Understanding Excess Emissions from Industrial Facilities: Evidence from Texas

Nikolaos Zirogiannis[*+], Alex J. Hollingsworth[+], David M. Konisky[+]

[+]School of Public and Environmental Affairs, Indiana University Bloomington, Bloomington, Indiana 47405, Unites States

[*]Corresponding author: 1315 East Tenth Street, Bloomington Indiana-47405, USA, nzirogia@indiana.edu, Tel: 812-856-8323

Draft: January 22, 2018
Forthcoming in *Environmental Science & Technology*

Understanding Excess Emissions from Industrial Facilities: Evidence from Texas

Abstract

We analyze excess emissions from industrial facilities in Texas using data from the Texas Commission on Environmental Quality. Emissions are characterized as excess if they are beyond a facility's permitted levels and if they occur during startups, shutdowns, or malfunctions. We provide summary data on both the pollutants most often emitted as excess emissions and the industrial sectors and facilities responsible for those emissions. Excess emissions often represent a substantial share of a facility's routine (or permitted) emissions. We find that while excess emissions events are frequent, the majority of excess emissions are emitted by the largest events. That is, the sum of emissions in the 96^{th}-100^{th} percentile is often several orders of magnitude larger than the remaining excess emissions (i.e. the sum of emissions below the 95^{th} percentile). Thus, the majority of events emit a small amount of pollution relative to the total amount emitted. In addition, a small group of high emitting facilities in the most polluting industrial sectors are responsible for the vast majority of excess emissions. Using an integrated assessment model, we estimate that the health damages in Texas from excess emissions are approximately $150 million annually.

1 Introduction

Hurricane Harvey made landfall in Texas on Friday August 25^{th}, 2017. Within just a few days, some of the state received over 50 inches of rainfall, inflicting both human casualties as well as devastating infrastructure and property damages. The hurricane caused many industrial facilities in the state to shut down operations and restart them after the rain and flooding subsided. From August 23^{rd} to September 19^{th}, shutdowns and startups (as well as other malfunctions related to Hurricane Harvey) resulted in excess emissions of 1,927 tons of criteria pollutants (616 tons of CO, 735 tons of VOCs, 435 tons of SO_2, 141 tons of NO_X) across the entire state of Texas[1]. These excess emissions represent 7% of criteria pollutant excess emissions from all Texas facilities in 2016 due to shutdowns, startups and malfunctions. Although these emissions resulted from a singular, extreme weather event, the same types of emissions (i.e. due to startups, shutdowns, and malfunction) occur on a regular basis during the routine operation of many industrial facilities.

The release of air pollutants is an expected byproduct of many industrial processes. In the United States, the Clean Air Act (CAA) requires that the Environmental Protection Agency (EPA) set ambient air quality standards for a variety of pollutants and emissions limits or technology-based standards for many others. Pollution sources are then required to meet source-specific targets set forth in government-issued permits. Overall, these permits aim to regulate normal operations – for example, the emissions that result from a power plant burning coal to generate electricity or a refinery processing crude oil to make gasoline. However, these types of facilities also often emit "excess emissions," defined by the EPA as emissions "*that occur during the startup, shutdown, malfunction or other modes of source operation, i.e., emissions that would be considered violations of the applicable emission limitation but for an impermissible automatic or discretionary exemption from such emission limitation*."[2].

Excess emissions can be difficult to control because they are attributed to unexpected or unavoidable circumstances. For example, when a facility has to shut down due to a power outage or natural disaster, and then restart its operations, emission levels will likely increase. This increase is due to that fact that pollution control devices require a constant and high temperature environment to function properly[3]. Similarly, if a control device malfunctions, emissions will increase until the device is repaired. In any case, excess emissions are violations of the CAA per EPA policy because they go beyond authorized limits delineated in a facility's permit[2].

In this paper, we use data obtained from the Texas Commission of Environmental Quality (TCEQ) to examine the pattern of excess emissions in the state. We focus on differences across industrial sectors for emissions of criteria pollutants, Volatile Organic Compounds (VOCs) as well as some key hazardous air pollutants (HAPs). In addition, we use topic modeling to delineate the typical causes of excess emissions events, and an integrated assessment model to approximate the monetized value of health damages from the excess emissions that cause either direct or indirect particulate matter (PM).

Between 2004-2015 excess emissions events in Texas resulted in releases of 104,202 tons of VOCs, which are equivalent to 7.5% of routine (permitted) emissions from all facilities reporting to the state's Emissions Inventory. For CO and SO_2, the relevant figures during the same time-period are 89,202 tons (2.03% of routine emissions) and 123,823 tons (1.84% of routine emissions), respectively. The causes of these events include, unplanned or scheduled startups and shutdowns that exceeded emissions thresholds, weather-induced power outages, and malfunctions due to poor maintenance. We estimate that the monetized value of health impacts, just from direct PM emissions and indirect PM caused by SO_2 and NO_x to be $148 million in 2015 alone.

The remainder of the paper is organized as follows. In the next section, we briefly discuss the relevant regulatory history regarding excess emissions, and examine recent changes in EPA policy that affect how the agency interprets excess emission. In section 3, we review the limited existing literature on excess emissions. Section 4 describes the excess emissions data we collected from the TCEQ, and analyzes patterns over time and across sectors, as well as the causes of excess emission events. Last, in section 5, we provide monetary estimates of health damages. We conclude with a discussion of the implications of our research for federal and state policies regarding excess emissions.

2 Policy background on excess emissions

Air pollution control in the United States is a shared responsibility between the federal government (i.e., the EPA) and the states. Under the CAA, the EPA sets maximum ambient concentration levels for six criteria air pollutants – CO, lead, NO_x, ozone, PM (2.5 and 10), and SO_2 – as part of the National Ambient Air Quality Standards (NAAQS) program. Based on these standards, the EPA then designates areas of the country meeting the NAAQS as in attainment, and areas failing to meet the NAAQS as nonattainment. Under the CAA, every state is required to submit a State Implementation Plan (SIP) that specifies how it will meet NAAQS requirements, particularly in nonattainment areas. Once the EPA approves a SIP, the state agency begins to implement its strategy, which includes issuing permits to stationary air sources that contain specific emission limits.

Since the outset of the SIP program in the 1970s, states have incorporated policies regarding excess emissions. Initially the EPA did not consider periods of startup, shutdown and maintenance of equipment as being part of a facility's "normal" operations. As a result, many SIPs included automatic exemptions and affirmative defense provisions that effectively shielded facilities from enforcement[4]. In a 1977 guidance, however, the EPA clarified that automatic exemptions were not allowed. Yet, even though the EPA was concerned that granting sources automatic exemptions for excess emissions would compromise the NAAQS, it did not make enforcement of the 1977 guidance a priority[5].

The EPA issued memoranda in 1982 and 1983 clarifying its policies on excess emissions. The agency encouraged states to use an "enforcement discretion approach" to address excess emissions resulting from accidental releases[5,6], whereby states could characterize such emissions as allowable if they were clearly due to an unavoidable malfunction. In cases where excess emissions resulted from *scheduled* startup, shutdown, and maintenance activities, the EPA memoranda made clear that these were considered part of the normal operation of a facility, since they were predictable events. Therefore, any excess emissions during scheduled events would be considered a violation of a facility's permit. This basic approach was reaffirmed and expanded upon in a 1999 memorandum in which the EPA delineated the parameters for a new "affirmative defense" approach that states could incorporate into their SIPs. If specific criteria were met, states could provide air sources with enforcement relief. There were important limitations, however. States could only apply affirmative defense provisions to requests for penalties (not requests for injunctive relief). Moreover, affirmative defense was disallowed for sources that had the potential to cause an exceedance of NAAQS, and to get around any limitations that derive from either New Source Performance Standards or National Emissions Standards for HAPs [7,8].

2.1 Current policy

For about two decades, the EPA did not take significant steps to systematically bring SIPs in line with its own interpretation of excess emissions regulations. In practice, this meant that many states had provisions in their SIPs that conflicted with stated EPA policy, including automatic exemptions and so-called director's discretion exemptions, which were cases where a state agency head explicitly excused what would otherwise have been a CAA violation. This changed in 2013 when the agency proposed a revision to its policy on excess emissions in response to a petition filed by the Sierra Club. The Sierra Club had argued that many states had provisions in their SIPs that did not observe EPA's guidance on automatic exemptions, including provisions endorsed by the EPA that limited the jurisdiction of federal courts to impose civil penalties. In its 2015 final ruling on the issue, the EPA found that 36 states had provisions in their SIPs with regards to excess emissions from startup, shutdown and malfunction (SSM) events that were "*substantially inadequate to meet CAA requirements*"[2]. The agency requested that these states revise their SIPs to, where relevant, eliminate automatic exemptions, director's discretionary exemptions, overly broad enforcement discretion, and all affirmative defense provisions. The elimination of affirmative defense provisions was particularly significant, since the EPA previously allowed them in some circumstances in instances of malfunctions.

This regulatory history illustrates the complexity and ambiguity surrounding excess emissions. The challenge comes at distinguishing between events that a facility could have reasonably prevented through careful planning and maintenance and those that are truly unavoidable. This complexity is confounded by the lack of data, since with the exception of Texas, Louisiana and Oklahoma, states do not systematically collect data on excess emissions in a fashion that distinguishes them from routine emissions or, if they do, do not make that information publicly available. Consequently, it is virtually impossible to systematically discern the magnitude of excess emissions relative to routine emissions as well as the extent to which excess emissions contribute to poor air quality and adverse health outcomes. In this paper, we explore the most comprehensive dataset on excess emissions, which is based on information reported by major air sources in Texas to the TCEQ. Before turning to these data, we first review the limited existing literature on excess emissions.

3 Literature Review

Scholars have examined excess emission events in two different ways. First, atmospheric scientists have investigated the degree to which excess emissions of VOCs and NO_x impair air quality. Several studies conducted in the Houston-Galveston region of Texas have found that these emissions, depending on atmospheric conditions, can result in elevated concentrations of ozone [9–11]. A separate set of studies have found that excess emissions can result in higher ambient concentrations of fine PM[12,13]. Furthermore, scholars have recorded large discrepancies between the emissions documented in the Texas emissions inventory and those measured directly through observational studies. The Texas Air Quality Study (Texas AQS) of 2000, for example, found that the state emission inventory underestimated the amount of highly reactive volatile organic compounds (HRVOC) emitted from petrochemical facilities by 1 – 2 orders of magnitude[14]. The 2006 Texas AQS found that the emission inventory underestimated emissions for ethene and propene by a factor of 10 and 11 respectively[14]. Collectively, these studies provide evidence that excess emissions can have measurable impacts to air quality.

Most relevant to our work is a second stream of research that has explored patterns of excess emissions across states and industries. A report by the Environmental Integrity Project (2004)[15] was an early attempt to document the way states keep track of and regulate excess emissions. Of the facilities it analyzed, EIP found that CO was the highest emitted pollutant (~48% of total excess emissions) followed by VOCs (~24%), SO_2 (~23%) and very low amounts of NO_x and H_2S. In addition, EIP[15] identified several natural gas plants that in 2003 released excess VOC emissions that were substantially higher than the routine emissions during the previous year. EIP reached similar conclusions in a subsequent report that analyzed data from Texas for the years 2014 and 2015[16]. In this study, EIP found that most of the excess emissions of SO_2 and VOCs deriving from malfunctions and maintenance activities were from oil and gas extraction sites, chemical manufacturing plants, oil refineries, and power plants.

Two other studies have focused their attention on excess emissions from Texas oil refineries. McCoy et al. (2010) [17] find that 96% of reported excess emissions pertained to criteria pollutants (in which they include VOCs), while 63% of the total number of emissions events were concentrated in four areas of Texas (namely Port Arthur, Corpus Christi, Houston and Texas City). The authors also calculated the ratio of upset over total emissions, and found that 30% of the refineries they analyzed had excess SO_2 emissions that exceeded 10% of their total emissions.

Ozymy and Jarrell (2011)[18] conducted a similar analysis for 18 Texas refineries for the 2003-2008 period pointing to the fact that a small number of large excess events released upwards of 500 thousand pounds of SO_2, CO, Propane and Isobutane. When comparing excess and routine emissions from the Toxics

Release Inventory (TRI) the authors find that a single excess event can overwhelm the total annual routine emissions of a facility for some toxic pollutants.

Our paper relies on similar data as much of this prior research, but provides a more comprehensive analysis. Specifically, we analyze excess emissions across sectors, facilities, multiple pollutants, and over more than a decade of time. This approach provides a more complete picture of the nature of excess emissions, in terms of the frequency of occurrence and magnitude of releases. Moreover, we use topic modeling to classify typical common causes of events, as well as an integrated assessment model to provide a monetary estimate of the health damages that can be attributed to excess emissions. We describe the data in the next section.

4 Data

The data used in this paper come from TCEQ's Emissions Inventory (EI) and Air Emissions and Maintenance Events (AEME) datasets [1,19]. The former includes annual totals for more than 2,000 pollutants from major sources (i.e., CAA Title V facilities) in Texas [20,21]. Facilities report the following annual amounts in the EI dataset: 1) Routine emissions (i.e., permitted emissions); 2) Emissions events (EE); and 3) Emissions attributed to scheduled startup, shutdown or maintenance (SMSS) events. Taken together, EE and SMSS emissions constitute the total amount of excess emissions. The TCEQ defines an Emissions Event as "*any upset event or unscheduled maintenance, startup or shutdown activity…that results in unauthorized emissions*" [22]. Emissions events result in releases from a stack as opposed to fugitive emissions that "*could not reasonable pass through the stack*" [22]. An SMSS event, is a scheduled event that is expected to exceed authorized emissions levels and for which a facility is required to provide prior notification and submit a final report to the TCEQ [22].

The TCEQ introduced a rule in 2003 that requires all facilities in the state (not just Title V facilities) to report EE and SMSS emissions within 24 hours of their occurrence provided they surpass an emissions threshold [23]. Upon receiving an initial report of an excess emission event from a facility, the TCEQ posts that information on its web-site making it immediately available to the public. The reporting facility has two weeks to submit a final report where it can provide updated information on the event. The compilation of those events across all years (i.e. 2002 until April of 2017) constitute the AEME dataset we obtained from the TCEQ. While facilities are required to report emissions events that exceed a "reportable" quantity in the AEME dataset, the same is not true for the EI dataset. There, Title V facilities are required to report emissions from both "reportable" and "non-reportable" events (i.e. events below the emissions threshold). Because emissions in the EI are reported at the end of the year, facilities might update the excess emissions information they provide to the EI. As a result, at times, there can be discrepancies in the annual sum of excess emissions between the AEME and EI datasets[24]. Finally, there is no information on routine emissions in the AEME dataset.

4.1 Excess Emissions from all facilities

Table 1 captures the magnitude and severity of excess emissions in Texas compiling information on EE and SMSS emissions for criteria pollutants, VOCs, and some important HAPs for the period 2004-2015 for all Title V facilities. Excess emissions of SO_2 during this period were 123,823 tons, followed by VOCs (104,202 tons) CO (89,202 tons), NOx (20,227 tons), PM10 (9,572 tons) and PM2.5 (6,070 tons). These are large amounts of pollution, accounting for the equivalent of up to 2% of routine (i.e., permitted) emissions for most pollutants, and 7.5% for VOCs. Among the VOCs with highest levels of excess emissions, propane is at the top with 12,081 tons, representing 16.1% of routine emissions. Amongst the 463 HAPs, hexane has the highest levels of excess emissions. For all pollutants depicted in Table 1 (with the exception of NO_x) the majority of excess emissions come from Emissions Events (as opposed to SMSS).

4.2 Excess emissions by industrial sector

As of 2004 there were 3,158 facilities from 231 industrial sectors reporting excess emissions in the EI dataset. Figure 1 displays on the vertical axis tons of excess emissions for criteria pollutants, VOCs and benzene (one of the most prevalent HAPs in excess emissions), while the horizontal axis captures the cumulative number of excess events from all facilities in a given industry. Each circle in the six panels of Figure 1 represents a different industrial sector. The area of each circle captures the ratio of excess over routine emissions for that sector over the 2004-2015 period. The numerical value of the ratio is shown in red for the five sectors with the highest number of cumulative excess emissions across all pollutants. For example, the petroleum refining sector released 18,109 tons of carbon monoxide (CO) excess emissions which represented 10% of the sector's routine emissions during 4,463 events. Of particular interest is the fact that the Natural Gas Liquids sector emitted 77,429 tons of SO_2 during 8,057 events for a ratio of 58% of excess over routine emissions. The refining sector is of particular interest since a very small number of facilities (28 refineries in 2015) release large amounts of excess emissions. The refining sector is the largest emitter of excess VOCs, second largest in excess $PM_{2.5}$, benzene and SO_2, third largest in CO and fourth in NO_x.

4.2.1 Excess emissions from refineries

In this section, we take a closer look at excess emissions from oil refineries, an industrial sector that emits disproportionately high levels of excess emissions per facility. Figure 2, provides information similar to that of Figure 1, only this time at the facility level for the top five polluting refineries. The vertical axis in Figure 2 plots routine emissions with each circle being a specific refinery, and the horizontal axis captures the total number of excess emissions events per facility over the 2002-2015 period. The size of each circle captures the ratio of excess over routine emissions. Figure 2 identifies the top five emitting refineries and uses green labels to distinguish them. These five refineries have the highest routine emissions and, often, the largest number of excess emissions events.

4.3 Distribution of excess emissions

One of the unique characteristics in the pattern of excess emissions is the skewness of their distribution. Figures S14-S17 (in the Supporting Information) plot the skewness parameter of the distribution of each pollutant by industry. In all cases, the values of the skewness parameter are indicative of a distribution that has a small number of events that emit large amounts of pollutants. This fact is further substantiated in Tables S4-S8 (in the Supporting Information) which show the percentiles of the excess emissions distributions by industry. In most industries the median events released less than 1 ton of a pollutant, while the maximum event often released over 1,000 tons. Figures S18-S34 (in the Supporting Information), highlight the fact that a few extreme events dominate the excess emissions distributions. The blue bars in each of those Figures show the total amount of excess emissions (by year) from the bottom of the distribution up to (and including) the 95th percentile. The red bars show the total amount of excess emissions from the 96th percentile to the top of the distribution. The green bars capture the amount released from the single largest event in each year. In the vast majority of cases in Figures S18-S34, the total excess emissions released from the top 5% of the distribution are *larger* than the total excess emissions released from all other events combined. This highlights the fact that several extreme events in each year dominate the distribution of excess emissions. A similar pattern has been documented by Brandt et al. (2016)[25] in the case of fugitive methane emissions from natural gas systems, where the largest 5% of leaks represent upwards of 50% of the total amount of leakage.

4.4 Important polluters

The skewness in the distribution of excess emissions can be traced to events occurring in a small number of facilities. Figure 3, shows total amounts of CO excess emissions for the six most polluting (in terms of excess CO emissions) refineries. The blue dotted line in each panel of Figure 3 traces the annual totals of CO excess emissions in each refinery, while the red solid line shows the number of excess emissions events in each refinery in each year. Those 6 refineries emitted 77% of the total excess CO released from all refineries in Texas between the period 2002-2016. In addition, 35% of all CO excess emissions events that occurred during the same period happened in those six refineries. Figures S35-S38 (in the Supporting Information) illustrate the top six polluting refineries for the remaining criteria pollutants. The Exxon Mobil refineries in Baytown and Beaumont are consistently among the top 6 most polluting refineries in four out of the five pollutants depicted in Figure 3 and Figures S35-S38. Tables S9-S12 (in the Supporting Information), indicate that in all of the top polluting industries, a few key facilities are responsible for the bulk of excess emissions. Detailed information on the top polluting facilities for other key industrial sectors is presented in Figures S39-S50 in the Supporting Information.

4.5 Causes of excess emissions events

An additional piece of information in the AEME dataset is a description of the cause of the excess emissions events, provided by each facility as part of their report to the TCEQ. We analyze these descriptions using a three-step approach that incorporates structural topic modeling (STM)[26]. First, we find common groupings of words that organize into distinct topics. Second, we determine how well this set of topics explains the observed excess emission descriptions. Finally, we determine the topics that are most likely related to unexpected weather events. The decision to label each topic is validated using a data driven process. Details about the three-step STM approach as well as the validation process are provided in Section 1 of the Supporting Information.

Figure 4 presents the top five topics by prevalence as well as the prevalence of force majeure weather related topics (e.g., lightning, flash floods, rain, hurricanes, thunderstorms, fires). The most common topics are related to plant shutdowns, flaring, TCEQ reporting terminology, malfunctions, and scheduled maintenance/repairs. On average, around 5% of event description text is related to shutdowns. Text related to weather induced accidents composes just over 10% of all comment text, indicating that while weather related accidents are an important source of emissions events, they are far from the dominant source.

5 Monetary estimates of health damages from excess emissions

While the adverse health effects of excess emissions are likely significant, their precise empirical estimation would require access to detailed data on mortality, morbidity, and pollution exposure. We use an integrated assessment model (IAM) to calculate a "back of the envelope" estimate of the monetary value of the health effects from excess emissions. To conduct this analysis, we first aggregate all the annual emissions of $PM_{2.5}$, SO_2, and NO_x from the EI dataset to the county level for the period 2004-2015. These annual county level emissions are put into the Estimating Air Pollution Social Impact Using Regression (EASIUR)[27,28] model, which predicts the total damage from a marginal increase in pollution from any county in the continental U.S. The damage estimates provided by the EASIUR model are from the perspective of the source county, where total damage from each source county's pollution is the aggregation of damage done by that source county on all receptor counties. The EASIUR model is based on the Comprehensive Air Quality Model with Extensions (CAMx) and it's damage predictions compare well with the results from other IAMs. The marginal damage estimate of the EASIUR model is based upon the impact of directly and indirectly emitted $PM_{2.5}$ on mortality. Estimates include damages that occur both locally and in downwind regions. In addition to varying across geographic space, predicted marginal damages vary with seasonal patterns in pollution transport, stack emission height, and pollutant type ($PM_{2.5}$, SO_2, NO_x, and NH_3).

We find that in 2015 excess emissions are responsible for at least $148 million in health damage annually, with approximately 10% of this damage coming from oil refineries. The EASIUR model uses a value of statistical life (VSL) estimate of $8.8 million implying that 16.82 deaths are caused per year by excess emissions. Damages vary across the state, and are concentrated in areas with more large, industrial facilities. Figure 5 displays the county-level damages estimated by EASIUR for 2015. Figures S52-S55 in the Supporting Information provide county level damages by pollutant by year. Figure S56 in the Supporting Information shows aggregate annual health damages by year from all sources and from the refining sector specifically. It is important to emphasize that our damage estimates represent only those mortality impacts due to direct and indirect PM2.5 emissions. Thus, they are intended to serve as a conservative, lower bound for potential health damages. Excluded health damages include all acute health events that do not lead to mortality and all pollution induced mortality that is not related to $PM_{2.5}$.

6 Conclusion

This study examines the significance of excess emissions, a category of air pollution that has received little attention in the scholarly literature. Our analysis shows that excess emissions are not exceptional, outlier events, but rather a regular feature of operations at industrial facilities. The data reported to the TCEQ show that these emissions can also be substantial in magnitude, raising important questions for future research about their effects on air quality and public health. In addition, in the most polluting industries, a small group of facilities are responsible for the vast majority of criteria pollutant excess emissions. Given the significant public health impacts, which we estimate to be at least $148 million annually in Texas alone, it is imperative that all states begin to systematically track excess emissions.

Excess emissions are also important from a policy perspective. As discussed above, the EPA has recently revised its policy on how excess emissions are regulated under the CAA. The agency has always regarded these emissions as a violation of a facility's permit obligations under the statute[2]. However, enforcement has largely been left to the states, and the EPA has determined that too often states have relied on policies and procedures that inappropriately shield firms from penalties. The EPA is in the process of reviewing many states' SIPs to ensure that treatment of excess emissions is consistent with EPA's interpretation. As is the case with many EPA policies, the Trump Administration is now reviewing the policy itself, which leaves the question of how excess emissions will be handled under the CAA in the future. Given the importance of these emissions, these policy decisions will be consequential.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website:

- Additional tables and Figures
- Details on the three step Structural Topic Modeling Approach

Acknowledgments

The authors thank Michael Brow, Richard Shultz, Dongping Xu as well as staff members at the TCEQ Emissions Assessment Section for providing helpful guidance on data obtained from the TCEQ. We acknowledge Shahzeen Attari, Kira Burkhart, Stephen Glaholt, Seth Johnson, Eric Schaeffer, Allison Youatt Schnable, Philip Stevens, and Coady Wing for helpful comments and suggestions, as well as three anonymous referees for their constructive feedback on an earlier version of the manuscript. This work has benefited from feedback of participants at the 2017 Northeast Agricultural and Resource Economics Association Annual Meeting, and research seminars at the School of Public and Environmental Affairs at Indiana University. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. CNS-0521433. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of the NSF or the individuals mentioned above.

References

1. Texas Commission on Environmental Quality. *Air Emissions and Maintenance Events Dataset*.; 2017.
2. Environmental Protection Agency. *State Implementation Plans: Response to Petition for Rulemaking; Restatement and Update of EPA's SSM Policy Applicable to SIPs; Findings of Substantial Inadequacy; and SIP Calls To Amend Provisions Applying to Excess Emissions During Periods of Startup, S*. Vol 40 CFR 52.; 2015:33840-33985.
3. Peterka A. EPA targets industry's defense for "off-the-book" emissions. *E&E News*. October 16, 2014.
4. Mehta K, Samuels M. The receding role of affirmative defense provisions in Clean Air Act regulations. *Trends*. 2015;46(3):1-4.
5. Environmental Protection Agency. *Policy on Excess Emissions During Startup, Shutdown, Maintenance, and Malfunctions*.; 1982.
6. Environmental Protection Agency. *Policy on Excess Emissions During Startup, Shutdown, Maintenance, and Malfunctions*.; 1983.
7. Sierra Club. Petition to Find Inadequate and Correct Several State Implementation Plans under Section 110 of the Clean Air Act Due to Startup, Shutdown, Malfunction, and/or Maintenance Provisions. 2011.
8. Environmental Protection Agency. *State Implementation Plans: Policy Regarding Excess Emissions During Malfunctions, Startup, and Shutdown*.; 1999.
9. Murphy CF, Allen DT. Hydrocarbon emissions from industrial release events in the Houston-Galveston area and their impact on ozone formation. *Atmos Environ*. 2005;39(21):3785-3798. doi:10.1016/j.atmosenv.2005.02.051.
10. Nam J, Kimura Y, Vizuete W, Murphy C, Allen DT. Modeling the impacts of emission events on ozone formation in Houston, Texas. *Atmos Environ*. 2006;40(28):5329-5341. doi:10.1016/j.atmosenv.2006.05.002.
11. Vizuete W, Kim B-U, Jeffries H, Kimura Y, Allen DT, Kioumourtzoglou M-A, Biton L, Henderson B. Modeling ozone formation from industrial emission events in Houston, Texas. *Atmos Environ*. 2008;42(33):7641-7650. doi:10.1016/j.atmosenv.2008.05.063.
12. Choi Y-J, Hyde P, Fernando HJS. Modeling of episodic particulate matter events using a 3-D air quality model with fine grid: Applications to a pair of cities in the US/Mexico border. *Atmos Environ*. 2006;40(27):5181-5201. doi:10.1016/J.ATMOSENV.2006.04.025.
13. Kulkarni P, Chellam S, Fraser MP. Tracking Petroleum Refinery Emission Events Using Lanthanum and Lanthanides as Elemental Markers for PM2.5. 2007;41(19):6748-6754. doi:10.1021/ES062888I.
14. Parrish DD, Allen DT, Bates TS, Estes M, Fehsenfeld FC, Feingold G, Ferrare R, Hardesty RM, Meagher JF, Nielsen-Gammon JW, Pierce RB, Ryerson TB, Seinfeld JH, Williams EJ. Overview of the Second Texas Air Quality Study (TexAQS II) and the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS). *J Geophys Res*. 2009;114(D7):D00F13. doi:10.1029/2009JD011842.
15. Environmental Integrity Project. *Gaming The System: How Off-the-Books Industrial Upset Emissions Cheat the Public Out of Clean Air*. Washington, D.C.; 2004.
16. Environmental Integrity Project. *Breakdowns in Air Quality: Air Pollution from Industrial Malfunctions and Maintenance in Texas*.; 2016. http://environmenttexas.org/reports/txe/breakdowns-air-quality (accessed on 8/19/17).
17. McCoy BJ, Fischbeck PS, Gerard D. How big is big? How often is often? Characterizing Texas petroleum refining upset air emissions. *Atmos Environ*. 2010;44(34):4230-4239. doi:10.1016/j.atmosenv.2010.07.008.
18. Ozymy J, Jarrell ML. Upset over air pollution: Analyzing upset event emissions at petroleum refineries. *Rev Policy Res*. 2011;28(4):365-382. doi:10.1111/j.1541-1338.2011.00502.x.
19. Texas Commission on Environmental Quality. *Emissions Inventory Dataset*.; 2016.

20. Texas Commission on Environmental Quality. *Air Permit Reviewer Reference Guide: Potential to Emit Guidance (APDG 5944)*.; 2008.
21. Texas Commission on Environmental Quality. Contaminant Code List. https://www.tceq.texas.gov/airquality/point-source-ei/psei.html. Published 2016. Accessed January 12, 2017.
22. Texas Commission on Environmental Quality. *Texas Administrative Code: General Air Quality Rules. Rule 101.1: Definitions*.; 2016.
23. Texas Commission on Environmental Quality. *Texas Administrative Code: General Air Quality Rules. Rule 101.201: Emissions Event Reporting and Recordkeeping Requirements*.; 2015.
24. Texas Commission on Environmental Quality. *Personal Communication with TCEQ Staff*.; 2017.
25. Brandt AR, Heath GA, Cooley D. Methane Leaks from Natural Gas Systems Follow Extreme Distributions. *Environ Sci Technol*. 2016;50(22):12512-12520. doi:10.1021/acs.est.6b04303.
26. Roberts ME, Stewart BM, Airoldi EM. A Model of Text for Experimentation in the Social Sciences. *J Am Stat Assoc*. 2016;111(515):988-1003. doi:10.1080/01621459.2016.1141684.
27. Heo J. Evaluation of Air Quality Impacts on Society: Methods and Application. February 2015. http://repository.cmu.edu/dissertations/503 (accessed on 9/5/17).
28. Heo J, Adams PJ, Gao HO. Public Health Costs of Primary $PM_{2.5}$ and Inorganic $PM_{2.5}$ Precursor Emissions in the United States. *Environ Sci Technol*. 2016;50(11):6061-6070. doi:10.1021/acs.est.5b06125.
29. QGIS Development Team. QGIS Geographical Information System. https://qgis.org/en/site/index.html. Published 2017. Accessed March 24, 2017.
30. Manson S, Schroeder J, Van Riper D, Ruggles S. IPUMS National Historical Geographic Information System: Version 12.0. *Minneap Univ Minnesota http//doi org/1018128 D (accessed 4/20/17)*. 2017.

Tables and Figures

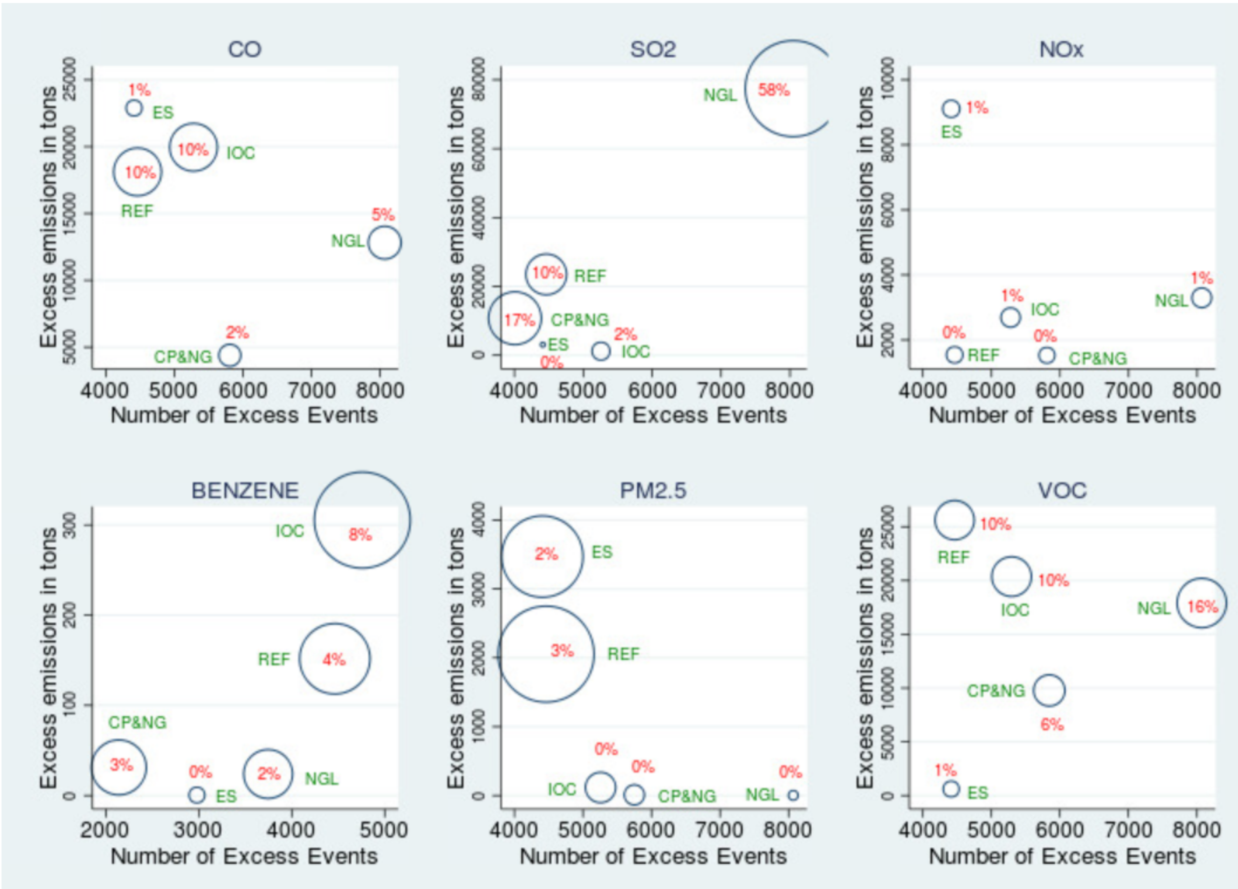
Table 1: Total amounts of excess emissions (in tons) from all facilities reporting to the Emissions Inventory of the TCEQ during the period 2004-2015. Source: Table compiled by the authors using data from TCEQ [19].

	Contaminant	Total excess emissions (tons)	Total excess / Total Routine	Emissions Events / Total excess emissions
Criteria Pollutants	Sulfur Dioxide	123,823	1.84%	82.91%
	Volatile Organic Compounds	104,202	7.50%	72.74%
	Carbon Monoxide	89,202	2.03%	57.18%
	Nitrogen Oxides	20,277	0.47%	43.16%
	Particulate Matter 10	9,572	1.37%	54.71%
	Particulate Matter 2.5	6,070	1.46%	53.77%
VOCs	Propane	12,081	16.09%	84.56%
	Propylene	6,527	19.70%	83.38%
	Isobutane	4,632	13.43%	82.47%
	Butene	734	16.26%	83.20%
HAPs	Hexane	2,150	8.42%	54.33%
	Toluene	840	4.53%	58.65%
	Benzene	776	5.53%	60.23%
	Xylene	318	2.06%	56.02%
	Formaldehyde	70	0.22%	66.40%

Note: Between 1990 and 2003, facilities reported: a) total excess emissions, that is, Emissions Events (EE) + emissions from Scheduled Maintenance Startup and Shutdown (SMSS) and b) routine emissions. Starting in 2004, amounts were reported separately for each of the three categories of emissions (i.e. routine, EE and SMSS emissions). We do not report data on lead because of low levels of emissions, and ozone which is not directly emitted. The column labeled “Total excess emissions (tons)” shows the total amount of EE+SMSS. The column labeled “Total excess over Total Routine” shows the ratio of Total excess emissions (EE+SMSS) over routine emissions. Finally, the column labeled “Emissions Events/Total Excess emissions” shows the ratio of emissions events (EE) over total excess emissions (EE+SMSS).

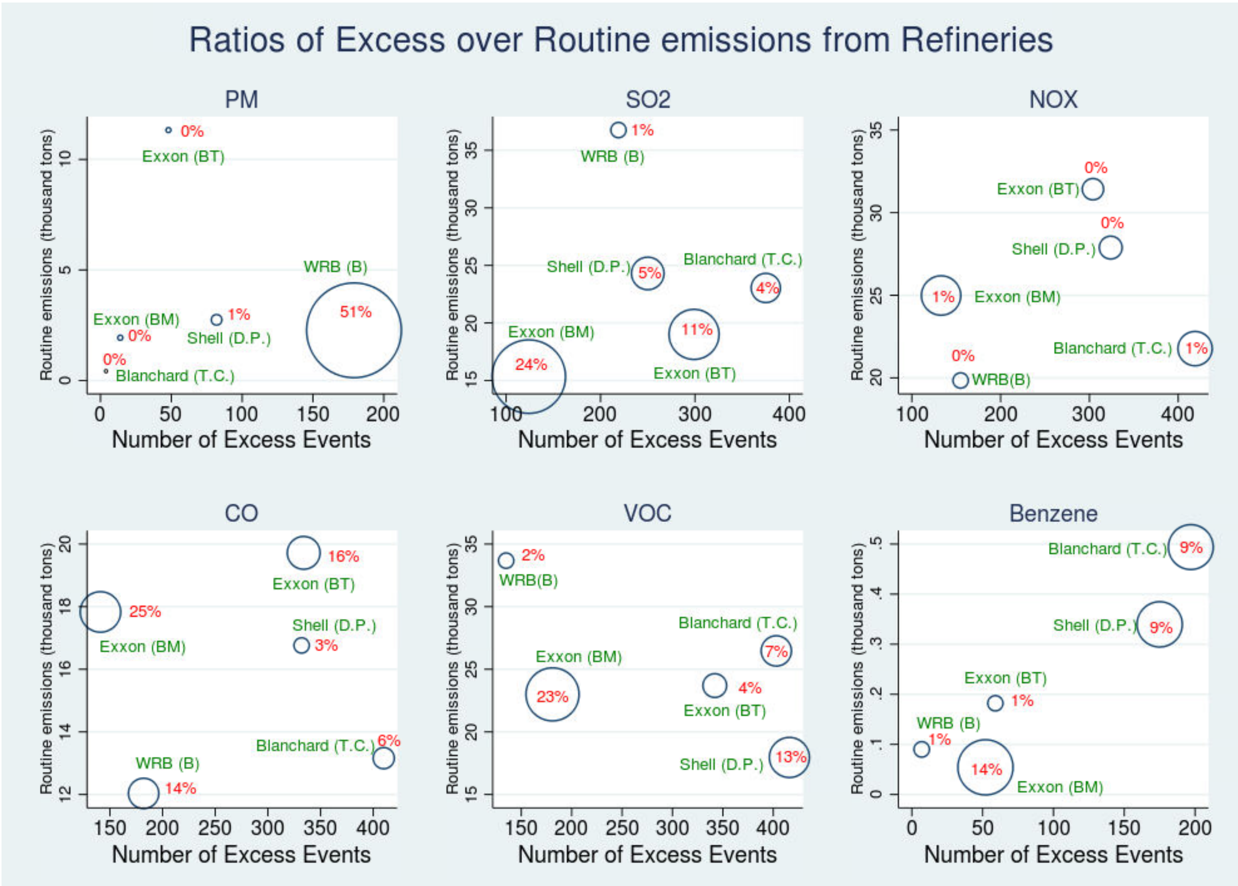
A more detailed version of this table is provided in Table S1 of the Supporting Information. Figures S1 and S2 of the Supporting Information, provide information on ratios of excess over routine emissions by year for criteria pollutants and important HAPs. Tables S2 and S3 of the Supporting Information provide data on excess emissions by industrial sector, while Figures S3-S8 provide time trends of excess emissions of criteria pollutants by industrial sector.

Figure 1: Ratio of excess emissions over routine (permitted) emissions (captured in red), total amount of excess emissions (on the vertical axis), and total number of excess emissions events (on the horizonral axis) from the top 5 polluting industries during the period 2004-2015. Source: Figure compiled by the authors using data from TCEQ [19].



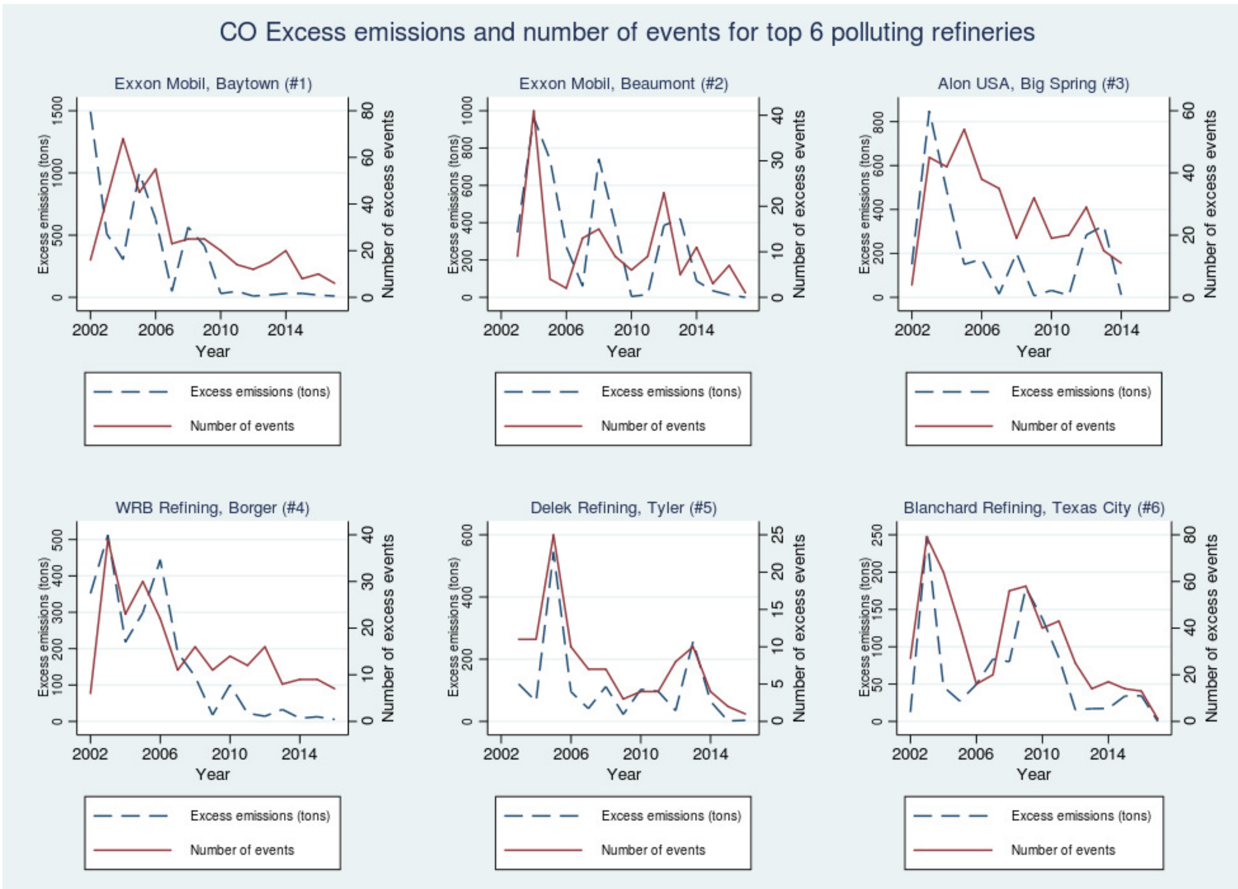
Note: The five industries with the highest amount of cumulative excess emissions across all pollutants (as identified in Table S2 of the Supporting Information) are highlighted in green. Those industries are: a) Crude Petroleum and Natural Gas (CP&NG), b) Natural Gas Liquids (NGL), c) Petroleum Refining (REF), d) Industrial Organic Chemicals (IOC), e) Electric Services (ES). The ratio of excess over routine emissions is captured in red and is also depicted by the area of each circle (the larger the circle, the higher the ratio of excess over routine emissions). Note that the areas of each circle are not comparable across the six panels of Figure 1, but are comparable within each of the six panels. The number of excess emissions events in the horizontal axis comes from the EI dataset and includes both reportable and non-reportable events by facility between 2006-2015 (there is no information on counts of events for prior years in the EI dataset).

Figure 2: Ratio of excess emissions over routine (permitted) emissions (captured in red), total amount of routine emissions (vertical axis), and total number of excess emissions events (horizontal axis) from the top 5 polluting refineries in Texas, during the period 2002-2015. Source: Figure compiled by the authors using data from TCEQ [1,19].



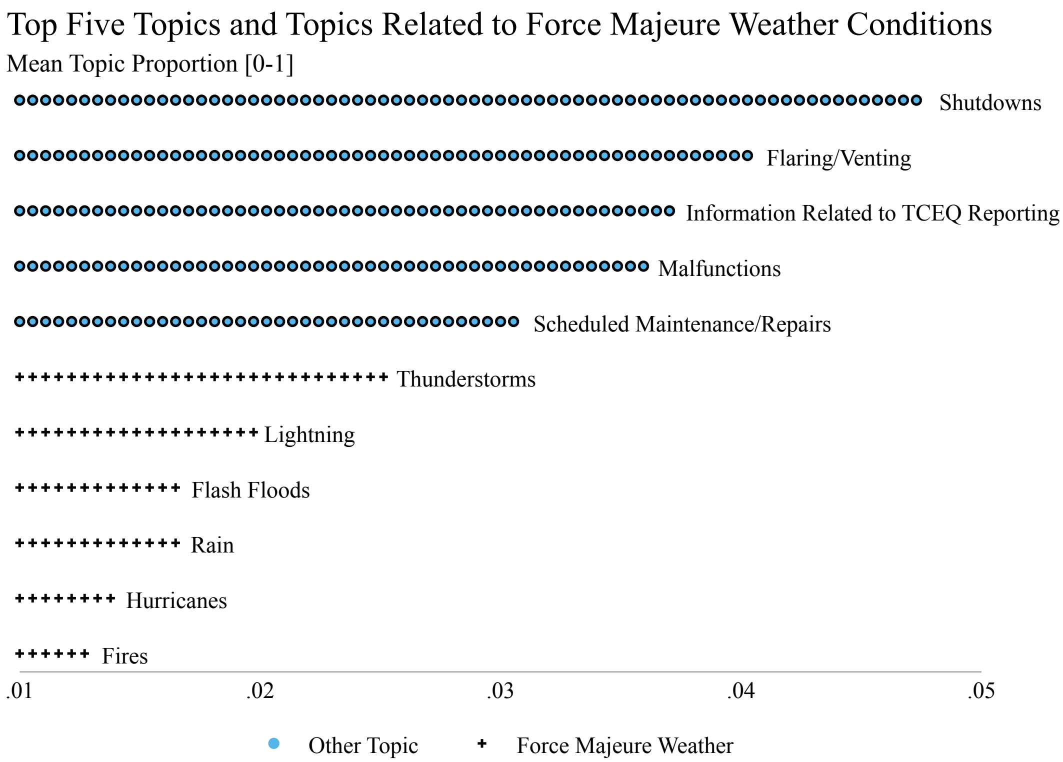
Note: The refineries with the highest amounts of cumulative excess emissions are highlighted in green. These are: a) Shell Oil, Deer Park (D.P.), b) Exxon Mobil, Baytown (BT), c) Blanchard, Texas City (T.C.), d) WRB, Borger (B), e) Exxon Mobil, Beaumont (BM). The ratio of excess over routine emissions is captured in red and is also depicted by the area of each circle (the larger the circle, the higher the ratio of excess over routine emissions). Note that the areas of each circle are not comparable across the six panels of Figure 2, but are comparable within each of the six panels. Figure S10 in the Supporting Information is a version of Figure 2 that includes all Texas refineries. Figures S11-S13 in the Supporting Information provide similar information for other key industrial sectors (namely, Crude Petroleum and Natural Gas, Industrial Organic Chemicals, and Natural Gas Liquids). Contrary to Figure 1, the number of excess emissions events in Figure 2 comes the AEME dataset and covers the period 2002-2015.

Figure 3: Excess emissions and number of excess emission events involving the release of Carbon Monoxide (CO) for the top 6 most polluting refineries. In each of the six panels, the amount of excess emissions for each facility is measured on the left axis and illustrated with the blue dotted line, while the number of excess emission events for each facility is measured on the right axis and illustrated with the red solid line. The name and location (city) of each facility are listed in the title of each panel. Those 6 facilities released 77% of all CO excess emissions across all refineries between 2002-2016. The total number of excess emissions events from those 6 refineries represent 35% of all excess emissions events from all refineries between 2002-2016. Source: Figure compiled by the authors using data from TCEQ[1].



Note: Figures S35-S38 in the Supporting Information present similar information as that depicted in Figure 3 for the remaining criteria pollutants on the top 6 polluting refineries. Figures S39-S50 in the Supporting Information have data on facilities in the most polluting sectors (namely, Crude Petroleum and Natural Gas, Natural Gas Liquids, and Industrial Organic Chemicals). Tables S9-S12 in the Supporting Information show summary statistics for the top six polluting facilities in each of the top four polluting sectors.

Figure 4: Results of Structural Topic Modeling. Source: Figure compiled by the authors using data from TCEQ [1]



Note: The prevalence and full set of all 50 topics estimated is available upon request.

Figure 5: Damages from Excess Emissions in 2015, by Texas County. Source: Figure compiled by the authors using data from TCEQ [19], EASIUR[27], QGIS[29] and Manson et al (2017)[30]