury.

2.4.3.4 Dissolved reactive phosphorus

For the last nutrient compound of interest, dissolved reactive phosphorus (DRP), the pattern of changes in DRP runoff (see Figure 2.29) is similar to the pattern seen for TP runoff. The range of changes in DRP runoff span $\leq$ -3 kg/km2/year to $\geq$ 6 kg/km2/year. Within the North Island, a larger fraction of basins in Manawatu-Wanganui have increased their DRP pollution compared to their TP pollution over time. There are also basins with increases in DRP runoff located in Waikato, Taranaki, Hawke's Bay, Gisborne, and Northland. Wellington continues to have either no change or minimal increase in DRP pollution.

In the South Island, the basins with the highest average increases in DRP runoff are located in Southland. There are also basins in Marlborough that have large increases in DRP runoff over time. Otago basins have close-to-zero or modest increases in DRP pollution over time. Lastly, basins that show improvement in reducing DRP runoff are located primarily in Southland, and scattered in parts of Marlborouth and Canterbury.

2.4.4 Linear trends in USMRB

2.4.4.1 Total nitrogen

In the USMRB, total nitrogen runoff linear trends range from $\leq$ 267,972 kg/km2/year to $\geq$ 240,412 kg/km2/year (see Figure 2.31). The 50th decile is -855 kg/km2/year to -72 kg/km2/year. Basins whose TN pollution runoff are decreasing on average are located closer along the major rivers, and especially clustered in the lower half of the USMRB. In particular, these basins are located in northern Arkansas, southeastern Missouri, and southern Illinois.

Regions where TN runoff is increasing over time are located throughout the USMRB. Some of the basins with increasing pollution are located along the norther portion of the Mississippi River, along the border to Iowa and Illinois. There are also basins located along the western portion of the Arkansas River that suffer from increasing TN pollution. Lastly, there are also basins in Montana, Indiana, Pennsylvania, and West Virginia with increasing TN runoff.

The remaining twenty percent of basins have little to no change on average in the amount of TN runoff that they generate. These basins tend to be located further away from the major rivers, but are spread out across the USMRB.

2.4.4.2 Total Kjedahl nitrogen

In the USMRB, total Kjedahl nitrogen (TKN) runoff linear trends range from $\leq$ -60,932 kg/km2/year to $\geq$ 52,704 kg/km2/year. The 50th decile is -316 to 0 kg/km2/year. Area normalized TKN pollution is decreasing in about 50 percent of the basins. The basins with the largest decrease in runoff tend to be located along the major rivers, especially near the juncture of the Mississippi RIver, Ohio River, and Arkansas River.

Basins with a modest decrease in TKN pollution over time are located throughout the USMRB, and tend to be further away from the major rivers.

Lastly, about 50 percent of basins actually experience an increase in TKN runoff. Basins with the highest average increase in TKN runoff over time are located throughout the country, and do not appear to be clustered in space. Regions that have basins with the highest increase include southern Wisconsin, parts of Iowa, parts of Illinois, parts of Indiana, parts of Ohio, parts of Missouri, and parts of Arkansas.

Lastly, western USMRB contain basins that exhibit increases in TKN pollution over time, although the level of pollution increase is largely within the 40th to 70th percentile.

2.4.4.3 Total phosphorus

For the last compound of interest, total phosphorus (TP), the average changes over time in TP runoff range from $\leq$ -18,024 kg/km2/year to $\geq$ 24,712 kg/km2/year (see Figure 2.32). The 50th percentile is -56 to 0 kg/km2/year.

About 50 percent of basins exhibit a decrease in TP runoff over time. While these basins tend to be clustered along the major rivers, unlike changes in TN and TKN, these basins are not all located near the juncture of the Mississippi, Arkansas, and Ohio rivers. Basins with the largest decrease in area normalized TP pollution loads over time are along the western

portion of the Arkansas River, the middle portion of the Mississippi River, and intermittently along the Ohio River.

The other half of basins exhibit an increase in TP runoff over time. Basins with the largest average increase are scattered throughout the USMRB. Regions with basins with increasing TP runoff are located in eastern Iowa, northwestern Illinois, central Illinois, central Missouri, and Indiana.

2.5 Discussion

Our model for localizing nonpoint source water pollution can be applied to watersheds around the world, ranging from continent-scale catchments to smaller regional catchments. Our results are useful for evaluating historical water pollution and establishing baseline pollutant loads, and measuring changes in agricultural runoff over time. Additionally, our approach is useful for forward-looking programs; by illuminating which basins are problematic polluters, we have highlighted areas that could benefit from locating additional water monitoring stations. Given that collecting water monitoring samples is a resource intensive enterprise, this allows water monitoring agencies to strategically site their water monitoring stations in more vulnerable areas.

2.5.0.1 Application to climate change

Mitigating nonpoint source water pollution becomes more pressing under climate change. In the USMRB, the Midwest has already experienced increased extreme precipitation and flooding events during spring months, around the start of the growing season. This has led to increased runoff as rain events carry recently applied fertilizer before they are able to percolate into the soil. This trend of increased extreme precipitation events is likely to continue. Additionally, as temperatures warm, agricultural production within the US will move northwards. Unless these new agricultural regions are able to reduce fertilizer runoff, this will lead to a spatial expansion of NPS pollution in freshwater systems, as pollution will begin in more upstream waterways and travel downstream.

Furthermore, climate change has also caused warming ocean temperatures and ocean acidification. As such, coastal systems are likely to be less resilient to the effects of eutrophication, as the existing linchpins of marine ecosystems, such as coral reefs or kelp forests, die off.

2.5.0.2 Implications of nonpoint source pollution worldwide

The environmental issue of eutrophication is not only limited to the water systems in the United States or New Zealand; this is a growing problem worldwide. However, in many regions there is lack of spatially and temporally resolved monitoring data on both historical and contemporaneous water quality trends. It is important that we increase the availability of high quality on-the-ground monitoring at key locations within the river network, and that we apply increasingly sophisticated machine learning and artificial intelligence techniques to satellite data to construct new water quality datasets. We have seen an explosion in

the use of satellite products for addressing air quality related issues; there needs to be a commensurate increase for water quality issues as well. Lastly, expanding the availability of water quality data is of great importance for low and middle-income countries as they increasingly adopt intensive agricultural farming techniques with more reliance on chemical fertilizer inputs or implement concentrated animal feedlot operations, which will result in increased non-point source pollution.

The granularity of our dataset can also shed light on the distributional impacts of water pollution. Our model can fill out our understanding of who the polluters are versus who the recipients of pollution are. This allows us to apply market based mechanisms under a Coasean framework so that recipients of pollution receive payments from the polluters for the negative social costs. Additionally, if governments were to tax the polluters for their negative environmental externalities, they could transfer this tax payment to the recipients as well. We are able to establish baseline level of areas that generate high quantities of nutrient pollution, but measuring low levels of pollution is important information too, since there is not enough research looking at health impacts of low exposure to water pollution. Lastly, the dataset output from our model is useful for water quality trading permit programs, in that high and low polluters can receive and trade permits based off of their water polluting activities.

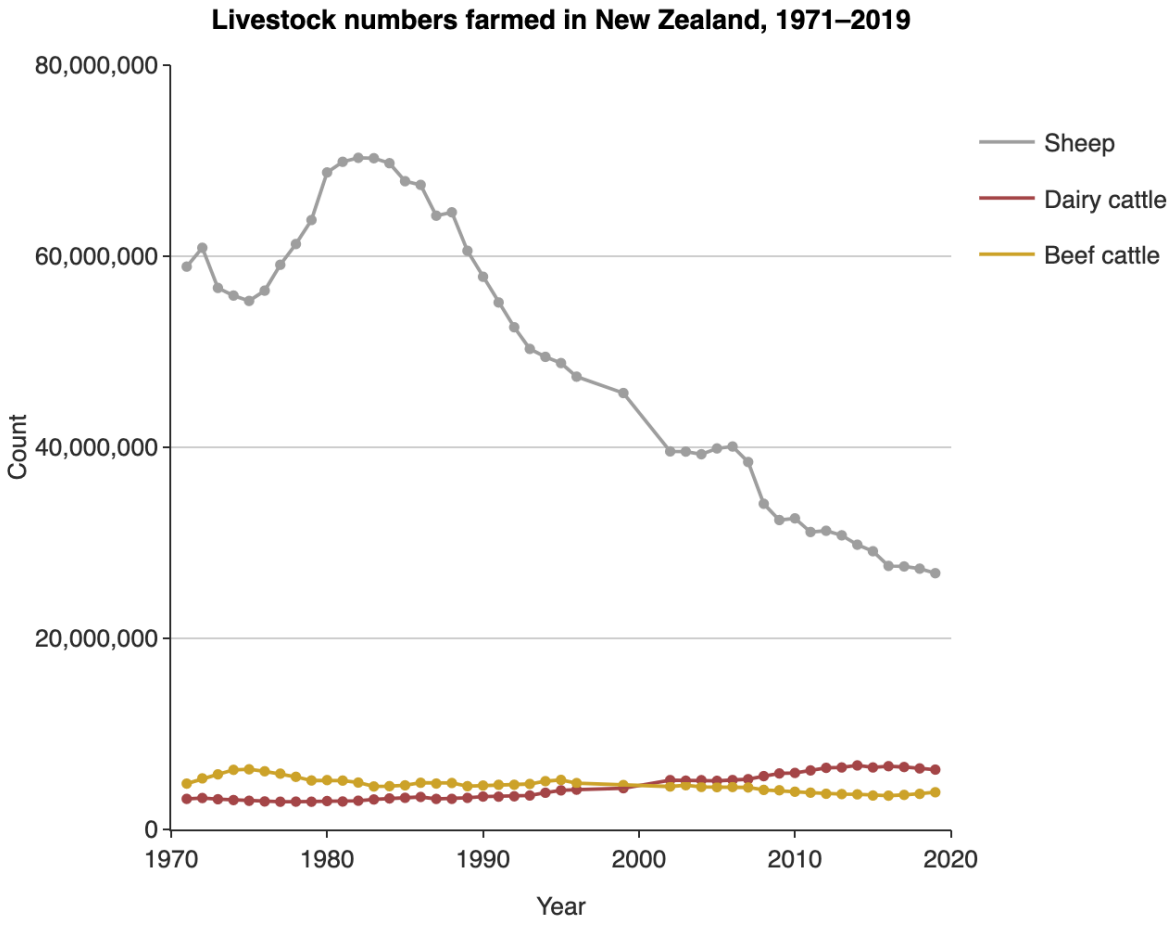


Figure 2.1: Timeseries of sheep, dairy cattle, and beef cattle count in New Zealand between 1971–2019. (figure and data source: Stats NZ)

Figures

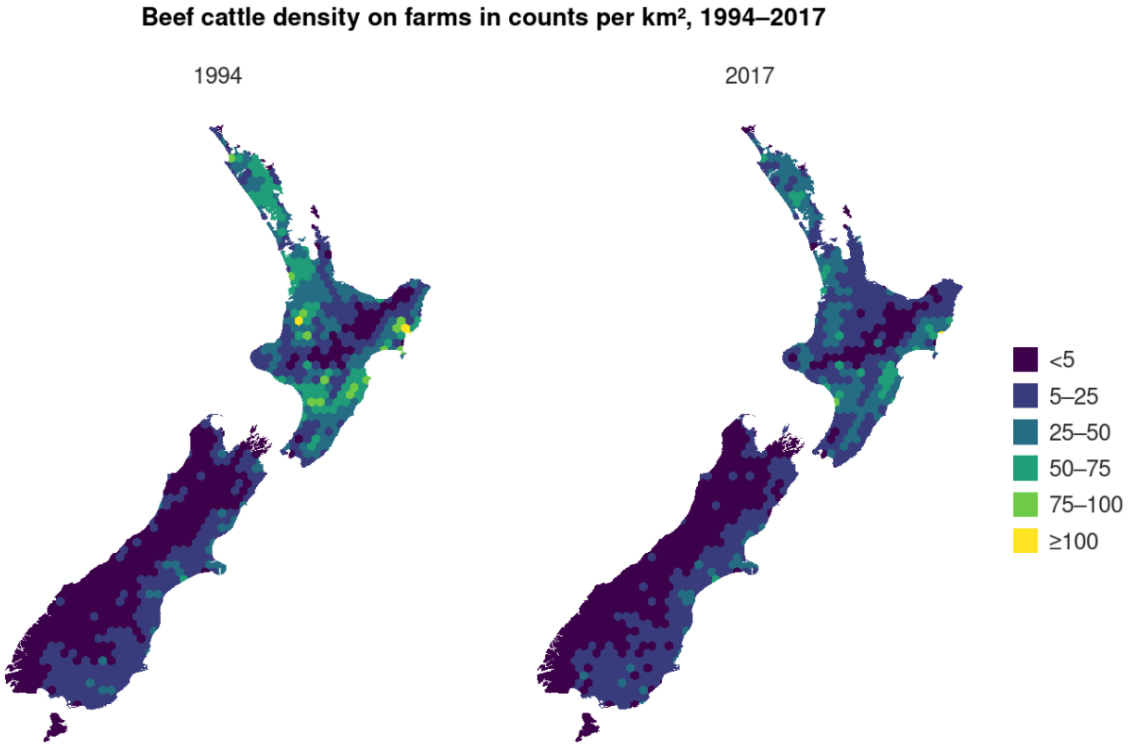


Figure 2.2: Spatial distribution of beef cattle density across New Zealand, comparing between 1994 and 2017. (figure and data source: Stats NZ)

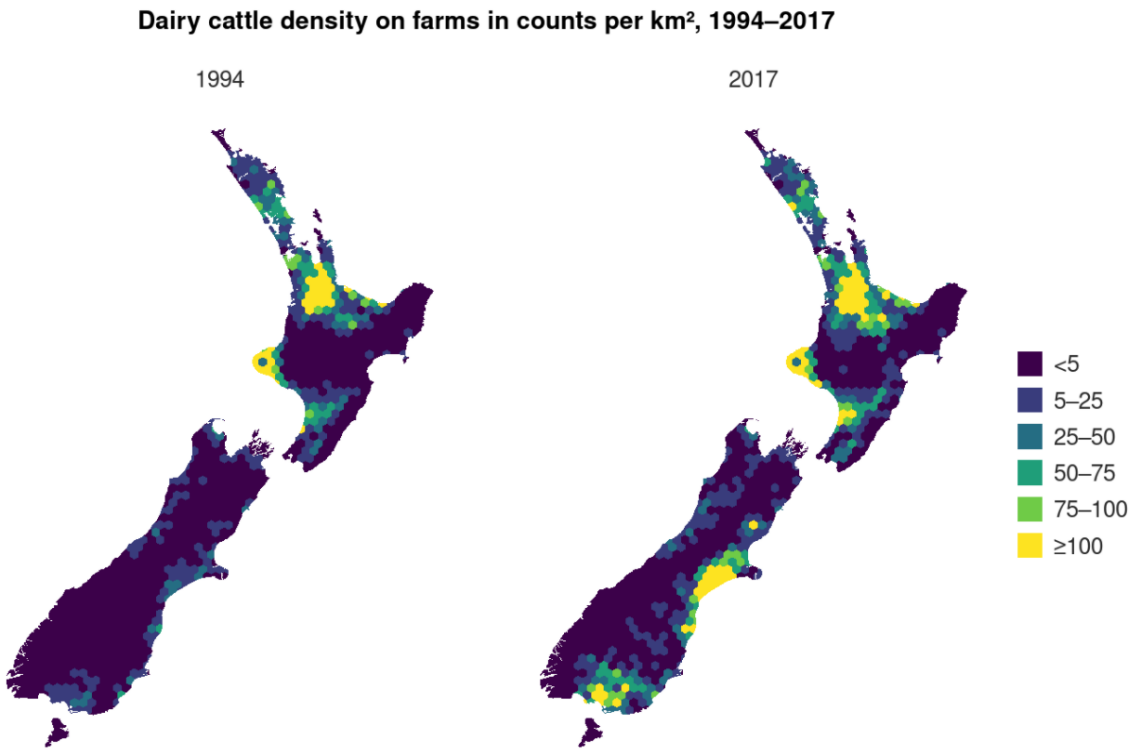


Figure 2.3: Spatial distribution of dairy cattle density across New Zealand, comparing between 1994 and 2017. (figure and data source: Stats NZ)

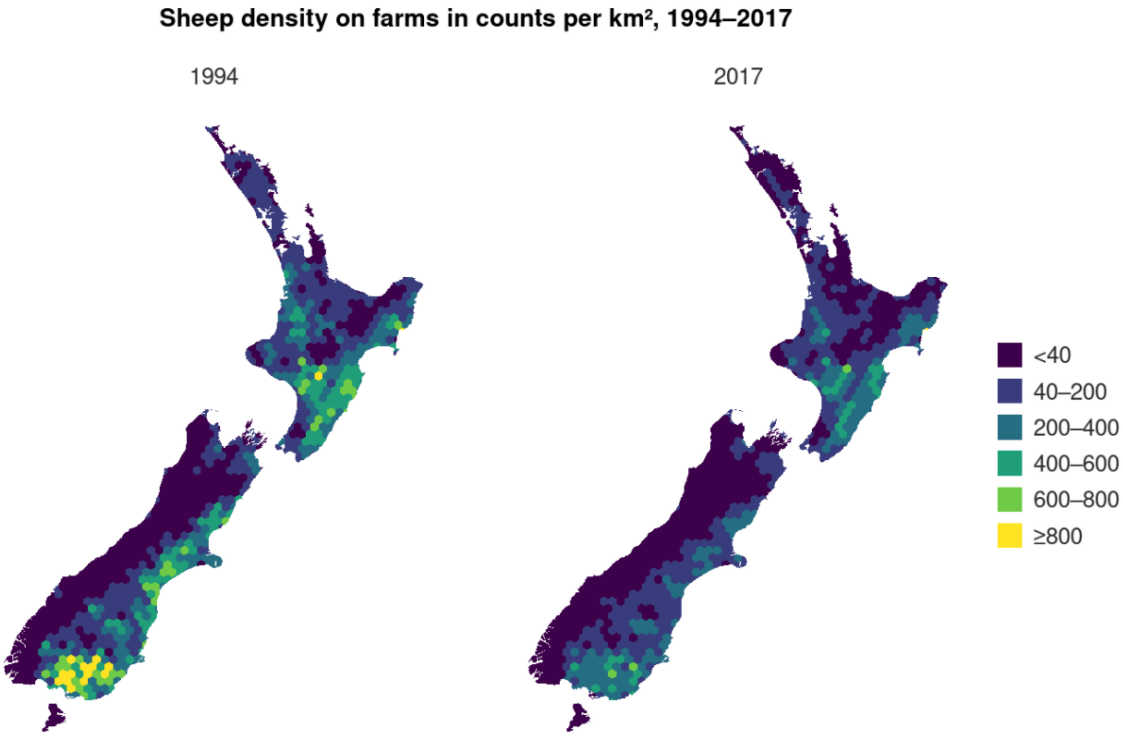


Figure 2.4: Spatial distribution of sheep density across New Zealand, comparing between 1994 and 2017. (figure and data source: Stats NZ)

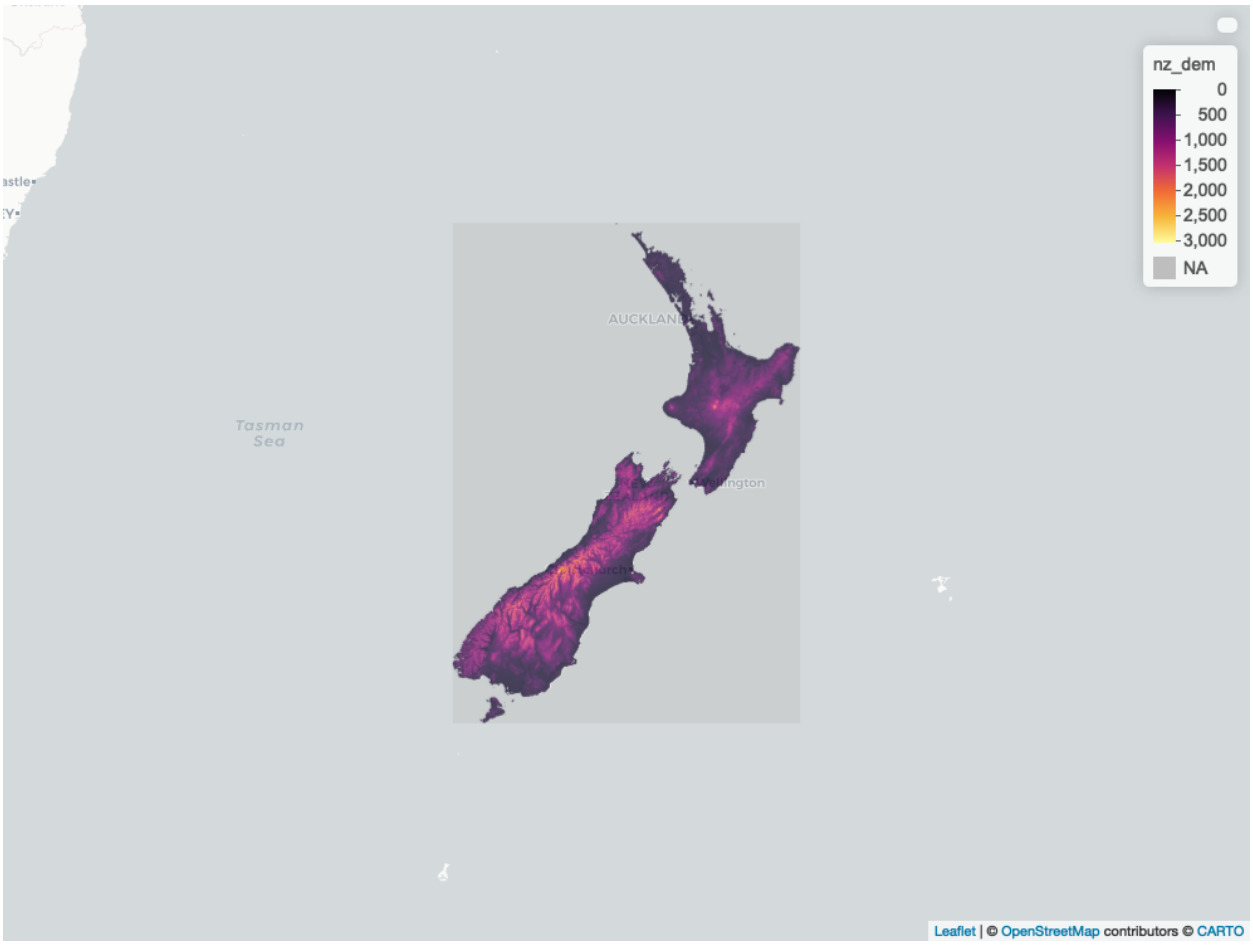


Figure 2.5: New Zealand digital elevation model (DEM) after hydrographic processing. The DEM resolution is 25 meters. Hydrographic processing of the DEM included filling sinks, carving rivers using river polylines, connecting river segments to ensure that all river pixels flowed into the ocean.

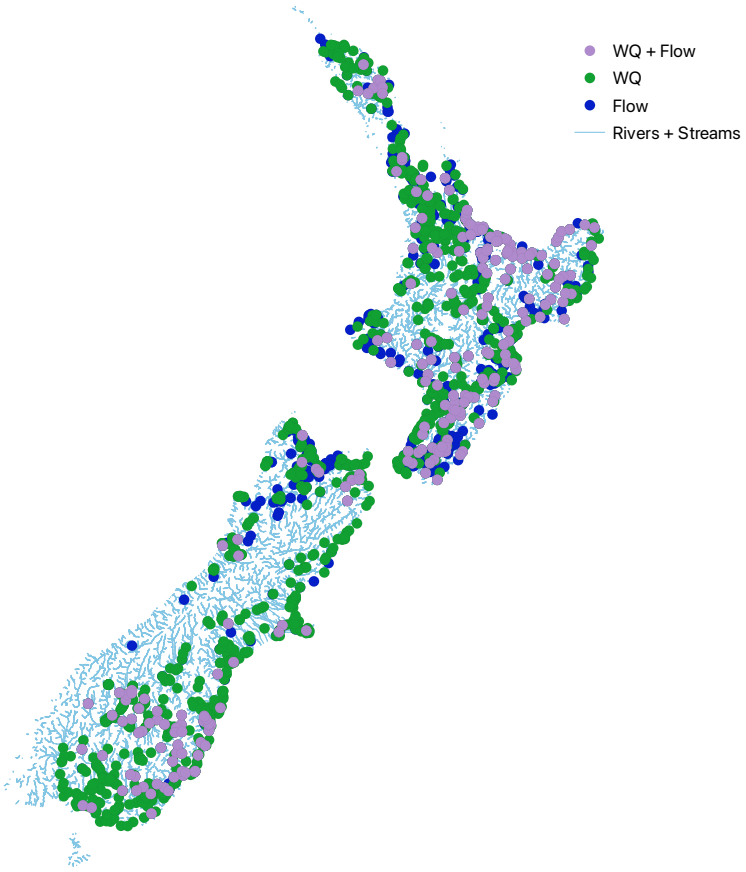


Figure 2.6: Water quality and flow monitoring stations in the United States Mississippi River Basin used in our modeling and analysis. These sites collected data at some point between January 1, 1981 - December 31, 2018. Water quality data, flow data, and station coordinates are obtained through environmental and water agencies from each regional council, as well as through LAWA.

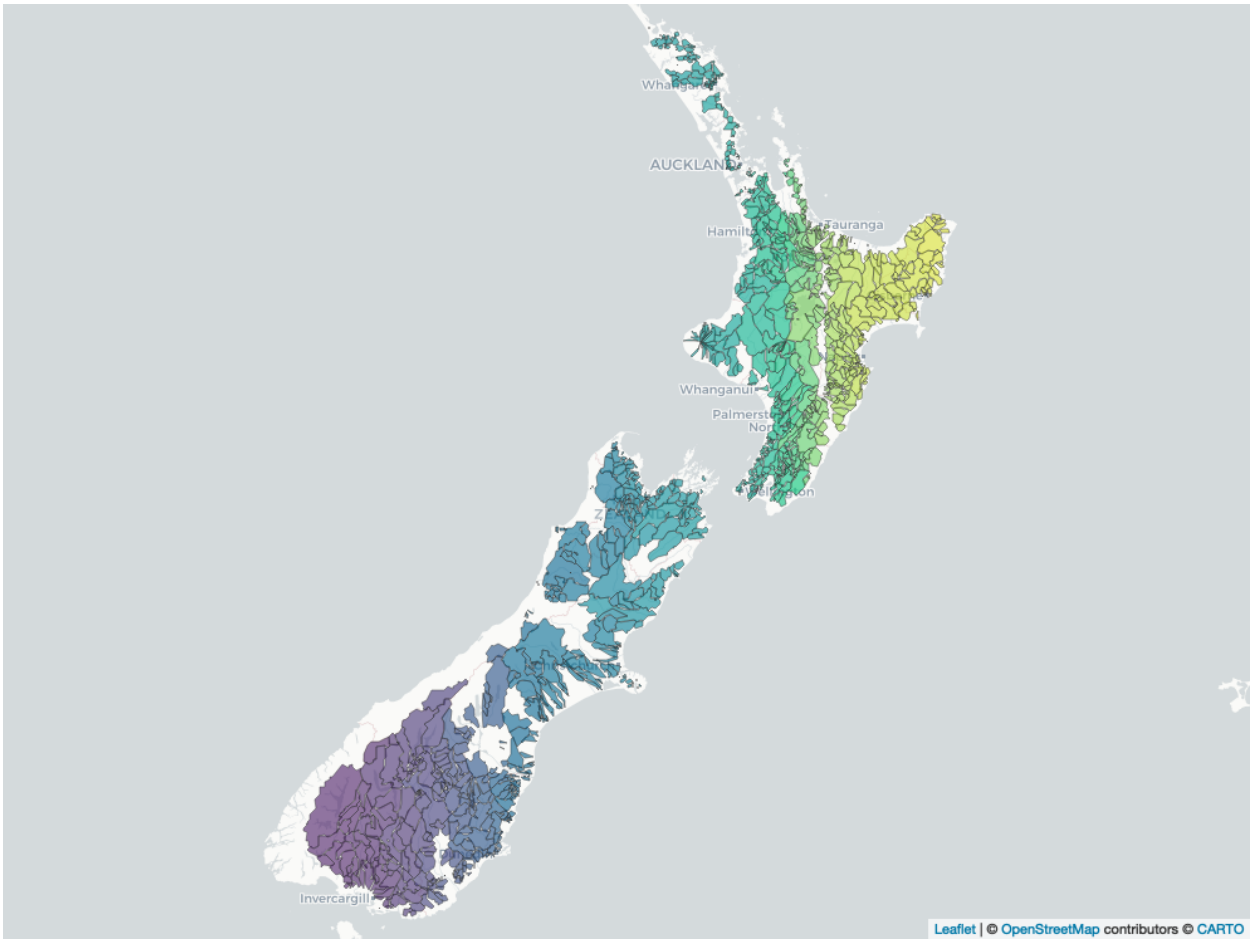


Figure 2.7: Catchments in New Zealand used in our modeling and analysis. Each of these catchments are associated with an existing water quality and/or river flow monitoring site.

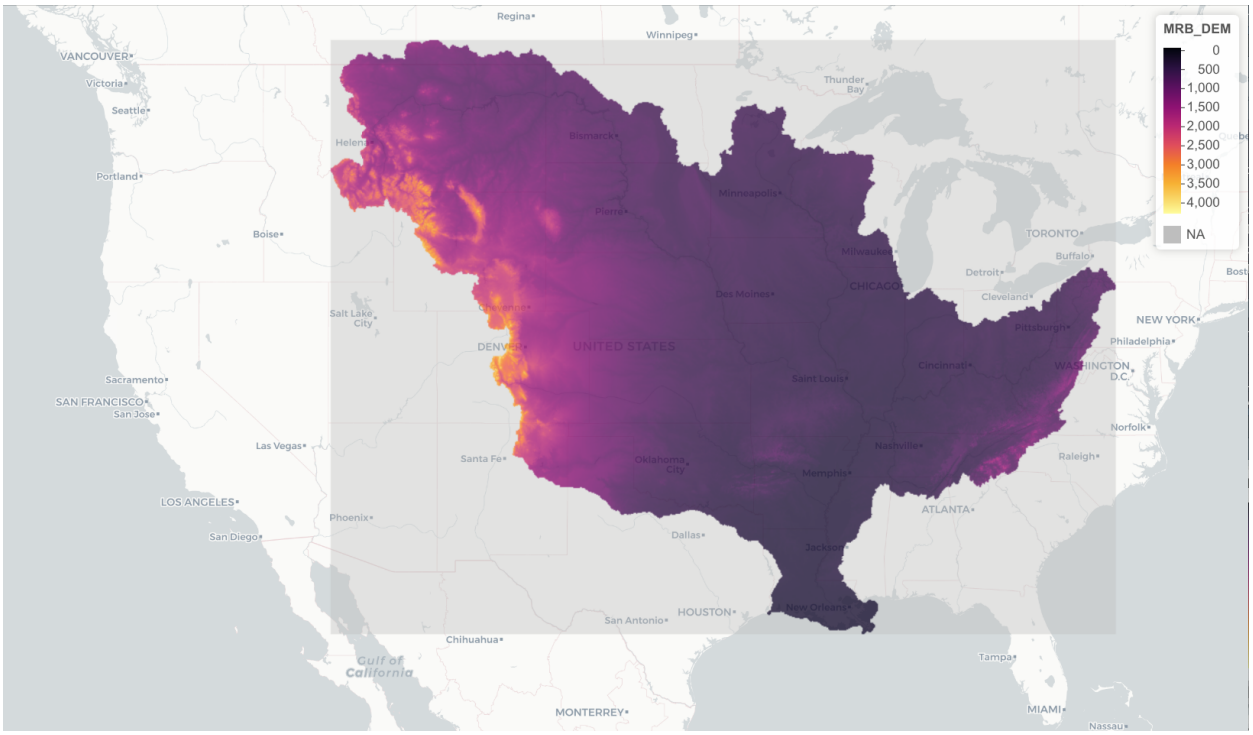


Figure 2.8: United States Mississippi River Basin digital elevation model (DEM) after hydrographic processing. The elevation model was created from data captured by the Shuttle Radar Topographical Mission (SRTM). The DEM resolution is 305 meters. Hydrographic processing of the DEM included filling sinks, carving rivers using river polylines, connecting river segments to ensure that all river pixels flowed into the ocean, and constructing a buffer around the boundaries of the USRMB to ensure that all pixels drained into the Mississippi River.

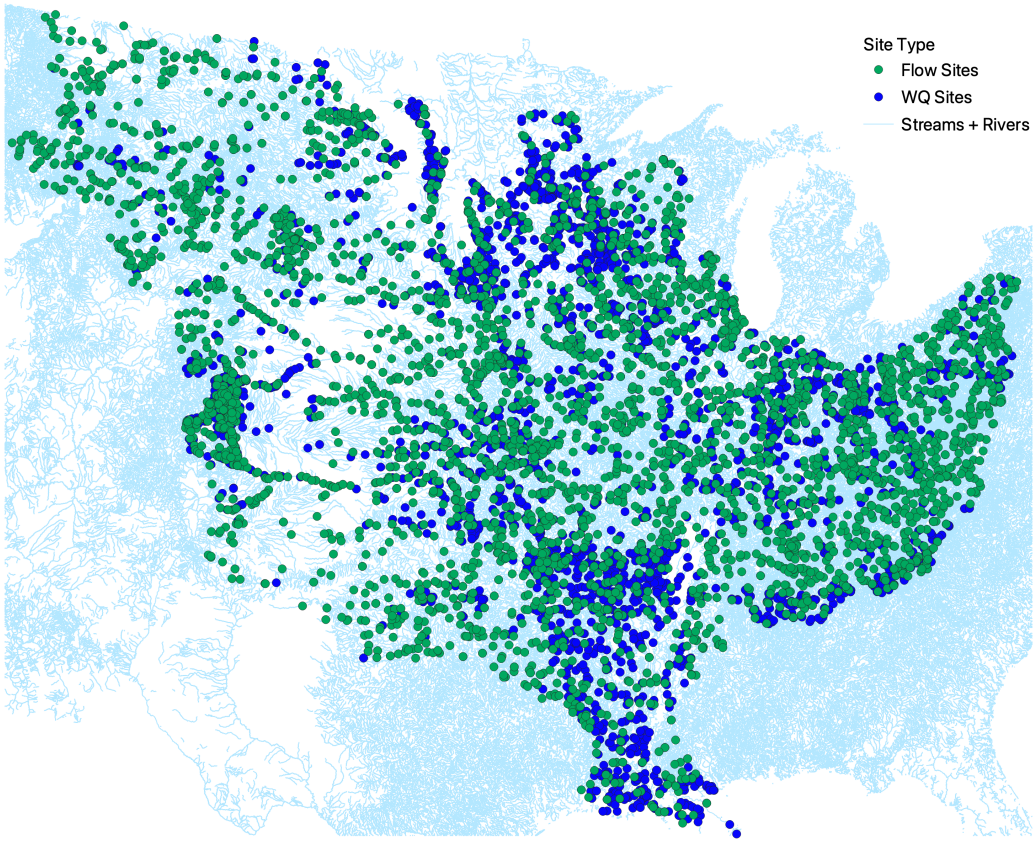


Figure 2.9: Water quality and flow monitoring stations in the United States Mississippi River Basin used in our modeling and analysis. These sites collected data at some point between January 1, 1981 - December 31, 2018. Water quality data and station coordinates are queried through the Water Quality Portal. Flow data are queried through the USGS Daily or Instantaneous Value Service.

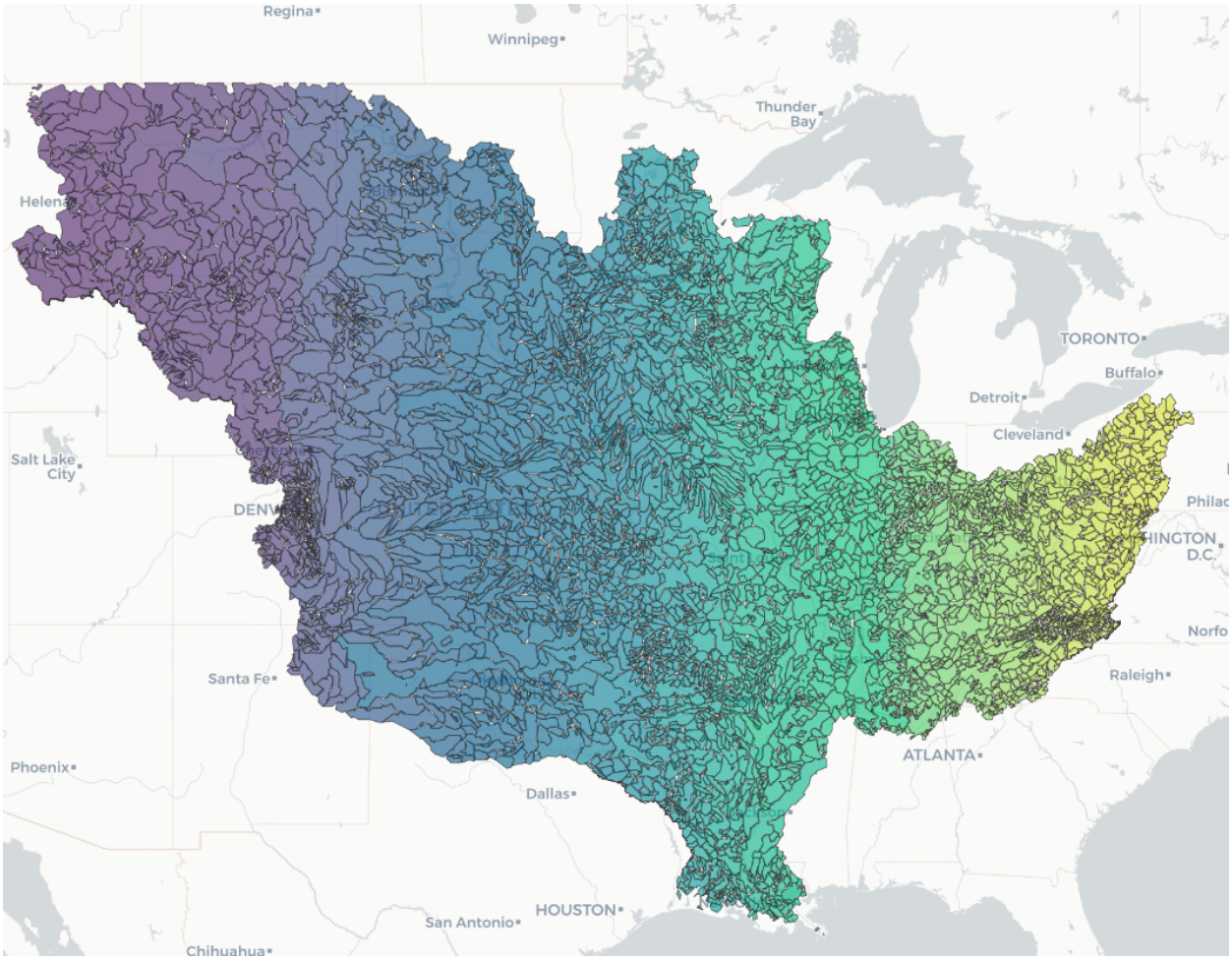


Figure 2.10: Catchments in the United States Mississippi River Basin used in our modeling and analysis. Each of these catchments are associated with an existing water quality and/or river flow monitoring site.

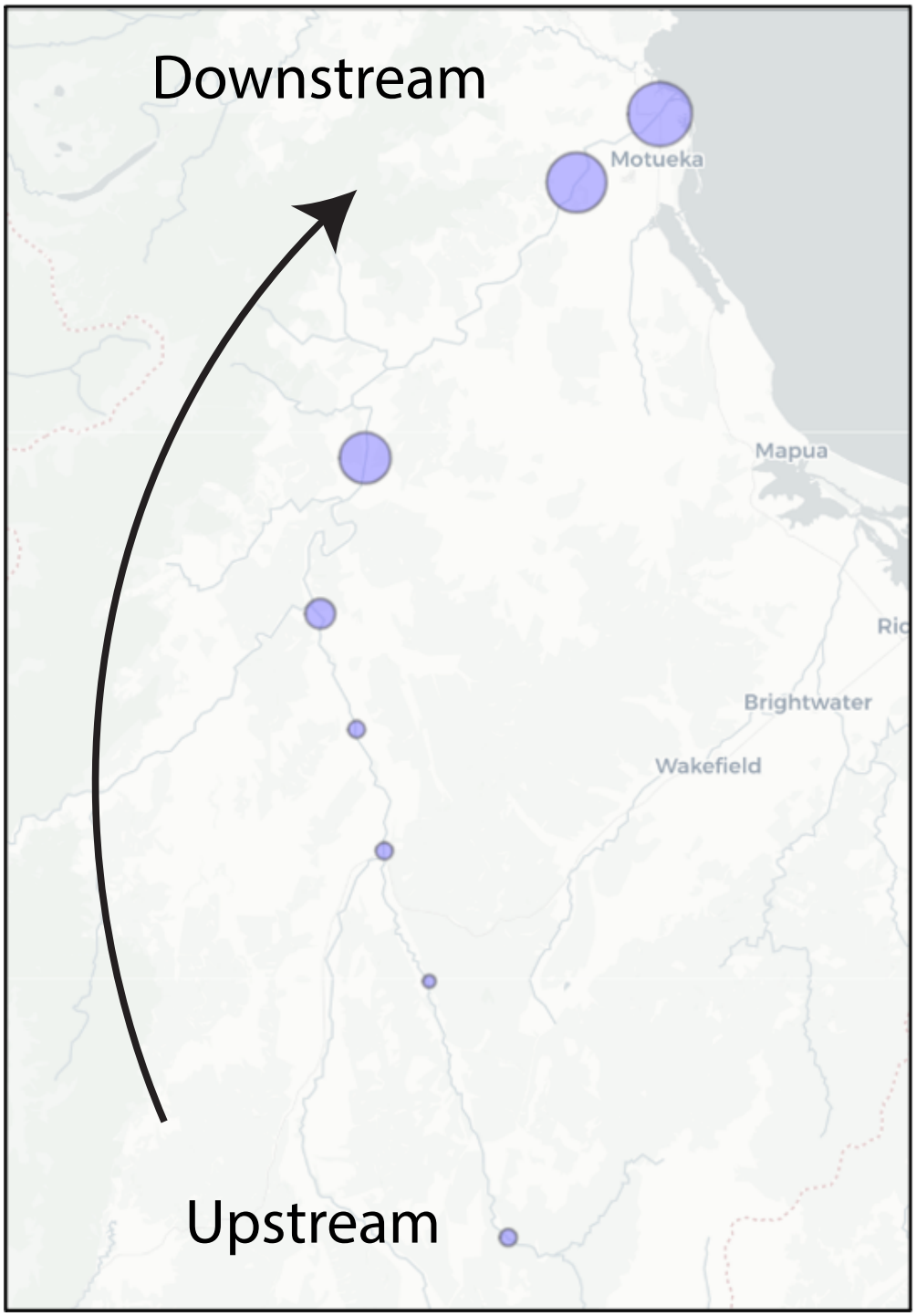


Figure 2.11: Example of a set of streamflow monitoring sites in New Zealand that have been georeferenced to a river segment. Using the digital elevation model, we are able to recover the up to downstream ordering of the monitoring sites.

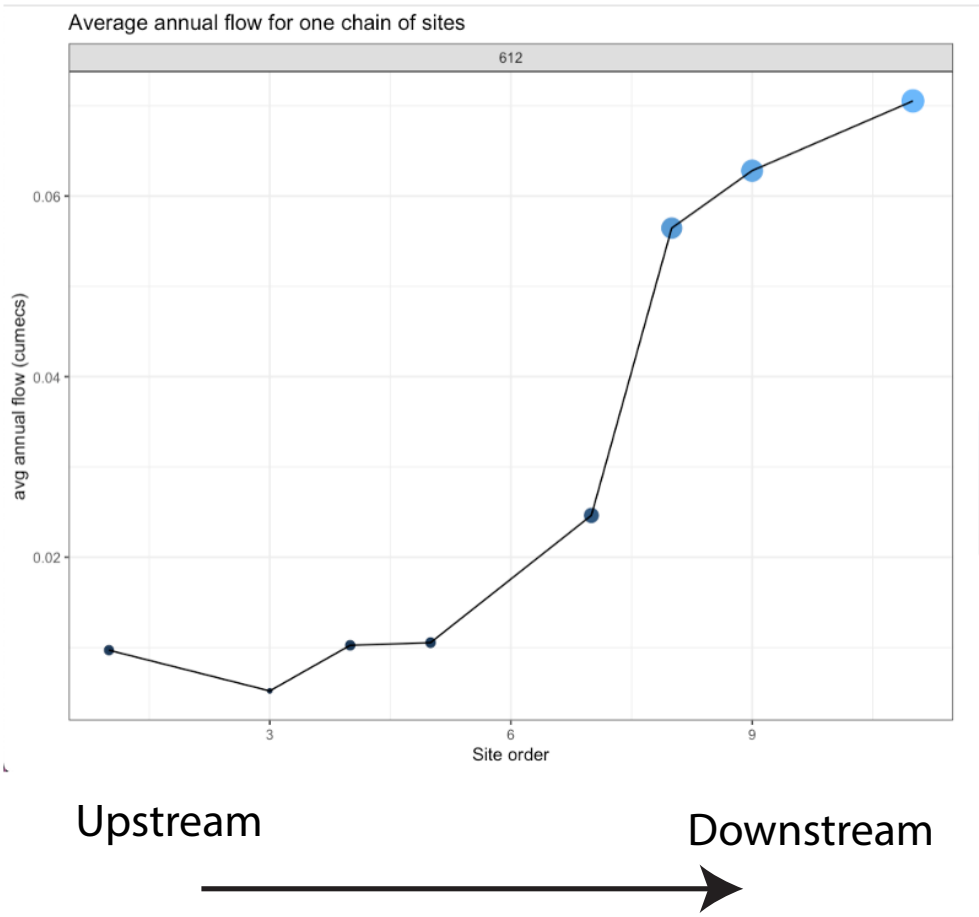


Figure 2.12: The average annual stream flow (in cubic meters per second) at the sites from Figure 2.11, ordered from upstream to downstream most site. This example demonstrates that we were able to georeference the water monitoring sites and recover their correct ordering. We were able to verify this through evaluating the average annual flowrate at each site, and observing that flowate on the whole monotonically increases from up to downstream sites along the same river.

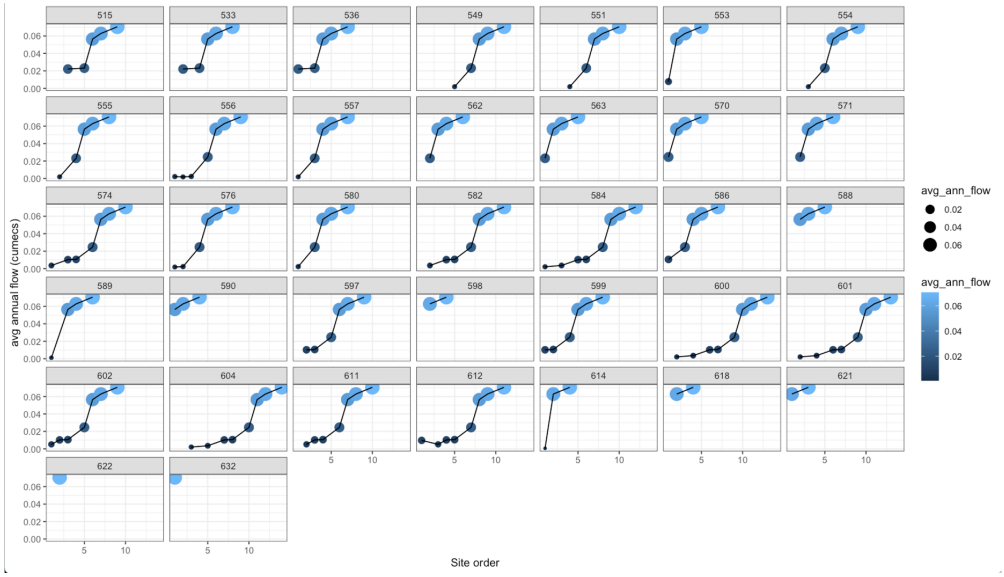


Figure 2.13: Example of monotonically increasing average annual flow rate across multiple chains of flowrate monitoring sites located along different river segments.

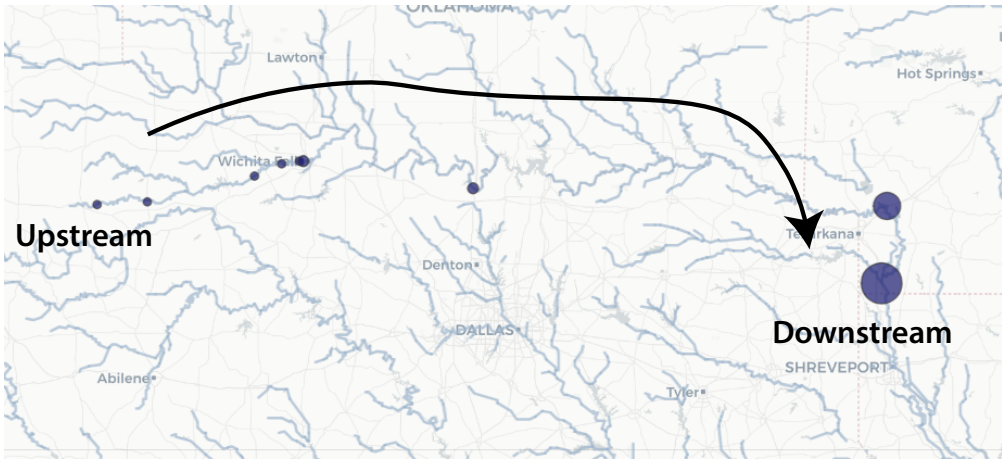


Figure 2.14: Example of a set of streamflow monitoring sites in the US Mississippi River Basin that have been georeferenced to a river segment. Using the digital elevation model, we are able to recover the up to downstream ordering of the monitoring sites.

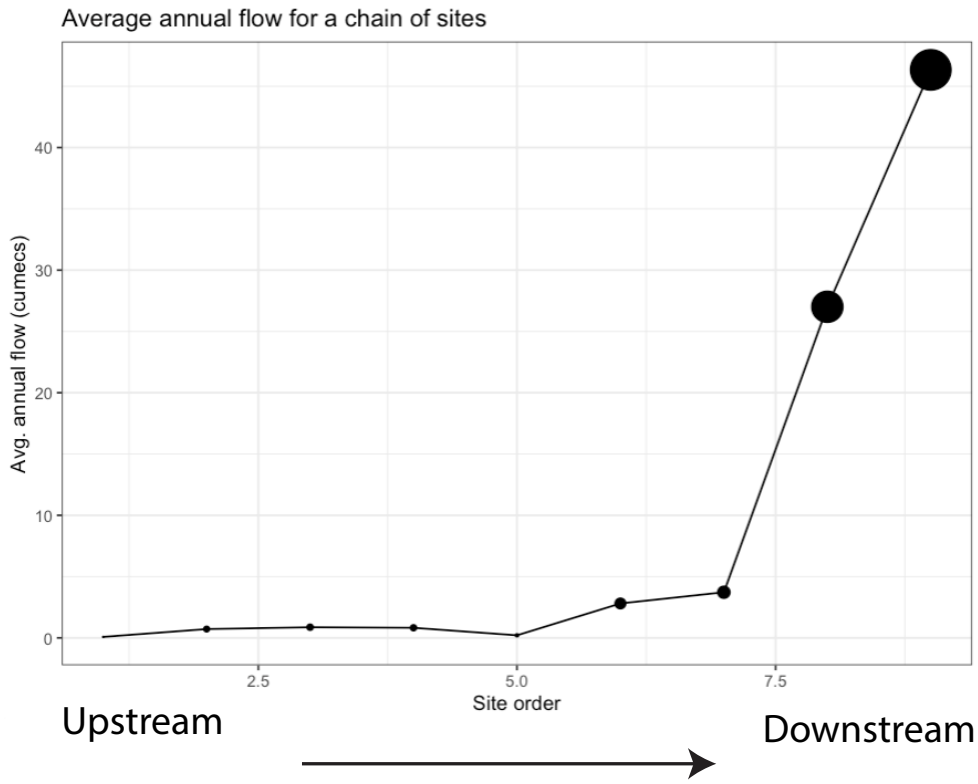


Figure 2.15: The average annual stream flow (in cubic meters per second) at the sites from Figure 2.14, ordered from upstream to downstream most site. This example demonstrates that we were able to georeference the water monitoring sites and recover their correct ordering. We were able to verify this through evaluating the average annual flowrate at each site, and observing that flowate on the whole monotonically increases from up to downstream sites along the same river.

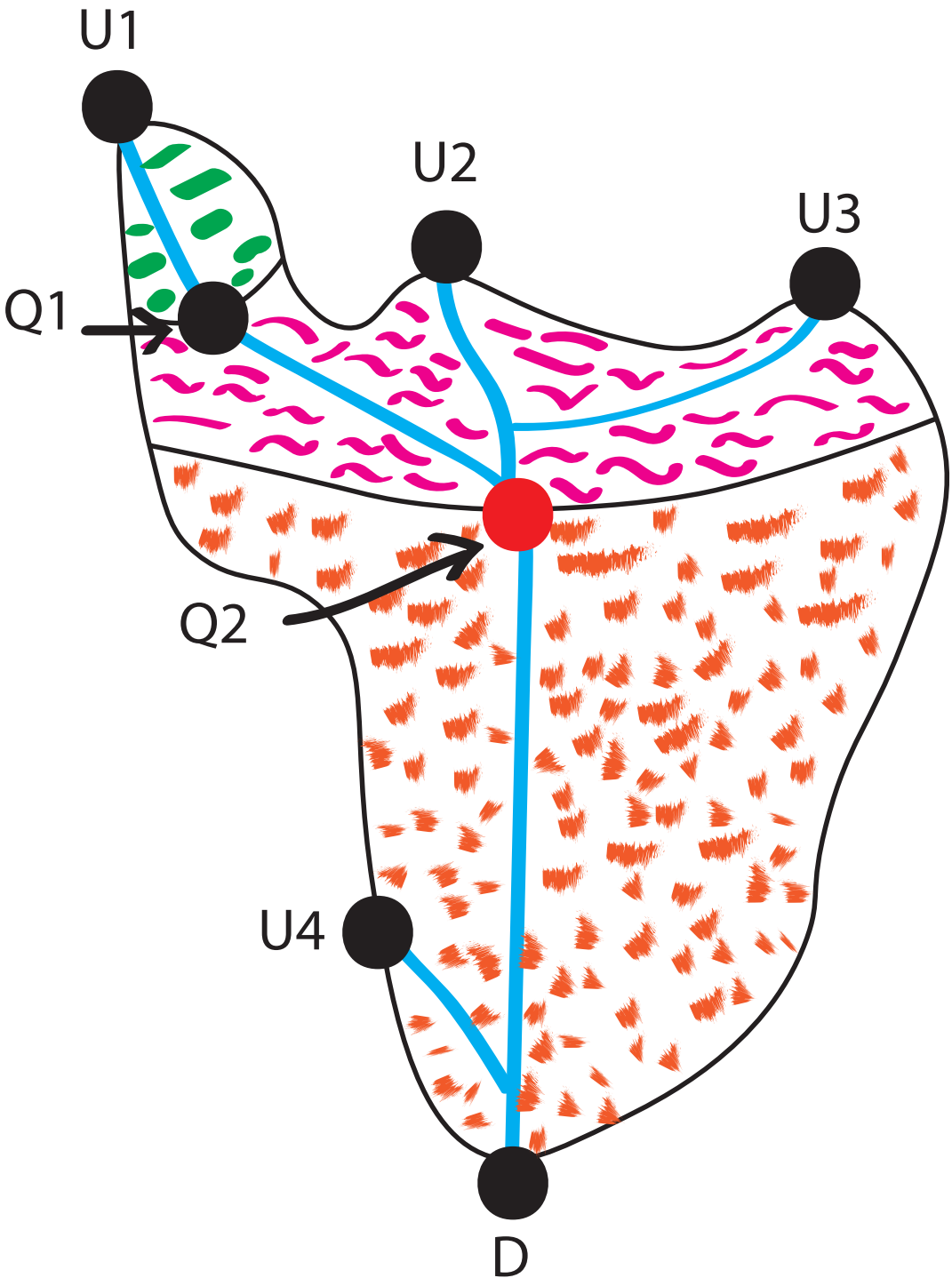


Figure 2.16: Example of spatial flow interpolation.

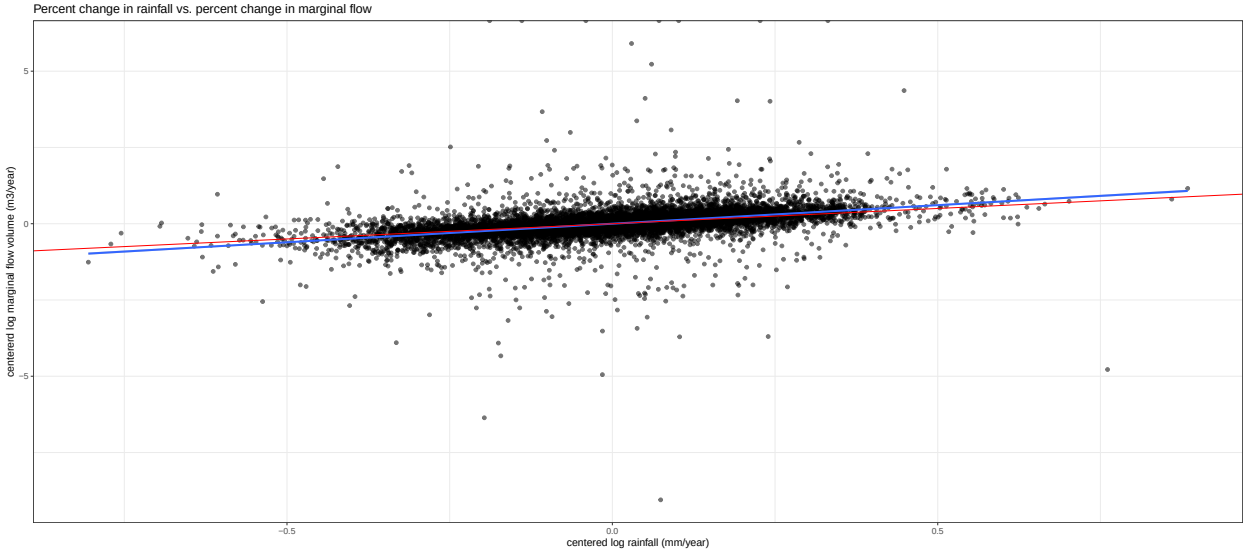


Figure 2.17: Scatter of annualized log rainfall (mm) versus log marginal flow by catchment in New Zealand (values have been centered).

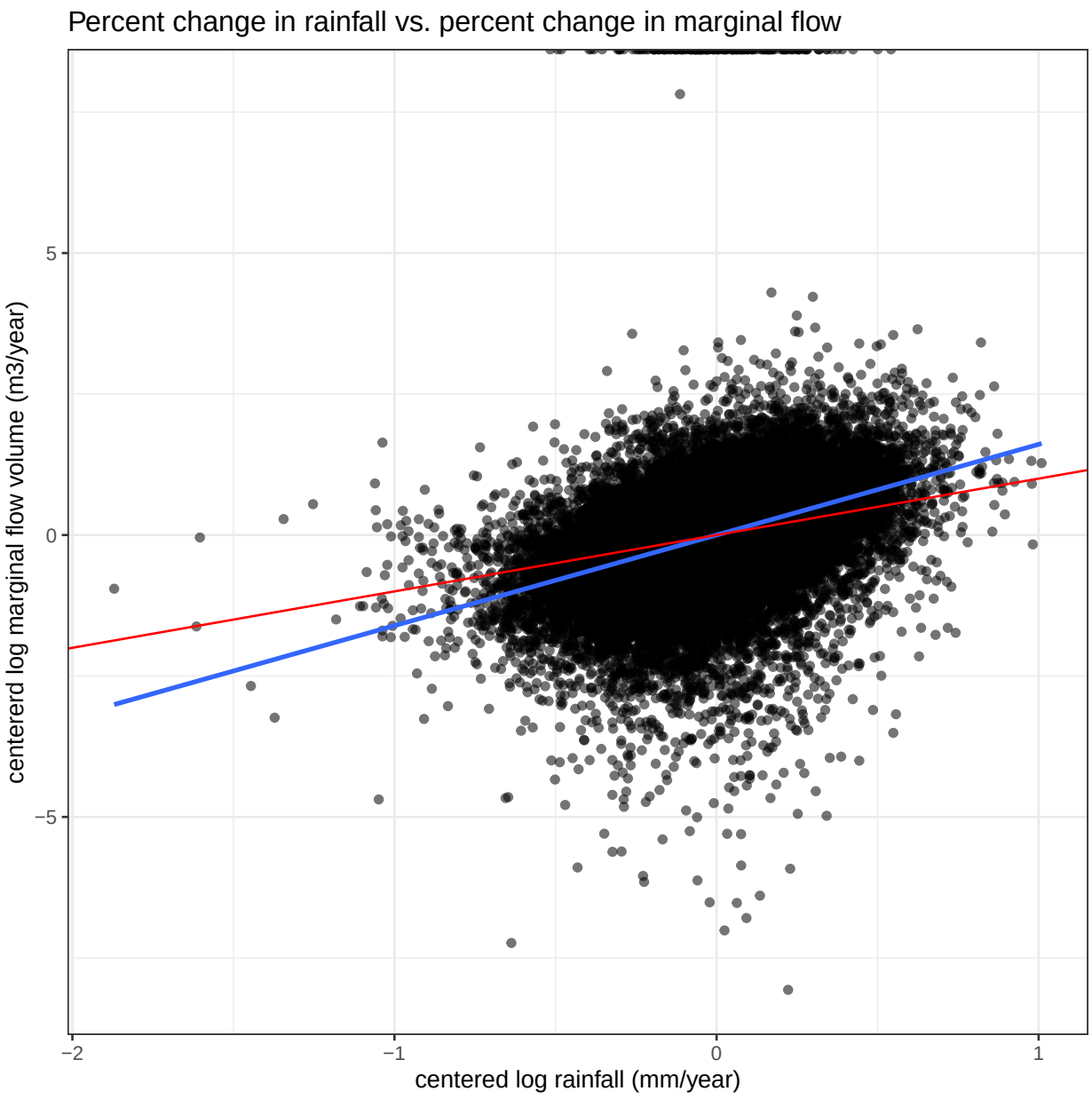


Figure 2.18: Scatter of annualized log rainfall (mm) versus log marginal flow by catchment in USMRB (values have been centered).

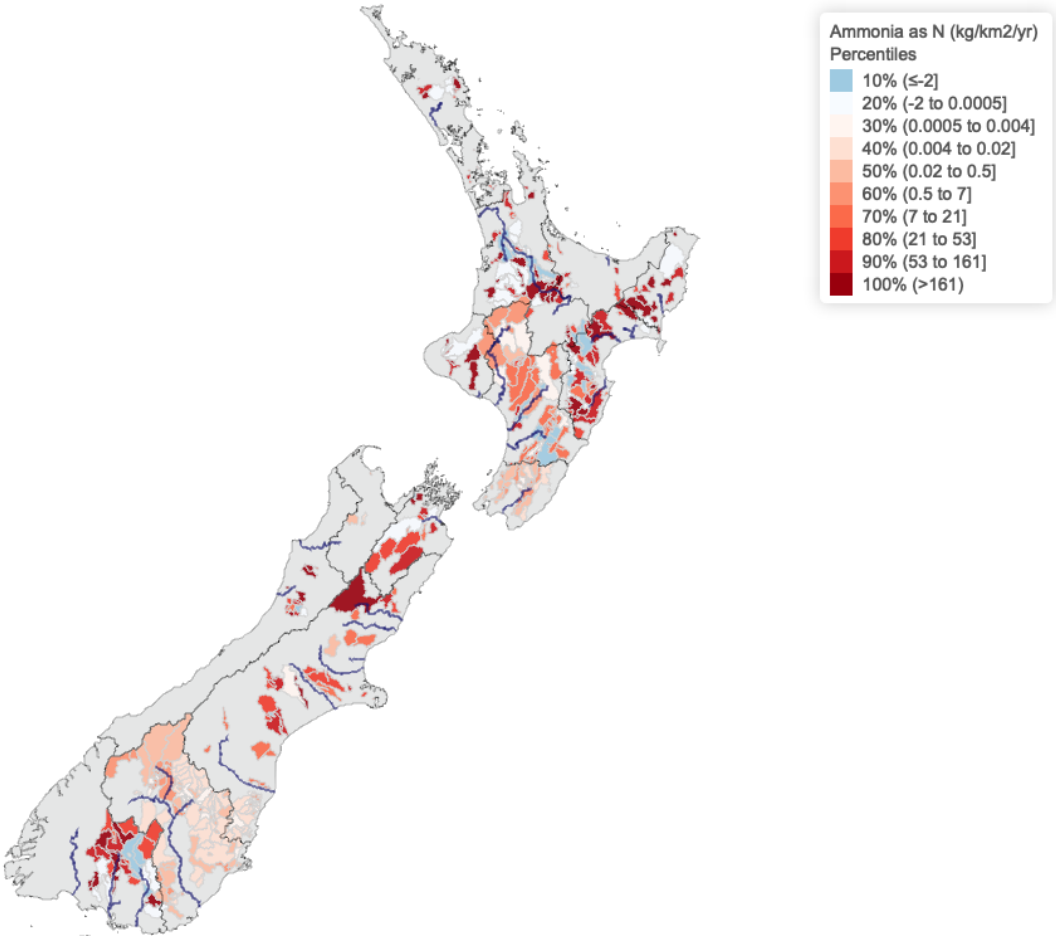


Figure 2.19: Average annual marginal ammonia (as nitrogen) pollutant loads by basin in New Zealand. Marginal loads are computed as the difference in measured annual pollutant loads at adjacent up and downstream water quality monitoring sites that measure this pollutant. The unit of analysis is a basin. A basin is colored blue if the amount of nonpoint source pollution attributed to it is on average negative. A basin is colored red if the amount of pollution attributed to it is on average positive. A basin is colored white if the amount of pollution attributed to it is close to zero. Regional council boundaries are demarcated by a black border. Major rivers in the country are marked in dark blue.

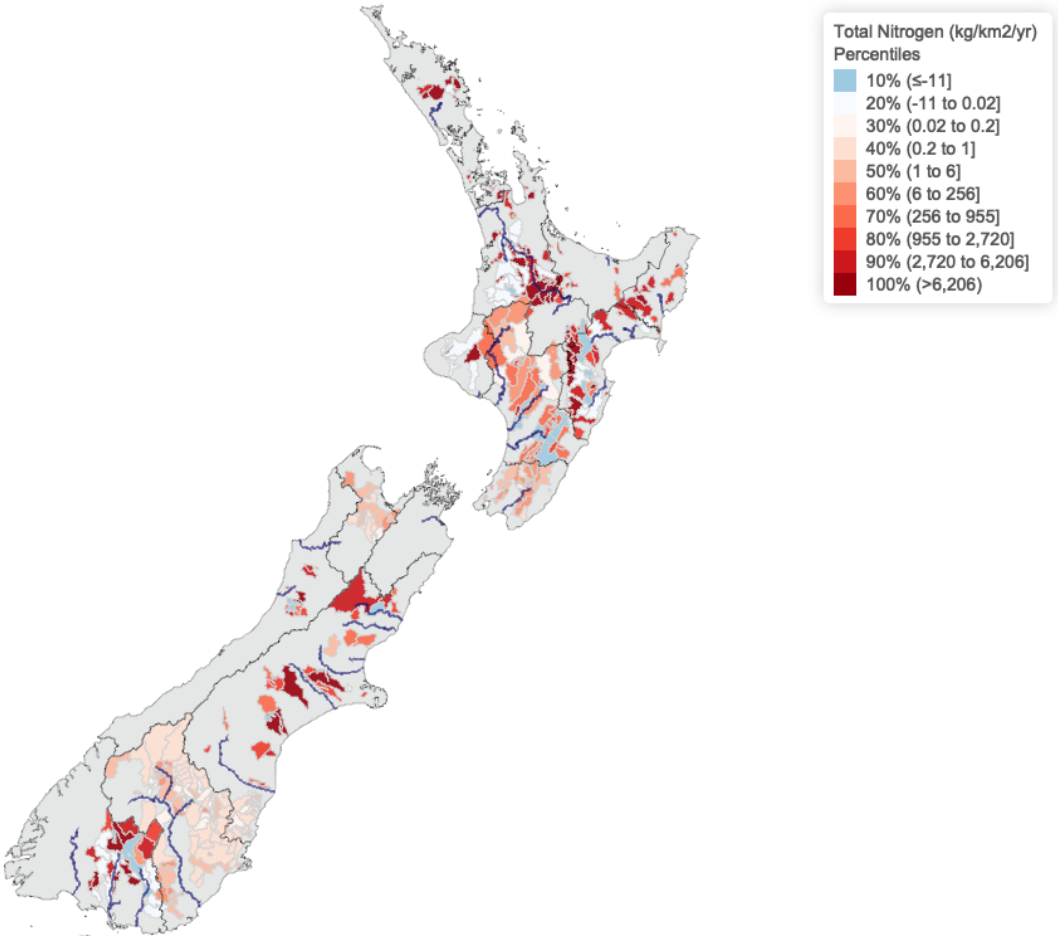


Figure 2.20: Average annual marginal total nitrogen pollutant loads by basin in New Zealand. Marginal loads are computed as the difference in measured annual pollutant loads at adjacent up and downstream water quality monitoring sites that measure this pollutant. The unit of analysis is a basin. A basin is colored blue if the amount of nonpoint source pollution attributed to it is on average negative. A basin is colored red if the amount of pollution attributed to it is on average positive. A basin is colored white if the amount of pollution attributed to it is close to zero. Regional council boundaries are demarcated by a black border. Major rivers in the country are marked in dark blue.

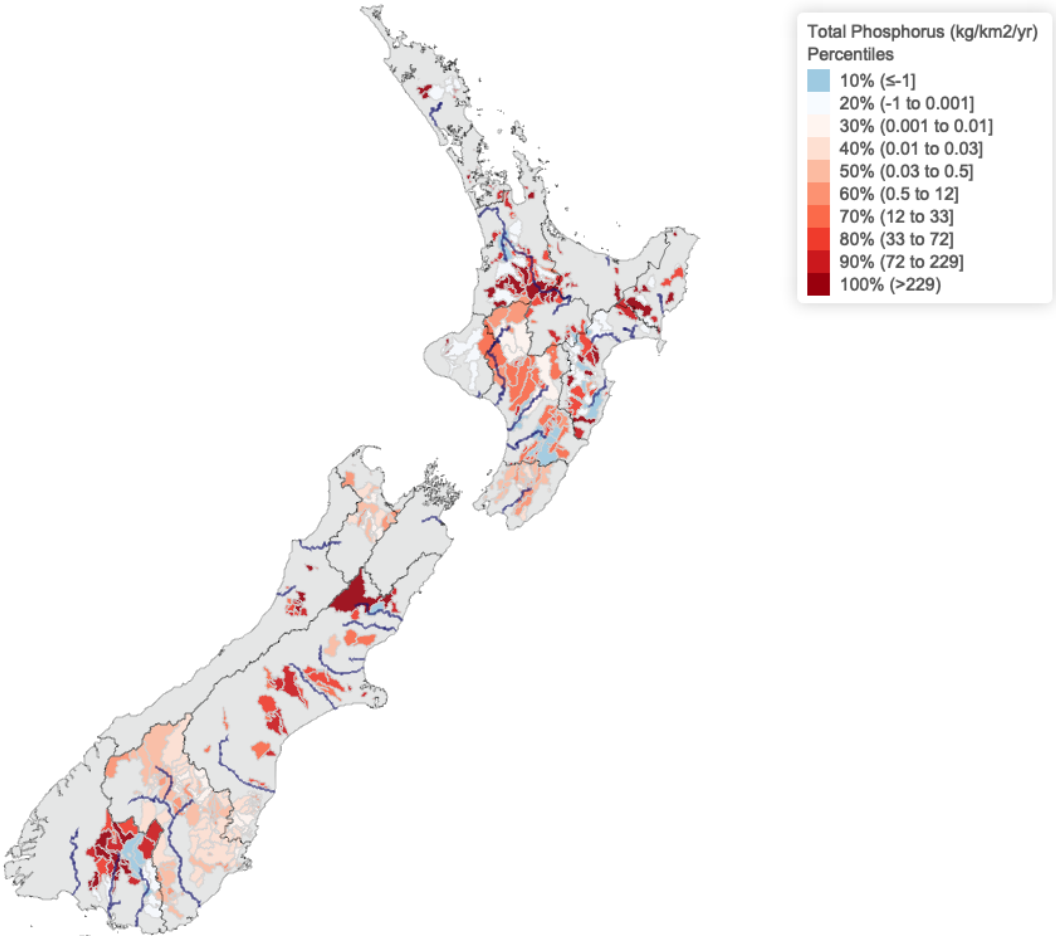


Figure 2.21: Average annual marginal total phosphorus pollutant loads by basin in New Zealand. Marginal loads are computed as the difference in measured annual pollutant loads at adjacent up and downstream water quality monitoring sites that measure this pollutant. The unit of analysis is a basin. A basin is colored blue if the amount of nonpoint source pollution attributed to it is on average negative. A basin is colored red if the amount of pollution attributed to it is on average positive. A basin is colored white if the amount of pollution attributed to it is close to zero. Regional council boundaries are demarcated by a black border. Major rivers in the country are marked in dark blue.

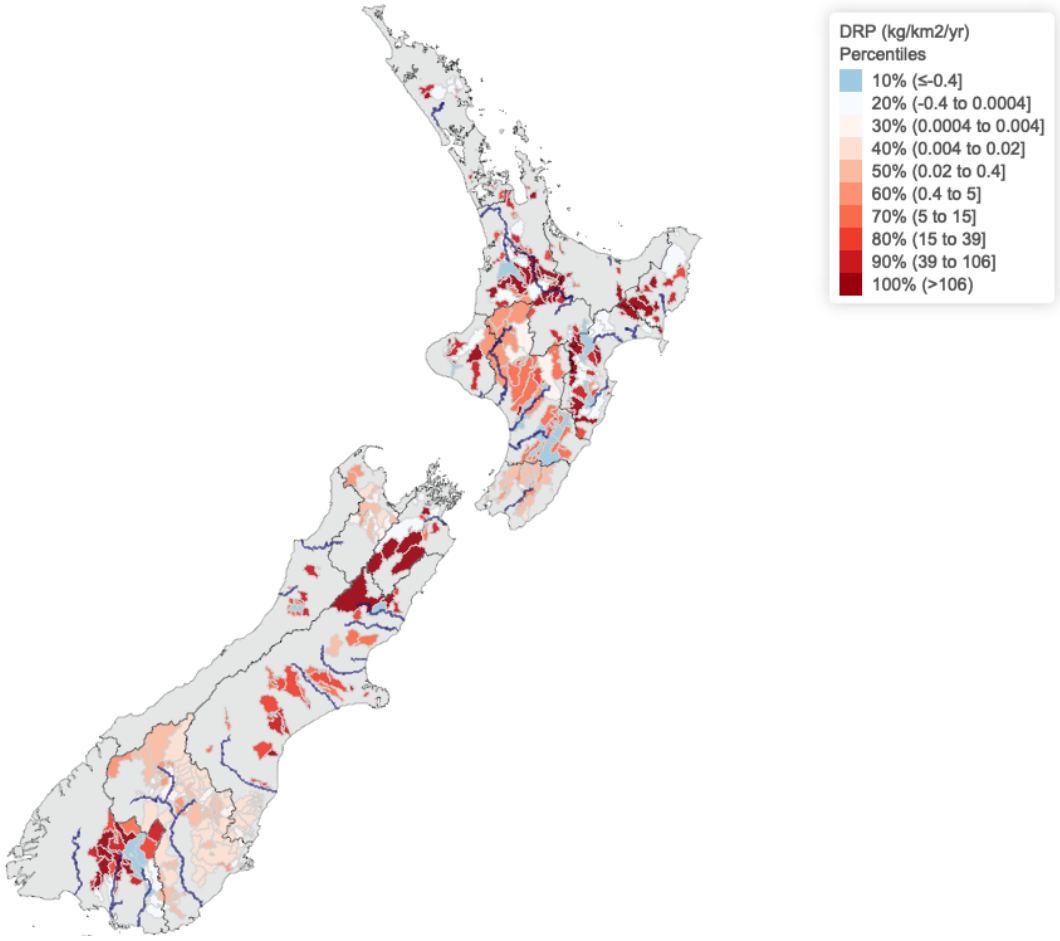


Figure 2.22: Average annual marginal dissolved reactive phosphorus pollutant loads by basin in New Zealand. Marginal loads are computed as the difference in measured annual pollutant loads at adjacent up and downstream water quality monitoring sites that measure this pollutant. The unit of analysis is a basin. A basin is colored blue if the amount of nonpoint source pollution attributed to it is on average negative. A basin is colored red if the amount of pollution attributed to it is on average positive. A basin is colored white if the amount of pollution attributed to it is close to zero. Regional council boundaries are demarcated by a black border. Major rivers in the country are marked in dark blue.

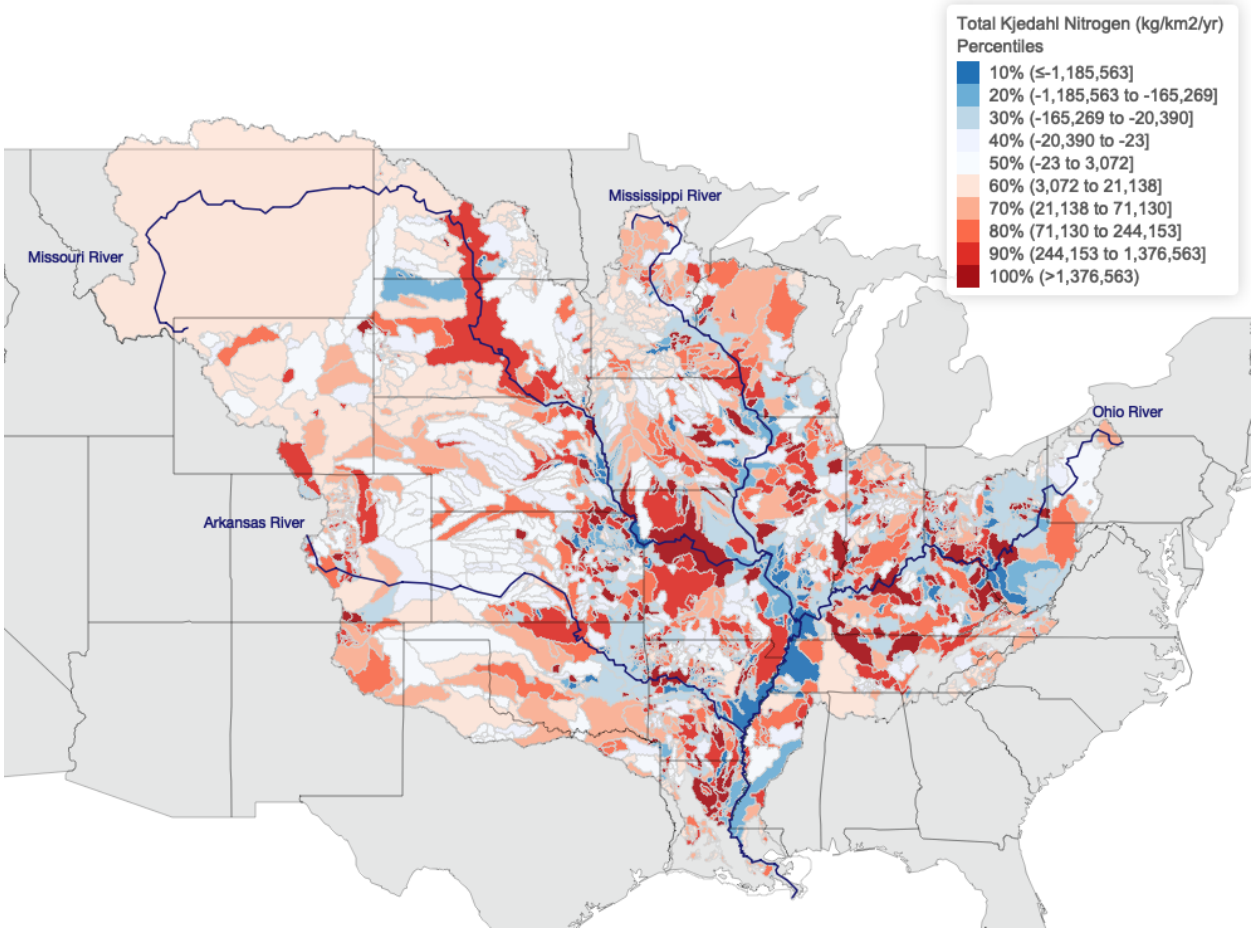


Figure 2.23: Average annual marginal total Kjedahl nitrogen pollutant loads by basin in the United States Mississippi River Basin. Marginal loads are computed as the difference in measured annual pollutant loads at adjacent up and downstream water quality monitoring sites that measure this pollutant.

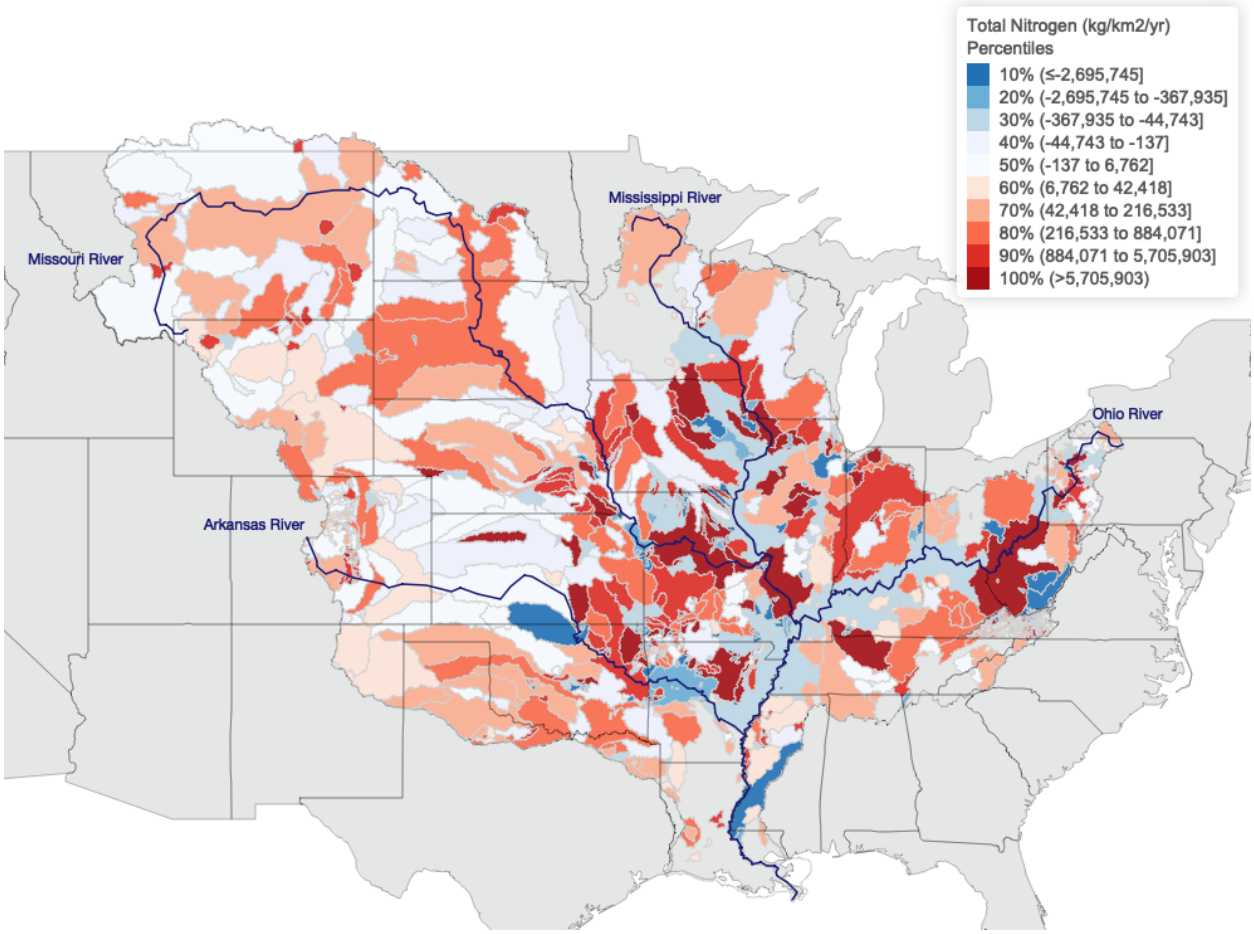


Figure 2.24: Average annual marginal total nitrogen (mixed forms) pollutant loads by basin in the United States Mississippi River Basin. Marginal loads are computed as the difference in measured annual pollutant loads at adjacent up and downstream water quality monitoring sites that measure this pollutant.

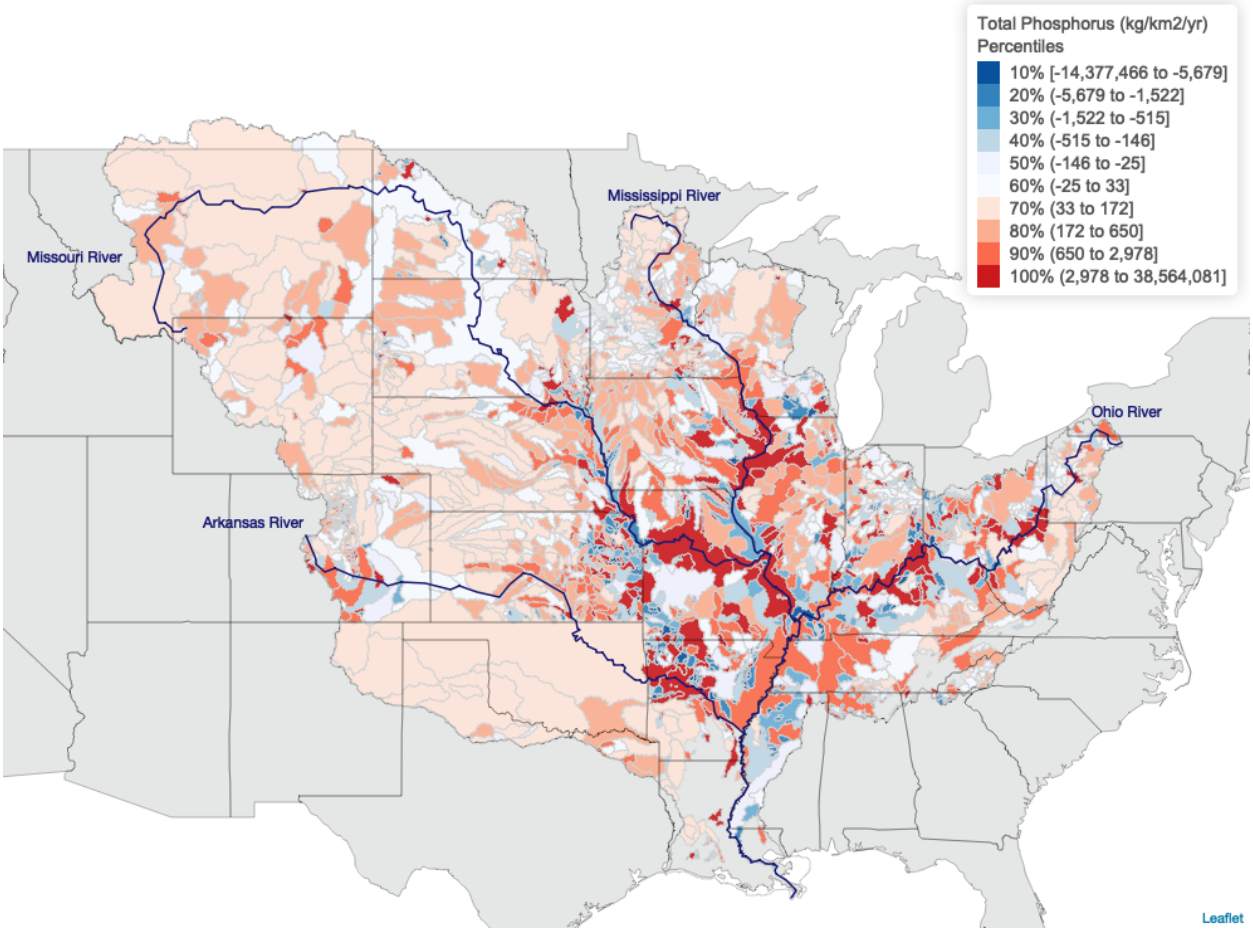


Figure 2.25: Average annual marginal total phosphorus (mixed forms) pollutant loads by basin in the United States Mississippi River Basin. Marginal loads are computed as the difference in measured annual pollutant loads at adjacent up and downstream water quality monitoring sites that measure this pollutant.

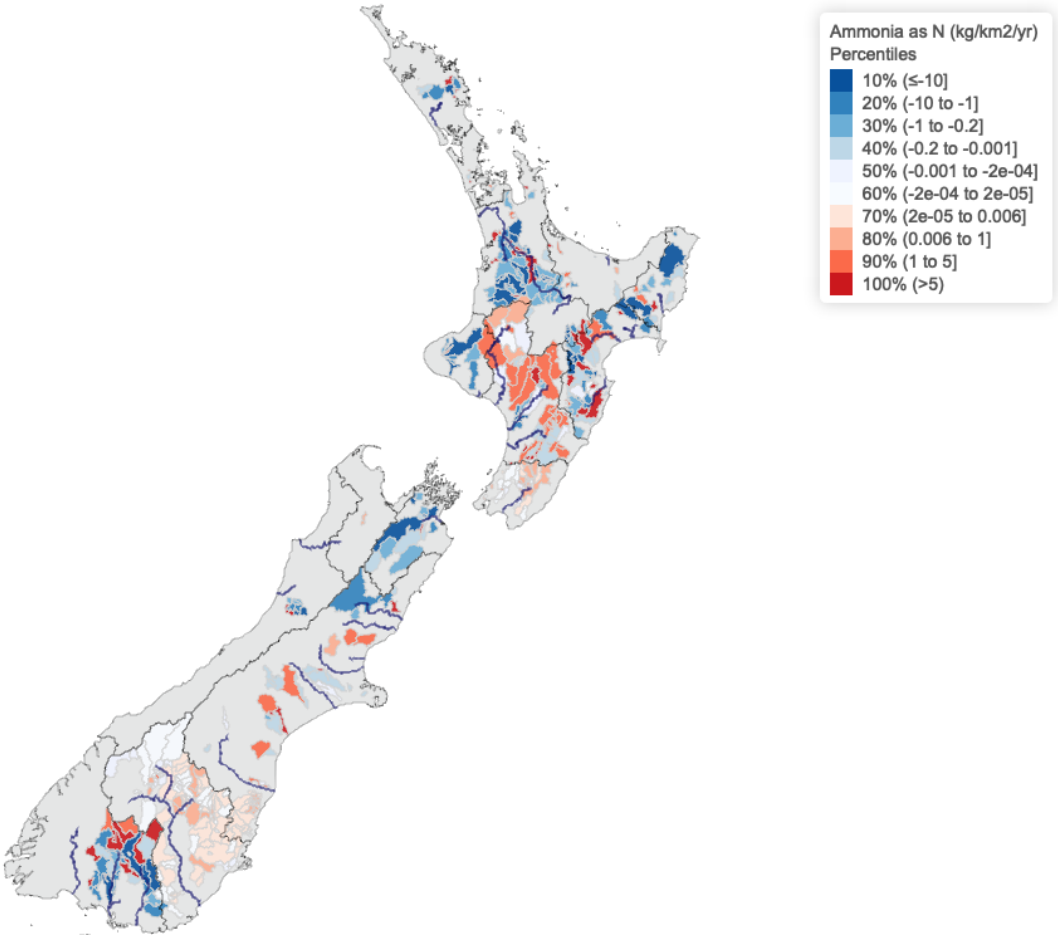


Figure 2.26: Trend in ammonia (as nitrogen) pollution at the basin-level in New Zealand between 1981-2018. A positive value indicates that the pollutant load is increasing on average over time in that basin, whereas a negative value indicates that the pollutant load is decreasing on average over time in that basin. The unit of analysis is a basin. A basin is colored blue if the amount of nonpoint source pollution attributed to it is decreasing over time. A basin is colored red if the amount of pollution attributed to it is increasing over time. A basin is colored white if the amount of pollution attributed to it not changing over time. Regional council boundaries are demarcated by a black border. Major rivers in the country are marked in dark blue.

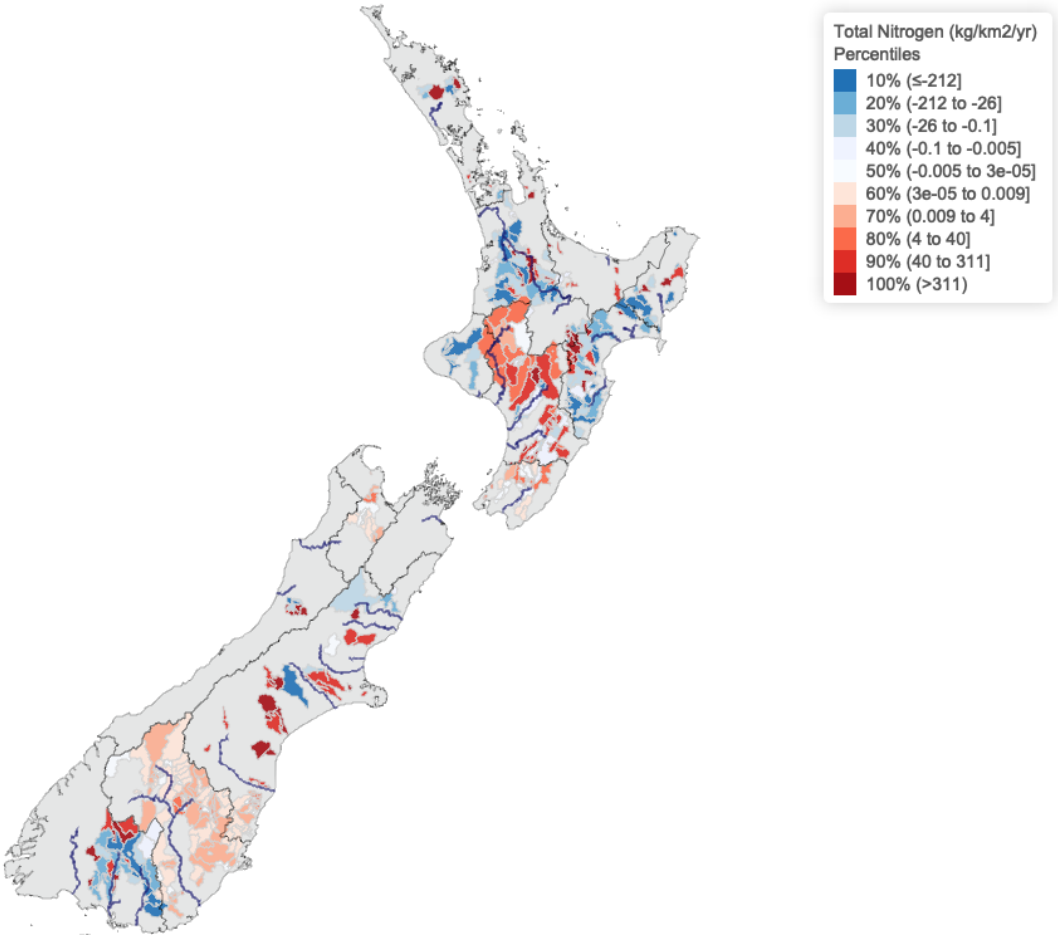


Figure 2.27: Trend in total nitrogen pollution at the basin-level in New Zealand between 1981-2018. A positive value indicates that the pollutant load is increasing on average over time in that basin, whereas a negative value indicates that the pollutant load is decreasing on average over time in that basin. The unit of analysis is a basin. A basin is colored blue if the amount of nonpoint source pollution attributed to it is decreasing over time. A basin is colored red if the amount of pollution attributed to it is increasing over time. A basin is colored white if the amount of pollution attributed to it not changing over time. Regional council boundaries are demarcated by a black border. Major rivers in the country are marked in dark blue.

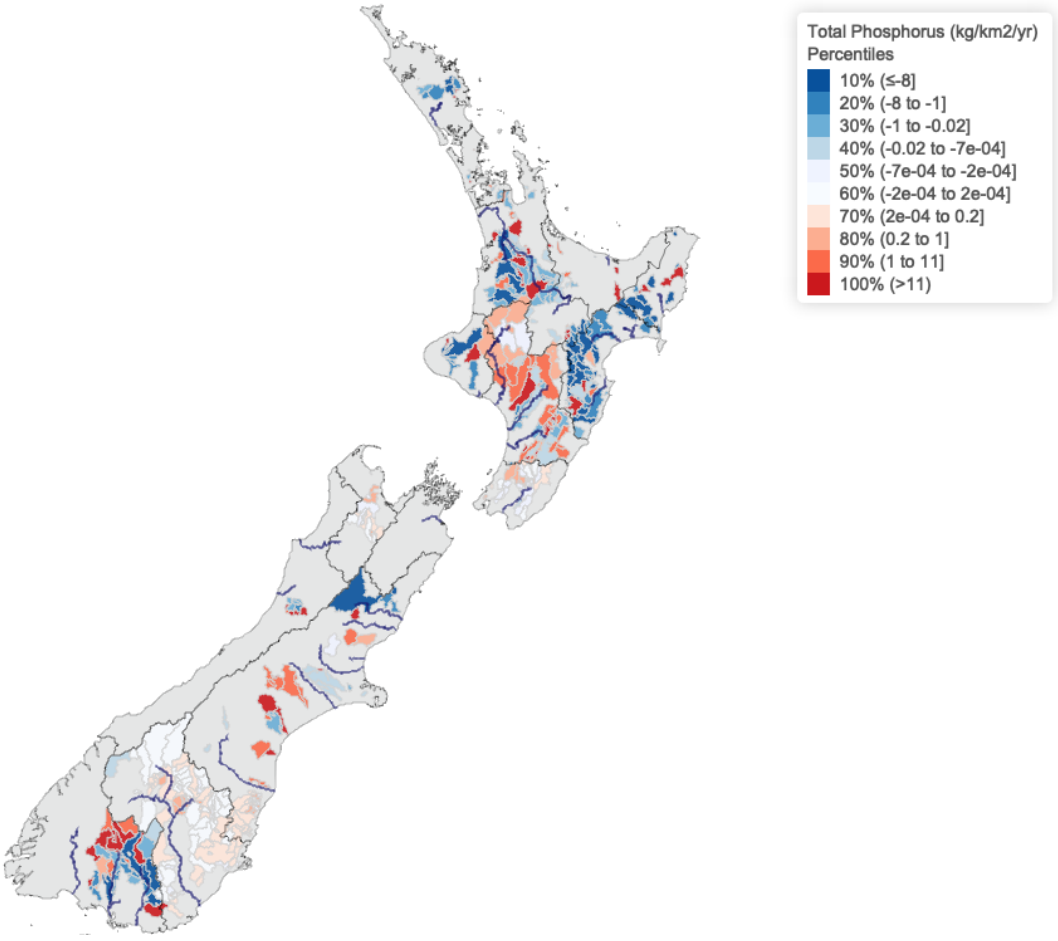


Figure 2.28: Trend in total phosphorus pollution at the basin-level in New Zealand between 1981-2018. A positive value indicates that the pollutant load is increasing on average over time in that basin, whereas a negative value indicates that the pollutant load is decreasing on average over time in that basin. The unit of analysis is a basin. A basin is colored blue if the amount of nonpoint source pollution attributed to it is decreasing over time. A basin is colored red if the amount of pollution attributed to it is increasing over time. A basin is colored white if the amount of pollution attributed to it not changing over time. Regional council boundaries are demarcated by a black border. Major rivers in the country are marked in dark blue.

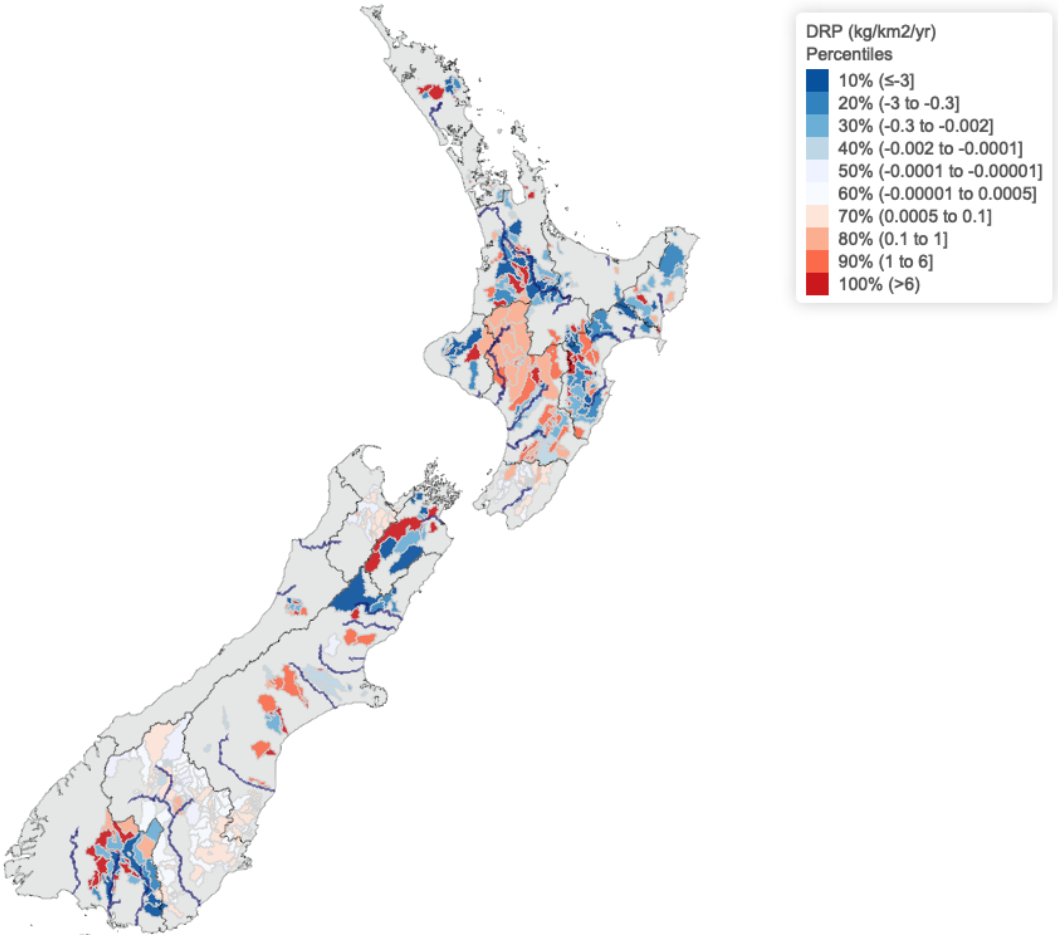


Figure 2.29: Trend in dissolved reactive phosphorus (DRP) pollution at the basin-level in New Zealand between 1981-2018. A positive value indicates that the pollutant load is increasing on average over time in that basin, whereas a negative value indicates that the pollutant load is decreasing on average over time in that basin. The unit of analysis is a basin. A basin is colored blue if the amount of nonpoint source pollution attributed to it is decreasing over time. A basin is colored red if the amount of pollution attributed to it is increasing over time. A basin is colored white if the amount of pollution attributed to it not changing over time. Regional council boundaries are demarcated by a black border. Major rivers in the country are marked in dark blue.

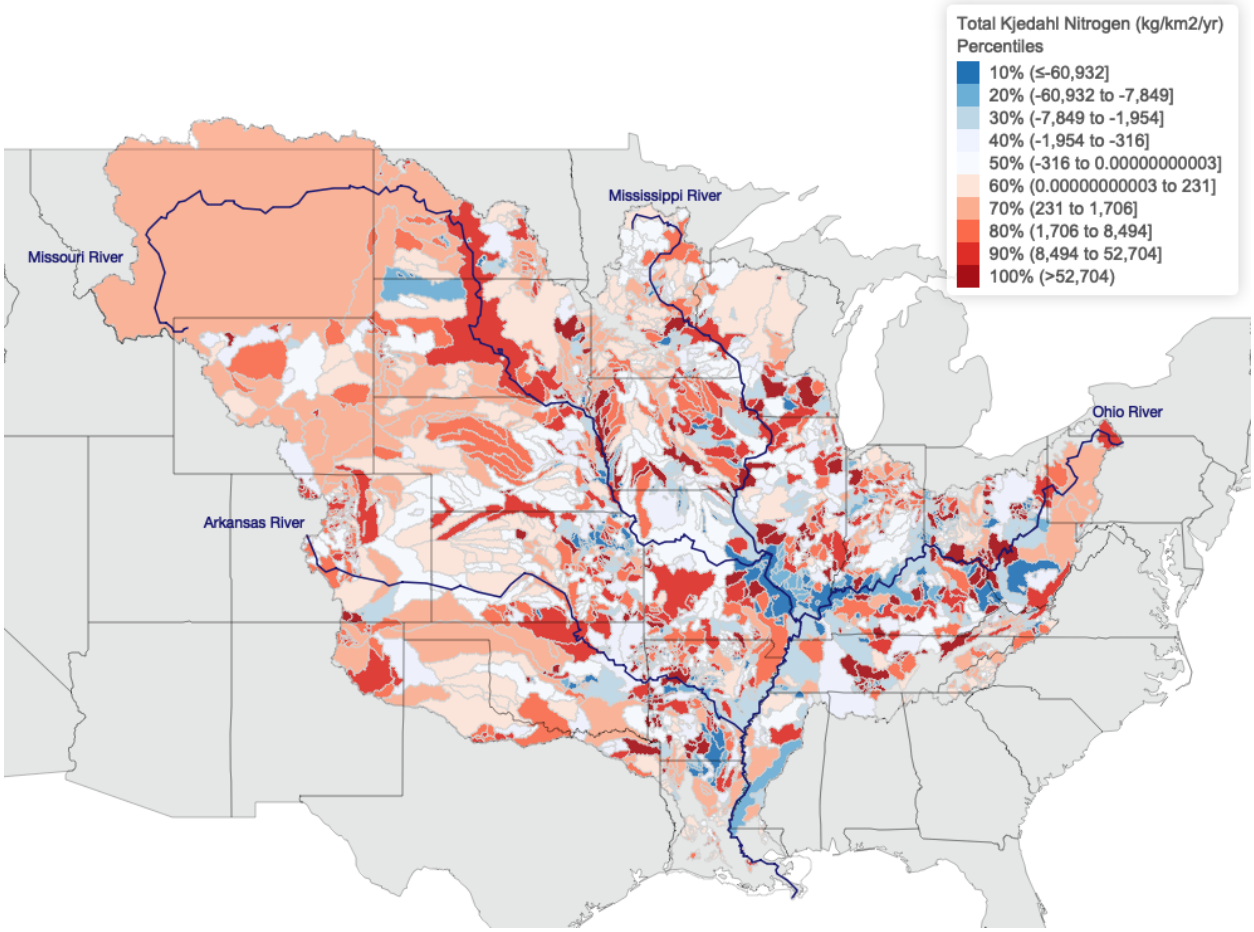


Figure 2.30: Trend in total Kjedahl nitrogen pollution at the basin-level in the United State Mississippi River Basin between 1981-2018. A positive value indicates that the pollutant load is increasing on average over time in that basin, whereas a negative value indicates that the pollutant load is decreasing on average over time in that basin.

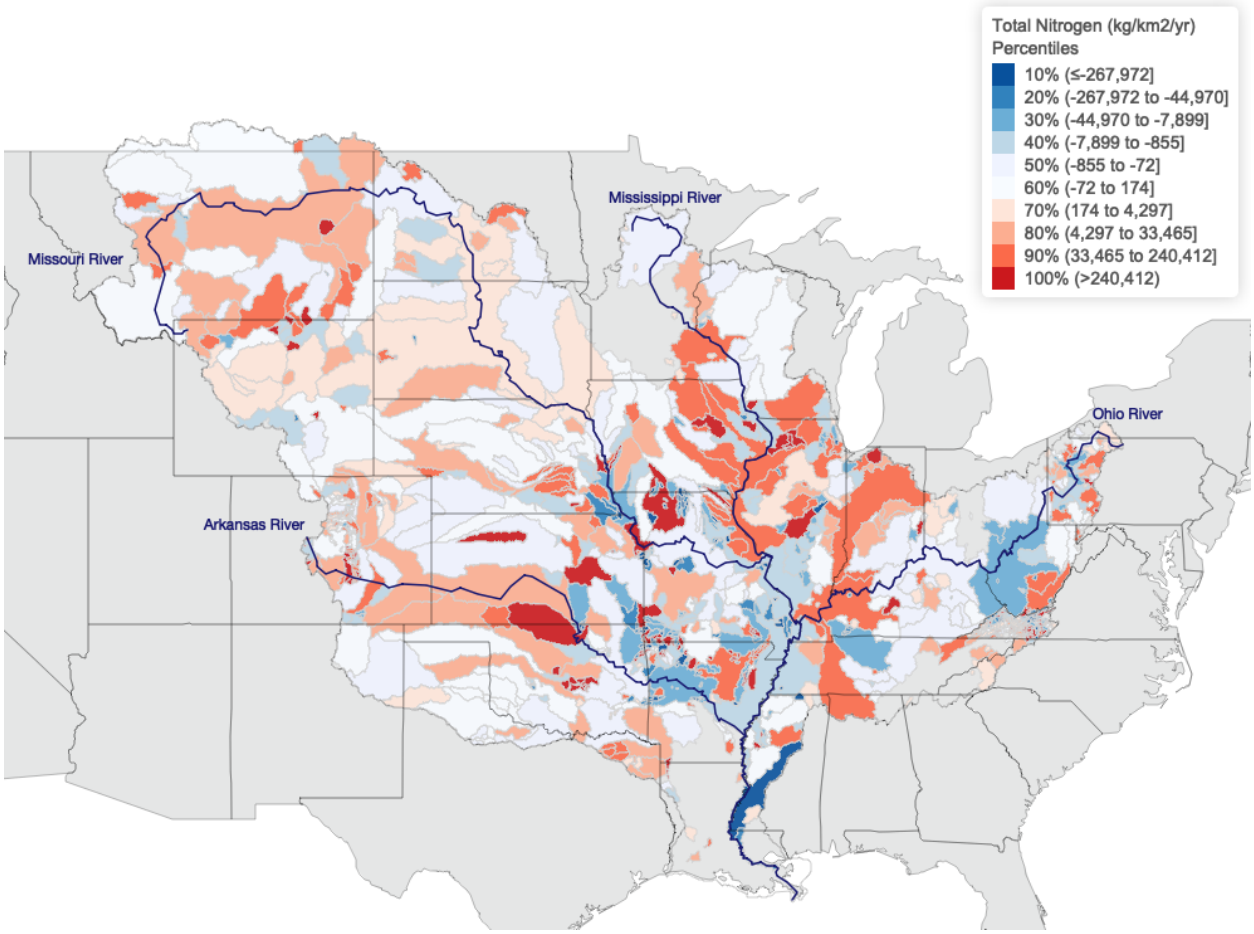


Figure 2.31: Trend in total nitrogen (mixed forms) pollution at the basin-level in the United State Mississippi River Basin between 1981-2018. A positive value indicates that the pollutant load is increasing on average over time in that basin, whereas a negative value indicates that the pollutant load is decreasing on average over time in that basin.

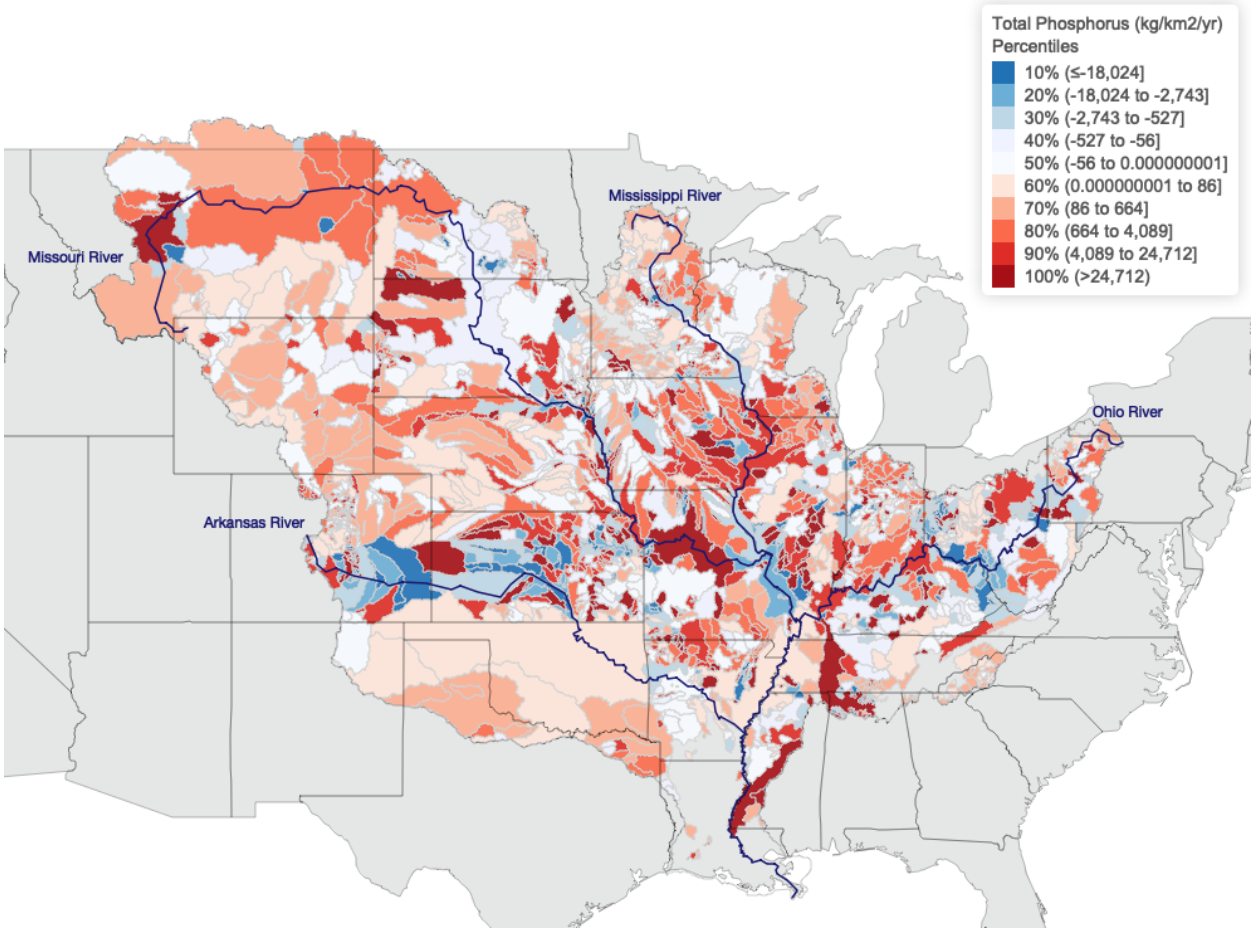


Figure 2.32: Trend in total phosphorus (mixed forms) pollution at the basin-level in the United State Mississippi River Basin between 1981-2018. A positive value indicates that the pollutant load is increasing on average over time in that basin, whereas a negative value indicates that the pollutant load is decreasing on average over time in that basin.

Tables

Table 2.1: Summary statistics of basin level annual cumulative (total) load (kg/year). Cumulative load is computed using the Beale Ratio Estimator method. This value represents the total runoff flux that passes by the water monitoring site in a given year. For the USMRB, we report summary statistics pertaining to total nitrogen (TN), total phosphorus (TP), and total Kjedahl nitrogen (TKN). In New Zealand, we report summary statistics pertaining to total nitrogen (TN), total phosphorus (TP), ammonia, and dissolved reactive phosphorus (DRP).

	Mean	**Median**	**Min.**	**Max.**
US Mississippi River Basin				
TN	5,320,139,493	4,076,451	69.85	1,080,299,586,525
TP	136,054,459	234,981.1	5.97	80,677,489,391
TKN	540,215,100	1,510,369	22.13	168,677,355,378
New Zealand				
TN	676,560.4	53,343,646	0	53,343,646
TP	24,460.36	2,031,396	0	2,031,396
Ammonia	18,841.18	3,282,781	0	3,282,781
DRP	11,551.18	248.06	0	1,047,885

Table 2.2: Summary statistics of basin level area normalized annual marginal load (kg/km2/year). Marginal load is computed by taking the difference between the cumulative annual load measured at a downstream water monitoring site from the cumulative annual load measured at its upstream water monitoring site. This value represents the total runoff flux that is attributed to the land mass (i.e. basin) flanked by the two monitoring sites in a given year. For the USMRB, we report summary statistics pertaining to total nitrogen (TN), total phosphorus (TP), and total Kjedahl nitrogen (TKN). In New Zealand, we report summary statistics pertaining to total nitrogen (TN), total phosphorus (TP), ammonia, and dissolved reactive phosphorus (DRP).

	Mean	**Median**	**Min.**	**Max.**
US Mississippi River Basin				
TN	-4,251,375	1,615.4	-17,254,328,770	42,649,647,658
TP	15,672,376	84.23	-32,155,029,552	467,261,587,689
TKN	3,894,974	597.69	-34,305,195,045	454,981,839,321
New Zealand				
TN	-17,010.96	151.4	-19,034,337	9,352,073
TP	3,448.23	11.04	-1,027,198	1,975,447
Ammonia	-741.3	3.46	-2,318,504	955,036.6
DRP	-635.76	3.29	-4,262,001	1,209,529

Chapter 3

Estimating the Corporate Tax Elasticity in Extractive Industries: Evidence from Multinational Oil Producers networks[1]

3.1 Introduction

A growing body of literature examines the trend of rapidly increasing wealth inequality, both within and across countries around the globe. In the United States, share of household wealth held by the ultra wealthy has followed a 'U-shape' over the past century. Households in the top 0.1% went from holding around 25% of total household wealth in 1925, dipping to 7% in 1978, and climbing back up to 23% in 2013 (Saez and Zucman 2016). This 'U-shape' trend also characterizes distributions of wealth in countries like France and Britain (Alvaredo, Atkinson, and Morelli 2017). There is broad consensus that long run income inequality is attributable to the accumulation of capital, rather than disparity in labor income, and that reversing the concentration of wealth requires governments to implement and enforce an effective corporate tax regime. However, accurately measuring wealth and inequality, and particularly the ownership of capital, remains challenging. In particular, the rise of globalization has allowed firms to utilize complicated corporate networks spanning multiple countries to bypass cumbersome tax regulations and obfuscate their true economic activity. Firms' use of tax evasionary practices, such as the artificial shifting of profits and use of shell corporations in tax havens, enable multinational corporations (MNCs) to reduce their global tax burden. The lack of reliable data renders it difficult for governments and economists to ascertain the amount of wealth held by wealthy individuals and corporations, and to assess whether they are complying with tax regulations.

[1]I am deeply grateful to James Sallee and Solomon Hsiang for their brilliant advice and patient mentorship on this project. I would like to also thank Michael Anderson, Ethan Ligon, Reed Walker, Gabriel Zucman, members of the 2015 ARE cohort, and members of the Global Policy Lab for their helpful advice. Lastly, I would like to thank participants at the University of California Berkeley Climate Lunch, University of California Berkeley WEB Breakfast, and University of California Berkeley ERE Seminar for their useful feedback.

In this paper, I will investigate the intersection of corporate tax evasion and natural resource extraction by focusing on the tax paying behavior of multinational oil corporations. The aim of this project is to quantify the magnitude of base erosion and profit shifting (BEPS) of oil firms through estimating how responsive their reported profits are with respect to statutory corporate tax rate differences. I will also look at how this magnitude differs between developed and developing oil producing countries.

I believe this area deserves special interest because while many of these corporations are multi-billion dollar firms that operate in developed and developing countries, exercise large control over the future of energy consumption, and use their political clout to lobby against climate change policies, little evidence is available on whether these firms are complying with government tax policy and paying corporate taxes or if they are illegally shifting their profits out of the country and using the returns to benefit their shareholders.

In this project, I will combine the existing methodology of estimating the semi-elasticity corporate tax rate incentive with respect to reported profits (Hines and Rice 1994; Huizinga and Laeven 2008) with the novel approach of using satellite imagery to observe annual oil-field by firm level production (in order to impute real economic activity of oil companies) in order to estimate the semi-elasticity with respect to true profits. The ability to compute the mismatch between declared profits with remotely sensed oil production will allow me to derive an estimate of the true magnitude of tax evasion by multinational oil firms. In future work, I will delve into the mechanisms oil companies employ to evade taxes, as well as explore payments made by oil companies in countries where the main tax regime is the use of production sharing contracts (PSCs), which have different financial implications for firms.

The rest of the prospectus will go as follows: in section 3.2, I will provide background context on the oil industry and applicable tax regimes worldwide. In section 3.3, I present descriptive facts about the relationship between oil production, tax and royalty payments, and regulatory enforcement capacity of governments where extraction occurs. In section 3.4, I review the literature. In section 3.5, I cover the theory underlying my empirical specification. In section 3.6, I present the data that I will use. In section 3.7, I present my main results and robustness checks. In section 3.8, I discuss caveats and future work. Lastly, in section 3.9, I conclude.

3.2 Background

As shown in Figures 3.1 and 3.2, crude oil extraction spans six continents and takes place in both how and low income countries. Additionally, the level of production does not appear to be correlated with the oil producing country's GDP per capita, where poorer and richer nations are just as likely to produce similar quantities of crude oil (see Figure 3.3). In 2007, oil production was dominated by Russia, Saudi Arabia, the United States, Iran, China, Mexico, Venezuela, Kuwait, the United Arab Emirates, and Norway. A decade later, in 2017, the dominant oil producers remained mostly the same, with Iraq, Canada, and Brazil entering the list of top ten largest crude oil producers (JODI 2018).

3.2.1 Oil and Gas Sector Fiscal Regimes

For assessing taxes on oil producing firms, countries either use a PSC or tax/concession regime, or a combination of both (Sunley, Baunsgaard, and Simard 2003; EY 2017). As you can see from Figure 3.4, most countries in Europe and North America rely primarily on a tax/concession regime, whereas African, Asian, and South American countries tend to implement a PSC regime or a combination of both regimes. Even though I am focusing only on corporate income taxes in this project and not accounting for the revenue governments collect via PSC regimes, corporate tax enforcement is still relevant because governments continue to tax oil producers' corporate profits even in countries where PSC is the primary regime (EY 2017).

3.2.1.1 Production sharing contract (PSC) regime

Under a PSC, the government owns the rights to the hydrocarbon reserves and contracts out the exploration and production of oil to a company. The company bears the exploration risk, but if it finds an economically viable reserve, then it splits the production (net of exploration and operational costs) with the government according to their contract. More specifically, the contracting company will receive a portion of total production as 'cost oil,' in order to cover its exploration, development, and production costs. The remaining oil is denoted as 'profit oil,' and is split between the government and contractor.

3.2.1.2 Tax/concession regime

In oil-producing countries where governments use a tax/concession regime, companies obtain a title to the petroleum resources and have to pay income taxes on their profits.

3.2.1.3 Corporate income tax policy

Corporate tax policy worldwide can be divided into two types of tax regimes: 'residence-based' and 'source-based' taxation. The former tax regime taxes corporations based off of where the firm is incorporated. In the latter tax regime, corporations pay taxes based off of where production occurs.

The United States is one of two countries in the world that has 'residence-based' taxation. In the residence-based tax regime, firms can deduct corporate taxes paid abroad against the income tax they have to pay domestically to avoid double taxation. However, because the United States allows for indefinite deferral of foreign profits, in practice the American tax regime is akin to a source-based tax system. Under a source-based tax system, profit-maximizing firms are incentivized to artificially shift profits made in high tax jurisdictions to low tax jurisdictions in order to reduce their global tax burden.

Firms in oil and gas industries have to pay income tax on their profits both in countries that use a PSC and a tax/concession regime. In fact, in many countries, oil companies may have to pay a higher tax rate than firms in non-extractive industries. Since income tax is levied on profits, multinational oil corporations may use various mechanisms to reduce their tax base through the artificial shifting of their profits to lower-tax jurisdictions. Many

countries have explicit laws regulating such practices, such as around transfer pricing or thin capitalization, so some of these methods employed by oil companies to reduce their global tax burden are illegal.

Transfer pricing is when the firm purposefully over/under charges on intra-group transactions in order to manipulate its reported profits. The intuition behind transfer pricing is that the MNC maximizes after-tax global profits by decreasing profits in a high tax jurisdiction (and vice versa in a low tax jurisdiction) by inflating production costs or lowering revenue. Other mechanisms that MNCs may use to shift profits include:

- Earnings stripping: A subsidiary in a high tax jurisdiction takes on a loan for operational expenses from another subsidiary located in a low tax jurisdiction. The high tax subsidiary is able to deduct the loan payments from its taxable profits. The loans made between firms of the same parent company can often be made at above market-rate interest rates, thereby increasing the deductible expenditure.

- Firms can overcharge for management fees, consultancy fees, capital or machinery, or manipulate other import/export prices in intra-group transactions.

- Price hedging: the use of price hedging mechanisms between related parties could either be used to manage legitimate risk or for transfer pricing purposes.

- Shell company: in countries where the corporate tax rate on the oil industry is higher than the rate on other industries, the company might create a shell company in the same country to make loans to the oil company. The oil company is able to deduct the loan payments from its taxable income and pay a lower tax rate on its profits through the shell company.

3.2.1.4 Royalties

Royalties can be either a unit levy (dependent on the quantity of oil extracted) or an *ad valorem* levy (dependent on the value of oil extracted). For *ad valorem* royalties, the value of the oil is often calculated on the sales price or the f.o.b. export price of oil net qualifying production costs. Additionally, the market price of oil used to calculate the *ad valorem* royalty is adjusted to reflect the quality of the oil as well transportation costs. Given that the sales price depends on firm production costs and oil quality (margins that may be difficult for the government to observe), the firm could potentially obfuscate true costs and product quality, thereby lowering the value of the oil and the amount of royalties it pays to the government.

Governments prefer royalties because it is levied at time of production, so they receive revenue early on in the oil firm's operation life cycle. However, governments may limit the royalty rate because firms can only deduct royalties from their taxable income but not as foreign tax credit against their income tax payable to their home government.

3.3 Descriptive Facts

In this section, I detail correlative associations relating oil production, tax payments, and institutional capacity in enforcing regulation. These relationships suggest the need for using causal identification techniques to identify whether oil firms are evading corporate taxes. Furthermore, these descriptive facts show a disparity in the capacities of different governments in collecting tax payments from oil producers, with lower income countries collecting less taxes and royalty payments than higher income countries, holding production constant.

As seen in Figure 3.5, there is a positive association between oil production in high income countries and the country's ranking on the World Justice Project's Rule of Law Index (Ponce 2018), a number that encapsulates regulatory enforcement in that country. However, this relationship is downward sloping for middle and lower-middle income countries. This suggests that less developed oil producing countries likely lack the capacity to detect tax evasion and effectively enforce tax policy.

This correlation is further illustrated in Figures 3.6 and 3.7, where I plot the association between national crude oil production against tax payments (Figure 3.6) and royalties (Figure 3.7) paid by oil companies to the oil-producing project countries. In these two subfigures, it is striking to observe that higher income oil-producing states receive higher tax payments than their lower income counterparts at similar levels of production.

The caveat is that the oil production data from JODI (2018) are estimates of national oil production, whereas the tax and royalty payments (Subhash 2018) only account for payments made by oil and gas firms listed on European stock markets, as opposed to the universe of oil companies that operate in that country. I only have payments data made by this subset of firms because they are required under a European Union law passed in 2013 to disclose their country-by-country payments to the EU government(Commission 2018). Under this regulation, which was implemented in mid-2014 (hence the paucity of data from that year), an oil and gas firm on a European market with multiple oil projects in various countries has to disclose the amount of payments it made to each of the governments where extraction occurred, for each of their oil projects in each country. This regulation only requires that the firms disclose their payments amount, but not the level of production in their various projects. As such, in this figures, I cannot compare company payments to company production across countries.

3.4 Contribution to Literature

This paper contributes to two main strands of literature: corporate tax evasion, and monitoring illicit economic activity using remote sensing. While there is a robust literature documenting the extent of corporate tax evasions by multinational corporations using various methods, my paper is one of the few that lies at the intersection of tax evasion and natural resources.

The majority of the existing literature on tax evasion finds evidence of artificial profit shifting in American and European MNCs, with the difference in the tax rates across subsidiaries located in different countries driving the direction of profit shifting (Hines and Rice

1994; Huizinga and Laeven 2008). Within this particular strain of literature, much of the advancement has come from moving from national-level cross-sectional data to micro-level panel data, allowing for a much more credible identification of BEPS behavior (Dharmapala 2014).

In a different approach to identifying BEPS behavior, Dharmapala and Riedel (2013) show that exogenous shocks to a parent firm's income are positively associated with pre-tax profits of affiliates in low tax jurisdictions. Furthermore, there is evidence demonstrating that MNCs primarily shift profits artificially through transfer pricing, or the manipulation of import/export prices in intra-group transactions, and that these prices differ from arm's length prices (Bertrand, Mehta, and Mullainathan 2002; Clausing 2003; Heckemeyer and Overesch 2013). Recent work documenting tax evasion in developing countries find that the corporate tax elasticity is a magnitude higher than that of firms in OECD countries (Besley and Persson 2013; Bachas and Soto 2015; Johannesen, Torslov, and Wier 2016).

Recent research finds evidence of profit shifting in the oil industry. Johannesen and Larsen (2016) find that the passage of the 2013 EU regulation requiring that extractive firms disclose their country-by-country payments causes their firm stock values to drop, suggesting that increasing financial transparency reduces opportunity for extractive firms to evade taxes. Wright and Zucman (2018) find that US MNCs in the oil sector have much higher returns than MNCs in other sectors and pay much lower tax rates to oil producing countries today than in the 1970s.

The second contribution is to the strand of literature that Zitzewitz (2012) dubs 'forensic economics,' or the use of economic tools to detect a variety of illicit activity. In particular, there are some papers that use remote sensing data to ascertain firm-level behavior in extractive industries. Saavedra and Romero (2017) use machine learning to classify land pixels as either 'mined' or 'not mined' in order to estimate the effect of a tax reform on illegal mining in Colombia. Within the oil and gas sector, Do et al. (2017) use oil flares to estimate the level of oil production in ISIS controlled territory. While both of these papers demonstrate the research questions satellite data can help answer, their work focuses only on a specific region. My paper will be the first to use satellite data to estimate global oil production at the oil concession spatial resolution. Additionally, my paper will contribute to the existing tax evasion literature through the construction of a novel dataset of true oil firm profits through the remotely sensed oil production dataset, thus allowing me to compare the response of reported versus true pre-tax profits with respect to tax rate differentials.

3.5 Model

I closely draw upon the approach from Hines and Rice (1994) and Huizinga and Laeven (2008). In this set-up, MNC M generates true profits P_i in country i, but it can also shift profits S_i using transfer pricing techniques from abroad into country i (where S_i can be greater than or less than 0). However the firm incurs shifting costs (both variable and fixed costs incurred by setting up profit shifting mechanisms, e.g. hiring financial consultants to carry out the profit shifting or setting up shell companies, and penalties paid to governments in expectation). The marginal cost of shifting is proportional to the ratio of shifting costs and

profits in country i, $\frac{S_i}{P_i}$, which we define as γ. The MNC M maximizes its worldwide after-tax profits with respect to the amount shifted subject to the constraint that the cumulative sum of shifted profits is weakly negative:

$$\mathcal{L} = \Sigma_{i=1}^{n}(1-\tau_i)\left(P_i + S_i - \frac{\gamma}{2}\frac{(S_i)^2}{P_i}\right) - \lambda\Sigma_{i=1}^{n}S_i \tag{3.1}$$

where τ_i is the statutory tax rate in country i and $\frac{\gamma}{2}\frac{(S_i)^2}{P_i}$ is the expenses incurred by transfer pricing.

After taking the first derivative, we can solve for the optimal amount S_i^* that the MNC will shift in or out of country i:

$$S_i^* = \left(\frac{P_i}{\gamma(1-\tau_i)}\right)\frac{\sum_{k\neq i}^{n}\left(\frac{P_k}{1-\tau_k}\right)(\tau_k - \tau_i)}{\sum_{k=1}^{n}\left(\frac{P_k}{1-\tau_k}\right)} \tag{3.2}$$

Note that S_i^*, the amount of profits shifted into country i, increases with tax differential $(\tau_k - \tau_i)$. This intuitively makes sense; as the tax difference grows larger, the firm has more to gain by shifting a larger amount of profits into or out of country i.

Furthermore, we can define the relationship between true profits P_i and reported profits P_i^r as:

$$P_i = P_i^r - S_i \tag{3.3}$$

Plugging Equation (3) into (2) and solving for P_i^r, we arrive at:

$$P_i^r = P_i\left[1 - \frac{1}{\gamma}\underbrace{\frac{1}{(1-\tau_i)}\frac{\sum_{k\neq i}^{n}\left(\frac{P_k}{1-\tau_k}\right)(\tau_k - \tau_i)}{\sum_{k=1}^{n}\left(\frac{P_k}{1-\tau_k}\right)}}_{C_i}\right] \tag{3.4}$$

Next, once we take logs from both sides, we can rewrite Equation (4) as:

$$p_i^r = p_i - \frac{1}{\gamma}C_i \tag{3.5}$$

where $p_i^r = \log(P_i^r)$ and $p_i = \log(P_i)$.

The variable of interest in this equation is C_i, or what Huizinga and Laeven (2008) call the composite tax variable. This variable captures the firm's incentive and opportunity to shift profits. As the tax differential between the other countries and i grows, the incentive for the firm to profit shift into i increases. Secondly, as its profits in country k grow, it becomes less costly for the firm to shift profits in and out of that country.

However, we only observe reported pre-tax profits in the firm-level financial data, but not true profits. So in order to estimate true profits p_i, we assume that the firm's production function is Cobb-Douglas given by $Q_i = cA_i^{\epsilon}L_i^{\alpha}K_i^{\psi}e^{u_i}$. If we assume perfect competition, then the firm's true profit is equal to Q_i minus the wage cost, or $P_i = Q_i - w_iL_i$. If we take logs of the profit then plug in the relationship into Equation (5), we can rewrite the equation as:

$$p_i^r = \beta_1 + \beta_2 a_i + \beta_3 l_i + \beta_r k_i - \hat{\gamma} C_i + u_i \tag{3.6}$$

We can now bring Equation (6) to the data and estimate the semi-elasticity of corporate tax differential with respect to reported profits with the following estimating equation:

$$p_{m,i,t}^r = \beta_1 + \beta_2 a_{m,i,t} + \beta_3 l_{m,i,t} + \beta_r k_{m,i,t} - \hat{\gamma} C_{m,i,t} + FE + \epsilon_{m,i,t} \tag{3.7}$$

for MNC M, in country i, in year t, with fixed effects FE.

The variable of interest is $\hat{\gamma}$, which is the semi-elasticity of the composite tax variable with respect to reported profits. The interpretation of $\hat{\gamma}$ is that it represents the percentage change in pre-tax income associated with a one-percentage point change in the composite tax rate differential.

3.6 Data

3.6.1 Data used in the established tax evasion analysis

In the first part of the paper where I employ Huizinga and Laeven (2008)'s approach to firms in the oil industry, I use the following data:

3.6.1.1 List of oil and gas firms listed on European stock exchanges

The list of companies in my paper are oil and gas companies listed on an European stock exchange, with the exchanges being: BME Spain, Euronext Amsterdam, Euronext Brussels, Euronext Paris, Helsinki, Italian Brosa, NASDAQ OMX Nordic-Copenhagen, Oslo Bors, SIX Swiss Exchange, Deutsche Borse - Frankfurt, and London (LSE). As mentioned in Section 3, these extractive companies that are listed on a European stock exchange (but not necessarily incorporated in an EU country) are subject to financial country-by-country payments disclosure requirements imposed by the European Commission. The actual list of companies is the same as the list used by Johannesen and Larsen (2016) and was obtained from email communication with the authors.

The summary statistics of all the firms can be found in Table 3.1. The summary statistics for the parent companies are in Table 3.2 and for the subsidiaries are in Table 3.3. As one might expected, the size of the parent companies are much larger than their subsidiaries. On average, MNCs have subsidiaries in about four different countries. However, the distribution of the number of countries an MNC as affiliates in is heavily right skewed, with the median

of MNCs having affiliates in only one foreign country and the largest MNC with subsidiaries in 33 other countries (see Table 3.4).

The composition of firms in my panel dataset are relatively stable over time. There are roughly 50-60 multinational corporations each year in my dataset between 2008 - 2017 (see Table 3.5). These MNCs are located in around 50 different countries each year (see Table 3.6), with the parent firms incorporated in around 20 different countries and the subsidiaries incorporated in around 50 different countries each year (see Tables 3.7 & 3.8, respectively).

3.6.1.2 Firm level annual financial data

I have firm-level annual data available for 2008 - 2017 from the Bureau van Dijk ORBIS database for firms that are on the list obtained from Johannesen and Larsen (2016), the ultimate holding company (*i.e.* parent company) of these firms, as well as all the subsidiaries where these ultimate holding companies have at least a 75% ownership share of. The financial variables that I obtain from ORBIS are: *'Earnings Before Interest and Taxes,' 'Fixed Assets,' 'Costs of Employees,' 'Number of Employees,' 'Total Operating Revenue,'* and *'Ratio of Debt to Equity.'* Furthermore, the database contains the country code for where the firm, its ultimate holding company, and the holding company's various affiliates are located.

3.6.1.3 Corporate tax rates

I have statutory country level corporate tax rates from TradingEconomics.com for 249 countries between 2008 - 2017. The summary statistics for the tax rates applicable to the firms in my dataset are in Table 3.9.

3.7 Results

As presented in Section 5, my main estimating specification is

$$p^r_{m,i,t} = \beta_1 + \beta_2 a_{m,i,t} + \beta_3 l_{m,i,t} + \beta_r k_{m,i,t} - \hat{\gamma} C_{m,i,t} + FE + \epsilon_{m,i,t} \tag{3.8}$$

where p, l, k are log of earnings before interest and taxes, log of either the number of employees or log of total employee compensation, and log of fixed assets, respectively. For a, I use two different measures, one employed in Huizinga and Laeven (2008), where they measure a as GDP per capita, which is specified at the country level. The second measure of a is used in Maffini and Mokkas (2011) where they estimate that the host country's statutory corporate tax rate has a significant impact on the MNC's measured TFP in that country. In their paper, they estimate a as:

$$log(A_{m,i,t}) = log\left(\frac{Y_{m,i,t}}{L_{m,i,t}}\right) - \alpha_K log\left(\frac{K_{m,i,t}}{L_{m,i,t}}\right) \tag{3.9}$$

The benefit of using this measure of a is that it is firm-specific.

In table 3.10, I use the log of total labor compensation for *Labor*, whereas in table 3.11, I use log of total number of employees. In both of these tables, I use the first measure of a, GDP

per capita. My two preferred specifications are specifications (3) and (4). In specification (3) I use year and MNC-country fixed effects and in specification (4) I use MNC and country-year fixed effects. In table 3.10, the coefficient $\hat{\gamma}$ for composite tax variable C in specification (3) is 0.98, and not statistically significant, and in specification (4) it is 1.744 and significant at the 10% level. In table 3.11, the same coefficients are 0.286 and -0.099, respectively, and neither are statistically significant.

Here in tables 3.12 and 3.13, I use the second measure of a, TFP from Maffini and Mokkas (2011). I enter each term separately into my regression, hence the variable TFP1 corresponds to $log\left(\frac{Y_{m,i,t}}{L_{m,i,t}}\right)$ and TFP2 is $log\left(\frac{K_{m,i,t}}{L_{m,i,t}}\right)$. Again, the difference between the two tables is that table 3.12 uses log of total labor compensation whereas table 3.13 uses log of number of employees. In table 3.12, the coefficients for $\hat{\gamma}$ in specification (3) and (4) are 0.582 and 0.523 respectively. In table 3.13, the coefficients for the same specifications are 0.681 and 0.210. In all four cases, none of the coefficients are statistically significant.

3.7.1 Robustness check

As a robustness check, I construct a tax differential variable that is equal to the difference between the tax rate in the parent and affiliate country that I use in place of the composite tax variable C. Here I don't take into account the MNC network and don't incorporate the tax rates of subsidiaries in other countries. Again, I use the two different specifications for labor and two different measures for a.

In this specification, we would predict that if the firm is profit shifting, then as the tax differential between the tax rate in parent and affiliate country increases, then the amount of pre-tax profits reported in the affiliate country would rise, so we would expect to see a positive coefficient.

In table 3.14 the coefficient on the tax rate difference in specification (3) is 0.048 and statistically significant at the 1% level and in specification (4) is 0.04 and is statistically significant at the 5% level. In 3.15, the coefficients are 0.030, and significant at the 5% level, and 0.008, but is not statistically significantly different than 0. Again, in these two tables we use the first measure of a.

In the last set of tables, we use the second measure of a. In specification (3), the coefficients on the tax rate differential variable is 0.041 in 3.16 and 0.039 in 3.17, and both are significant at the 1% level. In specification (4), the coefficient on the tax rate differential variable is not statistically significantly different than 0 in both of the aforementioned tables.

3.7.2 Interpretation of results

In the main specification, the coefficient of interest, $\hat{\gamma}$ can be interpreted as the semi-elasticity of the tax variable on pre-tax reported profit, or the effect of a one-percentage point increase in the composite tax variable on the percentage change in pre-tax reported profits. Of the four iterations of specification (4) in the main estimation, only one had a statistically significant results, where $\hat{\gamma}$ is equal to 1.744, meaning that a 10 percentage point increase in the composite tax variable (evaluated at the sample mean) would increase the reported

pre-tax profits in country i by 17.44%. While the other three iterations were not statistically significantly different than a null effect, the sign of the coefficients were at least the right direction.

3.8 Discussion

As stated in Section 3.7, in my main estimation in Table 3.10, I largely failed to reject the null hypothesis. In other words, I did not find statistically significant evidence that European oil firms were practicing profit shifting in response to changes corporate tax rates. The main caveat when interpreting these results are the challenges with the financial data in the ORBIS database.

First is potential measurement error in the independent variables, especially in measuring wages, assets, and other labor and capital data. If the data are measured imperfectly, then noise in the independent variables could attenuate the coefficient towards zero.

Secondly, the estimation strategy relies on obtaining information on the network structure between parent companies and their many subsidiaries, as well as which countries these companies are incorporated. Given that these firms have a financial incentive in reducing the transparency of their corporate network structure, it is plausible that the ORBIS database is either missing financial data on many subsidiaries, or that the subsidiaries are listed as being owned by an owner whose name is different than its true parent company (i.e. by a shell company). If that is the case, that it is plausible that the semi-elasticity that I estimated is only measured off of firms who have self-selected into the ORBIS database. These could be firms who do not engage in profit shifting, and thus attenuate the coefficient towards zero.

3.8.1 Future extension

In future extensions, I will incorporate remotely sensed oil flares data to estimate the semi-elasticity of true profits with respect to changes in corporate tax rates. The remote sensing approach relies on the geophysical process of oil production. When crude oil is extracted, natural gas that is naturally mixed into the oil at fairly constant proportions is released. In many parts of the world, firms do not have the ability to capture the natural gas to sell at the market. However, both because governments often curtail the venting of the methane gas and because venting poses a safety hazard to well operators, the oil producers choose to flare the natural gas (see Section 1.2 for additional context). Therefore, for regions where this is the case, a remotely detected observation of gas flares is a reliable indication of oil extraction and production.

As part of this extension, a crucial component of this analysis rests on the ability to estimate the relationship between volume of gas flared, and the quantity of crude oil extracted. This ratio will be region specific. In order to groundtruth the oil production, I will utilize Rystad Energy's UCube database, which contains field level oil production data.

The benefit of establishing a measure of oil production using remotely sensed data, instead of relying on the Rystad Energy UCube dataset is that (1) the database is subscription based, and (2) the production data are from reported sources. If I am able to establish

a reliable measure of field level oil production data, then I can disseminate the dataset that wouldn't require other researchers to expend funds to purchase the data. Secondly, while the production data that Rystad have are presumably reliable, it would be worth analzying whether there are discrepancies between reported production data and remotely sensed measures of oil production. Lastly, the ability to use remotely sensed data to ascertain oil production is very useful in countries where reporting may be less accurate, such as the estimation of oil production in ISIS controlled territories (Do et al. 2017).

Once I have established the relationship between radiant heat and oil production, then I will subset the oil flares by the oil lease concession polygon shapefiles (OpenOil 2018), allowing me to sum up the quantity of oil production (in barrels) at the concession by year level. Since the lease boundary shapefile contains the firm who has the rights to the lease, I will be able to connect the firm's oil production with the project level payments they make to the oil producing country from the country-by-country payments data.

Next, after I have estimated concession level oil production, I'll need to estimate concession level costs so that I can construct a measure of concession-level true profits. As a first pass, I plan on using Rystad Energy UCube databases' data on capital, operational, production, transportation and other costs. In the future, I'd like to determine other methods of imputing costs that do not rely on data from Rystad Energy.

One potential way of computing costs would be to estimate transportation costs by how proximal the oil concession is to other oil infrastructure, such as pipelines or refineries. This can be done by overlaying the oil concessions with the infrastructure shapefiles the GOGI Database (Rose et al. 2018).

Lastly, after I have constructed the estimate of oil production, then I will rerun the same specifications from Section 6, but using the log of true oil profits in place of log of earnings before interest and taxes. The result from this regression will indicate how much the firm's true economic activity changes with respect to changes in the tax incentive. The difference between this result and the result from above will indicate the level of profit shifting by oil companies.

3.9 Conclusion

The results from my main regression model estimating the semi-elasticity of pre-tax reported profits on measures of capital, labor, and total factor productivity indicate that there is perhaps a null effect or an effect of 17% of income shifted for a 10 percentage point increase in the tax incentive (evaluated at the sample mean). While I fail to reject the null hypothesis in the main model, the estimates are statistically significant at the 1% and 5% level in my robustness check, where the tax incentive variable was constructed as the simple difference between the corporate tax rate in the parent and affiliate country. This indicates that weighted corporate tax rates of other affiliates in other countries do not have a significant effect on the reported profits of affiliate i.

Another possible explanation of the null effect estimated in the main specification is that if the multinational corporation is shifting profits to shell companies located tax havens where the tax rates are very low or non-existent but financial data for the shell companies

are not available in the ORBIS database, then the way the composite tax variable C is constructed would not capture the multinational corporation's profit shifting incentive.

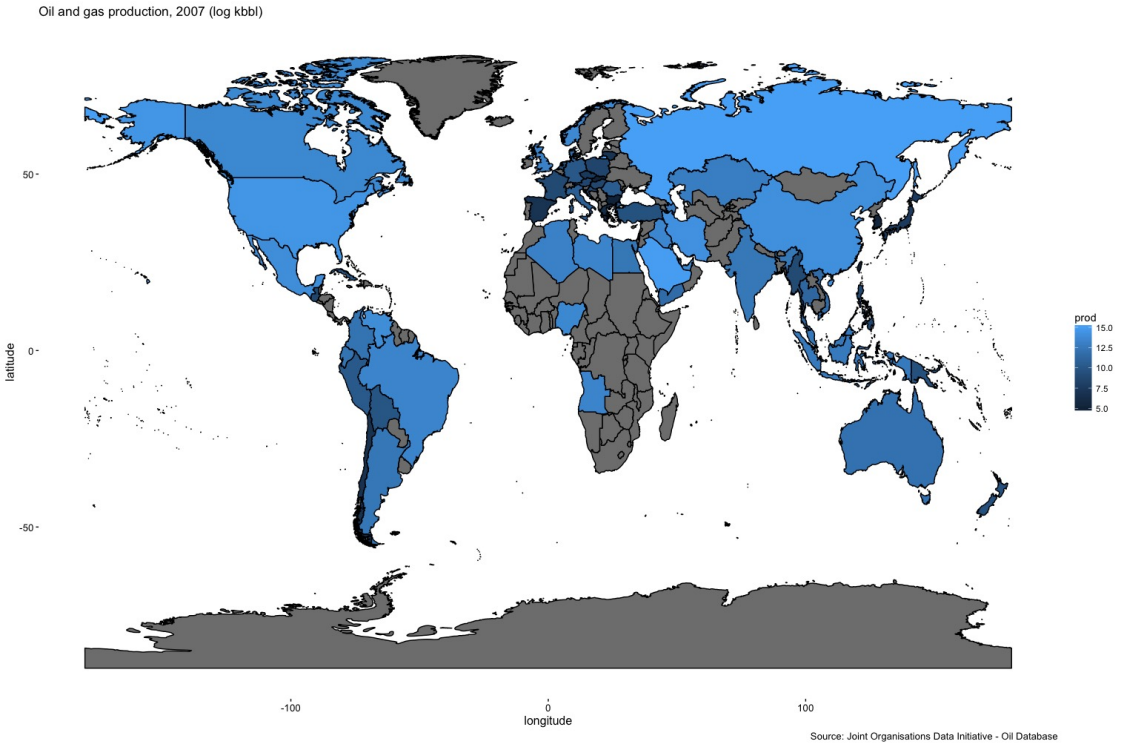


Figure 3.1: Log oil production (kbbd) in 2007

Figures

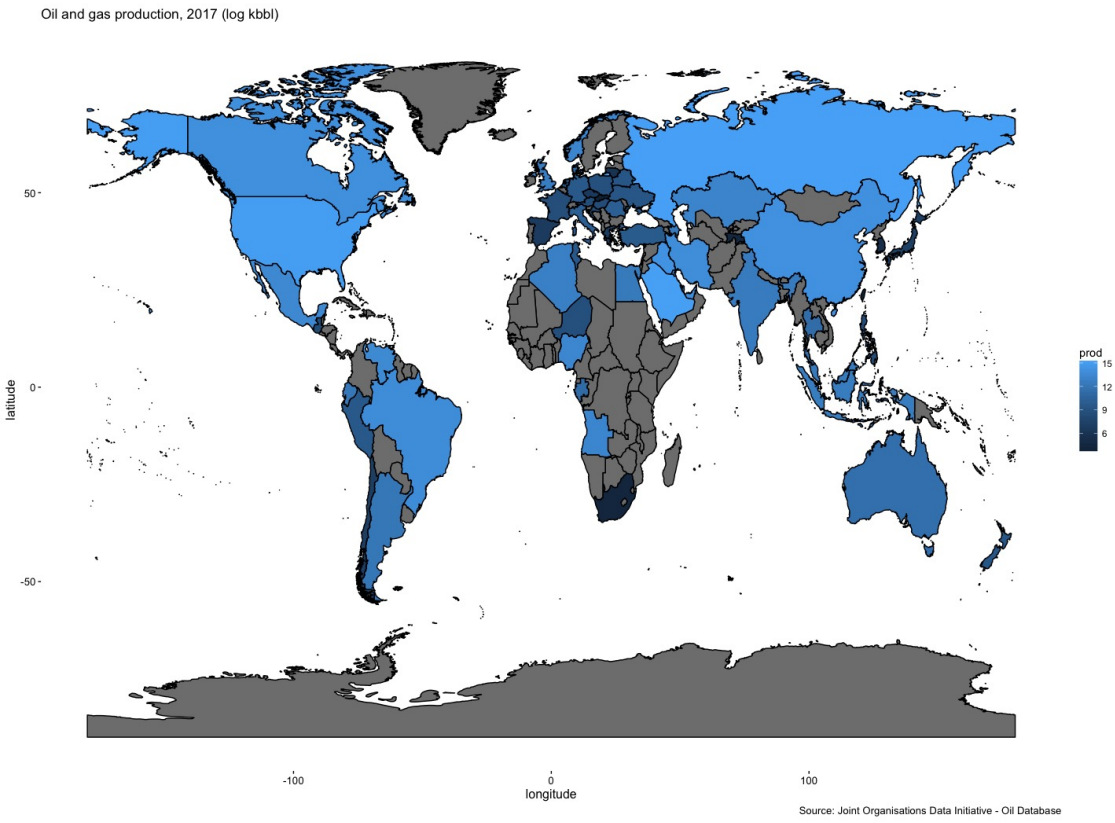


Figure 3.2: Log oil production (kbbd) in 2017

Figure 3.3: Log oil production against GDP per capita, 2016

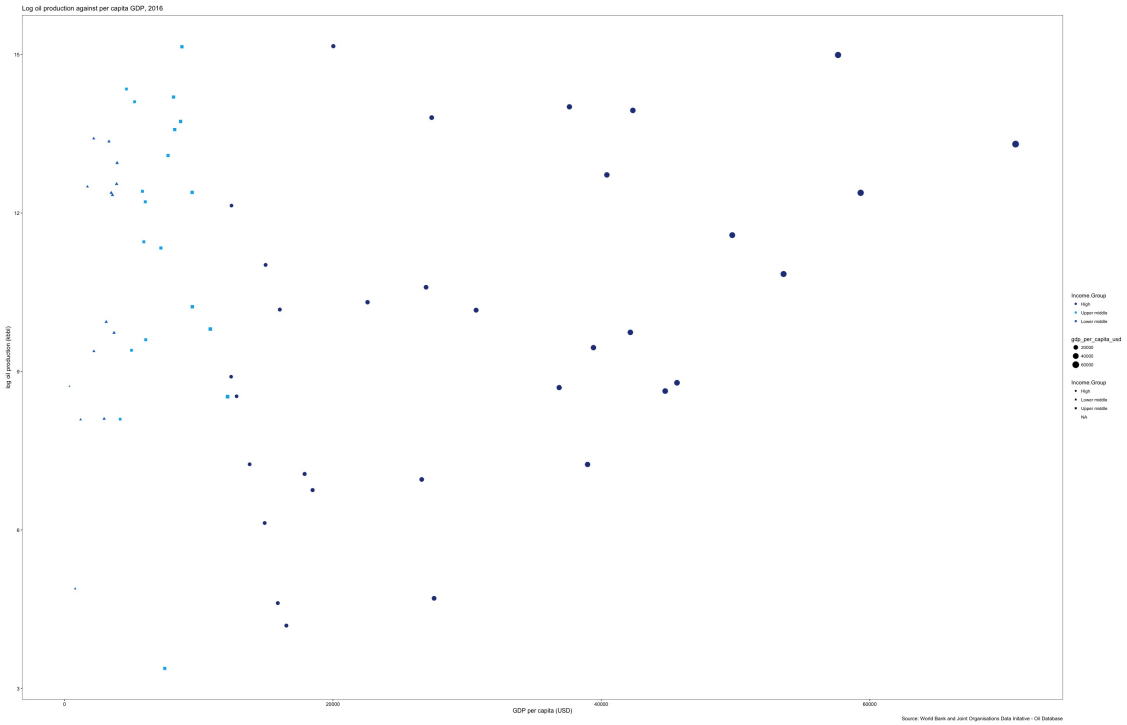


Figure 3.4: Oil and Gas Tax Regimes in 2017

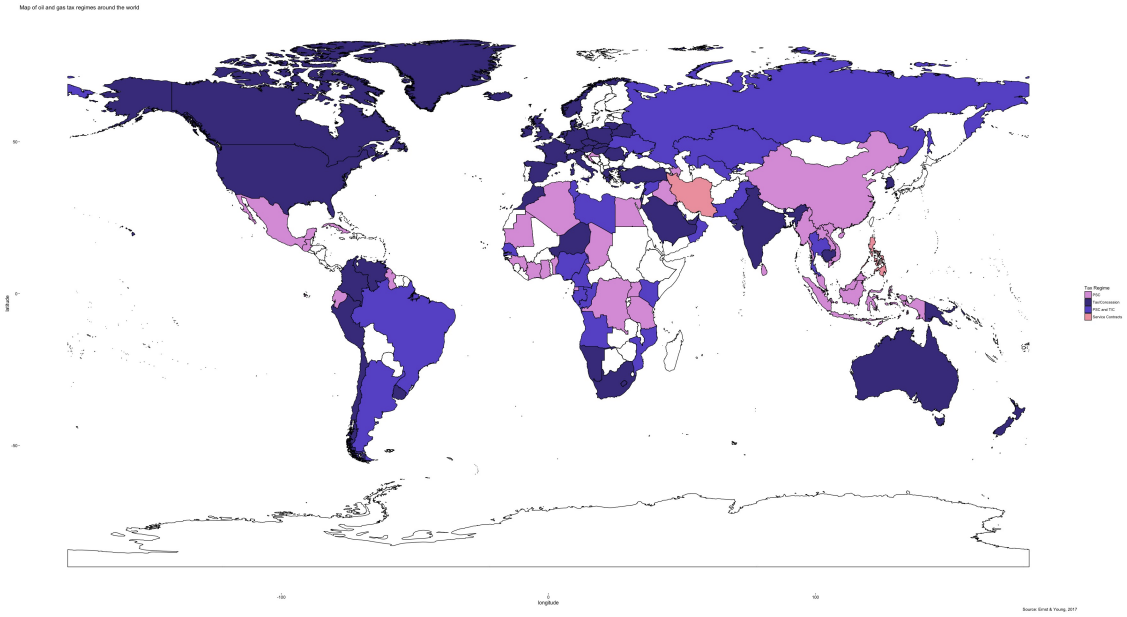


Figure 3.5: Oil production in 2017 vs. Rule of Law Index

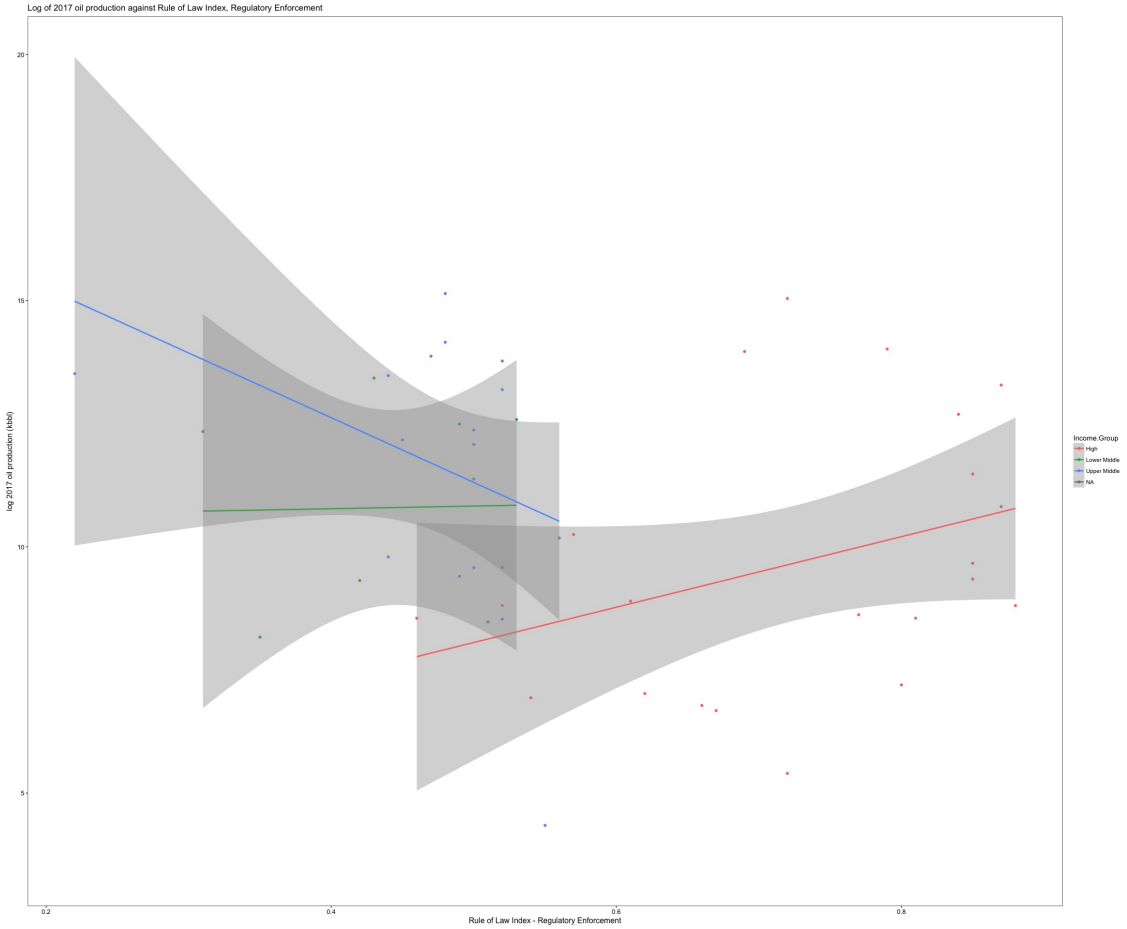


Figure 3.6: Oil production in 2017 vs. Rule of Law Index

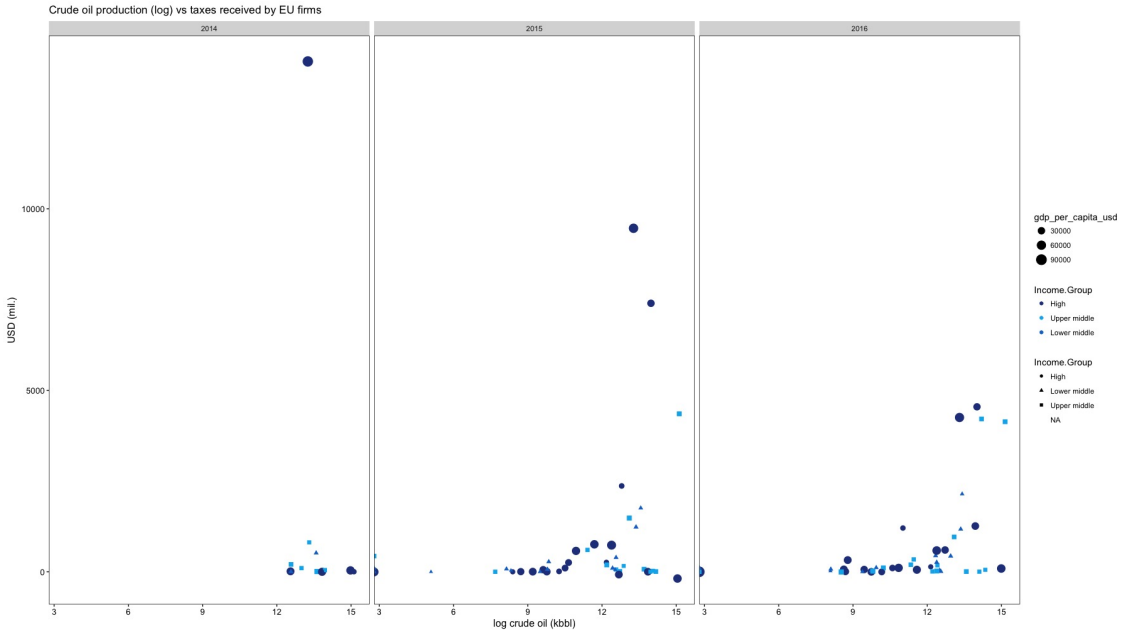
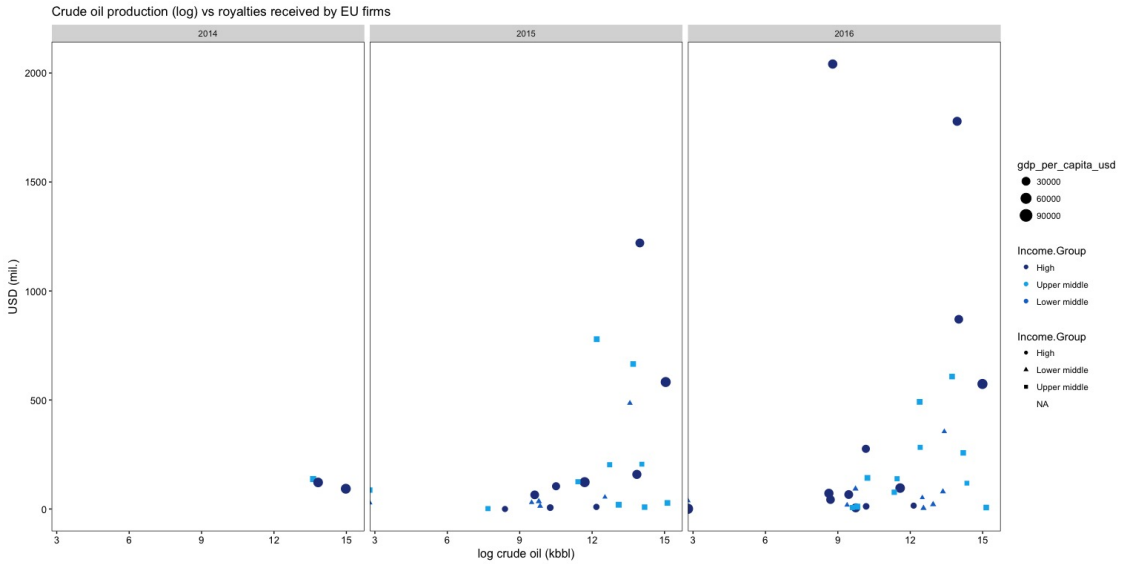


Figure 3.7: Oil production in 2017 vs. Rule of Law Index



Tables

Table 3.1: Summary statistics of all firms

	mean	median	sd	min	max	number
EBIT	10.72	10.63	3.12	0.00	18.01	2931
Capital	12.23	12.47	3.87	0.00	19.61	2931
Cost of Emp.	10.30	10.06	2.81	0.69	16.66	1669
Num. of Emp.	6.40	6.31	2.86	0.00	11.72	1735
C	-0.01	-0.00	0.08	-0.29	0.38	2931

Table 3.2: Summary statistics of parent companies

	mean	median	sd	min	max	number
EBIT	13.36	13.33	2.35	3.09	18.01	649
Capital	15.64	15.64	2.39	6.77	19.60	649
Cost of Emp.	12.85	13.59	2.34	6.85	16.61	1669
Num. of Emp.	9.07	9.72	2.16	0.69	11.59	1735
C	0.00	0.00	0.02	-0.29	0.08	649

Table 3.3: Summary statistics of subsidiaries

	mean	median	sd	min	max	number
EBIT	9.97	9.93	2.89	0.00	17.26	2282
Capital	11.26	11.55	3.66	0.00	19.61	2282
Cost of Emp.	9.24	9.24	2.24	2.20	14.98	1669
Num. of Emp.	5.22	5.16		0	11.37	1735
C	-0.01	-0.00	0.09	-0.26	0.38	2282

Table 3.4: Number of countries MNCs have subsidiaries in

mean	median	min	max	sd
4.33	2.00	1.00	33.00	5.92

Table 3.5: Number of MNCs by year

year	# MNCs
2008	47
2009	53
2010	57
2011	58
2012	58
2013	63
2014	59
2015	48
2016	47
2017	15

Table 3.6: Number of countries by year

year	# countries
2008	48
2009	50
2010	54
2011	53
2012	56
2013	55
2014	56
2015	52
2016	50
2017	23

Table 3.7: Number of countries where parent companies are incorporated

year	# countries
2008	20
2009	22
2010	25
2011	26
2012	27
2013	24
2014	21
2015	20
2016	20
2017	10

Table 3.8: Number of countries where subsidiaries are incorporated

year	# Countries
2008	43
2009	46
2010	49
2011	47
2012	51
2013	49
2014	51
2015	47
2016	47
2017	18

Table 3.9: Summary Statistics of Statutory Corporate Tax Rates

year	mean	median	min	max
2008	27.31	29.60	0.00	40.69
2009	26.39	28.00	0.00	40.69
2010	26.47	28.00	0.00	40.69
2011	26.46	26.00	0.00	40.69
2012	25.77	25.00	0.00	38.01
2013	25.66	25.00	0.00	38.01
2014	24.72	25.00	0.00	35.64
2015	24.87	25.00	0.00	35.00
2016	23.99	25.00	0.00	35.00
2017	24.74	24.00	9.00	55.00

Table 3.10: Estimation of the profit shifting equation

	(1)	(2)	(3)	(4)
Labor	0.310^{***}	0.315^{***}	0.410^{***}	0.315^{***}
	(0.018)	(0.024)	(0.040)	(0.024)
Capital	0.528^{***}	0.484^{***}	0.344^{***}	0.484^{***}
	(0.013)	(0.018)	(0.027)	(0.018)
Per capita income	-0.017	0.531	0.911^{***}	0.542^{*}
	(0.034)	(0.386)	(0.314)	(0.310)
C	-1.564^{***}	-1.599	-0.980	-1.744^{*}
	(0.402)	(1.066)	(0.898)	(1.057)
MNC dummies	N	Y	N	Y
Year dummies	Y	Y	Y	N
Country dummies	N	Y	N	N
MNC-country dummies	N	N	Y	N
Country-year dummies	N	N	N	Y
Observations	1,607	1,607	1,607	1,607
R^2	0.819	0.869	0.927	0.869
Adjusted R^2	0.817	0.857	0.913	0.857

Note: *p<0.1; **p<0.05; ***p<0.01
The dependent variable is log of EBIT.
Labor is log of total labor compensation.

Table 3.11: Estimation of the profit shifting equation

	(1)	(2)	(3)	(4)
Labor	0.331***	0.426***	0.408***	0.428***
	(0.020)	(0.025)	(0.036)	(0.025)
Capital	0.517***	0.415***	0.384***	0.415***
	(0.014)	(0.020)	(0.028)	(0.019)
Per capita income	0.122***	0.296	0.243	0.428*
	(0.039)	(0.275)	(0.246)	(0.238)
C	−0.561	0.135	−0.286	0.099
	(0.416)	(1.029)	(0.933)	(1.026)
MNC dummies	N	Y	N	Y
Year dummies	Y	Y	Y	N
Country dummies	N	Y	N	N
MNC-country dummies	N	N	Y	N
Country-year dummies	N	N	N	Y
Observations	1,644	1,644	1,644	1,644
R^2	0.834	0.890	0.933	0.890
Adjusted R^2	0.833	0.881	0.919	0.881

Note: *p<0.1; **p<0.05; ***p<0.01
The dependent variable is log of EBIT.
Labor is log of number of employees.

Table 3.12: Estimation of the profit shifting equation

	(1)	(2)	(3)	(4)
Labor	0.318***	0.321***	0.431***	0.323***
	(0.022)	(0.028)	(0.047)	(0.027)
Capital	0.532***	0.483***	0.434***	0.483***
	(0.015)	(0.021)	(0.034)	(0.021)
TFP1	0.00001***	0.00001***	0.00000	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
TFP2	−0.00000***	−0.00000***	−0.00000	−0.00000***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
C	−1.247***	−0.594	−0.582	−0.523
	(0.460)	(1.121)	(1.013)	(1.114)
MNC dummies	N	Y	N	Y
Year dummies	Y	Y	Y	N
Country dummies	N	Y	N	N
MNC-country dummies	N	N	Y	N
Country-year dummies	N	N	N	Y
Observations	1,288	1,288	1,288	1,288
R^2	0.836	0.892	0.930	0.892
Adjusted R^2	0.835	0.881	0.916	0.881

Note: *p<0.1; **p<0.05; ***p<0.01
The dependent variable is log of EBIT.
Labor is log of total labor compensation.

Table 3.13: Estimation of the profit shifting equation

	(1)	(2)	(3)	(4)
Labor	0.382***	0.485***	0.557***	0.489***
	(0.023)	(0.033)	(0.049)	(0.033)
Capital	0.477***	0.363***	0.356***	0.361***
	(0.017)	(0.025)	(0.036)	(0.025)
TFP1	0.0002***	0.0002***	0.0001***	0.0002***
	(0.00001)	(0.00002)	(0.00002)	(0.00002)
TFP2	−0.00000***	−0.00000**	0.00000	−0.00000**
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
C	0.168	−0.317	−0.681	−0.210
	(0.447)	(1.073)	(1.000)	(1.066)
MNC dummies	N	Y	N	Y
Year dummies	Y	Y	Y	N
Country dummies	N	Y	N	N
MNC-country dummies	N	N	Y	N
Country-year dummies	N	N	N	Y
Observations	1,270	1,270	1,270	1,270
R^2	0.848	0.901	0.932	0.900
Adjusted R^2	0.846	0.890	0.917	0.890

Note: *p<0.1; **p<0.05; ***p<0.01
The dependent variable is log of EBIT.
Labor is log of number of employees.

Table 3.14: Estimation of the profit shifting equation, using difference between parent and subsidiary tax rates

	(1)	(2)	(3)	(4)
Labor	0.314***	0.287***	0.489***	0.289***
	(0.024)	(0.031)	(0.053)	(0.031)
Capital	0.467***	0.402***	0.220***	0.402***
	(0.016)	(0.022)	(0.032)	(0.022)
Per capita income	0.042	1.187**	1.579***	0.737*
	(0.044)	(0.547)	(0.422)	(0.424)
Tax Rate Diff.	0.016***	0.040**	0.048***	0.040**
	(0.005)	(0.017)	(0.014)	(0.017)
MNC dummies	N	Y	N	Y
Year dummies	Y	Y	Y	N
Country dummies	N	Y	N	N
MNC-country dummies	N	N	Y	N
Country-year dummies	N	N	N	Y
Observations	1,068	1,068	1,068	1,068
R^2	0.695	0.790	0.899	0.789
Adjusted R^2	0.691	0.764	0.873	0.765

Note: *p<0.1; **p<0.05; ***p<0.01
The dependent variable is log of EBIT.
Labor is log of total labor compensation.

Table 3.15: Estimation of the profit shifting equation, using difference between parent and subsidiary tax rates

	(1)	(2)	(3)	(4)
Labor	0.318***	0.331***	0.253***	0.333***
	(0.027)	(0.036)	(0.047)	(0.035)
Capital	0.476***	0.399***	0.339***	0.397***
	(0.017)	(0.023)	(0.032)	(0.023)
Per capita income	0.195***	0.337	0.475	0.409
	(0.052)	(0.344)	(0.299)	(0.300)
Tax Rate Diff.	0.005	0.009	0.030**	0.008
	(0.005)	(0.016)	(0.013)	(0.016)
MNC dummies	N	Y	N	Y
Year dummies	Y	Y	Y	N
Country dummies	N	Y	N	N
MNC-country dummies	N	N	Y	N
Country-year dummies	N	N	N	Y
Observations	1,067	1,067	1,067	1,067
R^2	0.710	0.822	0.902	0.821
Adjusted R^2	0.707	0.799	0.873	0.800

Note: *p<0.1; **p<0.05; ***p<0.01
The dependent variable is log of EBIT.
Labor is log of number of employees.

Table 3.16: Estimation of the profit shifting equation, using difference between parent and subsidiary tax rates

	(1)	(2)	(3)	(4)
Labor	0.331***	0.255***	0.380***	0.258***
	(0.031)	(0.037)	(0.067)	(0.037)
Capital	0.465***	0.425***	0.315***	0.423***
	(0.019)	(0.026)	(0.041)	(0.026)
TFP1	0.00001***	0.00001***	0.00000	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
TFP2	−0.00000***	−0.00000***	−0.00000	−0.00000***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Tax Rate Diff.	0.010*	0.024	0.041***	0.022
	(0.006)	(0.017)	(0.015)	(0.017)
MNC dummies	N	Y	N	Y
Year dummies	Y	Y	Y	N
Country dummies	N	Y	N	N
MNC-country dummies	N	N	Y	N
Country-year dummies	N	N	N	Y
Observations	806	806	806	806
R^2	0.711	0.822	0.898	0.821
Adjusted R^2	0.706	0.796	0.869	0.797

Note: *p<0.1; **p<0.05; ***p<0.01
The dependent variable is log of EBIT.
Labor is log of total labor compensation.

Table 3.17: Estimation of the profit shifting equation, using difference between parent and subsidiary tax rates

	(1)	(2)	(3)	(4)
Labor	0.351***	0.378***	0.380***	0.382***
	(0.034)	(0.047)	(0.071)	(0.047)
Capital	0.438***	0.344***	0.327***	0.342***
	(0.020)	(0.030)	(0.041)	(0.030)
TFP1	0.0002***	0.0001***	0.0001***	0.0002***
	(0.00002)	(0.00002)	(0.00002)	(0.00002)
TFP2	−0.00000***	−0.00000***	0.00000	−0.00000***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Tax Rate Diff.	−0.005	0.018	0.039***	0.016
	(0.006)	(0.017)	(0.015)	(0.016)
MNC dummies	N	Y	N	Y
Year dummies	Y	Y	Y	N
Country dummies	N	Y	N	N
MNC-country dummies	N	N	Y	N
Country-year dummies	N	N	N	Y
Observations	788	788	788	788
R^2	0.726	0.833	0.894	0.832
Adjusted R^2	0.721	0.808	0.863	0.809

Note: *p<0.1; **p<0.05; ***p<0.01
The dependent variable is log of EBIT.
Labor is log of number of employees.

References

(EPA), US Environmental Protection Agency. 2015. *Nitrogen and Phosphorus Pollution in the Mississippi River Basin: Nitrogen and Phosphorus Pollution in the Mississippi River Basin: Findings of the Wadeable Streams Assessment.* Online, March.

Agerton, Mark. 2020. "Learning where to drill: Drilling decisions and geological quality in the haynesville shale."

Agerton, Mark, Ben Gilbert, and Gregory B Upton Jr. 2020. "The Economics of Natural Gas Flaring in US Shale: An Agenda for Research and Policy." PhD diss., Rice University.

Alexander, Richard B, Richard A Smith, Gregory E Schwarz, Elizabeth W Boyer, Jacqueline V Nolan, and John W Brakebill. 2008. "Differences in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River Basin." *Environmental science & technology* 42 (3): 822–830.

Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman. 2019. "Tax evasion and inequality." *American Economic Review* 109 (6): 2073–2103.

Alvaredo, Facundo, Anthony B Atkinson, and Salvatore Morelli. 2017. "Top wealth shares in the UK over more than a century."

Anchondo, Carlos. 2019. *Gas Flare Blips Are World's Biggest Methane Source—Report.*

Anderson, Soren T, Ryan Kellogg, and Stephen W Salant. 2018. "Hotelling under pressure." *Journal of Political Economy* 126 (3): 984–1026.

Arps, Jan J, et al. 1945. "Analysis of decline curves." *Transactions of the AIME* 160 (01): 228–247.

Bachas, Pierre, and Mauricio Soto. 2015. "Not (ch) your average tax system: corporate taxation under weak enforcement." *Unpublished manuscript.*

Bartik, Alexander W, Janet Currie, Michael Greenstone, and Christopher R Knittel. 2019. "The local economic and welfare consequences of hydraulic fracturing." *American Economic Journal: Applied Economics* 11 (4): 105–55.

Bertrand, Marianne, Paras Mehta, and Sendhil Mullainathan. 2002. "Ferreting out tunneling: An application to Indian business groups." *The Quarterly Journal of Economics* 117 (1): 121–148.

Besley, Timothy, and Torsten Persson. 2013. "Taxation and development." In *Handbook of public economics,* 5:51–110. Elsevier.

Best Practices for Submitting Nutrient Data to the Water Quality eXchange (WQX). https://www.epa.gov/sites/production/files/2017-06/documents/wqx_nutrient_best_practices_guide.pdf. Published: 2017-06-02.

Blundell, Wesley, and T Kokoza. 2018. "Natural gas flaring, respiratory health, and distributional effects."

Boomhower, Judson. 2019. "Drilling like there's no tomorrow: Bankruptcy, insurance, and environmental risk." *American Economic Review* 109 (2): 391–426.

Breitburg, Denise, Lisa A Levin, Andreas Oschlies, Marilaure Grégoire, Francisco P Chavez, Daniel J Conley, Véronique Garçon, Denis Gilbert, Dimitri Gutiérrez, Kirsten Isensee, et al. 2018. "Declining oxygen in the global ocean and coastal waters." *Science* 359 (6371).

Broussard, Whitney, and R Eugene Turner. 2007. "Agriculture and the Mississippi River Basin."

Bruckner, Monica. 2019. "The Gulf of Mexico Dead Zone." https://serc.carleton.edu/microbelife/topics/deadzone/index.html.

Cao, C, X Xiong, R Wolfe, F DeLuccia, Q Liu, S Blonski, G Lin, M Nishihama, D Pogorzala, H Oudrari, et al. 2013. "Visible Infrared Imaging Radiometer Suite (VIIRS) Sensor Data Record (SDR) User's Guide Version 1.2." *NOAA technical report NESDIS* 142:43.

Clausing, Kimberly A. 2003. "Tax-motivated transfer pricing and US intrafirm trade prices." *Journal of Public Economics* 87 (9-10): 2207–2223.

Commission, European. 2018. "Public country-by-country reporting." https://ec.europa.eu/info/business-economy-euro/company-reporting-and-auditing/company-reporting/public-country-country-reporting_en.

Covert, Thomas. 2015. "Experiential and social learning in firms: the case of hydraulic fracturing in the Bakken Shale." *Available at SSRN 2481321.*

Currie, Janet, Michael Greenstone, and Katherine Meckel. 2017. "Hydraulic fracturing and infant health: New evidence from Pennsylvania." *Science advances* 3 (12): e1603021.

Cushing, Lara J, Kate Vavra-Musser, Khang Chau, Meredith Franklin, and Jill E Johnston. 2020. "Flaring from Unconventional Oil and Gas Development and Birth Outcomes in the Eagle Ford Shale in South Texas." *Environmental health perspectives* 128 (7): 077003.

De Marchi, Scott, and James T Hamilton. 2006. "Assessing the accuracy of self-reported data: an evaluation of the toxics release inventory." *Journal of Risk and uncertainty* 32 (1): 57–76.

Dharmapala, Dhammika. 2014. "What do we know about base erosion and profit shifting? A review of the empirical literature." *Fiscal Studies* 35 (4): 421–448.

Dharmapala, Dhammika, and Nadine Riedel. 2013. "Earnings shocks and tax-motivated income-shifting: Evidence from European multinationals." *Journal of Public Economics* 97:95–107.

Diaz, Robert J, and Rutger Rosenberg. 2008. "Spreading dead zones and consequences for marine ecosystems." *science* 321 (5891): 926–929.

Do, Quy-Toan, Jacob N Shapiro, Christopher D Elvidge, Mohamed Abdel Jelil, Daniel Ahn, Kimberly Baugh, Jamie Hansen-Lewis, and Mikhail Zhizhin. 2017. "How much oil is the Islamic state group producing? Evidence from remote sensing." Technical report. The World Bank.

Do, Quy-Toan, Jacob N Shapiro, Christopher D Elvidge, Mohamed Abdel-Jelil, Daniel P Ahn, Kimberly Baugh, Jamie Hansen-Lewis, Mikhail Zhizhin, and Morgan D Bazilian. 2018. "Terrorism, geopolitics, and oil security: Using remote sensing to estimate oil production of the Islamic State." *Energy research & social science* 44:411–418.

Egwurugwu, Jude Nnabuife, and Arthur Nwafor. 2013. "Prolonged exposure to oil and gas flares ups the risks for hypertension." *American Journal of Health Research* 1 (3): 65–72.

Elliott, AH, RB Alexander, GE Schwarz, Ude Shankar, JPS Sukias, and Graham B McBride. 2005. "Estimation of nutrient sources and transport for New Zealand using the hybrid mechanistic-statistical model SPARROW." *Journal of Hydrology (New Zealand),* 1–27.

Elvidge, Christopher D, Morgan D Bazilian, Mikhail Zhizhin, Tilottama Ghosh, Kimberly Baugh, and Feng-Chi Hsu. 2018. "The potential role of natural gas flaring in meeting greenhouse gas mitigation targets." *Energy strategy reviews* 20:156–162.

Elvidge, Christopher D, Mikhail Zhizhin, Kimberly Baugh, Feng Chi Hsu, and Tilottama Ghosh. 2019. "Extending nighttime combustion source detection limits with short wavelength VIIRS data." *Remote Sensing* 11 (4): 395.

Elvidge, Christopher D, Mikhail Zhizhin, Kimberly Baugh, Feng-Chi Hsu, Tilottama Ghosh, et al. 2015. "Methods for global survey of natural gas flaring from visible infrared imaging radiometer suite data." *Energies* 9 (1): 1–15.

Environmental Protection Agency. 2014. "Load Estimation from Monitoring Data." Technical report.

EY. 2017. "Global oil and gas tax guide." Technical report. Ernst and Young.

Fetter, T Robert, et al. 2018. "Fracking, toxics, and disclosure."

Feyrer, James, Erin T Mansur, and Bruce Sacerdote. 2017. "Geographic dispersion of economic shocks: Evidence from the fracking revolution." *American Economic Review* 107 (4): 1313–34.

Fisman, Raymond, Peter Moustakerski, and Shang-Jin Wei. 2008. "Outsourcing tariff evasion: A new explanation for entrepôt trade." *The Review of Economics and Statistics* 90 (3): 587–592.

Fisman, Raymond, and Shang-Jin Wei. 2004. "Tax rates and tax evasion: evidence from "missing imports" in China." *Journal of political Economy* 112 (2): 471–496.

Fitzgerald, Timothy. 2015. "Experiential gains with a new technology: An empirical investigation of hydraulic fracturing." *Agricultural and Resource Economics Review* 44 (1203-2016-95583): 83–105.

Fowlie, Meredith, Edward Rubin, and Reed Walker. 2019. "Bringing satellite-based air quality estimates down to earth." In *AEA Papers and Proceedings,* 109:283–88.

Fu, Baihua, Wendy S Merritt, Barry FW Croke, Tony R Weber, and Anthony J Jakeman. 2019. "A review of catchment-scale water quality and erosion models and a synthesis of future prospects." *Environmental modelling & software* 114:75–97.

Garcia, Ana Maria, Anne B Hoos, and Silvia Terziotti. 2011. "A Regional Modeling Framework of Phosphorus Sources and Transport in Streams of the Southeastern United States 1." *JAWRA Journal of the American Water Resources Association* 47 (5): 991–1010.

Garcıa, Ana Marıa, Richard B Alexander, Jeffrey G Arnold, Lee Norfleet, Michael J White, Dale M Robertson, and Gregory Schwarz. 2016. "Regional effects of agricultural conservation practices on nutrient transport in the Upper Mississippi River Basin." *Environmental science & technology* 50 (13): 6991–7000.

Gelman, Andrew. 2008. "Scaling regression inputs by dividing by two standard deviations." *Statistics in medicine* 27 (15): 2865–2873.

Gopalakrishnan, Sathya, and H Allen Klaiber. 2014. "Is the shale energy boom a bust for nearby residents? Evidence from housing values in Pennsylvania." *American Journal of Agricultural Economics* 96 (1): 43–66.

Grainger, Corbett, Andrew Schreiber, and Wonjun Chang. 2016. "How states comply with federal regulations: strategic ambient pollution monitoring." Technical report. Working paper, University of Wisconsin-Madison.

Gray, Jim, David T Liu, Maria Nieto-Santisteban, Alex Szalay, David J DeWitt, and Gerd Heber. 2005. "Scientific data management in the coming decade." *Acm Sigmod Record* 34 (4): 34–41.

Gray, Wayne B, and Jay P Shimshack. 2011. "The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence." *Review of Environmental Economics and Policy* 5 (1): 3–24.

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman. 2020. "Tax Evasion by the Wealthy: Measurement and Implications." In *Measuring and Understanding the Distribution and Intra/Inter-Generational Mobility of Income and Wealth.* University of Chicago Press.

Gvakharia, Alexander, Eric A Kort, Adam Brandt, Jeff Peischl, Thomas B Ryerson, Joshua P Schwarz, Mackenzie L Smith, and Colm Sweeney. 2017. "Methane, black carbon, and ethane emissions from natural gas flares in the Bakken Shale, North Dakota." *Environmental Science & Technology* 51 (9): 5317–5325.

Al-Hamdan, Mohammad Z, Richard A Smith, Anne Hoos, Gregory E Schwarz, Richard B Alexander, William L Crosson, Jayanthi Srikishen, Maurice Estes Jr, James Cruise, Ashraf Al-Hamdan, et al. 2017. "Developing a Dynamic SPARROW Water Quality Decision Support System Using NASA Remotely-Sensed Products." In *AGU Fall Meeting Abstracts,* vol. 2017, EP11C–1574.

Hausman, Catherine, and Ryan Kellogg. 2015. "Welfare and distributional implications of shale gas." Technical report. National Bureau of Economic Research.

He, Guojun, Shaoda Wang, and Bing Zhang. 2020. "Watering down environmental regulation in China." *The Quarterly Journal of Economics* 135 (4): 2135–2185.

Heckemeyer, Jost, and Michael Overesch. 2013. "Multinationals' profit response to tax differentials: Effect size and shifting channels."

Hill, Elaine, and Lala Ma. 2017. "Shale gas development and drinking water quality." *American Economic Review* 107 (5): 522–25.

Hill, Elaine L. 2018. "Shale gas development and infant health: evidence from Pennsylvania." *Journal of health economics* 61:134–150.

Hines, James R., and Eric M. Rice. 1994. "Fiscal paradise: Foreign tax havens and American business." *The Quarterly Journal of Economics* 109 (1): 149–182.

Howard, Brian Clark. 2014. "Mississippi Basin Water Quality Declining Despite Conservation," April. https://www.nationalgeographic.com/news/2014/4/140411-water-quality-nutrients-pesticides-dead-zones-science/.

Huizinga, Harry, and Luc Laeven. 2008. "International profit shifting within multinationals: A multi-country perspective." *Journal of Public Economics* 92 (5-6): 1164–1182.

Jacobsen, Grant D. 2019. "Who wins in an energy boom? Evidence from wage rates and housing." *Economic Inquiry* 57 (1): 9–32.

JODI. 2018. "The Joint Oil Data Initiative-Oil World Database." https://www.jodidata.org/oil/.

Johannesen, Niels, and Dan Thor Larsen. 2016. "The power of financial transparency: An event study of country-by-country reporting standards." *Economics Letters* 145:120–122.

Johannesen, Niels, Thomas Torslov, and Ludvig Wier. 2016. "Are less developed countries more exposed to multinational tax avoidance? Method and evidence from micro-data." Technical report. WIDER Working Paper.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng. 2008. "Unobserved actions of mutual funds." *The Review of Financial Studies* 21 (6): 2379–2416.

Kanter, David R, Olivia Chodos, Olivia Nordland, Mallory Rutigliano, and Wilfried Winiwarter. 2020. "Gaps and opportunities in nitrogen pollution policies around the world." *Nature Sustainability* 3 (11): 956–963.

Karplus, Valerie J, Shuang Zhang, and Douglas Almond. 2018. "Quantifying coal power plant responses to tighter SO2 emissions standards in China." *Proceedings of the National Academy of Sciences* 115 (27): 7004–7009.

Keiser, David A, and Joseph S Shapiro. 2019. "Consequences of the Clean Water Act and the demand for water quality." *The Quarterly Journal of Economics* 134 (1): 349–396.

Krasovich, Emma, Peiley Lau, Jeanette Tseng, Kendon Bell, and Solomon Hsiang. 2021. "United States Mississippi River Basin Nitrogen- and Phosphorous-Based Water Quality Concentration from 1980 to 2018."

Lade, Gabriel E, and Ivan Rudik. 2020. "Costs of inefficient regulation: Evidence from the Bakken." *Journal of Environmental Economics and Management,* 102336.

Lange, Ian, and Michael Redlinger. 2019. "Effects of stricter environmental regulations on resource development." *Journal of Environmental Economics and Management* 96:60–87.

Lapworth, DJ, Nicole Baran, ME Stuart, and RS Ward. 2012. "Emerging organic contaminants in groundwater: a review of sources, fate and occurrence." *Environmental pollution* 163:287–303.

Lee, Ruiwen. 2020. "Essays on the Regulation and Remote Sensing of Natural Gas Flaring." PhD diss., Columbia University.

Ljungqvist, Alexander, Christopher Malloy, and Felicia Marston. 2009. "Rewriting history." *The Journal of Finance* 64 (4): 1935–1960.

Lyon, David, Mark Omara, Ritesh Gautam, Kate Roberts, Beth Trask, Colin Leyden, Isabel Mogstad, Daniel Zavala-Araiza, and Steven Hamburg. 2020. "Environmental Defense Fund Permian Basin Campaign: a science and advocacy-based approach to quantify and mitigate methane emissions from the oil and gas industry." In *EGU General Assembly Conference Abstracts,* 11640.

Maffini, Giorgia, and Socrates Mokkas. 2011. "Profit shifting and measured productivity of multinational firms." *Oxford Bulletin of Economics and Statistics* 73 (1): 1–20.

Malik, Naureen. 2019. "Snuffed-Out Flares Are Biggest Methane Offender, Satellites Show." *Bloomberg. com.*

Ministry for Primary Industries Manatu Ahu Matua. 2021. "Protecting freshwater health." https://www.mpi.govt.nz/funding-rural-support/environment-and-natural-resources/protecting-freshwater-health/.

Ministry for the Environment. 2021. "Resource Management (Measurement and Reporting of Water Takes) Amendment Regulations 2020." Technical report.

Mishra, Prachi, Arvind Subramanian, and Petia Topalova. 2008. "Tariffs, enforcement, and customs evasion: Evidence from India." *Journal of public Economics* 92 (10-11): 1907–1925.

Muehlenbachs, Lucija. 2015. "A dynamic model of cleanup: Estimating sunk costs in oil and gas production." *International Economic Review* 56 (1): 155–185.

Muehlenbachs, Lucija, Elisheba Spiller, and Christopher Timmins. 2015. "The housing market impacts of shale gas development." *American Economic Review* 105 (12): 3633–59.

Myers, Donna N. 2015. "Foundations of water quality monitoring and assessment in the United States." In *Food, Energy, and Water,* 21–92. Elsevier.

National Park Service. 2021. "Mississippi River Facts." https://www.nps.gov/miss/riverfacts.htm.

"National Policy Statement for Fresh Water Management 2014." 2017. Technical report. Ministry for the Environment.

"North Dakota Industrial CommissionOrder 24665 Policy/Guidance Version 041718." 2014.

Oliva, Paulina. 2015. "Environmental regulations and corruption: Automobile emissions in Mexico City." *Journal of Political Economy* 123 (3): 686–724.

Olmstead, Sheila M, Lucija A Muehlenbachs, Jhih-Shyang Shih, Ziyan Chu, and Alan J Krupnick. 2013. "Shale gas development impacts on surface water quality in Pennsylvania." *Proceedings of the National Academy of Sciences* 110 (13): 4962–4967.

OpenOil. 2018. "Open Oil Concession Map." http://maps.openoil.net/concessions/.

"Our Freshwater Report 2020." 2020. Technical report. Ministry for the Environment.

Ponce, Alejandro. 2018. "WJP Rule of Law Index 2017-2018." Technical report. World Justice Project.

Potash and Phosphate Institute. 1999. "Phosphorus and the environment: Better Crops with Plant Food." 83 (1): 37–39.

Read, Emily K, Lindsay Carr, Laura De Cicco, Hilary A Dugan, Paul C Hanson, Julia A Hart, James Kreft, Jordan S Read, and Luke A Winslow. 2017. "Water quality data for national-scale aquatic research: The Water Quality Portal." *Water Resources Research* 53 (2): 1735–1745.

Regional Guidance on Handling Chemical Concentration Data Near the Detection Limit in Risk Assessments. https://www.epa.gov/risk/regional-guidance-handling-chemical-concentration-data-near-detection-limit-risk-assessments. Accessed: 2020-02-15.

Reimer, Adam P, Riva CH Denny, and Diana Stuart. 2018. "The impact of federal and state conservation programs on farmer nitrogen management." *Environmental management* 62 (4): 694–708.

Richards, R. Peter. 2017. "" Technical report. National Center for Water Quality Research.

Rose, K., J. Bauer, V. Baker, A. Bean, J. DiGiulio, K. Jones, D. Justman, et al. 2018. "Development of an Open Global Oil and Gas Infrastructure Inventory and Geodatabase." Technical report NETL-TRS-6-2018. US Department of Energy, National Energy Technology Laboratory, Office of Fossil Energy, March.

Ruddy, Barbara C, David L Lorenz, and David K Mueller. 2006. "County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982-2001."

Saavedra, Santiago, and Mauricio Romero. 2017. "Local incentives and national tax evasion: The response of illegal mining to a tax reform in Colombia."

Saez, Emmanuel, and Gabriel Zucman. 2016. "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data." *The Quarterly Journal of Economics* 131 (2): 519–578.

Schwarz, GE, AB Hoos, RB Alexander, and RA Smith. 2006. "The SPARROW surface water-quality model: theory, application and user documentation." *US geological survey techniques and methods report, book* 6 (10): 248.

Shimshack, Jay P. 2014. "The economics of environmental monitoring and enforcement: A review." *Annual Review of Resource Economics* 6:339–60.

Smith, RA, J Brakebill, GE Schwarz, AW Nolin, J Shih, J Blomquist, RB Alexander, and M Macauley. 2012. "Use of MODIS Vegetation Data in Dynamic SPARROW Modeling of Reactive Nitrogen Flux." In *AGU Fall Meeting Abstracts,* vol. 2012, H11E–1244.

Sprague, Lori A, Gretchen P Oelsner, and Denise M Argue. 2017. "Challenges with secondary use of multi-source water-quality data in the United States." *Water research* 110:252–261.

Steck, Andrew L. 2018. "Industry dynamics with social learning: Evidence from hydraulic fracturing."

Stuart, D, RL Schewe, and M McDermott. 2014. "Reducing nitrogen fertilizer application as a climate change mitigation strategy: Understanding farmer decision-making and potential barriers to change in the US." *Land use policy* 36:210–218.

Subhash, Hari. 2018. "Resource Projects Initative: Extractive industries project payments." https://www.resourceprojects.org/.

Sunley, Emil M, Thomas Baunsgaard, and Dominique Simard. 2003. "Revenue from the oil and gas sector: Issues and country experience." *Fiscal policy formulation and implementation in oil-producing countries,* 153–183.

Tin, Myint. 1965. "Comparison of some ratio estimators." *Journal of the American Statistical Association* 60 (309): 294–307.

United States Department of Agriculture. 2021. "Mississippi River Basin Healthy Watersheds Initative." https://www.nrcs.usda.gov/wps/portal/nrcs/detailfull/national/programs/initiatives/?cid=stelprdb1048200.

United States Geological Service. 2016. "Great Lakes Restoration Initative: Edge-of-Field Monitoring." https://wim.usgs.gov/geonarrative/glri-eof/.

———. 2021a. "USGS Daily Values Site Web Service." https://waterservices.usgs.gov/rest/DV-Service.html.

———. 2021b. "USGS Instantaneous Values Web Service." https://waterservices.usgs.gov/rest/IV-Service.html.

Vollaard, Ben. 2017. "Temporal displacement of environmental crime: Evidence from marine oil pollution." *Journal of Environmental Economics and Management* 82:168–180.

World Bank. 2015. *Zero Routine Flaring by 2030.* World Bank.

———. 2020. *Global Gas Flaring Jumps to Levels Last Seen in 2009.* World Bank, July.

WQX/STORET Nutrient Data Review. https://acwi.gov/monitoring/ppt/indianapolis_112916/wqx-nutrient-best-practices-pp-nwqmc.pdf. Published: 2017-06-02.

Wright, Thomas, and Gabriel Zucman. 2018. "The Exorbitant Tax Privilege." *Working Paper.*

Wuepper, David, Solen Le Clech, David Zilberman, Nathaniel Mueller, and Robert Finger. 2020. "Countries influence the trade-off between crop yields and nitrogen pollution." *Nature Food* 1 (11): 713–719.

Young, Thomas C, Joseph V DePinto, and Thomas M Heidtke. 1988. "Factors affecting the efficiency of some estimators of fluvial total phosphorus load." *Water Resources Research* 24 (9): 1535–1540.

Yu, ChaoQing, Xiao Huang, Han Chen, H Charles J Godfray, Jonathon S Wright, Jim W Hall, Peng Gong, ShaoQiang Ni, ShengChao Qiao, GuoRui Huang, et al. 2019. "Managing nitrogen to restore water quality in China." *Nature* 567 (7749): 516–520.

Yuan, Lifeng, Tadesse Sinshaw, and Kenneth J Forshay. 2020. "Review of Watershed-Scale Water Quality and Nonpoint Source Pollution Models." *Geosciences* 10 (1): 25.

Zhang, Yuzhong, Ritesh Gautam, Sudhanshu Pandey, Mark Omara, Joannes D Maasakkers, Pankaj Sadavarte, David Lyon, Hannah Nesser, Melissa P Sulprizio, Daniel J Varon, et al. 2020. "Quantifying methane emissions from the largest oil-producing basin in the United States from space." *Science Advances* 6 (17): eaaz5120.

Zitzewitz, Eric. 2012. "Forensic economics." *Journal of Economic Literature* 50 (3): 731–69.

Zou, Eric. 2018. "Unwatched pollution: The effect of intermittent monitoring on air quality." *URL: https://static1. squarespace. com/static/56034c20e4b047f1e0c1bfca* 5:1535772291252.

Zucman, Gabriel. 2015. *The hidden wealth of nations: The scourge of tax havens.* University of Chicago Press.

Appendix A

Appendix: Who is responsible for damaging the commons? Identifying nonpoint source polluters in national-scale river networks

B.0.1 Methods

Using a backward solving spatially referenced approach, we localize annualized land based sources of nitrogen and phosphorous pollutant loads. We apply our method to two national-scale watersheds: the United States Mississippi River Basin and the country of New Zealand. While our model is applied to two distinct large watersheds, we are able to estimate nutrient flux at the catchment level. In our paper, we define a catchment as a region whose geographical boundary is determined by the location of downstream monitoring site, and local topography.

We breakdown our process of localizing annual pollutant loads into four main steps: (1) how the United States and New Zealand data were collected and collated; (2) how the water quality and streamflow data were cleaned; (3) how the digital elevation models were hydrographically processed and how the water monitoring sites were referenced; (4) and how the nutrient loads were estimated.

B.0.1.1 Data collection and collation, and data attributes

The USMRB water quality data are available through the Water Quality Portal (Read et al. 2017), a cooperative service jointly maintained by the United States Geological Survey (USGS), the Environmental Protection Agency (EPA), and the National Water Quality Monitoring Council. Water quality data collected between January 1, 1980 - December 31, 2018 at all water quality monitoring sites located within the boundaries of the USMRB (i.e. sites located within hydrologic unit codes 05, 06, 07, 08, 10, and 11) are queried using the R (version 3.6.0) dataRetrieval package version 2.7.6. Streamflow data are queried through USGS web services; data between January 1, 1981 - October 31, 2007 are downloaded from

the USGS Daily Values Site Web Service (United States Geological Service 2021a), and data between November 1, 2007 - December 31, 2018 were retrieved from the USGS Instantaneous Values Web Service (United States Geological Service 2021b).

The New Zealand water quality and streamflow data also span January 1, 1981 - December 31, 2018. We obtain the data via personal electronic communication with individual environmental and water agencies from each of the New Zealand Regional Councils in the country, and made available either through Excel worksheets and .csv files, or queried through the regional council's API. We supplemented the data with additional water monitoring site level data from Land Air Water Aotearoa (LAWA) (Land Air Water Aotearoa 2021), the New Zealand consortium of government and academic organizations that provide environmental data. Both the USMRB and New Zealand water quality data include the name of the water agency, datetime and location (coordinates and name of and/or unique identifier associated with the water sampling station), and other metadata associated with the water sample.

In addition to the water quality and stream flow data, we use topographical raster data, river and lake shapefiles, and gridded precipitation data. For the United States, we download the digital elevation model (DEM) using the elevatr package (package version 0.3.4); the processed elevation data contain 305.7 $\times$ 305.7 meter resolution data and were collected from the Shuttle Radar Topography Mission. The New Zealand 25 meter $\times$ 25 meter pixel resolution DEM is available through Land Information New Zealand (LINZ). The river polylines and lake polygon shapefiles are available from Esri and National Oceanic and Atmospheric Administration for the United States and LINZ for New Zealand. For the precipitation data, we obtain daily rainfall data at each of the water quality monitoring site locations in both the USMRB and in New Zealand between January 1, 1981 - December 31, 2018. The U.S. data are a reanalysis product available through PRISM.

B.0.1.2 Water data cleaning techniques

We clean the raw data according to hydrological best practices (Sprague, Oelsner, and Argue 2017; *Regional Guidance on Handling Chemical Concentration Data Near the Detection Limit in Risk Assessments; Best Practices for Submitting Nutrient Data to the Water Quality eXchange (WQX); WQX/STORET Nutrient Data Review*). For the water quality data, we process and harmonize the data using available water sample metadata so that the water quality concentrations are comparable across water agencies. Steps for cleaning the water quality data include: winsorizing outliers, imputing concentration values for non-detected values, removing duplicates, and discarding observations that did not meet a minimum number of water quality samples by chemical compound in a given quarter (see Section 2.2.4 for additional information on the data cleaning process). For the streamflow data, we discard (i) outlier data at a given flow site (which we attributed to sensor error on the flow monitor); (ii) annual-site pairs if there were less than XX flow observations in a given year; and (iii) negative observations. In order to estimate nutrient loads, we require flow measurements at both the same date and location as a water quality sample.

Temporal flow interpolation: We temporally interpolate for missing flow observations at a given flow monitoring site by using daily site level precipitation using a tree-based Cubist machine learning model from the caret package (version 6.0.86) in R. Our training set uses

ten folds. We also implement cross validation on the training object. Our cubist model regresses daily flow on day-of-year (doy), rainfall and its polynomials, and lagged rainfall (up to the 7th day) and lagged rainfall polynomials. Our cubist model is as such:

$$\text{flow} = \sum_{i=1}^{5} \gamma_i \text{doy}^i + \sum_{l=0}^{7} \sum_{p=1}^{4} \beta_{lp} \text{lag}_l \text{precip}^p + \epsilon \tag{A.1}$$

We evaluate model performance through minimizing root mean squared error. We then use this model to generate predicted daily flow values (in cumecs) at that water monitoring site for days that we do not have flow measurements.

Spatial flow interpolation: We spatially interpolate at water quality sites that do not collect flow observations by using both measured and temporally interpolated flow data at nearby flow monitoring sites and elevation data. For each water quality site that doesn't have flow rate data, we determine whether it lies within the catchment of a downstream flow site. If it does, then we know that the water that flows along the river into the downstream flow site first passes through the water quality site. This gives us an upper bound on the flow rate value and flow behavior at the water quality site. Next, we determine whether there are any flow sites directly upstream of the water quality site, because the flow rate at those upstream flow sites serve as a lower bound.

Figure 2.16 contains a toy example illustrating how we interpolate flow at a water monitoring site that does not measure flow data. In Figure 2.16, there are seven sites: Q_2 is the water quality site where we estimate interpolate flow rate; Q_2 is located within the catchment defined by downstream flow site D, and upstream flow sites U_1, U_2, U_3, and U_4. In other words, any water that flows through site D, first either flows through U_1, U_2, U_3, U_R and/or Q_2. Lastly, Q_1 is another water quality site that is located between U_1 and Q_2. The area upstream of Q_2 within the catchment are the pink and green regions, and the area downstream is the orange region.

We assume that locally, the volume of water added to the river (i.e. additional increase in flowrate) is linearly proportional to the area of the catchment. We can model this as:

$$\frac{x - \sum ux}{P^*} = c \tag{A.2}$$

$$\frac{x_D - (\sum ud + x)}{P - P^*} = c \tag{A.3}$$

where x is flow rate of Q_2, $\sum ux$ is flow rate of flow sites upstream of Q_2 (so U_1, U_2, U_3), and P^* is the area upstream of Q_2 (i.e. sum of the green and pink regions). x_d is the flowrate at D, $\sum ud$ is the flow rate of flow sites upstream of D (i.e. U_4), and P P^* is the area downstream of Q_2. c is the constant ratio between the marginal difference in flow rate and area of the catchment. By setting Equations A.2 and A.3 equal to each other, we can solve for x, the interpolated flow rate at our site of interest Q_2.

B.0.1.3 Geospatial data cleaning techniques

We clean the geospatial data with the goal of recovering sensible flow direction in the elevation data so that we can construct a realistic graph structure that represents how water quality monitoring sites are connected. For the hydrographic processing of the DEMs we modify functions from the Matlab (version 2020b) Topotoolbox library (version 2.3.1) to ensure accurate flow direction and that all raster pixels flow into the ocean pixels. The main processing steps are filling sinks, burning river locations into the elevation raster, and computing flow direction and flow accumulation to generate raster of stream pixels.

Next, in order to construct the connectivity matrix, we georeference the water monitoring sites. Given that the recorded site coordinates do not always spatially overlap with the stream pixels in the DEM, we snap water monitoring sites to the closest stream pixel with the highest flow accumulation within 250 meters of its original location for New Zealand sites, and 650 meters for U.S. sites. Our assumptions include: (1) the monitoring site in reality is more likely to monitor a larger stream, rather than a smaller, ephemeral stream; and (2) if the site coordinate is not within 250 meters (in New Zealand) or 650 meters (in USMRB) meters of a derived stream, then either we have incorrect topographical data in that region, and therefore we would have inaccurate flow direction in that area, or that that the coordinates were mis-recorded.

We then use flow direction along the stream network to construct the connectivity matrix across the snapped water monitoring sites in order to recover which sites were immediately up- and downstream of which other sites. We verify the connectivity by plotting average annual flow rate against the ordering of the sites along the stream network (see Figures 2.11, 2.12, 2.13, 2.14, and 2.15).

B.0.1.4 Modeling approach

Our model objective is to localize land-based nonpoint source nutrient runoff. Our main assumption is that the nitrogen- and phosphorus-based pollutant loads originate on land, are picked up by precipitation events, and flow downhill and enter waterways. The approach underlying our model can be broken down into four steps:

(1) We identify pairs of water quality monitoring stations (up and downstream monitors);

(2) We identify the adjacent land pixels that are upstream of the downstream monitor (i.e. if a pollutant particle were placed on that land pixel, and flowed into the waterway, it would flow past the downstream monitor) and are also downstream of the upstream monitor (i.e. that pollutant particle flows into the waterway at a point in the stream that is downstream of the upstream monitoring site). Define this set of land pixels as the catchment that is identified by the coordinates of the up and downstream site pairs;

(3) Then, we compute the pairwise difference of pollutant loads at the up- and downstream water quality monitoring sites;

(4) Interpret the pollutant load as the quantity of runoff that entered into the waterway from one of the pixels in the catchment. In other words, the catchment is responsible for generating this quantity of nutrient runoff.

Constructing time invariant basins:

The size and number of unique catchments are informed by the location and number of water quality monitoring sites, as well as the local topography. The number of sampling stations that monitor a given nutrient compound changes year to year, depending on which sites collect water quality samples and whether the water sample meets our data cleaning criteria. To overcome the challenge of time varying site locations, which affects the catchment boundaries, we utilize a linear regression approach to construct time invariant basins. In our paper, we define basins as these time invariant spatial units that are a subset of time varying catchments. Basin locations are determined by where water quality monitoring stations are located; for a given nutrient compound, there is one basin per unique water quality monitoring site that measures that nutrient. For catchments in a given year, their spatial extent is comprised of one or more basins, depending on the number of monitoring sites that measure water quality data that year. In addition to constructing the spatial extent of a given basin, we also estimate time invariant pollutant weights associated with that basin for that nutrient compound. This allows us to allocate a portion of the catchment's pollutant load in a given year to the basins nested within it in proportion to the basins' weights.

Descriptive statistics about catchments:

In the New Zealand context, catchments are generally smaller because the topography is more mountainous, and there are many streams that flow directly into the ocean. The average catchment in New Zealand is 100 km^2 (0 - 100 , 25th and 75th percentile). The average number of catchments in a regional council is 116 . Canterbury has the largest number of catchments, with 245 catchments; whereas Nelson has the fewest number, with 20 catchments. In contrast, in the United States Mississippi River Basin, the catchments are larger because parts of the USMRB are relatively flat, and all of the rivers connect to the Mississippi River and drain into the Gulf of Mexico. The average number of catchments in a state is 244 . Colorado has the largest number of catchments with 577 , and Michigan and South Carolina have the fewest number of catchments, each having only 1 catchment.

Computing annual pollutant loads:

We are interested in measuring the annual nutrient load at a given water quality monitoring site; however, given that the water quality measurements are sampled monthly, we employ an estimator method to fill in the missing data. At each water quality monitoring site, we use the Beale Ratio Estimator (BRE) (Tin 1965; Young, DePinto, and Heidtke 1988) method to estimate nutrient flux. The BRE method is an approach (Environmental Protection Agency 2014) used by the EPA and the National Center for Water Quality Research (Richards 2017) for estimating "true" monthly and annual stream pollutant loads. The estimator combines streamflow and water quality concentration data, and uses a bias correction factor to account for the undersampling of nutrient loads. The BRE assumes that nutrient loading into the waterway is positively correlated with stream discharge. The equation for computing the pollutant load is given by

$$\hat{y} = X \frac{\bar{y}}{\bar{x}} \frac{(1 + \theta \frac{s_{xy}}{\bar{x}\bar{y}})}{(1 + \theta \frac{s_{\bar{x}}^2}{\bar{x}^2})}$$

where

$\hat{y}$ is the estimated pollutant load during that time interval,

X is the average observed flow on all days within the time interval;

$\bar{y}$ is the average observed pollutant load on days with water quality measurements $y_i (i = 1, \ldots, n)$;

$\bar{x}$ is the average observed flow on days with on days with water quality measurement,

s_{xy} is the covariance between flow and load;

s_x^2 is the variance in flow;

and $\theta = \frac{1}{n} - \frac{1}{N}$, for n days with observations of load y_i and N days with observations of flow.

We use the BRE to compute annual pollutant loads at each water quality monitoring site that has a minimum of six water quality samples in a year.

Modeling robustness checks:

We verify the model, where the goal of this calibration is to discern whether our model accurately estimates pollutant loads across space and over time. The challenge is that we do not have an external data source to corroborate the movement of nitrogen- and phosphorous-based flux. Instead, we are able to test our approach through comparing the movement of stream discharge across the catchments to the rainfall measured at the monitoring site locations.

Using our model, we are able to estimate the difference in the annual stream discharge (m^3 per year) at adjacent flow monitoring stations. We can interpret this quantity of marginal flow as the volume of new water that has entered into the waterway in the river reach between the up- and downstream flow monitoring stations. Then, we can compare this quantity to rainfall at the monitoring station locations.

Given that rainfall is the quantity of new water that has entered into the system, the quantity of marginal flow should be positively correlated with precipitation at that location. We estimate a 1 percent increase in rainfall on average increases marginal flow by 1.22 percent in New Zealand, by 1.61 percent in the USMRB (see Figures 2.17 and 2.18)).

Modeling limitations:

There are a few limitations with our model and the conclusions we can draw from the model outputs. First, given that our goal is to localize the land-based sources of nonpoint source pollution using a data-driven rather than a mechanistic process model, our model does not separate out the contribution of different land use or management practices to nutrient loading. Secondly, we abstract away from processes that happen on the land and in the waterway within each catchment. Specifically, we do not model land-to-water/in-stream transport or decay parameters, and we abstract away from nutrient absorption into groundwater or atmospheric deposition of nitrogen. Lastly, the resolution of our model depends on data availability. The number of catchments, where they are located, and the number of annual observations, depend on where and how frequently water quality concentration and flow are measured.

Total versus marginal load descriptive statistics:

In the USMRB, we have the 858 unique observations for total nitrogen, 14,469 for total Kjeldahl nitrogen, and 12,400 for total phosphorus.

The average annual cumulative (total) load is 5,320,139,493 kg/km^2 for total nitrogen (69.85 min kg/km^2, 1,080,299,586,525 max kg/km^2), 540,215,100 kg/km^2 for total Kjeldahl

nitrogen (22.13 min kg/km^2, 168,677,355,378 max kg/km^2), and 136,054,459 kg/km^2 for total phosphorus (5.97 min kg/km^2, 80,677,489,391 max kg/km^2).

The average annual marginal load is is -4,251,375 kg/km^2 for total nitrogen (-17,254,328,770 min kg/km^2, 42,649,647,658 max kg/km^2), 3,894,974 kg/km^2 for total Kjeldahl nitrogen (-34,305,195,045 min kg/km^2, 454,981,839,321 max kg/km^2), and 15,672,376 kg/km^2 for total phosphorus (-32,155,029,552 min kg/km^2, 467,261,587,689 max kg/km^2).

In New Zealand, we have 3,931 unique observations for total nitrogen, 4,411 for ammonia, 4,554 for dissolved reactive phosphorus, and 4,826 for total phosphorus.

The average annual cumulative (total) load is 676,560.4 kg/km^2 for total nitrogen (0 min kg/km^2, 53,343,646 max kg/km^2), 18,841.18 kg/km^2 for ammonia (0 min kg/km^2, 3,282,781 max kg/km^2), 11,551.18 kg/km^2 for dissolved reactive phosphorus (0 min kg/km^2, 1,047,885 max kg/km^2), and 24,460.36 kg/km^2 for total phosphorus (0 min kg/km^2, 2,031,396 max kg/km^2).

The average annual marginal load is is -17,010.96 kg/km^2 for total nitrogen (-19,034,337 min kg/km^2, 9,352,073 max kg/km^2), -741.3 kg/km^2 for ammonia (-2,318,504 min kg/km^2, 955,036.6 max kg/km^2), -635.76 kg/km^2 for dissolved reactive phosphorus (-4,262,001 min kg/km^2, 1,209,529 max kg/km^2), and 3,448.23 kg/km^2 for total phosphorus (-1,027,198 min kg/km^2, 1,975,447 max kg/km^2).

B.0.1.5 Additional analyses: basin level linear trends

We are able to measure changes in nonpoint source pollutant loads over time using a regression based approach. We estimate the linear trend separately for each basin by nutrient compound pair:

$$\text{load}_{vit} = \alpha + \beta \text{t}_{vit} + \epsilon_{vit}$$

The regressand, *load*, is the area normalized quantity (kg/km^2) of runoff for that chemical compound v attributed to basin i in year t. The regressor, t, is equal to 1 for the earliest year in the sample, and increases linearly with each time-step. Our coefficient of interest is β; the interpretation of β is the average change in the area normalized pollutant load in that basin over time. $\beta > 0$ indicates that, on average, the quantity of area-normalized runoff attributed to the basin is increasing over time.