Essays on the Regulation and Remote Sensing of Natural Gas Flaring

Ruiwen Lee

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020
Abstract

Essays on the Regulation and Remote Sensing of Natural Gas Flaring

Ruiwen Lee

Natural gas flaring from oil production is a pervasive yet understudied environmental issue. Recently available satellite imagery of gas flares has increased public awareness and concern over the severity and ubiquity of the problem. In the US, the relatively recent combination of hydraulic fracturing and horizontal drilling sparked the shale boom, leading to hundreds of thousands of wells being drilled within a decade, often in close proximity to residential populations. A major oil state that has emerged from the shale boom is North Dakota. In 2014, state regulators introduced a policy to limit the percentage of produced gas that oil-extracting companies are allowed to flare. Like many other places where flaring takes place, flared volumes are reported by oil companies themselves. What was the effect of North Dakota’s regulation on gas flaring according to self-reported and satellite data? What was the effect of the regulation on self-reporting behavior? In such a tight oil setting, how well does the prevailing satellite product used to monitor gas flares perform? This dissertation uses new data and methodologies from several disciplines to study these important questions around gas flaring. The results find that the predominant satellite product does not perform well in the on-shore oil production context. While regulation has reduced flaring in a major oil state, the reduction is smaller than thought because of underreporting by oil well operators. Further, the underreporting is associated with political economy and corporate culture factors.
Table of Contents

List of Tables ........................................................................................................ v

List of Figures ......................................................................................................... vi

Acknowledgments ................................................................................................... viii

Dedication ................................................................................................................ x

Chapter 1: Evaluating Environmental Policy when Pollution Might Be Misreported: What Satellite Data Reveals About a Gas Flaring Regulation ........................................... 1

1.1 Introduction ........................................................................................................ 1

1.2 Gas flaring regulation in North Dakota ............................................................. 7

   1.2.1 Oil production and gas flaring .................................................................... 7

   1.2.2 Flaring policy and enforcement ................................................................. 8

   1.2.3 Satellite monitoring of gas flaring .............................................................. 10

1.3 Data ..................................................................................................................... 11

   1.3.1 Well data .................................................................................................. 11

   1.3.2 Satellite data ............................................................................................. 12

   1.3.3 Merging data sources ............................................................................... 12

1.4 Empirical strategy .............................................................................................. 13
Chapter 2: Results

2.5 Results

2.5.1 Operator-level treatment

2.5.2 Cell-level treatment

2.5.3 Comparing the operator-level and cell-level treatments

2.5.4 Effect of being in Fort Berthold Indian Reservation

2.5.5 Effect of oilfield incidents

2.5.6 Robustness checks

2.6 Conclusion

2.7 Figures and tables

Chapter 3: Detection and Temperature Estimation of Gas Flares with Nighttime Landsat OLI

3.1 Introduction

3.2 Data and study area

3.2.1 Landsat 8 data

3.2.2 Visible Infrared Imaging Radiometer Suite data

3.2.3 Study area

3.3 Methods

3.3.1 Estimating flare temperatures using Planck inversion

3.3.2 Matching OLI and VIIRS Nightfire detections

3.4 Results

3.4.1 Characterization of gas flares in tight oil production using Landsat 8

3.4.2 Radiance distributions of gas flares

3.4.3 Flare persistence and area
3.4.4 Temperature estimates of gas flares ................................. 86
3.4.5 Comparison with VIIRS .................................................. 87
3.5 Conclusion and discussion ............................................... 89
3.6 Figures & Tables ......................................................... 92

Appendix A: Planck inversion estimates and misfit ....................... 100

Bibliography ........................................................................... 108
### List of Tables

1.1 Summary statistics of key variables ........................................ 30
1.2 Average effect of operator-level treatment on flaring .................. 35
1.3 Effect of pre-drilling gas and oil prices on pre-policy cell flaring rate 40
1.4 OLS and IV estimates of cell-level treatment effect on flaring ........ 41
1.5 OLS and IV estimates of cell-level treatment effect on flaring rate ... 42

2.1 Summary statistics of key variables ........................................ 65
2.2 Average treatment effect on misreporting ............................... 67
2.3 IV estimates of treatment effect on misreporting ....................... 69
2.4 Difference in average treatment effect between cells within and outside the Fort Berthold Indian Reservation ....................... 70
2.5 Difference in average treatment effect comparing cells of greater versus lower occurrence of oilfield incidents ....................... 72
List of Figures

1.1 Gas flaring around the world .................................................. 27
1.2 Daytime satellite image of the Bakken .................................... 28
1.3 Flaring in the Bakken observed from space .............................. 29
1.4 Example section of flare cells .............................................. 31
1.5 Statewide aggregate flaring: self-reported and satellite-detected .... 32
1.6 Statewide aggregate production: oil and gas ............................. 32
1.7 Aggregate flaring: self-reported and predicted counterfactual ....... 33
1.8 Map of flares used in this study ............................................. 34
1.9 Effect on reported flaring over time ....................................... 36
1.10 Effect on satellite-based flaring over time ............................... 37
1.11 Effect on reported flaring rate over time ............................... 38
1.12 Effect on satellite-based flaring rate over time ....................... 39

2.1 Fort Berthold Indian Reservation .......................................... 64
2.2 Flare zones from hierarchical clustering ................................. 66
2.3 Effect on misreporting over time ........................................... 68
2.4 Difference in effect over time between cells within and outside the Fort Berthold Indian Reservation ......................... 71
Acknowledgements

My gratitude goes first to my husband, Jordan, who never stopped showing me patience, kindness, and love, and continued believing in me through my moments of weakness. I thank little Jed, whose arrival has expanded my perspective and increased my joy. I thank my Singaporean and American families for the support and concern they have shown us. In particular, I thank my mother for her sacrificial love in helping us for many months. I am especially grateful to my mother-in-law and father-in-law for their generous hospitality and for showering Jed with love through the difficult covid months. I thank my dear friends from Redeemer Presbyterian Church, whose fellowship and prayers strengthened me. I thank Jesus, my Lord and God, for giving up everything to be with us.

This doctorate journey would not have been possible if not for my dissertation sponsor, Geoff Heal, who gave generously of his time and energy to encourage and advise me through many research ideas. I am also extremely thankful for the invaluable guidance and encouragement from other advisors, Ignacia Mercadal, Wolfram Schlenker, and Chris Small.

I thank John Mutter and Tomara Aldrich, whose patient support over the years helped me to cross the finishing line.

I sincerely thank the other teachers, colleagues, and friends in the SusDev and Columbia community who have helped and encouraged me in various ways, including Mehdi Benatiya Andaloussi, Anthony D’Agostino, Eugenie Duguoa, Tim Foreman, Eyal Frank, Joséphine
Gantois, Matt Harrington, Justin Ho, Jaehyun Jung, Helen Kim, Kimberly Lai, Upmanu Lall, Mathias Lécuyer, Eddie Lin, Anouch Missirian, José Montiel Olea, Jeffrey Shrader, Ana Varela, Xueeting Wang, Jason Wong, and Alice Zhang.

For sharing data on VIIRS-detected gas flaring in North Dakota, I am grateful to Chris Elvidge, Mikhail Zhizhin, and others in the Earth Observation Group.
To all who groan together with our broken world.
Chapter 1: Evaluating Environmental Policy when Pollution Might Be Misreported: What Satellite Data Reveals About a Gas Flaring Regulation

1.1 Introduction

Self-reporting is commonly used in public policy. In the United States polluters are frequently required to self-report under environmental regulations, including nationwide policies like the Clean Air Act and the Clean Water Act (Shimshack 2014). Regulators rely on polluters to self-report because it is often too costly to monitor each emitter directly. Policies that require self-reporting typically also impose a conditional cost, such as a tax that varies by income level or penalty for exceeding a specified environmental standard. Thus, reporting parties have the incentive to misreport in order to avoid these costs. If self-reported information could be well-measured and easily verified, misreporting could be identified and penalized. However, policymakers might not require proper measurement for a variety of reasons. Environmental regulators, for example, might not want to impose more costs on polluters, or they might be satisfied with an existing method of estimating emissions. In such situations, emissions are measured imperfectly, which creates the opportunity for misreporting. Misreporting is typically challenging to identify, since self-reporting is used precisely because independent, external measures are unavailable. The validity of policy evaluations that rely on potentially misreported data is thus questionable.

In this chapter, I provide the first satellite data-based evaluation of a gas flaring regulation.
I study the impact of a new flaring policy introduced in North Dakota, a major oil producing state in the US, using novel satellite data of flared gas volumes. I compare the policy’s impacts when evaluated using data informed by satellite-based measurements against those obtained using only flaring volumes reported by oil well operators.

Gas flaring is a useful context for studying policies where data might be misreported because flaring is largely self-reported around the world (Elvidge et al. 2016). Flaring refers to the burning of unwanted natural gas extracted together with crude oil and is extremely pervasive worldwide (Figure 1.1). The World Bank-led Global Gas Flaring Reduction Partnership (GGFR) estimates that in 2018 worldwide flaring totaled more than 5 trillion cubic feet of natural gas, the volume used by Central and South America combined in an entire year (World Bank 2019b). This amount of natural gas burnt translates to more than 350 million metric tons of carbon dioxide equivalent emissions released into the atmosphere, which is about 1% of global carbon dioxide emissions from fossil fuel combustion (World Bank 2020). Flaring thus exacerbates climate change. Other negative externalities include air pollution from SO₂ and NOₓ emissions, impacts on human health and agriculture, and loss of royalties and tax revenue to mineral owners and governments. While gas flaring is not new to the oil production industry, the recent surge of onshore oil production makes flaring particularly concerning for two reasons. First, people live very close to flares, and, second, mineral owners tend to be households with less bargaining power and less resources to obtain information. In 2015, tight oil overtook conventional oil in terms of production share in the US, and is expected to continue as the larger source in years to come (EIA 2017). The US was the fourth largest flaring country in 2018, burning nearly 10 percent of gas flared globally (World Bank 2019a).

North Dakota provides a convenient setting to study the impact of gas flaring regulation. In mid-2014, North Dakota adopted a regulation to limit flaring. This new flaring policy provides a natural experiment setting as it impacted misreporting behavior only for operators who flared more
than the newly specified limit. North Dakota is also a particularly relevant setting for studying the effectiveness of gas flaring regulation. First, North Dakota was the top flaring state in the US before the regulation was introduced in mid-2014. In 2014, North Dakota flared 30% more gas than Texas, the largest oil-producing state in the country, even though North Dakota’s crude oil production was only 30% of Texas’. Second, North Dakota’s regulation was innovative and seen as a potential model for other states (Storrow 2015). Moreover, key components of North Dakota’s policy were adopted by the US Bureau of Land Management in its rule that sought to reduce gas flaring and venting on federal and native American lands (BLM Methane and Waste Prevention Rule 2016).¹

I make use of a novel satellite data product that provides estimates for flared gas volumes. The satellite product allows me to independently verify reported flaring. Satellite data is increasingly being used as an alternative, objective source of data for environmental monitoring. While some environmental regulators have begun incorporating satellite data in their monitoring and regulatory efforts, current applications have focused on ambient air quality measures such as aerosol concentrations (Duncan et al. 2014). The dataset I use is based on satellite imagery collected by the Visible Infrared Imaging Radiometer Suite (VIIRS) and processed by the US National Oceanic and Atmospheric Administration (NOAA). However, I find that this dataset, the VIIRS Nightfire product, is imperfect. First, it exhibits large periodic fluctuations that are not present in the reported oil and gas production and gas flaring data. I thus use machine learning prediction to deal with potential measurement error. Second, my co-authored work in chapter 3 shows that it fails to detect small flares. I check the robustness of this chapter’s results accordingly.

The issue of whether oil well operators are misreporting their flaring volumes is a subject of debate among concerned parties, including environmental groups and market observers. Both

¹Final rule published in November 2016. Took effect in January 2017 but immediately challenged in court. In February 2018, the Trump administration proposed a revision to the rule that would have effectively rolled back its key components. Currently, several key provisions in the original 2016 rule have been stayed, i.e. they are not in effect.
S&P Global (Collins 2018) and the Environmental Defense Fund (Leyden 2019) put forth articles noting that the aggregate flaring from the VIIRS Nightfire data exceeds the self-reported amounts at the state and sub-state regional levels. In response the Independent Petroleum Association of America cast doubt on the trustworthiness of the satellite data (Glennon 2019). The difficulty with drawing conclusion from a straightforward comparison of aggregate self-reported and satellite-detected flaring volumes is that the discrepancy between them is already large before the policy when operators did not have a real incentive to misreport. This suggests that the data could be suffering from inaccuracies that cannot be attributed to misreporting. This problem is addressed by a combination of three methods. First, I exploit the fact that the policy shock acts as a concrete incentive to misreport, such that the self-reported data before the policy can be assumed to be free from misreporting. Second, the difference-in-differences approach helps to remove level disparity between the two datasets. Third, I apply machine learning prediction in order to use the satellite data to generate counterfactual self-reported flaring in the absence of the policy.

I find that the policy reduced flaring significantly. Within a difference-in-differences framework, I assign treatment in two ways — at the operator-level and at the flare cell\(^2\) level. In the first approach, flare cells are treated if their operators were above the stipulated flaring limit before the policy was adopted. The results using the satellite-based data show that on average, each flare cell reduced flaring by 1.94 million cubic feet (MMcf) every month. This reduction is equivalent to greenhouse gas (GHG) emissions from 260,000 miles driven by an average passenger vehicle, according to the US Environmental Protection Agency’s greenhouse gas equivalencies calculator EPA (2018). The effect on the flaring rate, which is the proportion of gas flared out of gas produced, was a reduction of 11.2 percentage points. The policy impact evaluated using the flaring volumes reported by the oil well reporters is even larger. Flaring by each flare cell reduced by 3.21 MMcf

\(^{2}\)Given the satellite product’s coarse spatial resolution and that gas from multiple wells are typically flared together in the context of unconventional oil, my unit of analysis is at the level of a spatially proximate cluster of wells, which I term “flare cell”.

4
per month. The flaring rate decreased by an average of 16.8 percentage points, which is within and on the higher end of that found by Lade & Rudik (2019), a closely related study that uses only self-reported data.

The second approach uses each cell’s pre-policy flaring rate as the treatment variable since cells that flared a greater proportion of their gas before the policy are more likely to misreport afterwards. Since the cell flaring rate could be endogenously determined, I instrument it with the prices of gas and oil before wells in the cell are drilled. The effect on satellite-based flaring estimated using this approach is similar to that using the first approach; it is only 2% greater in magnitude. For reported flaring, this approach yields an estimated reduction that is 22% greater in magnitude than using the first approach. The treatment effect on reported flaring rates was a decrease of 16 to 17 percentage points. However, the decline was only about 9 to 11 percentage points when estimated with satellite-based flaring.

This chapter contributes to the literature in several ways. First, it adds to the literature on the regulation and economic impacts of natural gas flaring. In the North Dakota context, Lade & Rudik (2019) evaluate the state’s flaring regulation using data reported by oil well operators, and find that while the policy reduced the percentage of gas flared, it was less efficient than a tax on flaring would have been. Blundell & Kokoza (2018) use detailed patient-level data from North Dakota to show that gas flaring increased the number of hospital visits for respiratory health reasons. Other papers study the case of Nigeria, which currently ranks seventh out of all countries for amount of gas flared World Bank (2019a). Anejionu et al. (2015b) find that gas flaring increased air pollution in the Niger Delta. In another example, Dung, Bombom, and Agusomu (2008) find that flaring negatively affected the growth of two crops but was positive for one crop. They attribute these effects to the high temperatures near the flare. Their study, however, was limited to one gas flare site. More broadly, this chapter adds to the small, recent literature examining firm

---

3 The fraction of gas flared out of gas produced. Elsewhere in this chapter, a “rate” similarly refers to the proportion of gas produced.
behavior in the context of onshore unconventional oil and gas development (Covert (2015); Lange & Redlinger (2019); Lade & Rudik (2019)).

This chapter also adds to the burgeoning literature in economics research that uses satellite imagery as a data source. Donaldson & Storeygard (2016) review this literature and discuss wide-ranging applications in the areas of air pollution, forestry, agriculture, land use, development, and economic growth. Specifically, within environmental and energy economics, recent papers that use satellite data include Karplus et al. (2018), Benatiya Andaloussi (2018), Zhang (2018), and Zou (2018). To my knowledge, this study is the first to apply satellite data in the study of economic behavior in the oil production sector.

Methodologically, I apply machine learning techniques in a novel way to deal with potential measurement error. In recent literature, economists have been actively exploring and suggesting fruitful avenues for adopting machine learning tools for economics research; see for example Varian (2014), Mullainathan & Spiess (2017), and Athey & Imbens (2019). This chapter adds to this budding literature by illustrating a way machine learning tools are useful in empirical economic research. I use machine learning prediction to deal with potential measurement error in the self-reported and satellite-detected flaring data. I suspect that there is measurement error because the two data series are considerably different in the pre-policy period. The key advantage of using machine learning here is that assumptions on the measurement error are not needed.

The remainder of the chapter is organized as follows: Section 2 provides background information on gas flaring, focusing on the context of North Dakota and its new policy. Section 3 describes the data, and Section 4 explains the empirical framework and machine learning methods used. Section 5 presents the main results. Section 6 concludes.
1.2 Gas flaring regulation in North Dakota

1.2.1 Oil production and gas flaring

Oil was first discovered and produced in North Dakota in 1951. Production was modest thereafter, at less than 60 million barrels per year. In the mid-2000s, production surged dramatically when well drillers began combining the techniques of hydraulic fracturing and horizontal drilling. North Dakota produced more than 460 million barrels of oil in 2018, and has been the second largest oil producing state since 2012, according to data from the Energy Information Administration (EIA).

The major oil producing formations in North Dakota are the Bakken formation and the underlying Three Forks/Sanish formation, collectively referred to as the Bakken Petroleum System, (or simply, the Bakken). The Bakken is an oil-rich subsurface rock unit that occupies northwestern North Dakota, and extends into Montana to the west, and Saskatchewan and Manitoba to the north (Figure 1.2). It is a “tight oil” formation, where crude oil is trapped in rock with low permeability, often shale or sandstone. In 2013, the US Geological Survey (USGS) assessed that the Bakken and Three Forks formations still contained a combined 7.4 billion barrels of technically recoverable oil (USGS 2013). In North Dakota, producers are attracted primarily to the Bakken’s oil reserves, so the natural gas and natural gas liquids produced together with the oil are often burnt off upon extraction, or “flared”. This is unlike other shale formations that contain mainly gas, such as the Marcellus in Pennsylvania, where a producer’s main goal is to capture and sell the gas. Furthermore because oil and gas production is a relatively new industry in North Dakota, the infrastructure for capturing and processing gas is still inadequate. An effect of the policy was that operators connected their wells to gas capture infrastructure earlier than before (Lade & Rudik 2019). The main way gas is captured is through small pipelines (“gathering lines”). Alternative

4While the term “shale gas” is often used to describe gas produced from similar formations, the term “tight oil” is preferred over “shale oil”, a term that refers to another product in the oil industry, and far predates tight oil production.
methods of capturing the gas, such as compressing then transporting by truck, are available, but expensive\textsuperscript{5}, and rarely used in North Dakota.

Flaring has become a big concern in North Dakota since the tight oil boom. Western North Dakota used to be predominantly agricultural land, but now, gas flares are everywhere in sight. Satellite imagery captures the bright night lights from these flares, showing how widespread they are (Figure 1.3). Prior to the oil boom, the only bright lights that would have been seen at night are from the cities. Now, there are many more bright spots from the gas flares. Another telltale sign that these are gas flares is that they line up in neat rows, which is typical of onshore unconventional drilling. According to official data from EIA, between 2011-2014, North Dakota flared even more gas than Texas, even though Texas produced three times North Dakota’s oil during that period. North Dakota’s flaring rate hit a high of 36% in January 2014, when Texas flared just 12% of its gas produced.

1.2.2 Flaring policy and enforcement

Oil and gas production in the state is under the jurisdiction of the North Dakota Industrial Commission (NDIC). NDIC appoints the director of the Department of Mineral Resources (DMR), which houses the Oil and Gas Division (OGD). OGD is the state agency that administers oil industry regulations issued by NDIC. On 1 July 2014, NDIC issued Order 24665, which adopted limits for the proportion of produced gas each operator is allowed to flare every month (NDIC 2014). The monthly limit applies to each operator’s statewide\textsuperscript{6} gas production, and was scheduled to ratchet down over several years. By October 2014, the first compliance month, operators were permitted

\textsuperscript{5}A recent analysis by a consultancy on gas capture options concluded that the breakeven volume of gas ranged from 16-107 Mcf/day if gathering lines are used, or from 134-401 Mcf/day if the gas is first compressed then transported by trucks (ICF International 2016). Gas production for most wells fall to below 200 Mcf/day within 6 months, which means that the truck option is uneconomic.

\textsuperscript{6}While the regulation actually states that operators can comply by staying within the flaring limit at the well, field, county, or state level, the state level rate is the least restrictive, and hence is the binding limit on operators that this chapter considers.
to flare up to 26 percent of gas produced every month. The limit then decreased to 23 percent in January 2015. Operators who fail to meet the flaring limits will be penalized by having to curtail oil production. In reality, however, OGD rarely enforced production restrictions (Dalrymple 2017).

Flaring is poorly measured in North Dakota. Reported flared gas volumes are estimated by well operators (NDIC 2018c). To measure natural gas volumes, gas meters are needed; the concept is similar to measuring residential electricity consumption using household electricity meters. However, in North Dakota, meters are not required to measure flared volumes. An OGD staff responsible for auditing reports from operators estimates that 25-50 percent of all wells have gas meters for some kind of purpose, including for measuring sales to midstream gathering companies or allocation to a central storage tank. Gas that is sold is likely to be metered, since there is usually a counterparty, like the gas gathering company or another party further downstream, who has the incentive to ensure that it receives the correct volume of gas from the seller. This is not true for flared gas. Taking all this together implies that out of all wells in North Dakota, only half or less are metered in some way, and of the metered wells, it is likely only a tiny proportion are metered for flaring.

How then do OGD auditors verify reported flaring? The main method used appears to be twofold. The first step is to check that the reported gas production and oil production figures imply a gas-oil ratio (GOR) that is reasonable when compared to recent test measures of the GOR. The GOR is a metric widely used for understanding production conditions from oil wells and reservoirs. In North Dakota, operators test and report the GOR whenever they complete or recomplete a well, and when the producing pool appears to have reached bubble point. However, the GOR varies widely within the Bakken oil pool. In tests performed in 2014, the GOR ranged from 0 to 5121 standard cubic feet per barrel (scf/bbl), with a mean of 1070 scf/bbl and standard deviation of 576

---

7NDIC revised subsequent targets, which required producers to flare not more than 20 percent by April 2016, 15 percent by November 2016, 12 percent by November 2018, and finally, 9 percent by November 2020 (NDIC 2018b).

8The point when pressure in the reservoir drops to a level low enough for the gas dissolved in the oil to escape.
scf/bbl. This means that for every barrel of oil extracted, operators could be reporting 0 to 5000 cubic feet of gas produced. Just within one particular day, 1 June 2014, the fourteen GOR readings in the Bakken pool ranged from 327 to 1759. The second step in checking flared volumes is to verify that all the reported uses of the gas add up to the reported produced volume. Produced gas could be sold to the gathering company, flared, or consumed on site, e.g., as input to electricity generators. Operators report the amounts of gas produced, sold, and flared as separate fields to OGD, with any remainder assumed to have been consumed on site. As with flared gas, consumed gas is not required to be metered. Another rough level of checks is provided by field inspectors. They typically aim to check every well site at least every 90 days. In the absence of meters, however, it is unclear how an inspector can notice underreporting. The best inspectors could do is note whether a well was flaring during the inspection, so that the auditors could later see if the reported flaring is non-zero. In summary, given that flared gas is largely unmetered, operators can make estimates at best, and auditors can only discern if reported figures are out of the ballpark. In North Dakota, reporting false information for government records is a class A misdemeanor\footnote{North Dakota Code 12.1-11-05 — Tampering with public records.}, which carries a potential sentence of up to 1 year in jail and $2,000 in fines.

### 1.2.3 Satellite monitoring of gas flaring

While oil companies may know in general terms that satellite data on gas flaring exist, they are unlikely to have changed their flaring or self-reporting behavior in response to satellite monitoring. First, OGD does not use satellite data to verify reported flaring. Second, while NOAA has published studies on monitoring gas flaring using a few different satellite products, these have been focused on national and global applications, rather than regional or state-level (e.g. Elvidge \textit{et al.} 2009). Moreover, the spatial resolutions of the satellite sensors used to date in these studies, including VIIRS, are too low to pinpoint individual flares. A high spatial resolution product, such
as from Landsat, could observe individual flares (Lee & Small 2019). However, Landsat has low temporal frequency, which precludes its data from being used for monitoring gas flares on its own. Besides the spatial-temporal resolution trade-off, gas flares are best detected at shortwave infrared wavelengths at night, which other existing satellites may not be doing. Finally, that Zou (2018) was able to detect strategic polluting in response to intermittent monitoring by ground stations suggests that polluters may not know or care about being monitored via satellite.

1.3 Data

This section summarizes the data sources used in this chapter. The next section (1.4) describes how I generate key variables from the satellite data and the treatment variables pertaining to the two policy treatment assignments. The summary statistics of the key variables are reported in Table 1.1.

1.3.1 Well data

I collect data on well characteristics and monthly production between 2013-2015 from NDIC. This data includes each well’s key characteristics, including its location in terms of latitude and longitude, which oil pool it extracts from. Since the flaring regulation applies only to wells in the Bakken Petroleum System, I retain wells that extract from the Bakken, Three Forks, or Sanish pools. The data also contains monthly self-reported production figures: the amount of oil and gas produced, amount of oil and gas sold, and amount of gas flared\(^\text{10}\). 

\(^{10}\)While operators are to report gas vented and flared as a single figure, North Dakota prohibits venting (the direct releasing of unwanted natural gas to the atmosphere). State regulation requires well operators to equip each flare combustion point with “an automatic ignitor or a continuous burning pilot”, which ensures that all unwanted gas is burnt, rather than released. Enforcement officers conduct random onsite field inspections to ensure that no venting occurs. Raw natural gas is mostly made up of a mixture of gases that are highly flammable, including methane, propane, and butane. Especially because some of these gases are denser than air, the venting of natural gas increases the risk of explosions at the surrounding well site (EIA 2016b). Operators thus have the incentive not to vent, for safety reasons. Therefore, in this study, I assume that venting is negligible. Lade & Rudik (2019) assume the same.
1.3.2 Satellite data

I obtain satellite-based estimates of gas flare volumes from Elvidge et al. (2016). This is the source of the World Bank-led GGFR’s flaring data, and is the best remotely sensed gas flaring data available currently. These authors use imagery collected by the Visible Infrared Imaging Radiometer Suite (VIIRS), a sensor on board the Suomi National Polar-orbiting Partnership satellite, jointly operated by the US National Aeronautics and Space Administration (NASA) and NOAA. VIIRS records data every night, when combustion sources like flares are the only detectable features at the near- to shortwave infrared wavelengths. These are moderate resolution bands, with a spatial resolution of 750 m at nadir. Radiances of flare sources at these wavelengths are used to invert Planck’s law to yield flare temperatures and areas. Each flare’s temperature and area is then used to calculate its radiant heat, according to the Stefan-Boltzmann law. Finally, the radiant heat estimations are converted to corresponding flare volumes based on calibrations to flare volumes collected by a global database. The monthly estimates are aggregated from the valid daily estimates, normalized by the number of cloud-free daily estimates for each cell. The authors gauge that flare volumes are accurate to 9.5%.

1.3.3 Merging data sources.

This chapter’s analysis is conducted at the level of a spatially proximate cluster of wells, which I term “flare cell”. In tight oil production, multiple wells are commonly co-located on a single compact drilling site known as a well pad. Unwanted gas from the several wells occupying a well pad is typically combined and flared together at the same combustion point, such that satellite observation of flaring cannot be at the well level. Moreover, since VIIRS pixels span 750m-1650m in size, multiple flare sources could be detected within the same pixel. Elvidge et al. (2016) also provide the spatial outlines of each cell of flare sources (see Figure 1.4 for an example of an area
with flare cells demarcated). I thus assign each well to the cell it lies in, and aggregate well data to the flare cell level accordingly.

1.4 Empirical strategy

1.4.1 Machine learning prediction to handle discrepancy between reported and detected flaring

Figure 1.5 plots the total statewide self-reported and satellite-detected flaring by month. There is a clear discrepancy between the two, with the satellite data displaying pronounced periodic fluctuations that are not present in oil production and gas production levels (Figure 1.6). The periodic fluctuations in the satellite data could be driven by seasonal differences in cloud cover, snow cover, and vegetation cover (Levin 2017). Another factor could be longer daylight hours at high latitudes during the summer months, when there are less nighttime satellite observations (Elvidge et al. 2017). To make the two datasets comparable for the subsequent analysis, I use machine learning to capture the relationship between the two data series in the pre-policy period, before the policy would have changed operators’ self-reporting behavior. This approach allows me to avoid making strong assumptions on (the functional forms of) the errors in both datasets. Instead, this approach assumes more plausibly that the relationship between honest reporting and detection, which encompasses how both sets of data might be measured with error, was unchanged after the policy was introduced.

The goal of this approach is to use the satellite data to generate the counterfactual self-reported flaring in the absence of the policy. First, I train a prediction model with pre-policy data, using the satellite-detected flaring to predict reported flaring. Next, I apply the post-policy satellite data as input to the model to predict what the reported flaring would have been if the policy was not adopted. The assumption in taking this approach is that the relationship between
honest reporting and detection, which encompasses how both sets of data are measured with error, remained unchanged after the policy was introduced.

The prediction model’s features (independent variables) are satellite-detected flaring and other variables exogenous to operators’ decisions to flare — the flare cell’s latitude and longitude, and month of year dummies. The model’s target (dependent variable to be predicted) is self-reported flaring. The prediction algorithm I use is Xgboost (Chen & Guestrin 2016), or extreme gradient boosting, a tree-based ensemble algorithm currently widely used by data scientists for prediction tasks, given its high predictive power (Nielsen 2016). The training set consists of 70% of the pre-policy data randomly selected. I tune the model’s hyperparameters on the training set, using ten-fold cross-validation and 100 randomly selected sets of values in the hyperparameter space. The prediction error on the test set (the remainder 30% of the pre-policy data) is 0.009, about half the standard deviation of the training set. Figure 1.7 plots the actual and predicted self-reported flaring. Predicted reporting tracks actual reporting in the pre-policy, but not too closely, which suggests that the model is not overfitting. In the post-period, the difference between the two series increases. Henceforth, I refer to the counterfactual reported flaring predicted with satellite-detected flaring as “satellite-based flaring”.

1.4.2 Operator-level treatment

Since the new flaring limit effectively requires operators to comply on a state-wide basis (that is, the limit applies after aggregating production and flaring volumes from all wells for each operator), I first investigate the policy’s effect by assigning treatment at the operator-level.

1.4.2.1 Average treatment effect

North Dakota’s new flaring regulation provides a natural experiment setting to estimate the causal effect of the policy on misreporting. I use a difference-in-differences strategy to compare flaring
by operators originally above the policy threshold (treated group) to flaring by operators already within the threshold (control group) before and after the policy. Since NDIC adopted the regulation on 1 July 2014, I treat the 18 months before as the pre-policy (January 2013 to June 2014), and the 18 months from adoption as the post-period (July 2014 to December 2015). I calculate each operator’s monthly flare percentage by dividing the gas it flared by the gas it produced in each month. Treated operators are those who flare more than 23% for at least four of the twelve months immediately preceding the announcement (July 2013 to June 2014). Treated flare cells consist of at least one well belonging to a treated operator. In the final balanced panel, there are 859 flare cells, each with 36 months of observations. 236 of the flare cells are control, 623 are treatment. Flare cells consist of about 4 wells on average, and the mean age of wells is about 2.5 years. Figure 1.8 is a map of the control and treatment flare cells.

I estimate the average treatment effect using the following difference-in-differences model:

$$y_{it} = \beta \text{Treat}_i \times \text{Post}_t + X_{it}\theta + \alpha_i + \alpha_t + \epsilon_{it}$$  \hspace{1cm} (1.1)

where $y_{it}$ is the reported or satellite-based flare volume in levels, or as a proportion of gas produced, at cell $i$ in month $t$. $\text{Post}_t$ is an indicator variable that takes the value of one for months after the policy was adopted, and zero otherwise. $\text{Treat}_i$ is an indicator variable for whether cell $i$ is in the treated group. $X_{it}$ is a vector of time-varying cell characteristics, including oil and gas produced to control for scale, and the minimum, mean, median, and maximum age of the wells in the cell to control for the distribution of well ages. Cell fixed effects $\alpha_i$ control for time-invariant unobserved cell-specific characteristics, and month fixed effects $\alpha_t$ flexibly control for unobserved statewide changes that affect all flare cells in each month. The individual $\text{Treat}_i$ and $\text{Post}_t$ terms are respectively absorbed by the cell effects and month effects, so they do not appear in the equation. $\epsilon_{it}$ is the error term, which I cluster at the flare cell level to account for serial correlation. The coefficient of interest $\beta$ measures the average difference in policy impact on each measure of
flaring between the treated and control cells. A negative $\beta$ would indicate that treated cells flared less than control cells after the policy was adopted.

### 1.4.2.2 Testing for parallel pre-trends

The identifying assumption of this difference-in-differences framework is that in the absence of the policy, flaring trends would be the same for the treated and control groups. While this assumption cannot be directly tested, I perform a statistical check for the difference in pre-trends for the two groups using the following event study-style difference-in-differences specification that allows the treatment effect to vary month to month:

$$y_{it} = \sum_{s=-17}^{18} \beta_s \text{Treat}_i \times \lambda_s + \sum_{s=1}^{18} \beta_s \text{Treat}_i \times \lambda_s + X_{it} \theta + \alpha_i + \alpha_t + \epsilon_{it} \quad (1.2)$$

where $\lambda_s$ is an indicator variable that takes the value of one for observations $s$ months from June 2014, the last month before the policy was adopted, or the “base month”. The set of 35 coefficients $\beta_s$ measure the date-specific treatment effect relative to the base month: 17 for the months before the base month and 18 for the months after. The pre-policy coefficients ($s < 0$) are effectively results of placebo tests for treatment effects before the policy.

### 1.4.3 Cell-level treatment

Because operators produce from wells dispersed in various locations, operators are likely to respond to the policy differently across their wells. More specifically, operators likely treat wells with higher flaring rates differently from those of lower flaring rates. Thus, I also study the effect of the policy at the cell-level.
1.4.3.1 Cell flaring rate

Flare cells with higher flaring rates before the policy are most likely to be affected by the newly imposed flaring limit. The greater the pre-policy flaring rate, the greater the room for reduction, so operators are likely to focus their compliance efforts on these cells. My second approach to estimating the policy’s effect thus makes use of the pre-policy flaring rate of each cell as the treatment variable, which I calculate as an average over the pre-policy months. I model this relationship as follows:

\[ y_{it} = \beta^C CellRate_i \times Post_t + X_{it}\theta + \alpha_i + \alpha_t + \epsilon_{it} \]  

where the regressor \(CellRate_i\) is the average pre-policy flaring rate of cell \(i\), and all other variables are the same as in Equation 1.1. The coefficient of interest is \(\beta^C\), which estimates the difference in policy effects between a cell with pre-policy flaring rate of 0 and another cell with flaring rate of 1, i.e. 100%. To estimate \(\beta^C\) consistently, \(E[y_{it}\mid\epsilon|X_{it},\alpha_i,\alpha_t] = 0\) must hold. However, the requirement could be violated in this setting, biasing \(\beta^C\). For example, there could be reverse causality, whereby flaring in the pre-policy influences \(CellRate_i\). I thus use an instrumental variables approach to estimate the treatment effect.

1.4.3.2 Instrumenting with oil and gas prices

I use gas and oil prices prior to drilling a well to instrument for the pre-policy cell flaring rate. When gas prices are high, producers are more motivated to capture the gas\(^{11}\). When oil prices are high, oil producers respond by starting to drill for oil, but may not care about capturing the associated gas.\(^{12}\) For pre-drilling oil and gas prices to be valid instruments for pre-policy cell flaring rates, they should not affect misreporting except through the cell flaring rate. As Anderson

\(^{11}\)Lade & Rudik (2019) model the firm’s gas capture problem using gas price as the marginal benefit of capturing the gas.

\(^{12}\)Anderson et al. (2018) show that while oil production is insensitive to oil prices, drilling responds strongly.
et al. (2018) find, oil price strongly influences when operators start drilling new wells, but does not affect production from wells that are already producing. This is likely to be the case for gas price as well because most of the gas from a well is produced during the first few months a well starts producing. The oil and gas prices affect production, but are unlikely to affect misreporting behavior, except through whether the operator captures the gas, and thus through the cell flaring rate. The exclusion restriction is thus reasonable.

In my dataset, the mean duration between the spud date (when drilling first begins) to the first month of production is about 5 months. I thus use the oil and gas prices six months before production to construct the instrumental variables. First, for each well, I obtain the West Texas Intermediate spot oil price and the Henry Hub spot gas price six months before production began for that well. I then aggregate these prices to the cell-level by calculating the gas production-weighted average prices of all wells in each cell for each of the pre-policy months. The instruments are thus defined as:

\[
WeightedPrice_i = \frac{1}{18} \sum_{t=2013m1}^{2014m6} \frac{\sum_w \log \text{(LaggedPrice}_w) \times \text{GasProd}_{wt}}{\sum_w \text{GasProd}_{wt}}
\]

I use these price variables to instrument for pre-policy cell flaring rate in the following first stage estimation:

\[
CellRate_i \times Post_t = \gamma_g p_i^{Gas} \times Post_t + \gamma_o p_i^{Oil} \times Post_t + X_{it} \theta + \alpha_i + \alpha_t + \varepsilon_{it} \tag{1.4}
\]

The first stage regression estimates how well the oil and gas price instruments predict the pre-policy cell flaring rate. The second stage regression then uses these predicted cell flaring rates to estimate how the policy differentially affects cells with varying flaring rates.
1.5 Results

1.5.1 Operator-level treatment

1.5.1.1 Average treatment effect

Table 1.2 presents the results for the average treatment effect on flaring, both reported and according to the satellite-based data, as estimated in Equation 1.1. The dependent variable in columns (1) and (2) is flaring in levels (MMcfd) while the dependent variable in columns (3) and (4) is the flaring rate. Columns (1) and (3) display the effect on reported flaring while columns (2) and (4) do so for the satellite-based flaring data. All the coefficients of interest, on the interaction term, are negative and statistically significant at the 1% level. This shows that the policy did indeed decrease flaring. However, the magnitude of the policy’s impact differs greatly between that indicated by the reported flaring and that according to the satellite-based flaring. The average monthly decrease in flaring corresponding to the reported flaring is 3.2 MMcfd, which is 65% higher than the 1.9 MMcfd monthly decrease according to the satellite-based flaring. Similarly, the average decline in reported flaring rate every month after the policy is 16.8 percentage points, which is 50% higher than the 11.2% decline per month in the satellite-based flaring rate. Lade & Rudik (2019) estimate that the reported flaring rate decreased by 4 to 17 percentage points. The treatment effect on the reported flaring rate estimated here thus falls within (and is on the higher end of) the range they estimated. This lends further support to my estimates, which are obtained using a different unit of observation, flare cells, rather than individual wells.

1.5.1.2 Treatment effect over time

Figure 1.9 and Figure 1.10 show, respectively, the treatment effects on the reported flaring and satellite-based flaring over time. These are the results from the event study-style difference-
in-differences specification in Equation 1.2. As before, the points in the figures represent the estimated average treatment effect in each period, $\beta_s$, and the gray bands are the 95% confidence intervals. In both figures, the coefficients fluctuate around zero and are not statistically significant before the adoption of the flaring policy. This lends support to the parallel pre-trends identifying assumption in both cases. After the date the policy was introduced, the coefficients for reported flaring become consistently negative, statistically significant, and display a clear downward trend. In contrast, the treatment effect on the satellite-based flaring was less dramatic. While the post-policy coefficients are mostly negative, many are not statistically significant. Further, the change appears to be a level shift after the enforcement date rather than a consistent downward trend from the adoption date onwards.

The month-by-month treatment effects for the reported flaring rate and satellite-based flaring rate have similar patterns as their flaring level counterparts. Figure 1.11 shows the results for the reported flaring rate while Figure 1.12 shows the results for the satellite-based flaring rate. Again, for both dependent variables, the coefficients do not display a clear trend before the policy but clearly decrease afterwards. The coefficients representing the monthly treatment effects on the reported flaring rate become negative and statistically significant after the policy was adopted, and follow a clear downward trend. While the coefficients for the satellite-based flaring rate also become negative after the policy was adopted, they are of a smaller magnitude and seem to level off after the initial months of the policy.

1.5.2 Cell-level treatment

1.5.2.1 Effect of pre-drilling gas and oil prices on pre-policy cell flaring rate

Table 1.3 presents the results from estimating the first stage regression in Equation 1.4. The effective F-statistics reported are robust to heteroskedasticity, autocorrelation, and clustering (Olea & Pflueger 2013). Column (1) reports the estimation outcome when the dependent variable is misre-
porting in levels or IHS-transformed misreporting. The coefficient on the gas price instrument is statistically significant at the 1% level. As expected, a higher gas price leads to a lower flaring rate. A 1% increase in the pre-drilling gas price results in the pre-policy cell flaring rate decreasing by 0.0016 percentage points. The oil price instrument has an opposite effect of almost the same magnitude. When the pre-drilling oil price increases by 1%, the pre-policy cell flaring rate increases by 0.0017 percentage points. The coefficient on the oil price instrument, however, is only statistically significant at the 5% level. The effective F-statistic is 14.49 at the 5% significance level, which is greater than the critical value of 6.98 for the two-stage least squares (TSLS) estimator with a weak instrument threshold $\tau$ of 5%, thus rejecting the weak instruments hypothesis.\(^{13}\)

Columns (2) and (3) report the first stage results when the dependent variable is misreporting rate. As in other estimations when misreporting rate is the dependent variable, the regression is weighted by gas produced. As before, the coefficient on the gas price instrument is negative and statistically significant at the 1% level. Its magnitude is larger — a 1% increase in the pre-drilling gas price results in the pre-policy cell flaring rate decreasing by 0.002 percentage points. The oil price instrument has a positive coefficient of a smaller magnitude. When the pre-drilling oil price increases by 1%, the pre-policy cell flaring rate increases by 0.0012 percentage points. Although the oil price instrument is not statistically significant, the effective F-statistic of 10.21 lies between the TSLS critical values at $\tau=10\%$ (8.192) and at $\tau=5\%$ (11.999). Column (3) reports an alternative specification using only the gas price instrument. The coefficient on the gas price instrument does not change much. Although the effective F-statistic becomes higher, it is lower than the new TSLS critical value at $\tau=10\%$ (23.109). I thus use both the gas and oil price instruments in the second stage regressions, including the one that uses misreporting rate as dependent variable.\(^{14}\)

Table 1.4 shows the second stage IV results of the treatment effect on flaring volumes, as spec-

\(^{13}\)In a specification using only the gas price instrument but not the oil price instrument, the effective F-statistic is higher, but lower than the TSLS $\tau=5\%$ critical value.

\(^{14}\)The coefficient of interest does not change much whether both instruments or only the gas instrument is used.
ified in Equation 1.3, together with the corresponding OLS estimates. The coefficient of interest is on $Cell\ rate \times Post$, the interaction term between the pre-policy cell flaring rate and the post-policy indicator. This coefficient measures the difference in flaring between a cell that did not flare any of its gas before the policy (i.e. has pre-policy flaring rate of 0) to a cell that flared all its gas before the policy (i.e. has a pre-policy flaring rate of 1). Columns (1) and (2) report the OLS and IV estimates for reported flaring, while columns (3) and (4) report the corresponding estimates for the satellite-based flaring. All coefficients of interest are negative and significant at the 1% level. This provides further evidence that the policy did reduce flaring. The OLS results suggest that reported flaring declined by an average of 6.94 MMcf per month while showing that satellite-based flaring decreased by only 3.83 MMcf per month on average. The IV results are very different. They indicate that reported flaring decreased by 20.59 MMcf each month while satellite-based flaring was reduced by 10.38 MMcf. The magnitude of the IV estimates are almost three times that of their corresponding OLS estimates for both reported (2.96 times) and satellite-based flaring (2.71 times). This suggests that the interaction term between the cell flaring rate and the post-policy indicator is indeed endogenous in Equation 1.3. Moreover, its endogeneity results in a downward bias in its coefficient $\beta_C$ if Equation 1.3 were estimated by OLS, i.e., without instrumenting for the interaction term.

Table 1.5 shows the corresponding set of results for the flaring rate. As in the previous table, columns (1) and (2) respectively present the OLS and IV second-stage estimates for reported flaring. Similarly, columns (3) and (4) respectively report the OLS and IV second-stage estimates for the satellite-based flaring. All coefficients on the interaction term are negative and significant at the 1% level, showing that the policy reduced flaring rates. However, the estimated average treatment effects on the satellite-based flaring are smaller in magnitude than those on the reported flaring. The IV estimations show that the policy led to an 84 percentage points reduction in the reported flaring rate but that the decline in the satellite-based flaring rate was only about 50 per-
centration points. Here, the coefficients estimated using IV are also greater in magnitude than their OLS counterparts, indicating that the OLS estimation is downward biased.

### 1.5.3 Comparing the operator-level and cell-level treatments

How do the effects estimated using the operator-level and cell-level treatments compare? Cells in the treatment group (according to the operator-level treatment) have a mean pre-policy flaring rate of 35%, while cells in the control group have a mean pre-policy flaring rate of 16%. This means that the difference in the pre-policy flaring rates between an average cell in the treatment group and an average cell in the control group is 19 percentage points. I first consider how the two treatments square for reported flaring. The IV estimation using the cell-level treatment showed that a 1 percentage point increase in the pre-policy cell flaring rate results in an effect on reported flaring of -0.206 MMcf. A 19 percentage point increase then translates to an effect of -3.912 MMcf, which is around 22% greater in magnitude than -3.210 MMcf, the effect estimated using the operator-level treatment.

Next, I carry out the same thought process for satellite-based flaring. The estimation using the instrumented cell-level treatment showed that a 1 percentage point increase in the pre-policy cell flaring rate leads to an effect on satellite-based flaring of -0.104 MMcf. A 19 percentage point in the pre-policy cell flaring rate thus corresponds to an effect of -1.976 MMcf, which is around 2% greater in magnitude than -1.941 MMcf, the effect obtained using the operator-level treatment. Therefore, for both reported and satellite-based flaring, the operator-level treatment yields more conservative estimates for the treatment effect.

For the satellite-based flaring rate, the cell-level treatment effect estimated corresponds to a decrease in 9.4 percentage points in the operator-level treatment approach, which is 15% smaller in magnitude than the 11.2 percentage point reduction estimated using the operator-level treatment approach. For the flaring rate based on reported data, the estimated cell-level treatment effect
corresponds to a decline of 16.0 percentage points in the operator-level treatment approach, which is very close to the actual estimate of 16.8 percentage points.

1.5.4 Robustness checks

1.5.4.1 Detection limit of the VIIRS Nightfire satellite product

Lee & Small (2019) explore if nighttime Landsat data can be used to study gas flares. We find that Landsat’s OLI sensor, being of finer spatial resolution than VIIRS, can resolve individual gas flares in North Dakota, where oil wells are densely-spaced. Our finding agrees with the way data is treated in this chapter, by aggregating well-level data to the cell-level, and analyzing self-reporting behavior at the cell-level. We also find that many of these flares are not detected by VIIRS, and do not appear in the Nightfire data. This chapter’s analysis is nevertheless likely to be representative of the Bakken. First, 80 percent of statewide reported flaring is from wells whose flares appear in the Nightfire data. Second, the extent of misreporting smaller flares is limited by the zero lower bound for reported flaring.

Next, I consider the small flares that do appear in the Nightfire data, i.e. those studied in this chapter. For some of the flare cells in some months, Nightfire reports zero flaring even though the self-reported flaring is positive and non-zero. To check the effect of these “under-detections”, I conduct a sensitivity analysis by re-estimating Equation 1.1 to obtain the average treatment effect using several different values for the flaring variables for this set of under-detections. When I set satellite-based flaring to missing for these observations, the effect on satellite-based flaring decreases to -2.89 MMcf. This effectively measures the effect on larger flares, and suggests that larger flares saw a greater reduction in flaring. When satellite-based flaring is set to the same value as self-reported flaring for these observations, the effect decreases by 9% to -2.12 MMcf.
1.6 Conclusion

This study is the first to evaluate a gas flaring regulation using satellite data. I find evidence that North Dakota’s new regulation did reduce gas flaring but not by as much as self-reported data indicates. I make use of a novel satellite product to provide an objective, independent verification of the flaring self-reported by oil well operators. Since the self-reported and satellite data do not align before the policy, I use machine learning prediction to deal with the discrepancy which could have arisen from measurement inaccuracies. I use the policy as a natural experiment to estimate treatment effects within a difference-in-differences framework in two ways. First, I assign treatment at the operator-level. Second, I assign treatment at the flare cell-level level, instrumenting the treatment variable with oil and gas prices.

The average reduction in reported flaring by each flare cell is between 3.2 to 3.9 MMcf per month. The corresponding decline in satellite-based flaring is considerably smaller, about an average of 1.9 to 2.0 MMcf for each cell per month. According to reported data, flaring rates declined by about 16 to 17 percentage points. However, the decline was only about 9 to 11 percentage points when estimated with satellite-based flaring volumes.

Flaring from onshore oil production is likely to keep increasing. The EIA expects production from unconventional oil in the US to grow in the near term (EIA 2017). Significant unconventional oil reserves are yet untapped around the world, including in China, the EU, Russia, and Argentina. It is thus critical that we regulate flaring effectively. This chapter shows that a regulation like North Dakota’s, which imposes a flaring percentage limit on oil producers, does reduce flaring. However, that the policy impacts were smaller according to the satellite-based data suggests that oil well operators may have misreported the amount of gas they flared.

While satellite data is useful in environmental monitoring and compliance because of its objectivity and low cost, researchers should be cognizant that satellite data is not perfect. Even
though the specific satellite product used in this chapter is currently the predominant satellite
dataset for gas flare volumes, it displays large cyclical fluctuations that do not mirror produc-
tion or pre-policy flaring trends. This study presented a way of using machine learning prediction
to enable information from the data to be used for policy evaluation, despite flaws in the data.

Further research could use the identification strategy from this study to estimate the envi-
ronmental and health impacts of flaring. This would improve our understanding of flaring’s welfare
impacts and potentially help to isolate its effects from those of other operations in onshore uncon-
ventional oil production, such as drilling, hydraulic fracturing, truck traffic, or waste disposal. If,
for example, it is found that gas flaring comprises a large portion of the overall impacts, a straight-
forward remedy would be more stringent gas capture policies and enforcement, without needing
to restrict onshore production.
1.7 Figures & Tables

Figure 1.1: Gas flaring around the world

Source: Skytruth.org
This map visualizes global flaring activity detected by satellite. According to the World Bank-led Global Gas Flaring Reduction Partnership, which sponsors part of the satellite data processing effort, the country that flared the most gas in 2018 was Russia, followed by Iraq, Iran, the US, and Algeria. Flaring in the Bakken is clearly seen on the US-Canada border.
Figure 1.2: Daytime satellite image of the Bakken

Source: US state boundaries from US Census Bureau, Canadian province boundaries from Statistics Canada, backdrop from Google Maps.
Daytime satellite image centered on western North Dakota, this paper’s study area. The most prominent geographical feature is the Missouri River, which courses through the main oil production area in North Dakota.
Figure 1.3: Flaring in the Bakken observed from space

Source: NASA
Nighttime satellite image of approximately the same area shown in Figure 1, captured by the VIIRS day-night band. Bright lights outside the cities (Williston, Minot, Bismarck, and Dickinson) are from gas flares at oil wells. The organized pattern of flares reflect the neat rows of well pads typical of onshore drilling.
Table 1.1: Summary statistics of key variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Treatment variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator-level treatment variable</td>
<td>0.73</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
<td>30,924</td>
</tr>
<tr>
<td>Cell pre-policy flaring</td>
<td>0.30</td>
<td>0.29</td>
<td>0</td>
<td>1</td>
<td>30,924</td>
</tr>
<tr>
<td><strong>B. Flaring variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-reported flaring (MMcf)</td>
<td>5.92</td>
<td>16.30</td>
<td>0</td>
<td>371</td>
<td>30,924</td>
</tr>
<tr>
<td>Satellite-detected flaring (MMcf)</td>
<td>7.30</td>
<td>21.79</td>
<td>0</td>
<td>400</td>
<td>30,924</td>
</tr>
<tr>
<td>Satellite-based (predicted) flaring (MMcf)</td>
<td>6.41</td>
<td>15.24</td>
<td>-2</td>
<td>371</td>
<td>30,924</td>
</tr>
<tr>
<td><strong>C. Control variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean well age (month)</td>
<td>30.27</td>
<td>16.27</td>
<td>-23</td>
<td>125</td>
<td>30,923</td>
</tr>
<tr>
<td>Median well age (month)</td>
<td>28.31</td>
<td>17.97</td>
<td>-28</td>
<td>155</td>
<td>30,923</td>
</tr>
<tr>
<td>Age of newest well (month)</td>
<td>20.13</td>
<td>17.26</td>
<td>-28</td>
<td>125</td>
<td>30,923</td>
</tr>
<tr>
<td>Age of oldest well (month)</td>
<td>45.51</td>
<td>23.33</td>
<td>-4</td>
<td>155</td>
<td>30,923</td>
</tr>
<tr>
<td>Gas prod (IHS; MMcf)</td>
<td>3.04</td>
<td>1.38</td>
<td>0</td>
<td>7.3</td>
<td>30,924</td>
</tr>
<tr>
<td>Oil prod (IHS; Mbbl)</td>
<td>2.91</td>
<td>1.31</td>
<td>0</td>
<td>6.55</td>
<td>30,924</td>
</tr>
</tbody>
</table>

Variables of gas and oil volumes have units in million cubic feet (MMcf) and in thousand barrels (Mbbl) respectively. Variables controlling for age of wells in a flare cell have units in months. The age of a well is calculated as current month minus the month in which oil production is non-zero. For example, a well that began producing oil in January 2013 has age of two months in March 2013. Negative well age indicates that well appears in the reported data before oil production, likely because gas production can begin before oil production.
Figure 1.4: Example section of flare cells

Source: Flare cell boundaries from Elvidge et al, against Google Maps backdrop. Repeat satellite observations of flaring are available at the flare cell resolution. Well pads, the ground surface on which wells are drilled, are visible as rows of small brown rectangles in this sample area of the study area in North Dakota. Overlaid with the flare cell boundaries, this map section shows that each flare cell consists of flaring from multiple wells.
Figure 1.5: Statewide aggregate flaring: self-reported and satellite-detected

Total gas flared in billion cubic feet (Bcf) from flares in final dataset, according to data reported by well operators (blue dashed line) and from satellite detection (orange solid line).

Figure 1.6: Statewide aggregate production: oil and gas

Total gas produced in billion cubic feet (Bcf) and total oil produced in million barrels (MMbbl) from flares in final dataset.
Figure 1.7: Aggregate flaring: self-reported and predicted counterfactual

Total gas flared in billion cubic feet (Bcf) from flares in final dataset, from data reported by well operators (blue dashed line) and counterfactual reporting predicted by machine learning (orange solid line).
Figure 1.8: Map of flares used in this study

Source: Points plotted by author against Google Maps backdrop
Points indicate centroids of all 832 flares in the final dataset used for the analyses. Control flares are in blue, and treatment flares are in red.
Table 1.2: Average effect of operator-level treatment on flaring

<table>
<thead>
<tr>
<th></th>
<th>Flaring (1)</th>
<th>Flaring (2)</th>
<th>Flaring rate (3)</th>
<th>Flaring rate (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reported</td>
<td>Satellite-based</td>
<td>Reported</td>
<td>Satellite-based</td>
</tr>
<tr>
<td>Treat × Post</td>
<td>-3.210***</td>
<td>-1.941***</td>
<td>-0.168***</td>
<td>-0.112***</td>
</tr>
<tr>
<td></td>
<td>(0.531)</td>
<td>(0.486)</td>
<td>(0.019)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Oil prod</td>
<td>0.227***</td>
<td>0.230***</td>
<td>0.001***</td>
<td>0.003***</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.039)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Gas prod</td>
<td>0.162***</td>
<td>0.108***</td>
<td>-0.047***</td>
<td>-0.260***</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.035)</td>
<td>(0.016)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Gas prod (asinh)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean well age</td>
<td>0.537***</td>
<td>0.441***</td>
<td>0.005</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.179)</td>
<td>(0.165)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Median well age</td>
<td>-0.127*</td>
<td>-0.129**</td>
<td>-0.002</td>
<td>-0.003**</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.060)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Age of newest well</td>
<td>-0.239***</td>
<td>-0.172***</td>
<td>-0.010***</td>
<td>-0.009***</td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(0.059)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Age of oldest well</td>
<td>-0.209***</td>
<td>-0.179**</td>
<td>-0.002</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.078)</td>
<td>(0.003)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Flare FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean</td>
<td>5.920</td>
<td>6.408</td>
<td>0.236</td>
<td>0.254</td>
</tr>
<tr>
<td>No. cell-months</td>
<td>30,923</td>
<td>30,923</td>
<td>30,531</td>
<td>30,531</td>
</tr>
<tr>
<td>No. cells</td>
<td>859</td>
<td>859</td>
<td>859</td>
<td>859</td>
</tr>
</tbody>
</table>

Standard errors clustered at flare cell level are in parentheses.

* \( p < 0.1 \), ** \( p < 0.05 \), *** \( p < 0.01 \)

Average treatment effect of policy estimated using difference-in-differences, as specified in Equation 1.1. Dependent variable is flared volume in million cubic feet (MMcf) for columns (1) and (2), and is the flaring rate, that is the proportion of total gas produced that is flared, for columns (3) and (4).
Figure 1.9: Effect on reported flaring over time

Treatment effect in each month relative to the base month, as specified in Equation 1.2. Point estimates are in red, with their 95% confidence intervals in gray. Variable on y-axis is reported flared volume in million cubic feet (MMcf).
Treatment effect in each month relative to the base month, as specified in Equation 1.2. Point estimates are in red, with their 95% confidence intervals in gray. Variable on y-axis is amount of satellite-based flaring in million cubic feet (MMcf).
Treatment effect in each month relative to the base month, as specified in Equation 1.2. Point estimates are in red, with their 95% confidence intervals in gray. Variable on y-axis is the reported flaring rate, i.e. the proportion of gas produced that is flared, according to self-reported data.
Treatment effect in each month relative to the base month, as specified in Equation 1.2. Point estimates are in red, with their 95% confidence intervals in gray. Variable on y-axis is the satellite-based flaring rate, i.e. the proportion of gas produced that is flared, according to the satellite-based data.
Table 1.3: Effect of pre-drilling gas and oil prices on pre-policy cell flaring rate

<table>
<thead>
<tr>
<th></th>
<th>(1) Flaring</th>
<th>(2) Flaring rate</th>
<th>(3) Flaring rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas price × Post</td>
<td>-0.00016***</td>
<td>-0.00020***</td>
<td>-0.00021***</td>
</tr>
<tr>
<td></td>
<td>(0.00003)</td>
<td>(0.00005)</td>
<td>(0.00005)</td>
</tr>
<tr>
<td>Oil price × Post</td>
<td>0.00017**</td>
<td>0.00012</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00008)</td>
<td>(0.00011)</td>
<td></td>
</tr>
<tr>
<td>Mean well age</td>
<td>-0.00132</td>
<td>-0.00416*</td>
<td>-0.00417*</td>
</tr>
<tr>
<td></td>
<td>(0.00242)</td>
<td>(0.00233)</td>
<td>(0.00233)</td>
</tr>
<tr>
<td>Median well age</td>
<td>0.00026</td>
<td>0.00111</td>
<td>0.00117</td>
</tr>
<tr>
<td></td>
<td>(0.00096)</td>
<td>(0.00093)</td>
<td>(0.00092)</td>
</tr>
<tr>
<td>Age of newest well</td>
<td>0.00162*</td>
<td>0.00342***</td>
<td>0.00347***</td>
</tr>
<tr>
<td></td>
<td>(0.00095)</td>
<td>(0.00092)</td>
<td>(0.00092)</td>
</tr>
<tr>
<td>Age of oldest well</td>
<td>0.00105</td>
<td>-0.00097</td>
<td>-0.00087</td>
</tr>
<tr>
<td></td>
<td>(0.00265)</td>
<td>(0.00267)</td>
<td>(0.00266)</td>
</tr>
<tr>
<td>Oil prod</td>
<td>0.00018</td>
<td>-0.00022</td>
<td>-0.00021</td>
</tr>
<tr>
<td></td>
<td>(0.00024)</td>
<td>(0.00016)</td>
<td>(0.00016)</td>
</tr>
<tr>
<td>Gas prod</td>
<td>-0.00021</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gas prod (asinh)</td>
<td></td>
<td>0.01408</td>
<td>0.01416</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.01197)</td>
<td>(0.01199)</td>
</tr>
<tr>
<td>F_{eff}</td>
<td>14.49</td>
<td>10.21</td>
<td>16.59</td>
</tr>
<tr>
<td>Flare FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>No. cell-months</td>
<td>30,923</td>
<td>30,531</td>
<td>30,531</td>
</tr>
<tr>
<td>No. cells</td>
<td>859</td>
<td>859</td>
<td>859</td>
</tr>
</tbody>
</table>

Standard errors clustered at flare cell level are in parentheses.

* p < 0.1, ** p < 0.05, *** p < 0.01

First stage results of IV estimation, specified in Equation 1.4. Column(1) presents results when dependent variable is flaring in levels (MMcf). Columns (2) and (3) report the first stage results when the dependent variable is flaring rate; these regressions are weighted by gas production. Since the oil price instrument is not statistically significant in the case of the flaring rate, I use only the gas price instrument for the second stage. The results do not change much, since the coefficient on the gas price instrument does not change much. F-statistics reported are of the Montiel-Pflueger robust test for weak instruments.
Table 1.4: OLS and IV estimates of cell-level treatment effect on flaring

<table>
<thead>
<tr>
<th></th>
<th>Reported</th>
<th>Satellite-based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) OLS</td>
<td>(2) IV</td>
</tr>
<tr>
<td>Cell rate × Post</td>
<td>-6.936***</td>
<td>-20.588***</td>
</tr>
<tr>
<td></td>
<td>(1.287)</td>
<td>(5.996)</td>
</tr>
<tr>
<td>Oil prod</td>
<td>0.229***</td>
<td>0.231***</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>Gas prod</td>
<td>0.160***</td>
<td>0.159***</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>Mean well age</td>
<td>0.537***</td>
<td>0.512***</td>
</tr>
<tr>
<td></td>
<td>(0.184)</td>
<td>(0.194)</td>
</tr>
<tr>
<td>Median well age</td>
<td>-0.117*</td>
<td>-0.106</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>Age of newest well</td>
<td>-0.232***</td>
<td>-0.204***</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>Age of oldest well</td>
<td>-0.186***</td>
<td>-0.164***</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Flare FEs</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean</td>
<td>5.92</td>
<td>5.92</td>
</tr>
<tr>
<td>No. cell-months</td>
<td>30,923</td>
<td>30,923</td>
</tr>
<tr>
<td>No. cells</td>
<td>859</td>
<td>859</td>
</tr>
</tbody>
</table>

Standard errors clustered at flare cell level are in parentheses.

* p < 0.1, ** p < 0.05, *** p < 0.01

Results from estimating Equation 1.3, using OLS (columns (1) and (3)) and instrumenting for pre-period cell flaring rate (columns (2) and (4)). The dependent variable for columns (1) and (2) is reported flared volume in MMcf. The dependent variable for columns (3) and (4) is satellite-based flared volume in MMcf. All coefficients on the treatment interaction term are negative and statistically significant, indicating that the policy decreased flaring.
Table 1.5: OLS and IV estimates of cell-level treatment effect on flaring rate

<table>
<thead>
<tr>
<th></th>
<th>Reported</th>
<th>Satellite-based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) OLS</td>
<td>(2) IV</td>
</tr>
<tr>
<td>Cell rate $\times$ Post</td>
<td>$-0.667^{***}$</td>
<td>$-0.840^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.032)$</td>
<td>$(0.142)$</td>
</tr>
<tr>
<td>Mean well age</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>$(0.003)$</td>
<td>$(0.003)$</td>
</tr>
<tr>
<td>Median well age</td>
<td>-0.001</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>$(0.001)$</td>
<td>$(0.001)$</td>
</tr>
<tr>
<td>Age of newest well</td>
<td>$-0.008^{***}$</td>
<td>$-0.008^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.001)$</td>
<td>$(0.001)$</td>
</tr>
<tr>
<td>Age of oldest well</td>
<td>-0.003</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>$(0.002)$</td>
<td>$(0.002)$</td>
</tr>
<tr>
<td>Oil prod</td>
<td>$0.001^{***}$</td>
<td>$0.001^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.000)$</td>
<td>$(0.000)$</td>
</tr>
<tr>
<td>Gas prod (asinh)</td>
<td>$-0.039^{***}$</td>
<td>$-0.036^{***}$</td>
</tr>
<tr>
<td></td>
<td>$(0.012)$</td>
<td>$(0.012)$</td>
</tr>
<tr>
<td>Flare FE$\bar{s}$</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month FE$\bar{s}$</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean</td>
<td>0.24</td>
<td>0.24</td>
</tr>
<tr>
<td>No. cell-months</td>
<td>30,531</td>
<td>30,531</td>
</tr>
<tr>
<td>No. cells</td>
<td>859</td>
<td>859</td>
</tr>
</tbody>
</table>

Standard errors clustered at flare cell level are in parentheses.

$^*$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Results from estimating Equation 1.3, using OLS (columns (1) and (3)) and instrumenting for pre-period cell flaring rate (columns (2) and (4)). The dependent variable for each regression is listed at the top of the column. All coefficients on the treatment interaction term are positive and statistically significant, indicating that the policy decreased flaring rates.
Chapter 2: Does Self-Reporting Measure Up?

Environmental Misreporting in the Bakken

2.1 Introduction

In this chapter, I build on the previous chapter to study misreporting behavior. Misreporting of pollution directly undermines the effectiveness of environmental policy. At the same time, misreporting is difficult to study because alternative, independent measures are usually unavailable.

This chapter makes use of the satellite-based flaring data developed in the previous chapter to quantify misreporting of gas flaring in North Dakota. I exploit the introduction of North Dakota’s new gas flaring policy in a difference-in-differences framework to provide the first empirical evidence of misreporting as a result of a new energy/environmental standard. I present the first causally estimated quantification of misreporting of flared volumes. Quantifying the misreporting allows me to further study factors driving heterogeneity in misreporting. I find evidence that misreporting was much greater within a native Indian reservation, and that locations with more oilfield incidents such as oil spills also misreport more.

Gas flaring is a useful context in which to study misreporting since flaring is typically self-reported by oil well operators worldwide (Elvidge et al. 2016). The Bakken, a major tight oil formation in North Dakota, provides a convenient setting to study how a new standard affects self-reporting. North Dakota is the second largest US state in terms of oil production. Self-reporting of flaring has been required in North Dakota since even before the tight oil boom in the mid-2000s. In mid-2014, North Dakota adopted a regulation to limit flaring. The new flaring policy provides a
natural experiment setting as it increased the incentive to misreport only for operators who flared more than the specified limit. Because North Dakota does not require gas meters to measure flared gas, relying instead on an existing estimation method, measurement of the self-reported volumes is imperfect, providing well operators the opportunity to misreport.

I find that misreporting was significant. As in the first chapter, I use the same operator-level treatment and cell-level treatment approaches within a difference-in-differences framework to estimate the policy’s effect on misreporting. Using the operator-level treatment, I estimate that treated cells flared 1.27 million cubic feet (MMcf) of natural gas more than control cells on average every month. This amount of misreported flaring corresponds to greenhouse gas (GHG) emissions from 170,000 miles driven by an average passenger vehicle, according to the US Environmental Protection Agency’s greenhouse gas equivalencies calculator EPA (2018). Moreover, the misreporting rate increases by 5 percentage points because of the policy. Using the cell-level treatment, the estimated misreporting is 52% greater than that from the first approach.

The chapter then investigates heterogeneity in the misreporting in two ways in order to shed light on two policy-relevant questions. First, for policymakers to tackle misreporting, they need to ask: what drives misreporting? Second, with limited enforcement resources, environmental regulators should ask: is there a way to target areas such that the most misreporting can be prevented or discovered? To answer these questions, I focus on two factors that could explain heterogeneity in misreporting.

The first dimension of heterogeneity I consider is a political economy one. How does misreporting differ between flare cells inside versus outside a native American reservation in North Dakota? The federal government has jurisdiction over the oil and gas reserves within the reservation while the state does outside. The western half of the Fort Berthold Indian Reservation is part of the Bakken. The flaring policy applies uniformly to wells in the Bakken, and thus makes no exception to wells in the reservation. However, as with other native reservations in the country,
there are important ways Fort Berthold, as with other native Indian reservations in the US, is administered differently from other parts of the state it is located in. A key difference in this context is that approval for pipelines required to capture gas is granted by the federal government rather than by the North Dakota regulatory body. Delay by the US Bureau of Indian Affairs in permitting pipelines has allegedly resulted in much more flaring in the reservation (Holdman 2017). Also, operators might be more inclined to misreport where there is greater difficulty to comply.

The second factor driving heterogeneity which I study is whether the occurrence of oilfield incidents, such as oil spills or brine spills, is associated with misreporting. The underlying hypothesis is that “bad actors” who engage in one type of undesirable activity might tend to misbehave in other ways as well. In this context, we can also think of well operator firms as having a certain corporate culture which manifests as the tendency for incidents to occur as well as having the inclination to misreport.

To study the heterogeneity in misreporting, I use a triple differences strategy to exploit variation between groups of neighboring flare cells. However, I do not group cells according to administrative divisions like counties and townships, which is the typical approach in empirical work, because many cells do not lie completely within such divisions. Instead, I group them into “flare zones” using a machine learning technique called hierarchical clustering. This technique groups cells that are closest in terms of geographical distance, so that every cell eventually belongs to a cluster with some of its closest neighbors.

The results show, firstly, that there is much greater misreporting within the native American reservation. I find that treated flare cells in the reservation misreported 6.94 MMcf more than those outside. Thus, an unfortunate and unintended consequence of the policy is that misreporting leads to tribal mineral owners losing out more in terms of royalties. Secondly, locations with more pre-policy oilfield incidents, such as oil and brine spills, had greater misreporting. Cells with more than the median number of pre-policy incidents misreported is associated with 1.77 MMcf more
per month. A policy implication of this finding is that while misreporting is difficult to discern, regulators could focus their scarce enforcement resources for verifying reporting honesty within the reservation and in areas that experience more oilfield incidents. Both of these criteria are more easily observed than misreported flaring and are hence useful for enforcement targeting.

This chapter contributes to the literature in several ways. First, it adds to the scant evidence of environmental misreporting, and it is the first to document misreporting as a response to a new environmental standard. It is also the first to do so using an independent yet direct measure of emissions, rather than relying on the monitoring of ambient conditions. An earlier study that shows environmental misreporting is De Marchi & Hamilton (2006). They argue that polluters inaccurately self-reported in the Toxics Release Inventory using data from EPA pollution monitors to show that ambient concentrations do not always match the supposed reductions reported by polluters. Karplus et al. (2018) use satellite data on ambient air quality to find evidence of misreported emissions by power plants in China after an existing standard was tightened. In the broader literature of cheating to comply with environmental policy, Zou (2018) uses satellite data of ambient conditions in a setting where pollution levels are monitored only intermittently to show that polluters strategically increase emissions on unmonitored days.

This chapter provides empirical support for an earlier theoretical model of polluter self-reporting to comply with an environmental policy where a standard is enforced. In the Harford model, a polluting firm maximizes its expected profits by choosing its levels of output, actual pollution, and reported pollution, given that it faces separate penalties for exceeding the regulation’s standard and for underreporting. Under some plausible assumptions, the model predicts that if the regulator increases the fine for violating the standard, misreporting will increase.

1De Marchi & Hamilton (2006) also show that two types of chemicals were expected to follow Benford’s Law, since their monitored levels did. The reported levels, however, did not follow a similar distribution, indicating that these two chemicals might not have been accurately reported. I attempt a similar approach, but the reported flaring does not follow Benford’s Law in the pre-policy, and neither do any of the other production variables, like gas produced and gas sold.
The scenario in this chapter is a special case where the fine was previously zero, and becomes positive. The results show that operators underreported, thus agreeing with Harford’s prediction.

As with the previous chapter, this one also adds to the literature that uses satellite imagery for economics research and to the literature looking at the regulation and impacts of gas flaring. Additionally, it speaks to the broader literature on environmental monitoring and enforcement; see Shimshack (2014) for a recent review. Empirical studies have shown that observable characteristics like facility size, age, industry, and ownership affect compliance to environmental regulations (Stranlund 2013). This chapter provides evidence for two other factors that affect compliance — political economy, determined by whether wells are located in a native American reservation, and corporate culture, manifested in the occurrence of oilfield incidents. Such evidence could help regulators target their limited enforcement resources to maximize compliance.

This chapter applies machine learning in a novel way to create spatial clusters of the units of observations. Recent economics literature has enthusiastically embraced and incorporated a range of machine learning methods, such as those described in Varian (2014), Mullainathan & Spiess (2017), and Athey & Imbens (2019). This chapter illustrates another way machine learning tools can be useful in empirical economic research. I use hierarchical clustering to spatially group the flare cells in order to use between-variation to study heterogeneity in treatment effects. Hierarchical clustering is used by Mercadal (2018) to define markets in order to study the competitive structure of electricity markets. In empirical economics research, units of observations are often grouped according to their administrative designation like village or state, typically for the purpose of clustering standard errors. However, if many units of observations do not fall neatly within available administrative boundaries, as is the case in this chapter, they will have to be dropped or manually assigned to groups. Hierarchical clustering allows all the observations to be retained and assigns them to groups using an algorithm that bases its decisions on the observations’ proximity to one another.
The remainder of the chapter is organized as follows: Section 2 provides background information on the factors of misreporting heterogeneity studied in this chapter. Section 3 describes the additional data employed, and Section 4 explains the empirical framework and machine learning methods used. Section 5 presents the main results. Section 6 concludes.

2.2 Fort Berthold Indian Reservation and Oilfield Incidents

2.2.1 Fort Berthold Indian Reservation

The Fort Berthold Indian Reservation is located in western North Dakota and spans both sides of the Missouri River. The reservation is in the eastern section of the oil production region in the Bakken (Figure 2.1) and is one of the top oil-producing areas in the state. At the end of 2019, the reservation produced about a quarter of the state’s total oil output (EIA 2020). The reservation is home to the Mandan, Hidatsa, and Arikara Nation, also known as the Three Affiliated Tribes. Some tribe members welcomed the oil and gas development, hoping the royalties would improve economic and social conditions in the area (Volcovici 2017). The total area of the reservation is almost one million acres, of which half is trust land; that is, the federal government holds legal title of half of the reservation and, in principle, is required to exercise control over the land for the benefit of a Native American individual or tribe. This is in contrast to fee land, which the tribe has acquired legal title for. Regardless of whether land in the reservation is trust land or fee land, its natural resources cannot be developed without approval from the federal government at some point in the process (DOI 2019). Unfortunately, on this reservation and others in the country, native American tribes and their members distrust the federal government given its past “mistakes and misdeeds” (Crane-Murdoch 2012).

The flaring policy adopted by the state of North Dakota in 2014 (see Chapter 1) also applies to wells within the reservation. This means that flaring from a well in the reservation
counts towards its operator’s percentage. According to the North Dakota regulatory authority, the new regulation applies uniformly to wells on and off the reservation. However, as with other Indian reservations in the country, Fort Berthold is managed by the Bureau of Land Management (BLM), a federal agency within the US Department of the Interior. While the BLM reportedly has an informal “gentleman’s agreement” with the NDIC to enforce the flaring regulations in the reservation, BLM’s “overstressed” staff may not be able to execute these added responsibilities, as they are already pressured by a backlog of existing duties (DOE 2014; Nemec 2014). For example, according to media reports, the lengthy federal permitting process at the Bureau of Indian Affairs to approve pipeline right-of-ways has delayed the construction of gas capture infrastructure in the reservation, resulting in flaring rates higher than elsewhere in the state (Holdman 2017; Dalrymple 2018). In 2013, 46% of the reservation’s gas production was flared, compared to 29% from wells in other state and private land in North Dakota (CATF 2014). Gas from the reservation was also flared at a higher rate than from other public lands in the US (27%).

### 2.2.2 Oilfield incidents

In North Dakota, environmental incidents that occur in the oilfields and that are exempt under the Resource Conservation and Recovery Act (RCRA) are called oilfield environmental incidents. These incidents generally include fluids or gases produced or used during the production process, such as crude oil and brine, and they are reported to the Department of Mineral Resource’s Oil and Gas Division. Incidents that are not RCRA-exempt, or general environmental incidents, are reported to the Department of Health’s Environmental Health Section. Both datasets are available on from the Department of Environmental Quality’s website. To study whether environmental incidents is a factor for heterogeneity in misreporting, I focus on the oilfield incidents because most of them are associated with a particular well while the general incidents tend to be caused after the upstream process, such as vehicle accidents during transportation or at processing plants.
2.3 Data

In addition to the data used in chapter 1, this chapter uses additional data to study heterogeneity in misreporting. The summary statistics of these new key variables are reported in Table 2.1.

2.3.1 Fort Berthold Indian Reservation

To determine which cells lie within the reservation, I obtain the geospatial file of the reservation’s boundary from the US Census Bureau.

2.3.2 Oilfield incidents

I obtain data on oilfield incidents from the North Dakota Department of Environmental Quality. Typical oilfield incidents include oil spills or brine spills resulting from pipeline leaks, other equipment malfunction, or during transportation or storage. For incidents that are related to a particular well, the well number is recorded. I match these incidents to each well, and aggregate them to the flare cell level.

2.4 Empirical strategy

2.4.1 Treatment effect on misreporting

I use the two approaches — operator-level treatment and cell-level treatment — as described in chapter 1’s Empirical strategy section to estimate the treatment effect on misreporting.
2.4.1.1 Operator-level treatment

I estimate the average treatment effect on misreporting using the same difference-in-differences model as in chapter 1. The estimating equation, repeated here for convenience from Equation 1.1, is:

$$y_{it} = \beta \cdot \text{Treat}_i \times Post_t + X_{it} \theta + \alpha_i + \alpha_t + \epsilon_{it}$$

(2.1)

where $y_{it}$ is now the misreported flare volume in levels, transformed using the inverse hyperbolic sine (IHS) function\(^2\), or as a proportion of gas produced, at cell $i$ in month $t$. Other variables are the same as in Equation 1.1. As before, I control for cell and month fixed effects. $\epsilon_{it}$ is the error term, again clustered at the flare cell level to account for serial correlation during estimation. The coefficient of interest $\beta$ measures the average difference in policy impact on misreporting between the treated and control cells. A positive $\beta$ would indicate that treated cells misreported more than control cells after the policy was adopted.

As before, I test for parallel pre-trends and obtain month-by-month treatment effects using the same estimating equation in Equation 1.2, repeated here for convenience:

$$y_{it} = \sum_{s=-17}^{-1} \beta_s \cdot \text{Treat}_i \times \lambda_s + \sum_{s=1}^{18} \beta_s \cdot \text{Treat}_i \times \lambda_s + X_{it} \theta + \alpha_i + \alpha_t + \epsilon_{it}$$

(2.2)

where $y_{it}$ is now the misreported flare volume. As before, $\lambda_s$ is an indicator variable that takes the value of one for observations $s$ months from the base month, June 2014. The set of coefficients $\beta_s$ measure the date-specific treatment effect relative to the base month.

\(^2\)Transforming data using the IHS function was first proposed by Johnson (1949). It allows the coefficient to be interpreted as if a natural logarithm function were used instead. The advantage of using the IHS transformation here is that, unlike the more familiar natural logarithm function, the IHS can be used on negative values, which the misreporting variable contains.
2.4.1.2 Cell-level treatment

I use the same approach as in chapter 1 to estimate effects from the cell-level treatment. The estimating equation, repeated here for convenience from Equation 1.3, is:

\[ y_{it} = \beta^C \text{CellRate}_i \times \text{Post}_t + X_{it}\theta + \alpha_i + \alpha_t + \epsilon_{it} \]  

(2.3)

where the dependent variable \( y_{it} \) now represents the measures of misreporting described for Equation 2.1. As before, I instrument the average pre-policy flaring rate of each cell, \( \text{CellRate}_i \), with pre-drilling gas and oil prices in the same first stage estimation in Equation 1.4. This allows \( \beta^C \) to be estimated without bias that could arise from threats to internal validity in Equation 2.3. For example, there could be reverse causality, whereby misreporting before the policy influences \( \text{CellRate}_i \).

2.4.2 Heterogeneous effects

To investigate factors of heterogeneity in the treatment effect on misreporting, I use a triple differences framework. I make use of between-cell variation within groups of spatially proximate flare cells.

2.4.2.1 Group neighboring cells into flare zones using hierarchical clustering

Since pre-policy incident count does not vary across time for a particular flare cell, I use variation between flares to measure the effect of incident count on misreport volume. To reduce bias from unobserved heterogeneity since this is effectively a cross-sectional estimation, I limit the between-variation to groups of neighboring flare cells. Cells are grouped spatially, and group level fixed effects are used in the estimation. In empirical economics research, units of observation are usually grouped at an administrative level, such as state or county, for the purposes of obtaining between-
unit variation or clustering standard errors. In the context of North Dakota, possible groupings
are at the county or township levels. However, since many flare cells do not fall neatly within
these artificial boundaries, such cells would have to be dropped or assigned to one of the counties
or townships they overlap with based on some kind of decision rule. Instead, I use hierarchical
clustering to group the flare cells.

The hierarchical clustering algorithm I use is agglomerative, that is, each cell begins as
its own cluster. The closest cells are then iteratively grouped together. Since the clustering is on
a spatial basis, I use geographical distance as the cut-off threshold. I choose 20 km as the cut-off
distance, and use average linkage clustering. Taken together, this means that for any two of the
final clusters, the average of the pairwise distances between each cell in the first cluster and another
cell in the second cluster is more than 20 km apart. Figure 2.2 shows the resulting 38 groups, or
“flare zones”. The estimation results are robust to a range of cut-off distances; I discuss these in
the section on robustness checks.

2.4.2.2 Fort Berthold Indian Reservation

To determine which cells lie within the reservation boundary, I use geographic information system
(GIS) spatial techniques to join the cell locations with the reservation boundary coordinates. I
then create a binary variable that indicates whether a flare cell is within the reservation. Five cells
straddle the reservation boundary, so they are dropped for this part of the analysis, leaving 854
cells. 100 cells (about 12%) lie within the reservation. I use a triple differences strategy to see
if the level of misreporting is different between flare cells located within and without the Fort
Berthold Indian Reservation. The estimating equation is:

\[ y_{ijt} = \gamma_1 \text{Treat}_i \times \text{Post}_t \times \text{Berthold}_i + \gamma_2 \text{Treat}_i \times \text{Post}_t + \gamma_3 \text{Treat}_i \times \text{Berthold}_i + \gamma_4 \text{Post}_t \times \text{Berthold}_i + \gamma_5 \text{Treat}_i + \gamma_6 \text{Berthold}_i + X_{ijt}\theta + \alpha_j + \alpha_t + \epsilon_{ijt} \]  

(2.4)
where the new subscript \( j \) denotes flare zones created by hierarchical clustering. Accordingly, \( y_{ijt} \) is the misreporting at cell \( i \), which belongs to zone \( j \), in month \( t \). \( Berthold_i \) is an indicator variable taking the value of one if flare cell \( i \) is in the reservation, and zero otherwise. I include flare zone fixed effects, \( j \), and month fixed effects, \( t \). The coefficient on the triple interaction term, \( \gamma_1 \), quantifies the differential effect the policy has on flare cells within versus outside the reservation. If \( \gamma_1 \) is positive, misreporting as a result of the policy is greater within the reservation.

I test for parallel pre-trends by allowing the treatment variable \( Treat_i \) to vary by date, in the following equation:

\[
y_{ijt} = \sum_{s=-17}^{18} \gamma_s Treat_i \times Berthold_i \times \lambda_s + \sum_{s=1}^{18} \gamma_s Treat_i \times Berthold_i \times \lambda_s + X_{ijt} \theta + \alpha_j + \alpha_t + \epsilon_{ijt} (2.5)
\]

where \( \lambda_s \) is an indicator variable that takes the value of one for observations \( s \) months from the base month. The set of 35 coefficients \( \gamma_s \) measure the date-specific treatment effect relative to the base month. The pre-policy \( \gamma_s \) (\( s < 0 \)) act as falsification tests for pre-policy treatment effects.

### 2.4.2.3 Oilfield incidents

I use a triple differences approach to find out how the occurrence of oilfield incidents are related to misreporting. First, I construct a cell level metric of the total number of incidents in the pre-policy. Next, I create an indicator variable that takes the value of one if the cell has greater than the median number of pre-policy incidents. I then estimate the heterogeneity driven by oilfield incidents using the following equation:

\[
y_{ijt} = \delta_1 Treat_i \times Post_t \times MoreInc_i + \delta_2 Treat_i \times Post_t + \delta_3 Treat_i \times MoreInc_i \\
+ \delta_4 Post_t \times MoreInc_i + \delta_5 Treat_i + \delta_6 MoreInc_i + X_{ijt} \theta + \alpha_j + \alpha_t + \epsilon_{ijt} (2.6)
\]
where the new subscript \( j \) denotes flare zones created by hierarchical clustering. As before, \( y_{ijt} \) is the misreporting at cell \( i \), which belongs to zone \( j \), in month \( t \). \( MoreInc_i \) is an indicator variable that has two specifications: (a) it takes the value of 1 if flare cell \( i \) has more than the median number of incidents in the pre-policy, and 0 otherwise, or (b) it takes the value of 1 if flare cell \( i \) has more than the median number of incidents in the pre-policy period normalized by the mean age of wells in that cell, and 0 otherwise. I include the second specification to check if the age of wells affects the occurrence of oilfield incidents. All other variables are the same as in Equation 2.4. The coefficient of interest is \( \delta_1 \), which measures the differential treatment effect between cells of more or less pre-policy incidents. If \( \delta_1 \) is positive, it would suggest that where incidents tend to occur, misreporting is likely to be higher.

Similar to previous sections, I test for parallel pre-trends by allowing the treatment variable to vary by date, using the following equation:

\[
y_{ijt} = \sum_{s=-17}^{-1} \delta_s \text{Treat}_i \times MoreInc_i \times \lambda_s + \sum_{s=1}^{18} \delta_s \text{Treat}_i \times MoreInc_i \times \lambda_s + X_{ijt} \theta + \alpha_j + \alpha_t + \epsilon_{ijt} \tag{2.7}
\]

where \( \lambda_s \) is an indicator variable that takes the value of one for observations \( s \) months from the base month. The set of 35 coefficients \( \delta_s \) measure the date-specific treatment effect relative to the base month. The pre-policy \( \delta_s (s < 0) \) act as falsification tests for pre-policy treatment effects.

### 2.5 Results

#### 2.5.1 Operator-level treatment

**2.5.1.1 Misreporting**

The results for the average treatment effect estimated in Equation 2.1 are shown in Table 2.2. The three columns report results using different dependent variables, which are all metrics for
misreporting. All the coefficients of interest are positive and statistically significant at the 1% level, indicating that misreporting increased after the policy. Column (1) reports the effect in levels. Treated flare cells misreported 1.269 MMcf more flaring per month as compared to control cells after the policy was introduced. This is economically significant. According to EIA data, the average US household used about 63.9 Mcf of natural gas in 2017. The estimated volume of monthly misreported gas flared could have supplied a household with 20 years’ worth of natural gas.

The dependent variable in column (2) is misreporting transformed using the inverse hyperbolic sine (IHS) function, which allows the coefficient to be interpreted as if a natural logarithm function were used instead. Before transforming the misreporting variable, I first estimate the scale parameter by maximizing the concentrated log-likelihood function, as proposed by Burbidge et al. (1988). The coefficient on the interaction term in column (2) is 0.325, which translates to a 38% increase in misreporting as a result of the policy. Column (3) reports the estimation results with the dependent variable being misreporting rate, that is, misreporting as a proportion of gas produced. This regression is weighted by each cell’s gas production in each month, so that cells with very little production do not have an outsized influence on the estimate. The coefficient of interest shows that the policy increased the production-weighted misreporting rate by 5.6 percentage points. Given that pre-policy reported flaring rates in North Dakota were about 30%, and the policy sought to limit flaring to 23% by January 2015, the increase in misreporting rate by 5.6 percentage points is sizeable. The next section provides another perspective on the effect on the misreporting rate.

Results for the event study-style difference-in-differences specification in Equation 2.2 are presented in Figure 1.7. The points represent the $\beta_s$ estimates, and the gray band is their 95%
confidence intervals. For the months prior to the policy, the treatment effect is mostly not statistically significant, which indicates that pre-trends were parallel. This provides support to the parallel trends identifying assumption. The effect increases after the policy was adopted, rising sharply in September and October 2014, which are respectively the month right before operators need to comply, and the first month of compliance. The rise in misreporting even before enforcement began could be because operators try to present a smooth decline in flaring, which is likely to be more credible to the regulators. Interestingly, misreporting declines after October 2014, fluctuates around a lower level, then increases sharply again. Since flaring volumes naturally vary over time, I interpret this variation as different operators choosing to misreport a different amount each month, depending on how far they are from compliance.

2.5.2 Cell-level treatment

The results from the first stage estimation are the same as those described in chapter 1’s section 1.5.2.1. Table 1.5 presents the OLS and IV second stage estimations of the effect pre-policy cell flaring rate has on misreporting, specified in Equation 1.1. The coefficient of interest is on the interaction term between pre-policy cell flaring rate and the post-period indicator. This coefficient quantifies the difference in misreporting between a cell that has a pre-policy flaring rate of 0 to a cell that has a pre-policy flaring rate of 1. Columns (1), (3), and (5) report the OLS estimates respectively for misreporting in levels, IHS-transformed misreporting, and misreporting rate, while columns (2), (4), and (6) report the corresponding IV estimates. All coefficients on cell flaring rate are positive and significant at the 1% level, except that of the IV estimate when the dependent variable is misreporting rate, which is significant at the 10% level. The three IV coefficients are larger than their corresponding OLS coefficients, indicating that endogeneity of the cell flaring rate biases the OLS estimation downwards.

From Column (2), we see that increasing the pre-policy flaring rate by 1 percentage point
increases misreporting by 0.102 MMcf. The coefficient of interest in Column (4) is 2.881, indicating that when the pre-policy flaring rate increases by 1 percentage point, misreporting increases by about 17%. We see from column (6) that a 1 percentage point increase in the pre-policy flaring rate increases the misreporting rate by 0.34 percentage points.

2.5.3 Comparing the operator-level and cell-level treatments

How do the effects estimated using the operator-level and cell-level treatments compare? Cells in the treatment group (according to the operator-level treatment) have a mean pre-policy flaring rate of 35%, while cells in the control group have a mean pre-policy flaring rate of 16%. This means that the difference in the pre-policy flaring rates between an average cell in the treatment group and an average cell in the control group is 19 percentage points. The IV estimation using the cell-level treatment showed that a 1 percentage point increase in the pre-policy cell flaring rate increases misreporting by 0.102 MMcf. A 19 percentage point increase then translates to greater misreporting of 1.938 MMcf, which is around 52% greater than 1.269 MMcf, the effect estimated using the operator-level treatment. Therefore, the operator-level treatment is more conservative.

2.5.4 Effect of being in Fort Berthold Indian Reservation

Table 2.4 reports the estimated effects on misreporting of cells being located within the reservation. The coefficients of interest, on the triple interaction term, are all positive and statistically significant at the 1% level. This indicates that misreporting was higher within the reservation. Specifically, column (1) shows that on average, cells in the reservation misreported by 6.940 MMcf more than cells outside. This magnitude is more than five times that of the overall average treatment effect (1.269 MMcf) previously estimated, showing that the increase in misreporting by cells in the reservation is disproportionately high. The coefficient in column (2) is 1.913, indicating that as a result of the policy, misreporting increased by 5.8 times more in the reservation than outside. Column
(3) shows that the misreporting rate increased by 25 percentage points more for cells in the reservation. This is a sizeable difference, considering that the production-weighted mean misreporting rate across all cells and across the entire study period is only about 2%, as indicated in the bottom panel of Table 2.4 for Column (3).

Figure 2.4 displays the results from the event study-style difference-in-differences specification in Equation 2.5. In the pre-policy, there are several months during which the misreporting was much lower in the reservation, but the variation was very large at the same time. Overall, there is no pattern in the pre-policy coefficients, indicating that pre-trends do not drive the estimated heterogeneous effects. The coefficients increase clearly after the flaring policy was adopted, and appear to level out after December 2014.

2.5.5 Effect of oilfield incidents

Table 2.5 reports the estimated effect that higher occurrence of oilfield incidents has on misreporting. The two columns use the same dependent variable, misreporting, but two different definitions of the incident variable, MoreInc. In both specifications, the coefficient on the triple interaction term is positive and statistically significant at the 5% level. In column (1), the coefficient of interest shows that misreporting increased by 1.765 MMcf more for flare cells with more than the median number of pre-policy oilfield incidents. The coefficient in column (2) is very similar, indicating that this estimate remains robust after normalizing the number of incidents by the mean age of wells.

Figure 2.5 plots the series of estimated $\delta_t$ from Equation 2.7 for each time period. The pre-policy coefficients do not indicate different pre-trends between the cells with a greater or smaller number of pre-policy oilfield incidents. After the policy was adopted, the coefficients are all positive and visibly higher than in the pre-policy.
2.5.6 Robustness checks

2.5.6.1 Detection limit of the VIIRS Nightfire satellite product

I conduct a similar sensitivity analysis as in the previous chapter to consider the effect that under-detections may have on misreporting. I do so by re-estimating Equation 2.1 to obtain the average treatment effect using several different values for the misreporting variable for this set of under-detected cell-months. When I set misreporting to missing for these cell-month observations, the effect increases by 47% to 1.88 MMcf. When misreporting is set to 1, the effect decreases by 14% to 1.09 MMcf. For misreporting set to 10 and 100, the effect increases to 1.70 (33%) and 7.16 MMcf (460%) respectively.

2.5.6.2 Sensitivity to distance cut-off used in hierarchical clustering algorithm

To test the robustness of the heterogeneity results to the 20 km distance cut-off used to determine the flare zones, I re-run the triple difference estimations using different flare zone definitions based on a range of distance cut-offs between 10 km and 50 km. The heterogeneity in treatment effect caused by a cell being located in the Fort Berthold Indian Reservation is extremely robust, varying from 6.938 to 6.988 MMcf. To recall, the effect estimated using 20 km as cut-off was 6.940 MMcf, at the lower end of the range.

The oilfield incident-driven heterogeneity is also very robust to the flare zone definition. Cut-offs between 10 km and 50 km yield estimates from 1.743 to 1.797 MMcf. The effect estimated using 20 km as cut-off was 1.765 MMcf, in the middle of the range.
2.6 Conclusion

This study is the first to find evidence of misreported gas flaring and to provide a causally estimated quantification of misreported flare volumes. I find that flaring was misreported by oil companies in the Bakken after regulators in North Dakota adopted a new policy to limit gas flaring. Using the satellite-based flaring variable developed with machine learning prediction in the previous chapter, I find within a difference-in-differences framework that the amount of misreporting was large and significant. Moreover, I use a triple differences approach to show that misreporting was much greater within federally (versus state) administered lands, and that locations with more oilfield incidents also saw greater misreporting.

The average misreporting from each flare cell is 1.27 MMcf per month. To put this result in perspective, the average self-reported flaring for each cell was about 6.21 MMcf per month after the policy. In other words, if not misreported, the self-reported flaring would be about 20% higher. The misreporting resulted in the post-policy reduction in flaring percentage being overstated by 5.6 percentage points, which is 33% of the supposed 16.8% decline in reported flaring rates. Misreporting was not evenly distributed. Within the native Indian reservation, misreporting was 6.94 MMcf greater on average per cell each month. Moreover, cells where more than the median number of incidents occurred also saw more misreporting, by 1.77 MMcf per month.

Misreporting of flaring has several welfare implications. First, when flaring is underreported, mineral owners lose out on royalties, and the state loses out on revenue from production taxes. A back-of-the-envelope calculation yields a $0.57 million monthly loss in royalties and production tax revenues from the study’s sample of 623 treated flare cells.\footnote{Over the three years of this study, the average Henry Hub natural gas spot price was $3.10/Mcf. Mineral owners typically received a royalty rate of about 20%, which translates to $0.62/Mcf. The state of North Dakota levies an average gross production tax for natural gas of $0.10/Mcf. Taking all this together, the loss each month for the 700 flare cells in my sample was about $0.57 million (623 flare cells × 1269 Mcf × $0.72/Mcf). Over the 18 post-period months, the total is $10.25 million.} Second, misreporting
means that the external costs from flaring are understated. Like all hydrocarbons, carbon dioxide, a greenhouse gas, is released when natural gas is burnt. Just for the flare cells in the sample, the estimated monthly misreported volume translates to greenhouse gases emitted from more than 110,000 passenger cars driven for a month (EPA 2018). While natural gas sold to end-users is almost entirely methane and thus burns cleanly, the combustion of raw natural gas directly from the wellhead releases a range of environmental pollutants like sulfur oxides ($SO_X$), nitrogen oxides ($NO_X$), and particulate matter (PM). These pollutants cause externalities including impacts on human health (Blundell & Kokoza 2018), ecosystems (Sarma et al. 2018), crop yield (Dung et al. 2008), as well as light pollution (Chepesiuk 2009) and noise pollution (Emam 2015).

Preventing misreporting is crucial to effectively regulating gas flaring. A straightforward solution to prevent misreporting is to require that operators install meters to measure flaring. According to an industry supplier of flare gas meters, an innovative new gas meter product could cost $3,750 to own and operate for five years (Olin 2018). This works out to $750 per year, ignoring the cost of capital. If the price of gas is $2/Mcf, approximately the current spot price, which is on the low end of historical prices, the amount of gas that needs to be captured from each well is 375 Mcf, a tiny fraction of the amount of gas each well produces after five years.\(^5\)

According to the Bureau of Indian Affairs, the federal Indian trust responsibility is a legal obligation under which the United States “has charged itself with moral obligations of the highest responsibility and trust” toward Indian tribes (Seminole Nation v. United States, 1942). The findings in this chapter suggest that policymakers need to better consider how oil and gas resources are being managed in the Fort Berthold Indian Reservation, as well as on other federal lands. According to a study by the US Department of Housing and Urban Development, the Native Americans on North Dakota reservations are disproportionately poor (Dalrymple 2016). Forty percent (40%) of them live in poverty, compared to 31% of all US tribes. Ensuring that they

\(^5\)Oil and natural gas production declines over a well’s lifetime. In my sample, the mean volume of gas produced in the 60th month is 2,348 Mcf.
receive the royalties due to them for their oil and gas minerals would help alleviate the situation.

Enforcement targeting helps regulators achieve greater compliance (Friesen 2003). The findings in this chapter propose that the occurrence of oilfield incidents, such as oil spills or brine spills, could be explored as a way to target locations for verifying the accuracy of self-reported flaring. This is potentially useful not only in North Dakota, but elsewhere that flaring is self-reported.

On a broader level beyond gas flaring, this chapter provides evidence that polluters use misreporting as a strategy to comply on paper when their emissions are not well measured. This reinforces the importance of measurement, reporting, and verification, or MRV, in global climate discussions.

Further research can apply the prediction model to data in other states to investigate misreporting there. Moreover, future research can investigate if these findings also apply to native American reservations with oil and gas development and to oilfield incidents in other oil-producing states.
2.7 Figures and tables

Figure 2.1: Fort Berthold Indian Reservation

Source: Boundary of Fort Berthold Indian Reservation from US Census Bureau, against Google Maps backdrop
As with other native reservations and federal lands, subsurface minerals in the Fort Berthold reservation are under the jurisdiction of the federal government.
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berthold dummy</td>
<td>0.12</td>
<td>0.32</td>
<td>0</td>
<td>1</td>
<td>30,744</td>
</tr>
<tr>
<td>No. incidents before policy</td>
<td>4.83</td>
<td>5.79</td>
<td>0</td>
<td>35</td>
<td>30,924</td>
</tr>
<tr>
<td>Misreporting (MMcf)</td>
<td>0.49</td>
<td>9.67</td>
<td>-187</td>
<td>167.33</td>
<td>30,924</td>
</tr>
</tbody>
</table>

For the comparison between flare cells within and outside the Fort Berthold Indian Reservation, cells straddling the reservation’s boundary are dropped. Misreported flaring is in units of million cubic feet (MMcf).
Hierarchical clustering assigns the 859 flare cells to 38 flare zones, overcoming the issue of flare cells straddling borders of commonly used administrative divisions like counties or townships. Triple differences estimations make use of the variation between flare cells in the same flare zone.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treat × Post</td>
<td>1.269***</td>
<td>0.325***</td>
<td>0.056***</td>
</tr>
<tr>
<td>(0.394)</td>
<td>(0.070)</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>Oil prod</td>
<td>0.004</td>
<td>0.001</td>
<td>0.001***</td>
</tr>
<tr>
<td>(0.032)</td>
<td>(0.003)</td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>Gas prod</td>
<td>-0.054*</td>
<td>-0.008***</td>
<td></td>
</tr>
<tr>
<td>(0.028)</td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gas prod (asinh)</td>
<td></td>
<td>-0.213***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>Mean well age</td>
<td>-0.096</td>
<td>-0.007</td>
<td>-0.003</td>
</tr>
<tr>
<td>(0.131)</td>
<td>(0.016)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>Median well age</td>
<td>-0.003</td>
<td>0.003</td>
<td>-0.001</td>
</tr>
<tr>
<td>(0.055)</td>
<td>(0.008)</td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>Age of newest well</td>
<td>0.067</td>
<td>0.017***</td>
<td>0.000</td>
</tr>
<tr>
<td>(0.049)</td>
<td>(0.006)</td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>Age of oldest well</td>
<td>0.030</td>
<td>-0.004</td>
<td>-0.001</td>
</tr>
<tr>
<td>(0.049)</td>
<td>(0.009)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>Flare FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean</td>
<td>0.488</td>
<td>0.372</td>
<td>0.019</td>
</tr>
<tr>
<td>No. cell-months</td>
<td>30,923</td>
<td>30,923</td>
<td>30,531</td>
</tr>
<tr>
<td>No. cells</td>
<td>859</td>
<td>859</td>
<td>859</td>
</tr>
</tbody>
</table>

Standard errors clustered at flare cell level are in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Average treatment effect of policy estimated using difference-in-differences, as specified in Equation 2.1. Dependent variable is misreport flaring in million cubic feet (MMcf).
Treatment effect in each month relative to the base month, as specified in Equation 2.2. Point estimates are in red, with their 95% confidence intervals in gray. Variable on y-axis is amount of misreporting in million cubic feet (MMcf).
Table 2.3: IV estimates of treatment effect on misreporting

<table>
<thead>
<tr>
<th></th>
<th>Misreport</th>
<th>Misreport (IHS)</th>
<th>Misreport rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>OLS</td>
<td>IV</td>
<td>OLS</td>
</tr>
<tr>
<td>Cell rate × Post</td>
<td>3.109***</td>
<td>10.205***</td>
<td>1.109***</td>
</tr>
<tr>
<td></td>
<td>(0.731)</td>
<td>(3.861)</td>
<td>(0.134)</td>
</tr>
<tr>
<td>Oil prod</td>
<td>0.003</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.032)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Gas prod</td>
<td>-0.053*</td>
<td>-0.052*</td>
<td>-0.007**</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.028)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Gas prod (asinh)</td>
<td>-0.096</td>
<td>-0.082</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td>(0.134)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Mean well age</td>
<td>-0.007</td>
<td>-0.012</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.055)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Median well age</td>
<td>0.064</td>
<td>0.049</td>
<td>0.016**</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.052)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Age of newest well</td>
<td>0.020</td>
<td>0.009</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.062)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Age of oldest well</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flare FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean</td>
<td>0.49</td>
<td>0.49</td>
<td>0.37</td>
</tr>
<tr>
<td>No. cell-months</td>
<td>30,923</td>
<td>30,923</td>
<td>30,923</td>
</tr>
<tr>
<td>No. cells</td>
<td>859</td>
<td>859</td>
<td>859</td>
</tr>
</tbody>
</table>

Standard errors clustered at flare cell level are in parentheses.

* p < 0.1, ** p < 0.05, *** p < 0.01

Results from estimating Equation 2.3, using instrumented pre-period cell flaring rate. Dependent variables for column pairs (1) and (2), (3) and (4), and (5) and (6) are, respectively, the amount of misreporting in million cubic feet (MMcf), IHS-transformed misreporting, and misreporting rate. All coefficients on the treatment interaction term are positive and statistically significant, indicating that the policy increased misreporting. OLS estimates are smaller than IV estimates, suggesting that endogeneity leads to downward bias.
Table 2.4: Difference in average treatment effect between cells within and outside the Fort Berthold Indian Reservation

<table>
<thead>
<tr>
<th></th>
<th>(1) Misreport</th>
<th>(2) Misreport (IHS)</th>
<th>(3) Misreport rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post × Treat × Berthold</td>
<td>6.940***</td>
<td>1.913***</td>
<td>0.252***</td>
</tr>
<tr>
<td></td>
<td>(1.426)</td>
<td>(0.340)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>Post × Treat</td>
<td>0.437</td>
<td>0.129*</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>(0.394)</td>
<td>(0.068)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Post × Berthold</td>
<td>-3.156***</td>
<td>-1.179***</td>
<td>-0.172***</td>
</tr>
<tr>
<td></td>
<td>(0.984)</td>
<td>(0.294)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Treat × Berthold</td>
<td>-3.977***</td>
<td>-1.022***</td>
<td>-0.138***</td>
</tr>
<tr>
<td></td>
<td>(0.940)</td>
<td>(0.172)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Treat</td>
<td>-0.729***</td>
<td>-0.367***</td>
<td>-0.033***</td>
</tr>
<tr>
<td></td>
<td>(0.215)</td>
<td>(0.055)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Berthold</td>
<td>-0.164</td>
<td>0.332*</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>(1.028)</td>
<td>(0.172)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Zone FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Well age controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Production controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean</td>
<td>0.488</td>
<td>0.372</td>
<td>0.019</td>
</tr>
<tr>
<td>No. cell-months</td>
<td>30,743</td>
<td>30,743</td>
<td>30,353</td>
</tr>
<tr>
<td>No. cells</td>
<td>854</td>
<td>854</td>
<td>854</td>
</tr>
</tbody>
</table>

Triple differences estimation, as specified in Equation 2.4, measures difference in treatment effect between flare cells within and outside the reservation. Dependent variables for columns (1), (2) and (3) are, respectively, the amount of misreporting in million cubic feet (MMcf), IHS-transformed misreporting, and misreporting rate. All the coefficients of interest, on the triple interaction term, are positive and significant, indicating that misreporting was greater within the reservation.
Difference in treatment effect in each period relative to the base month, as specified in Equation 2.5. Point estimates are in red, with their 95% confidence intervals in gray. Cells in the reservation misreported much more than those outside after the policy.
Table 2.5: Difference in average treatment effect comparing cells of greater versus lower occurrence of oilfield incidents

<table>
<thead>
<tr>
<th></th>
<th>(1) Misreport</th>
<th>(2) Misreport</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post × Treat × MoreInc</td>
<td>1.765**</td>
<td>1.769**</td>
</tr>
<tr>
<td></td>
<td>(0.830)</td>
<td>(0.812)</td>
</tr>
<tr>
<td>Post × Treat</td>
<td>0.317</td>
<td>0.316</td>
</tr>
<tr>
<td></td>
<td>(0.338)</td>
<td>(0.339)</td>
</tr>
<tr>
<td>Post × MoreInc</td>
<td>0.197</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>(0.552)</td>
<td>(0.526)</td>
</tr>
<tr>
<td>Treat × MoreInc</td>
<td>-1.076***</td>
<td>-1.066***</td>
</tr>
<tr>
<td></td>
<td>(0.321)</td>
<td>(0.317)</td>
</tr>
<tr>
<td>Treat</td>
<td>-0.621***</td>
<td>-0.625***</td>
</tr>
<tr>
<td></td>
<td>(0.227)</td>
<td>(0.225)</td>
</tr>
<tr>
<td>MoreInc</td>
<td>0.555**</td>
<td>0.554**</td>
</tr>
<tr>
<td></td>
<td>(0.265)</td>
<td>(0.256)</td>
</tr>
<tr>
<td>Zone FEs</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Well age controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Production controls</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean</td>
<td>0.488</td>
<td>0.488</td>
</tr>
<tr>
<td>No. cell-months</td>
<td>30,923</td>
<td>30,923</td>
</tr>
<tr>
<td>No. cells</td>
<td>859</td>
<td>859</td>
</tr>
</tbody>
</table>

Triple differences estimation, as specified in Equation 2.6, reveals heterogeneity in treatment effect driven by the frequency of oilfield incidents, such as oil spills or brine spills. The dependent variable for each regression is listed at the top of the column. In column (1), MoreInc is an indicator variable, which takes the value of 1 if a cell has more than the median number of oilfield incidents in the pre-period, and 0 otherwise. The coefficient on the triple interaction term estimates the additional misreporting that occurs in a treated flare cell with more pre-period incidents. In column (2), the number of pre-period incidents is first normalized by the mean well age in the base month, such that MoreInc indicates whether a cell has more than the median normalized number of oilfield incidents in the pre-period. The result is robust to well age normalization. Cells with more pre-period oilfield incidents misreport more.
Figure 2.5: Difference in treatment effect over time between cells with higher occurrence versus lower occurrence of oilfield incidents

Difference in treatment effect in each period comparing cells with more than the median number of pre-period oilfield incidents, such as oil spills and brine spills, as specified in Equation 2.7. Point estimates are in red, with their 95% confidence intervals in gray. Cells with more pre-period oilfield incidents misreport more after the policy.
Chapter 3: Detection and Temperature Estimation of Gas Flares with Nighttime Landsat OLI

with Chris Small

3.1 Introduction

Ever since the tight oil\(^1\) boom began in the mid-2000s, the night skies of North Dakota are no longer dark. Bright flares light up what used to be a predominantly agricultural landscape as everyday millions of cubic metres of raw natural gas are burnt away (NDIC 2019). When crude oil is produced, natural gas is extracted at the same time. If it is not economic to capture and sell the gas, the gas is disposed of by combustion. This process is known as gas flaring and is a large-scale polluting activity carried out worldwide. The World Bank estimates that in 2017 more than 140 billion cubic metres (bcm) of such natural gas was flared globally.

As the public becomes increasingly concerned about gas flaring, both governments and non-governmental organizations are paying greater attention to the monitoring of gas flaring. Satellite imagery has emerged as a low-cost, objective tool for monitoring gas flaring. Currently, the predominant satellite product used to monitor global gas flaring is based on imagery collected by the Visible Infrared Imaging Radiometer Suite (VIIRS), a sensor on board the Suomi National Polar-orbiting Partnership satellite. VIIRS has a spatial resolution of hundreds of metres (hectometre resolution) and is thereby adequate for monitoring large gas flares typical of conventional

---

\(^1\) Tight oil refers to crude oil extracted from rock formations of low permeability, typically shale or sandstone. Tight oil is thus a category of unconventional oil.
oil production. Gas flares associated with tight oil production, however, are smaller and spaced more closely together. This suggests that a different satellite product of higher spatial resolution could possess advantages for flare detection in tight oil settings and thereby complement the use of VIIRS.

This paper is the first to investigate the use of nighttime Landsat imagery in detecting gas flares. It is also the first to focus on satellite imagery of gas flares from tight oil production. We use imagery collected by one of Landsat 8’s main sensors, the Operational Land Imager (OLI), which has a spatial resolution of tens of metres (decametre resolution). We find that Landsat 8 OLI better identifies gas flares associated with tight oil production than VIIRS. We then use the Landsat 8 OLI images to characterize the gas flare population in North Dakota. We further show that these images can be used to estimate plausible flare temperatures, which are useful for approximating flare volumes, as has been done using VIIRS and other satellite products.

Combining horizontal drilling with hydraulic fracturing is a relatively new technological advancement in oil and gas production that has dramatically increased production from tight oil formations (EIA 2016c). In 2015, tight oil production surpassed conventional non-tight oil and is expected to continue to make up the majority share of oil production in the USA (EIA 2017). Production from tight oil formations, e.g. of shale, typically involves drilling closely spaced wells. Production from each well is high initially before declining steeply in its first one to two years to a consistent low level (EIA 2016a). After the initial years of production in a tight oil region, the number of older wells, which produce less and hence flare less, will be greater than the number of new wells added. This suggests that studies of gas flares in tight oil formations like the Bakken could benefit from satellite products of a higher spatial resolution more than in settings of conventional oil production.

The Bakken, a major tight oil formation located mainly in North Dakota, offers a useful setting to study how satellite products compare when used to detect small, closely spaced gas flares.
Although the Bakken has long been known to contain vast amounts of oil deposits, production only took off with the advent of shale drilling technology, resulting in the development of many closely spaced wells within a relatively short span of several years. Moreover, the percentage of natural gas produced, known as the flaring rate, is very high in North Dakota. The flaring rate may be particularly high in North Dakota given the speed at which the industry has developed, delays in permitting processes for gas gathering infrastructure, and the relatively low price for gas, which together might make it uneconomic to capture and sell the gas. According to the EIA in 2018, North Dakota’s flaring rate was higher than any other state and was several times greater than that of Texas, which is the largest oil-producing state in the country (EIA 2019).

To compare VIIRS’ and Landsat 8 OLI’s detection capabilities, we align their images spatially such that their respective hotspots overlap (co-registration). This facilitates a direct visual comparison of the two sets of detections. We find that many of the hotspots visible in the OLI image cannot be seen in the VIIRS image. This indicates that VIIRS does not resolve many OLI detections. To characterize the gas flare population in this tight oil setting, we examine the flares’ radiance distributions in several ranges of recorded wavelengths (bands), their size distribution, and their temporal persistence. We find that flares’ maximum radiance (peak radiance) increases with flare size (area). We also find that most flares are small and persist across months. To estimate temperatures of gas flares, we perform Planck inversion, i.e. we find the best fit Planck curve for hotspots detected by OLI. The resulting temperature estimates lie within the reasonable range for typical gas flares. However, the misfits from the curve-fitting are high for some flares.

This paper contributes to the literature on the use of satellite imagery in detecting, measuring, and monitoring gas flaring. Previous studies have looked at the usefulness of other satellite products for this purpose. Elvidge et al. (2009) first estimated national and global gas flare volumes with satellite observations of night light. They used 15 years of night light imagery from the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS), and
estimated that global flared volume was between 140 to 170 bcm annually over the 1994-2008 study period. Subsequent studies used imagery collected by the Along Track Scanning Radiometers (ATSR) at thermal infrared wavelengths (Casadio et al. 2012), by the Moderate Resolution Imaging Spectroradiometer (MODIS) at thermal infrared wavelengths (Anejionu et al. 2015a), and by VIIRS at shortwave infrared wavelengths (Elvidge et al. 2016). All these products are adequate for monitoring large flares at the national or regional level because they image a wide surface area on earth (wide swaths) and thus provide high-level coverage of large areas frequently, imaging the same location almost every day (near daily revisit periods). However, the usual trade-off between the width of the area imaged and the spatial resolution of the image limits the spatial resolution of these sensors to >750 m at most, at the point over which the satellite passes directly (nadir). Points approaching either end of the image width (swath edges) have significantly diminishing spatial resolution. Day time Landsat imagery has been shown capable of detecting large gas flares (Anejionu et al. 2014; Chowdhury et al. 2014; Fisher & Wooster 2018) but its 16-day lapse between observing the same location again (16-day revisit period) precludes its use for operational monitoring. (Elvidge et al. 2015) show that nighttime Landsat shortwave infrared and thermal infrared bands can be used to differentiate between flaming and smouldering combustion phases of peat fires in Indonesia. However, to our knowledge, nighttime Landsat imagery has not been used for identification of small gas flares in settings where wells are too closely-spaced to be resolved by imagery of spatial resolution in the order of hundreds of metres.

The remainder of the paper is organized as follows: Section 2 introduces the study area and the satellite images of it used in this paper. Section 3 explains how we estimate temperatures of flares detected by Landsat 8 OLI and how we match OLI and VIIRS detections in order to compare their temperature estimates. Section 4 presents our results from the following analyses: the characterization of gas flares in the study area using Landsat, flare temperature estimation using Landsat, and the comparison between Landsat and VIIRS for gas flare detection. Section 5
concludes.

3.2 Data and study area

3.2.1 Landsat 8 data

Landsat 8 (L8) image acquisitions are downloaded from the United States Geological Survey’s (USGS) EarthExplorer website. L8 data used in our primary analysis are from the optical bands (1-7), which are collected by the Operational Land Imager (OLI). For our main analysis, we use nine relatively cloud-free nighttime L8 acquisitions in 2016-2017. Of the nine cloud-free nocturnal L8 acquisitions, five are from south of the Missouri river (path/row 126/217 on 9 May 2016, 26 June 2016, 14 September 2016, 16 October 2016, and 3 December 2016), and four are from north of the river (path/row 126/218, on 9 May 2016, 14 Sep 2016, 16 October 2016, and 10 April 2017). In the appendix, we also use the first thermal infrared band, collected by L8’s Thermal Infrared Sensor (TIRS) to check if it can be used to detect flares; we find that it cannot.

3.2.2 Visible Infrared Imaging Radiometer Suite data

Daily data from the Visible Infrared Imaging Radiometer Suite (VIIRS) is downloaded from the National Aeronautics and Space Administration’s (NASA) Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC). We obtained VIIRS Day/Night Band (DNB) monthly composites (Elvidge et al. 2017) produced by the National Oceanic and Atmospheric Administration’s (NOAA) Earth Observation Group (EOG). We also downloaded the EOG’s VIIRS Nightfire data (Elvidge et al. 2016), a VIIRS-derived product that reports the detections of combustion sources at night and some of their characteristics, including geographical coordinates and estimated temperatures.
3.2.3 Study area

Our study focuses on western North Dakota, where thousands of oil and gas wells are in operation (NDIC 2018a). Figure 3.1 is an index figure showing the study area within the study time frame of 2016-2017. The main pair of images is of the study area in the day time and was acquired by Landsat 8 on 4 May 2016. The image on the left is a composite of the shortwave infrared (displayed in red), near infrared (green), and visible green (blue) wavelengths. Areas in green are vegetation while those in shades of red are largely burnt cropland. Areas in dark blue or black are water bodies. The grayscale image on the right is of the first thermal infrared band from the same acquisition. Because land surfaces are warmer than water surfaces in the daytime, they appear lighter in the thermal image. The most conspicuous geographical feature in this area is the Missouri River, which roughly divides the study area into two halves, north and south.

In Figure 3.1, the inset image in the top left corner is composited from VIIRS’ Day/Night Band (DNB) data from different dates, except for the Missouri River, which is filled in with blue. Each colour channel represents the average luminance in a particular month in 2016 (red: December, green: September, blue: May). We see a wide range of colours with no clear spatial pattern. This illustrates how flaring varies across both time and space in the region.

3.3 Methods

3.3.1 Estimating flare temperatures using Planck inversion

In the OLI images, flares with radiances significantly larger than those from noise sources (noise floor) generally peak in either of the shortwave infrared bands (band 6 or 7), suggesting that it may be possible to estimate their temperatures by inverting the Planck distribution function for black-body radiation. Planck’s law describes the spectral-energy distribution of radiation emitted by a
blackbody\(^2\) as a function of its temperature. By finding the best-fit Planck distribution function to the observed radiances, the corresponding temperature that would have resulted in that distribution curve can be obtained. While Fisher & Wooster (2018) have shown that the mid-infrared-radiance method can be modified for higher temperature sources with peak emissions at shortwave infrared wavelengths to return unbiased estimates of Fire Radiative Power (FRP), we apply the Planck inversion approach used by Elvidge et al. (2013) for consistency with the Nightfire flare temperature estimates.

Since flare radiance peaks in the shortwave infrared bands, we first select 154 bright hotspots (clusters of pixels) clearly visible in the first shortwave infrared (SWIR1) band (band 6). We then select the pixel with the highest SWIR1 radiance within each hotspot for analysis. In this paper, we refer to these 154 pixels as flares for simplicity. Using the online atmospheric correction parameter calculator described in Barsi et al. (2003), we find that atmospheric transmissivity\(^3\) during acquisition times was fairly high, ranging from 0.80 to 0.96. We thus use Top of Atmosphere (ToA) radiances. Also, in the context of VIIRS data, Elvidge et al. (2016) confirm that atmospheric effects are small at the shortwave infrared wavelengths.

When the pixels imaging a flare or near a flare have very high radiances that exceed the upper limit of a sensor’s detection capacity, i.e. when saturation occurs, the cubic convolution resampling method used to process L8 images can result in the saturated pixels appearing dark within the cluster of bright pixels (J. Barsi, personal communication, July 18, 2017), rather than taking on the maximum digital number (DN) value, which represents the radiance sensed by the sensor. Schroeder et al. (2016) and Fisher & Wooster (2018) also describe pixel saturation in the OLI shortwave infrared bands in the contexts of active fire detection and fire radiative power

\(^2\)A theoretical, ideal object that absorbs all incident electromagnetic radiation. In thermal equilibrium, a blackbody will emit the same radiation as it absorbs.

\(^3\)Atmospheric transmissivity refers to the fraction of radiation coming into the earth that passes through the atmosphere to the ground surface. It is a unitless parameter taking values within the range of \([0,1]\), where 1 means incoming radiation fully reaches the ground.
retrieval respectively. We use the quality assessment band provided with the L8 Collection 1 Level-1 product to check for saturation. Of the 154 flares selected for study, 73 are saturated in at least one band, leaving 81 flares for temperature estimation.

We invert the Planck function for blackbody radiation to estimate temperature and emissivity using OLI’s first seven bands — four bands in the visible spectrum and three in the infrared — for each flare. Planck’s law is given by:

\[ L_\lambda = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1} \]  

(3.1)

where \( L_\lambda \) is the spectral radiance at wavelength \( \lambda \), \( T \) is the absolute temperature of the body, \( h \) is the Planck constant, \( c \) is the speed of light, and \( k_B \) is the Boltzmann constant.

The inversion was implemented using the fminsearch function in Matlab. The misfit is the Euclidean norm of the distance between the observed flare radiances and the Planck distribution radiances, summed over the optical bands:

\[ m = || \hat{L} - L || \]  

(3.2)

where \( m \) is the misfit, \( L \) is the vector of observed radiances from the seven optical bands, and \( \hat{L} \) is the corresponding vector of radiances resulting from the inversion. We then use the mean squared misfit to calculate the standard root mean square error (RMSE):

\[ r = \sqrt{\frac{1}{7} m^2} \]  

(3.3)

Since RMSE could increase mechanically with observed radiance, which differs by an order of magnitude across the pixels, we check if normalizing the RMSE affects our results. We also perform the Planck inversion using only the three infrared optical bands since the higher noise
floor and greater susceptibility to atmospheric scattering of the visible bands could compromise the inversion. Comparisons of the two sets of inversion results and the description of the RMSE normalization are given in the appendix.

3.3.2 Matching OLI and VIIRS Nightfire detections

In order to compare the flare temperatures estimated from OLI and VIIRS detections, we match them up into pairs. Since VIIRS pixels are larger than OLI pixels, we match each OLI flare with an encompassing VIIRS pixel that corresponds to a Nightfire detection. The size of VIIRS pixels ranges from 750 m at nadir to 1600 m at the swath edge, so we draw circles with radii of 375 m and 800 m centred on each Nightfire detection’s geographical coordinates. We superimpose these circles on the L8 flares to identify the corresponding Nightfire detections. Wells in the study area are often closely spaced, so many of our L8 flares have multiple, sometimes overlapping, Nightfire detection circles nearby. In such cases, we select the Nightfire detection with the closest location to the L8 flare for comparison. Because VIIRS pixels are large, multiple flares may lie within a single pixel, in which case VIIRS records the average radiance of these flares. Thus, we retain pairs of OLI and Nightfire detections only if we do not observe large neighbouring hotspots in OLI’s first shortwave infrared band that lie within 800 m of the same overlaying Nightfire detection.

3.4 Results

3.4.1 Characterization of gas flares in tight oil production using Landsat 8

Emitting at high temperatures from around 1300 K to 2700 K, gas flares are clearly observable at shortwave infrared wavelengths where their emitted radiance peaks. Flares can sometimes be observed in the thermal infrared wavelengths against the cool ground surface at night. Figure 3.2(a) compares the full resolution L8 optical (left) and thermal (right) imagery from a section around
the Missouri River. The optical image is a composite of the near infrared and the two shortwave infrared bands (5-7) while the thermal image displays the first thermal infrared band (10) in grayscale. Emitting sources can be seen clearly in the optical infrared composite. To verify that these are flares associated with oil production, we examine twelve (groups of) hotspots, labelled A through L on the optical infrared composite. Metre-resolution day time visible images of these hotspot locations in Figure 3.2(b) show clearly that oil wells or gas processing facilities exist at these sites, and that gas flares can indeed be seen in most of them. The decametre spatial resolution of Landsat images allows flares near each other to be distinguished (for example, locations G and K in Figure 3.2(b)).

### 3.4.2 Radiance distributions of gas flares

The radiance distributions for the full flare population illustrate the large number of flares and the wide range of sizes and radiances. In Figure 3.3, the four scatterplots in the diagonal (top left to bottom right) each show the bivariate distributions of flare radiance and area from one of four bands: visible red (band 4), near infrared (band 5), and shortwave infrared (bands 6-7) for the dense flare region around the Missouri River (area within red box in Figure 3.1). We refer to the size (or area) of a flare in the context of the spatially contiguous area of pixels with radiance levels above the noise floor for that spectral band. While the flares themselves are obviously smaller than 30 m diameter, the combination of scattering, ground reflection and adjacency effect results in signatures spanning several pixels in most cases. Flare area is calculated as the total area of each spatially contiguous (Rook’s case) cluster of pixels segmented using the band-specific noise floor as a threshold. To avoid spurious detections, only clusters 4 pixels and larger are considered.

The six scatterplots above the diagonal in Figure 3.3 provide pairwise comparisons of the radiances in the four bands. The three scatterplots above the diagonal in the first row are for band 4 against bands 5, 6, and 7, the two scatterplots above the diagonal in the second row are for band
5 against band 6 and 7, and the single scatterplot above the diagonal in the third row shows band 6 against band 7. Each plotted point could represent more than one pixel; the warmer its colour, the more pixels it represents. Because the images used are acquired at night, most pixels record noise. The maximum radiance from noise sources, i.e. the noise floor, is clearly evident for each band. The wavelength at which radiance peaks (peak radiance wavelength) in each band can also be clearly seen.

Below the diagonal are six graphs, each of which plots the radiance spectra recorded at one individual pixel. These pixels represent six of the largest flares in the selected area around the Missouri River (area within red box in Figure 3.1). The spectrum of each flare is shown with its best fit Planck distribution curves (as described in section 3.1) on three dates in 2016 (indicated by three different colours). Although the fits are shown on semilog plots for clarity, the temperature and emissivity estimates were made from untransformed at-sensor radiances.

Together, the plots in Figure 3.3 show the varying noise floors and peak radiances for four OLI bands, the asymptotic increase in peak radiance with flare area, and the generally plausible fits obtained for flares with peak radiances spanning 3 orders of magnitude.

### 3.4.3 Flare persistence and area

The temporal persistence and size distribution of the flare population are illustrated by comparing three of the L8 image acquisitions from 2016 (on 9 May, 14 September, and 16 October) for a densely drilled area south of the Missouri River (path/row 126/217). In Figure 3.4, the top panel shows flares from this region (indicated by the red box in Figure 3.1). This is a temporal persistence map, derived from the decision tree classification (inset) of the segmented areas of spatially contiguous pixels that have radiances above the noise thresholds (described in section 4.1.1). Pixels in blue, green, or red appear on only one of the three dates. Those in yellow, magenta, or cyan occur on two dates. White pixels are present on all three dates. The wide-ranging colours
and areas of contiguous pixels in this image thus illustrate the large number of flares of varying size and their persistence over time.

The three scatterplots in Figure 3.4 show how flare area compares between each date pair. Each scatterplot shows the subset of flares detected in the same location on both dates. As in Figure 3.3’s scatterplots, warmer-coloured points represent more pixels. Distributions of flare areas, including those present on only one or two dates, are obtained using the decision tree classification (inset top). It is evident that the vast majority of flares detected on multiple dates are at the small end of the size distribution, with a few flares either growing or shrinking between dates, and almost no flares larger than \(10^4\) m\(^2\) on any two dates. The scatterplots are truncated at \(10^5\) m\(^2\) to show the low end of the distribution more clearly.

At the bottom left of Figure 3.4 are rank-size plots of the flare area distributions of the full study area. On all three dates, the flare population scales similarly, despite the smaller number of flares on 9/14. The rank-size plots have slopes > -1, indicating that the distributions are dominated by large numbers of small flares.

### 3.4.4 Temperature estimates of gas flares

Applying the Planck inversion process (described in section 3.1) on the L8 flares yields temperature estimates between 1500-2200 K, within the expected range for gas flares. However, a range of misfits are observed. Figure 3.5 presents examples of some of the best fits and Figure 3.6 shows examples of some poor fits. RMSE of the fitted curves are generally an order of magnitude lower than peak radiance. Distributions of temperature, emissivity, and misfit are given in the appendix.
3.4.5 Comparison with VIIRS

3.4.5.1 Difference in detection capability

Comparison of near-coincident acquisitions of L8 and VIIRS illustrates the benefit of decametre resolution in areas with high densities of smaller flares. Figure 3.7 consists of three panels, each showing the same selected area south of the Missouri River on 9 May 2016 (area within red box in Figure 3.1). This figure compares an L8 OLI infrared band composite (top panel) with a composite of two OLI shortwave infrared bands and the VIIRS DNB co-registered (centre), and a composite of the VIIRS thermal infrared (in red), mid-infrared (green), and shortwave infrared bands (blue). OLI imaged the study area at 04:25 GMT (23:25 local time) and VIIRS imaged it at 10:24 GMT (05:24 local time). Also shown in the bottom panel are locations of NOAA Nightfire product detections for the same night.

The OLI infrared composite (top) is displayed as Log10 radiance with a linear contrast stretch applied separately for each band, between each band’s non-zero minimum and maximum radiance. The broad diagonal stripes in the background are noise from the sensor. The largest flares stand out against the noise background with the second shortwave infrared (SWIR2) band (red) imaging a larger halo than the near infrared band (blue) as a combined result of the lower SWIR2 noise floor (discussed in section 4.1.1) and that flares emit more in the shortwave infrared range. Smaller flares recede into the noise floor in the OLI infrared composite (top) but are clearly visible in the OLI shortwave infrared + VIIRS DNB composite (centre) when the OLI bands are stretched from the noise floor to maximum radiance (so that the image does not display the noise). VIIRS DNB records in the visible and near infrared range. Thus, at night, it detects not only gas flares but also electric lights. For example, the city of Williston located in the northwest corner of the composite appears as a cluster of bright VIIRS DNB pixels without bright OLI pixels. In areas with both VIIRS DNB and OLI detections, the VIIRS DNB bright pixels are likely recording
radiation from both gas flares and electric lights at the oil well pad. When the composite is viewed at full resolution, it is immediately apparent that OLI resolves many more small flares than VIIRS DNB and that many of the 750 m VIIRS DNB pixels image multiple flares. In contrast, the VIIRS composite (bottom) resolves many fewer flares in the shortwave infrared (blue) and only the few largest flares (or the heated ground surface surrounding the flare) in the thermal infrared (red). Because radiance from gas flares peak in the shortwave infrared wavelengths, this composite shows the flares detected by VIIRS. The NOAA Nightfire product (green dots) detects all of the flares visible in the shortwave infrared band of the VIIRS composite, as well as several others that are not obvious in the full range linear stretch. It is clear that OLI’s 30 m resolution provides a far more detailed depiction of the flare population in areas with densely-spaced wells and small flares.

3.4.5.2 Difference in estimated gas flare temperatures

Flare temperatures estimated from L8 OLI tend to be higher than NOAA’s VIIRS Nightfire estimates. We were only able to use a small number of flares for this comparison because for many of the L8 OLI flares, Nightfire either does not detect them or does not provide temperature estimates. Figure 3.8 plots Nightfire temperature estimates against L8 OLI temperature estimates for the 16 flares in our study with corresponding Nightfire detections on the same dates. In Figure 3.8, the red points represent isolated flares — L8 flares that only have one overlaying Nightfire detection and no neighbouring hotspots. The green points represent flares that have multiple Nightfire detections nearby, or small neighbouring hotspots within the overlaying Nightfire detection. Whether considering only the isolated flares (red points in Figure 3.8) or all flares with corresponding Nightfire detections (red and green points in Figure 3.8), we observe that the L8 estimates are more often higher than the Nightfire estimates. This is consistent with VIIRS viewing a much larger area on the ground at any given instant, i.e. having a much larger instantaneous field of view (IFOV), compared to L8 OLI. Because a flare is smaller than a VIIRS pixel, it is a subpixel feature. Each
VIIRS pixel records the average radiation from all features within the pixel. In this setting, each VIIRS pixel thus averages out the high radiation from flares with the low radiation from the ground surface.

If the OLI’s smaller IFOV yields more accurate temperature estimates than VIIRS, this may suggest that sensor spatial resolution and spatial density of wells could be important factors for further investigation in estimations of flare volume based on flare temperature. For instance, both Anejionu et al. (2015a) and Elvidge et al. (2016) use statistical methods to obtain gas flare volumes from estimated gas flare temperatures. Elvidge et al. (2016) first calculate the radiant heat (RH) from the estimated temperature and emissivity using the Stefan-Boltzmann law. Thereafter, they use linear regression to establish the relationship between flare RH to flare volume for gas flares whose volumes are known (obtained from a proprietary database). Anejionu et al. (2015a) directly find the relationship between flare radiance and flare volume, also using linear regression. They do so for a set of flares whose volumes are recorded at sample flow stations. Future work that uses similar techniques to develop a comprehensive dataset of estimated flare volumes from Landsat could provide a complementary source of calibration for existing satellite products using lower-resolution imagery.

3.5 Conclusion and discussion

Production from tight oil formations have become the dominant source of crude oil in the US. Gas flaring is a part of the crude oil production process (both conventional and unconventional) where natural gas is disposed of by combustion after it is extracted with the crude oil. This is a significant polluting activity and is conducted around the world. While previous studies have used satellite imagery to study gas flaring, they have centered on the context of conventional oil production. This paper instead focuses on the use of satellite imagery to detect the smaller, more densely spaced gas
flares typically found in tight, a type of unconventional, oil production. We investigate the use of nighttime Landsat 8 OLI imagery, which has a spatial resolution an order of magnitude higher than the predominant satellite product, VIIRS, in detecting gas flares.

We find from examining Landsat OLI imagery of gas flares associated with oil production from the Bakken formation in North Dakota that most of the gas flares are indeed small, densely-spaced, and persistent. Moreover, many of these flares are not detected by either VIIRS DNB or the VIIRS Nightfire product. This suggests that the majority of gas flared from this and other tight oil formations, where wells tend to be drilled close to each other, is not detected by hectometre-resolution sensors like VIIRS.

We also show that temperatures of flares detected in Landsat OLI imagery can be estimated using Planck inversion. These temperatures could subsequently be used to estimate flare volumes such as how previous studies have done via statistical calibration. If small flares in tight oil settings are generally persistent on time scales of months, as suggested by the bivariate distributions in Figure 3.4, intermittent Landsat OLI acquisitions could provide complementary estimates of small flare volumes undetected by hectometre-resolution sensors like VIIRS. Moreover, because the distance between flares (fired from different sources) is greater than 30 m, detections within a pixel on different dates can reliably be treated as the same flare. This is not true with products of hectometre resolution, because new flares could be started within the same area viewed by the satellite detector (IFOV) at a later date. Temporal analyses of individual flares that are closely spaced would therefore benefit greatly from decametre-resolution data provided by Landsat.

If nighttime acquisitions of Landsat imagery become operational in the future, those acquisitions could be useful in identifying gas flares in densely-drilled tight formations like those in North Dakota and Texas. Landsat could supplement studies of flares that use lower-resolution operational imagery. For higher frequency analysis, nighttime Landsat data could also be combined with images captured by other decametre-resolution optical sensors, such as those aboard Sentinel
2 and Worldview-3 to effectively shorten revisit periods. As production of oil resources from tight formations increases worldwide (EIA 2015), and monitoring these emissions increases in importance, observations of smaller, more closely-spaced gas flares using decametre-resolution imagery may become a necessary supplement to lower-resolution sensors.

Large numbers of undetected small flares could be contributing significantly to greenhouse gas emissions, particularly as tight oil production has increased globally. Masnadi et al. (2018) find that flaring accounts for a significant, yet reducible, fraction of the carbon intensity associated with the production of fossil fuels. They report that the World Bank Global Gas Flaring Reduction (GGFR) partnership has found a nearly continuous increase in global gas flaring between 2010 and 2016. However, these estimates are based, in part, on VIIRS observations provided by NOAA, suggesting that they may be underestimates if the contribution of large numbers of small flares are not detected. If satellite monitoring of flaring is to be used operationally going forward, it will be important to account for the contribution of smaller flares not detected by hectometre-resolution sensors like VIIRS.
3.6 Figures & Tables

Figure 3.1: Index map of study area

Daytime Landsat 8 scenes and nighttime VIIRS composite of study area, in western oil-rich North Dakota. Landsat 8 scenes are from path 34, rows 26 and 27, on 4 May 2016; false colour composite on the left and thermal band in grey scale on the right. Red box indicates illustrative area in Figure 2. The VIIRS image is a composite of the DNB from three different months in 2016 (Red: December, Green: September, Blue: May).
Figure 3.2: Validation of selected flares detected by Landsat OLI

(a) Illustrative area from path/row 126/217, on 14 Sep 2016. Individual gas flares can be identified in a composite of NIR, SWIR1, SWIR2 bands (left). Selected flares, labelled A-L, are validated using Google Maps in Fig. 3(b). Thermal band (right) shows the landscape south of the Missouri River. Large flares can be seen. (b) Validation of flares identified in Fig. 3(a) using Google Maps (data from Google, DigitalGlobe, USDA Farm Service Agency, CNES/Airbus, Landsat/Copernicus). Nighttime Landsat OLI resolution is sufficiently fine to distinguish flares from adjacent well pads (F, G, J, K). C is labelled as a gas plant in Google Maps.
Bivariate radiance distributions for 9 May 2016 Landsat 8 overpass with example Planck distribution fits. The area-radiance distributions of individual flares for OLI bands 4-7 along diagonal show peak radiance increasing asymptotically with area of flare halo detected above noise floor. Band pair correlations (above diagonal) show almost all flares peaking in band 6 (1.60 μm). Planck fits to single pixel spectra of hottest flares generally give plausible temperatures but wildly varying emissivities.
Figure 3.4: Flare distributions in 2016

Smaller flares are more persistent across three dates in 2016, but the total area of flares decreases with increasing persistence. Very few large flares persist beyond one date. Rank-size distributions of individuals scale consistently across dates, dominated by smaller flares.
Figure 3.5: Examples of good Planck curve fits

Path/row 126/217 pixel #2 on 20160626

Temperature = 1903K  RMSE = 0.02x10^{-3}
Emissivity = 0.08

Path/row 126/217 pixel #9 on 20160914

Temperature = 1716K  RMSE = 0.01x10^{-3}
Emissivity = 0.02

Path/row 126/218 pixel #12 on 20160509

Temperature = 1835K  RMSE = 0.01x10^{-3}
Emissivity = 0.04

Path/row 126/217 pixel #20 on 20160509

Temperature = 1860K  RMSE = 0.02x10^{-3}
Emissivity = 0.18

Path/row 126/217 pixel #6 on 20161203

Temperature = 1831K  RMSE = 0.02x10^{-3}
Emissivity = 0.23

Path/row 126/218 pixel #6 on 20170410

Temperature = 1985K  RMSE = 0.05x10^{-3}
Emissivity = 0.15
Figure 3.6: Examples of poor Planck curve fits

Path/row 126/217 pixel #11 on 20160626

\[
\text{Radiance (W/cm}^2\text{sr)}
\]
\[
\begin{array}{c}
0 \\
0.2 \\
0.4 \\
0.6 \\
0.8 \\
1 \times 10^{-3}
\end{array}
\]
\[
\begin{array}{c}
400 \\
800 \\
1200 \\
1600 \\
2000 \\
2400
\end{array}
\]
\[
\text{Wavelength (nm)}
\]

\[\text{Temperature} = 2047K \quad \text{RMSE} = 0.04 \times 10^{-3}\]
\[\text{Emissivity} = 0.07\]

Path/row 126/217 pixel #23 on 20160509

\[
\text{Radiance (W/cm}^2\text{sr)}
\]
\[
\begin{array}{c}
0 \\
0.2 \\
0.4 \\
0.6 \\
0.8 \\
1 \times 10^{-3}
\end{array}
\]
\[
\begin{array}{c}
400 \\
800 \\
1200 \\
1600 \\
2000 \\
2400
\end{array}
\]
\[
\text{Wavelength (nm)}
\]

\[\text{Temperature} = 1891K \quad \text{RMSE} = 0.13 \times 10^{-3}\]
\[\text{Emissivity} = 0.32\]

Path/row 126/217 pixel #9 on 20161203

\[
\text{Radiance (W/cm}^2\text{sr)}
\]
\[
\begin{array}{c}
0 \\
0.2 \\
0.4 \\
0.6 \\
0.8 \\
1 \times 10^{-3}
\end{array}
\]
\[
\begin{array}{c}
400 \\
800 \\
1200 \\
1600 \\
2000 \\
2400
\end{array}
\]
\[
\text{Wavelength (nm)}
\]

\[\text{Temperature} = 1983K \quad \text{RMSE} = 0.04 \times 10^{-3}\]
\[\text{Emissivity} = 0.07\]

Path/row 126/218 pixel #5 on 20170410

\[
\text{Radiance (W/cm}^2\text{sr)}
\]
\[
\begin{array}{c}
0 \\
0.2 \\
0.4 \\
0.6 \\
0.8 \\
1 \times 10^{-3}
\end{array}
\]
\[
\begin{array}{c}
400 \\
800 \\
1200 \\
1600 \\
2000 \\
2400
\end{array}
\]
\[
\text{Wavelength (nm)}
\]

\[\text{Temperature} = 1921K \quad \text{RMSE} = 0.05 \times 10^{-3}\]
\[\text{Emissivity} = 0.09\]

Path/row 126/218 pixel #8 on 20160914

\[
\text{Radiance (W/cm}^2\text{sr)}
\]
\[
\begin{array}{c}
0 \\
0.2 \\
0.4 \\
0.6 \\
0.8 \\
1 \times 10^{-3}
\end{array}
\]
\[
\begin{array}{c}
400 \\
800 \\
1200 \\
1600 \\
2000 \\
2400
\end{array}
\]
\[
\text{Wavelength (nm)}
\]

\[\text{Temperature} = 1718K \quad \text{RMSE} = 0.18 \times 10^{-3}\]
\[\text{Emissivity} = 0.3\]
Comparison of Landsat 8 and VIIRS on 9 May 2016. Flare radiance varies 1000x above the OLI noise floor. Many more flares are detected by VIIRS DNB than IR products, but many flares detected by OLI are not detected by VIIRS DNB.
Comparison of temperature estimates from the VIIRS Nightfire product and Landsat 8 data. Both sets of estimates are obtained from Planck inversion. Landsat estimates tend to be higher than Nightfire estimates.
Appendix A: Planck inversion estimates and misfit

Our results from the Planck inversion do not change significantly if we leave out the visible bands, which have higher noise floors and are subject to greater atmospheric scattering. Figure A1(a) shows that the RMSE distributions for inverting with all optical bands and with only the optical IR bands are similar. Normalizing the RMSE by the magnitude of observed radiances does not affect our results either. RMSE increases in peak radiance observed (Figure A1(b)), and is unrelated to estimated temperature (Figure A1(c)).

While both emissivity and temperature are estimated from the curve fit-optimizing procedure, the two variables appear unrelated (Figure A2). The estimated emissivity ranges from 0.005-0.715. This wide range of emissivity is consistent with the findings of previous studies that indicate flame emissivity depends on the thickness of the flame, and composition of the fuel source (Àgueda et al. 2010). Alternatively, as noted in Elvidge et al. (2016), since flares are sub-pixel sources, they appear as graybodies. The emissivity could therefore indicate the flare size relative to the IFOV of the sensor. One or both of these factors could explain the wide range of emissivity values estimated in our study.
Figure A.1

(a) RMSE histograms of Planck curve fits using all OLI optical bands (1-7) and optical IR bands only (bands 5-7). (b) RMSE increases in peak radiance observed. (c) RMSE is not related to estimated temperature.
Estimated temperatures and emissivities from Planck inversion using all OLI optical bands. Emissivity is not related to estimated temperature.
Bibliography


