A Spatial Analysis of Beef Production and its Environmental and Health Impacts in Texas

by

Amanda Rompala

A Thesis Presented to the
FACULTY OF THE USC DORNSIFE COLLEGE OF LETTERS, ARTS AND SCIENCES
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

May 2023

To my partner and family, who support me in everything I do, and to my pets, who were literally by my side throughout my whole graduate school experience

Acknowledgements

I am grateful to Dr. Darren Ruddell, who gave me direction and assistance in the early stages of planning my thesis. I am grateful to my advisor, Dr. Elisabeth Sedano, for the support and guidance I needed. I am grateful to Dr. Jennifer Swift and Dr. Robert Vos, who served on my committee and assisted me when I needed it. I am grateful for the data provided to me by the Texas Commission on Environmental Quality.

Table of Contents

List of Tables

List of Figures

Abbreviations

AERMOD	American Meteorological Society/Environmental Protection Agency Regulatory Model
AICc	Akaike's information criterion
ANOVA	Analysis of variance
APSI	Air Pollution Scientific Initiative
CAFO	Concentrated animal feeding operation
CARB	California Air Resources Board
CFD	Computational fluid dynamics
CSA	Community Supported Agriculture
EPA	Environmental Protection Agency
EPCRA	Emergency Planning and Community Right-to-Know Act
FAC2	Fraction of two
FB	Fractional bias
GHI	Global horizontal irradiance
GIS	Geographic information systems
LEfSe	Linear discriminant analysis Effect Size
MG	Geometric mean bias
MODIS	Moderate Resolution Imaging Spectroradiometer
NAAQS	National Ambient Air Quality Standards
NASA	National Aeronautics and Space Administration
NASS	National Agricultural Statistics Service
NMP	Nutrient management plan

NMSE	Normalized mean square error
NPDES	National Pollutant Discharge Elimination System
NREL	National Renewable Energy Laboratory
OLS	Ordinary least squares
PGT	Pasquill-Gifford-Turner
PM	Particulate matter
PM 2.5	Particulate matter less than 2.5 μm
PM 10	Particulate matter less than 10 μm
POWER	Prediction of Worldwide Energy Resource
RMSE	Root mean square errors
SCIPUFF	Second-Order Closure Integrated Puff
SEDAC	Socioeconomic Data and Applications Center
TCEQ	Texas Commission on Environmental Quality
TPWD	Texas Parks and Wildlife Department
US	United States
USDA	United States Department of Agriculture
VIF	Variance inflation factor
WIND	Wind Integration National Dataset

Abstract

About 99% of beef production in the United States occurs in concentrated animal feeding operations (CAFOs). CAFOs fatten huge quantities of cows in very dense, contaminated conditions before the cows go to slaughter. CAFOs use innovative technology like hormones, pharmaceuticals, and advanced feed to prepare cows for slaughter as fast as possible to maximize profits. These procedures come with negative impacts, like the emission of multiple airborne pollutants and the poor treatment of cows. This thesis focuses on the spatial distribution of CAFOs and the airborne contaminants they release. Prior research on airborne contaminant dispersion from CAFOs has found that air quality is poor for communities where CAFOs exist. This thesis identifies 221 CAFO facilities in Texas and uses meteorological variables with a Gaussian plume distribution model to estimate the 2016 PM 2.5 concentrations released from each CAFO. Linear regression is employed to compare the results to measured PM 2.5 concentrations. The Gaussian plume equation is found to estimate PM 2.5 from CAFOs accurately. Close to the beef production facilities, the estimated concentrations reach up to about 759 $\mu g/m^3$, which is higher than the national annual average standard range of 12 - 15 $\mu g/m^3$. Finally, this thesis makes recommendations for a healthier beef industry.

Chapter 1 Introduction

Concentrated animal feeding operations (CAFOs), the most commonly used method of beef production in the United States (US), generate large amounts of air pollution and are unhealthy for communities surrounding the production facilities. This study performed a spatial analysis of the most predominant mode of beef production in Texas and estimated the PM 2.5 concentrations they release using a Gaussian plume dispersion model. The first section of this chapter introduces beef production and its environmental harm in the US, the second section explains the health impacts of beef production, the third section discusses governmental regulation of CAFOs, the fourth section suggests mitigation strategies for a healthier beef industry, the fifth section describes the study area, and the last section briefly introduces this project as well as this thesis document.

1.1 Introduction to the US Beef Industry and its Environmental Impacts

From the mid-1800s to the 1970s, a balance between agriculture and beef production methods existed in the US; however, in the 1970s, food production methods became more industrial (Chiles and Fitzgerald 2018). The earlier food production methods provided an impact-neutral, low-pollutant food production industry. Specifically, in rural communities, beef production was traditionally integrated with agricultural production to benefit farmers, ranchers, society, and the environment by mimicking natural ecological processes. Compared to large-scale beef production, rotating animal and crop agriculture creates fewer pollutant emissions, relies on biologic rather than synthetic fertilizer, and yields higher food production rates (Gurian-Sherman 2008). The introduction of industrial agricultural practices scaled up operations to

increase efficiency and lower costs. With industrial practices, beef is produced year-round in confined conditions, and the integration between agriculture and beef production is lost.

Current beef production methods treat cows like goods produced in a factory. These factories are called CAFOs, which are large-scale, efficient industrial facilities that raise a single animal species and do not grow crops. In the US, the primary way that beef production occurs is through feedlots which are a type of CAFO where cattle are densely packed and fed for short periods. Cows inhabit feedlots for approximately 50-100 days to get fattened before they go to slaughter. Before the cows go to feedlots, a beef cow's life starts on a cow-calf operation pasture. In cow-calf operations, a herd of beef cows is maintained to raise calves. The calves live on these farms for three to seven months while they wean. After the calves are weaned, they either stay in a calf-cow operation to breed or go to a feedlot (USDA 2022).

With the high demand for beef in America, the technology and methods utilized in feedlots are constantly changing to be more efficient by using industrial logic to guide the process. These changes include incorporating growth-promoting techniques, varying the farmed cow breeds, and altering the location of the cows. The growth-promoting techniques include using feed additives, hormone implants inserted beneath the skin of cows, and several antibiotics (Drouillard 2018).

These techniques maximize beef output and minimize cost because the cows breed more often, gain weight faster, and avoid diseases to which the close quarters and corn feed makes them more vulnerable; however, they create harmful trade-offs. With about 99% of the beef produced in the US in CAFOs (Walton and Jaiven 2020), considering the adverse environmental impacts they cause is important. According to the Sierra Club (2021), a small CAFO produces urine and manure equal to the amount produced by 16,000 humans. Additionally, the production

byproducts of feed, excrement, and water runoff are often full of disease, chemicals, pharmaceuticals, heavy metals, and other pollutants. These byproducts are a massive amount of toxic waste that must be effectively treated (Sierra Club 2021). As winds blow past the CAFOs, particles released from untreated waste piles are transported to nearby communities.

As more information about the harms of factory farming becomes known, consumers are demanding more niche non-hormone, non-antibiotic, organic grass-fed beef, and the beef industry is highly adaptive and responds to consumer demands rapidly (Drouillard 2018). This more natural alternative to factory farming is grass-fed cattle ranching. With grass-fed cattle ranching, the cows can eat grass and graze in open pastures, which satisfies the public demand for healthy, sustainable beef. Pasture-raised farming occurs without the antibiotics, hormones, or diseases associated with crowded feedlot conditions. With growing interest in using grass-fed, pasture-raised cows, consideration of the associated economic and environmental problems is necessary.

A perception that grass-fed beef production is more eco-friendly than CAFO production is widespread; however, transitioning to an industry more reliant on grass-fed beef could bring environmental and economic harm. While it is true that current grass-fed beef production is more eco-friendly, expanding production also expands the geographic footprint and the number of cows needed to produce beef. Much land must be converted to pasture to produce enough beef to sustain the current demand. There would also have to be an increase in the number of cows in the country since the cows take longer to reach slaughter weight. Hayek and Garrett (2018) calculate that instead of the thirteen million cows in CAFOs in the US in 2012, 21 million cows would be required to produce the same amount of beef in pastures. This is a 61.5% increase in the US beef cow inventory. The required growth in the cow population is due to grass-fed cows taking

several months longer to reach the appropriate weight for slaughter than they do on feedlots (Hayek and Garrett 2018; Riely 2011). Having more cows would require more food to be grown for the cows to eat and would create more manure. Moreover, converting land to pasture may create competition with the food supply of other crops or disrupt the ecosystem in those areas.

In addition to the environmental effects of grass-fed beef production, grass-fed beef is more expensive than factory-produced beef. Riely (2011) explains that pasture-raised beef is a niche market where ranchers must be near urban locations to sell to health-oriented consumers. Because grass-fed beef is often sold in farmer's markets or through Community Supported Agriculture (CSA) programs, living close to cities with large populations is essential for ranchers. CSA programs allow consumers to purchase a subscription to a particular farm, and in return, they receive local, fresh, seasonal produce from the farm each week throughout the farming season. Additionally, grass-fed beef is highly susceptible to drought since the cows feed primarily on grass. Pasture-raised, grass-fed beef is more expensive and riskier for the ranchers than producing beef in feedlots (Riely 2011).

It is a strain on ranchers to expand grass-fed beef production operations because it challenges them to acquire enough money and supplies to be profitable. Gwin (2009) studied grass-fed beef producers in the US who tried to scale up their operations. The study finds that increased production creates supply-chain issues with the inputs needed to increase production. These inputs include increasing the number of cows, feed, land, infrastructure, and finance capital. The issues are with the supplies necessary to run the farm and produce the beef. Raw materials like cattle, feed, and land are difficult to assemble and afford. The ranchers also faced problems with acquiring enough capital to expand. It is not easy to find the delicate balance

between beef production speed and consistency and environmental quality, humane treatment, human health, and profitability (Gwin 2009).

1.2 Health Impacts of CAFOs

The shift from an agrarian to an industrial model of food production brings economic efficiency but environmental and health costs. The efficiency-adding practices generate toxins that are released into the air and are dangerous to breathe. There are over 168 types of gases that emanate from the CAFO waste. Some of the pollutants included in these gases are volatile organic compounds like ammonia, methane, nitrate, nitrous oxide, carbon dioxide, and hydrogen sulfide and other pollutants like pathogens, antibiotics, and hormones (Chase and Grubinger 2014; Sierra Club 2021; Yuan et al. 2017). Additionally, CAFOs produce pollutant-filled dust (Purdy, Clark, and Straus 2007). The particles suspended in the air are particulate matter (PM). PM are divided based on their aerodynamic diameter; particulate matter less than ten μm is PM 10, and particulate matter less than 2.5 μm is PM 2.5 (Rittner et al. 2020). PM 10 can irritate the eyes, nose, and throat if breathed. The finer PM 2.5 are more dangerous than coarser PM 10 because it can get into the deep part of the lungs or even the bloodstream, irritating the respiratory system and causing swelling, respiratory distress, and long-term illness (CARB n.d.).

The pollutants released by beef production are harmful and extensive because of the high density of cows in a small space. The manure produced by the cows is densely piled and releases hundreds of contaminants and particles into the air. With such dense manure concentrations, the air is toxic to breathe (Mitloehner and Schenker 2007). The amount of manure produced in CAFOs is so large that it is difficult to treat it effectively. When oxygen cannot reach the interior of the manure piles, the biological material cannot break down, and the production of dangerous gases occurs. These gases include methane, ammonia, and hydrogen sulfide (Sierra Club 2021).

In addition to the manure, cows often die in CAFOs and release bacteria, fungi, and pathogens into the environment (Sierra Club 2021).

Research quantifies the harm caused to communities living near the CAFOs who breathe the airborne pollutants released by CAFOs. Schultz et al. (2019) looked at respiratory health and allergic diseases among residents near dairy CAFOs in the Upper Midwest of the US in 2019. Schultz et al. (2019) examine the association between distance to CAFO and the prevalence of physician-diagnosed nasal and lung allergies, asthma, and decreased lung function. They find that residential proximity to a CAFO, specifically within 1.5 miles, is positively associated with these problems compared to residents three miles from a CAFO (Schultz et al. 2019).

In addition to studying the residents living near CAFOs, another population group worth considering are the children who attend school and spend much of their time near CAFOs. The Sigurdarson and Kline (2006) case study looks at the impact of CAFOs on schools by examining students' respiratory health at a school one-half mile from a CAFO and a school distant from a CAFO in Iowa in 2006. They found that asthma prevalence and severity were significantly higher in the school near the CAFO. In the school close to the CAFO, the asthma prevalence was 19.7%, and at the school far from the CAFO, the asthma prevalence was 7.3% (Sigurdarson and Kline 2006). Understanding where CAFO contaminants travel is vital to mitigate the respiratory health effects of those spending time in high-concentration areas.

Further research on the link between CAFO contaminants and respiratory health outcomes is limited. Quantifying air emission rates from open beef CAFOs is often tricky due to their heterogeneous surfaces, wide source geometry, and variable emission rates (Bonifacio et al. 2012). A heterogeneous surface means the landscape has many different land cover types. The wide source geometry means particle emissions come from many sources across a large polygon

extent rather than a single point source. Finally, a variable emission rate means the particle emission is not a constant, steady flow. More research is needed to quantify the environmental impacts, but the research discussed in this thesis indicates a negative impact.

1.3 Governmental Regulation of CAFOs

Stakeholders must consider the substantial problems with how CAFOs operate and release pollutants into the environment, where the particles concentrate, and their impacts on human health; however, no federal agency collects accurate and exhaustive data on CAFOs (Handan-Nader and Ho 2019). Environmental laws that limit the pollutants emitted from beef industry facilities exist, but the laws are ineffective. The CAFO industry has influence over its regulation, and state and federal laws do not oversee the beef production industry (Moses and Tomaselli 2017). For example, Congress passed the Title III of the Superfund Amendments and Reauthorization Act, which includes the Emergency Planning and Community Right-to-Know Act (EPCRA). EPCRA was designed to protect public health at the local level by requiring CAFOs to report some chemical pollutants, including ammonia. However, the federal Environmental Protection Agency (EPA) does not regulate air pollution from CAFOs, so the pollutants are disclosed but not regulated (Sierra Club 2021).

Laws such as the Clean Water Act, which reduces the allowable amount of water pollutants released from CAFOs, regulate some waterborne pollutants (Clark 2020), but getting the government to implement the laws is challenging (Sierra Club 2021). Other federal regulations also limit waterborne pollutant discharge, but for CAFOs, there is no enforcement of the regulations. For example, National Pollutant Discharge Elimination System (NPDES) permits require water pollutant discharge reporting and elicit considerable fines for violations (Clark 2020). The EPA estimates that 60% of CAFOs do not hold NPDES permits (Handan-

Nader and Ho 2019). They presume the under-permitting is likely due to federal courts rejecting the EPA's proposal to cover potential polluters rather than just active polluters. Potential polluters are CAFO facilities that have been closed, but their manure and storage facilities were not properly closed. These facilities no longer require an NPDES permit since they are not operational CAFOs, but there is still potential for them to discharge pollutants (Handan-Nader and Ho 2019).

Even if waterborne pollution were effectively limited, CAFOs would still pose environmental concerns since reducing water pollution increases air pollution. The dust and waste from CAFOs were being washed with sprinklers, making the pollutants run off in the water. If the pollutants do not exit through the water, they are left to disperse through the air.

With limited governmental regulation, communities suffer and cannot get help (Moses and Tomaselli 2017). The Texas Commission on Environmental Quality (TCEQ) received at least one hundred complaints from 2008 to 2017 about fecal dust and odor, but it did not address this problem (Collins 2020). Often ambiguity exists in whether the state or federal government is responsible for monitoring CAFO air pollution and enforcing regulations (Sierra Club 2021).

1.4 Mitigation Strategies for a Healthier Beef Industry

The long-term health of communities near beef production facilities depends on overcoming air quality challenges. It requires better CAFO monitoring and further research to understand how different animals and practices affect the environment (Bonifacio et al. 2012). Yuan et al. (2017) measure the spatial distribution of airborne pollutants from CAFOs in northeastern Colorado to compare the composition of pollutants among several types of CAFOs like dairy farms, beef feedlots, sheep feedlots, and chicken farms. Yuan et al. (2017) find that the contaminants vary based on the types of animals and their waste and the type of feed storage and

handling techniques (Yuan et al. 2017). Gathering information and raising awareness is essential for effective policy and production change.

Research suggests effective policy changes that would limit the airborne pollutants emitted by CAFOs. Donham et al. (2007) suggest making rules for acquiring CAFO permits more stringent, issuing permits based on the animal density allowed in a given watershed, and requiring environmental impact statements, public meetings at the local level, and more monitoring and regulation (Donham et al. 2007). Additionally, Gurian-Sherman (2008) suggests developing new regulations for the Clean Air Act that would reduce the emission of air pollutants from CAFOs. He thinks this would limit air pollution from manure since the current Clean Air Act does not address these pollutants. Gurian-Sherman (2008) argues that there should be more government support for smaller farms (Gurian-Sherman 2008).

Concerning beef production processes, one change that would lower the air pollutants generated by beef production is better manure management. The Clark (2020) article discusses the advantages of feeding cows more digestible food, moving manure to off-site locations, and implementing nutrient management plans (NMPs) to better manage manure. NMPs are part of the CAFOs' permitting process and optimize a CAFO's productivity while conserving nutrients and protecting the environment. For example, NMPs specify proper manure storage and management methods, the proper diet of the cows, and suitable irrigation practices. The improper management of cow manure is the most significant cause of pollution from CAFOs (Clark 2020). Understanding beef production, tracking the contaminants it releases, and finding strategies to make beef production less damaging to the ecosystem is essential.

1.5 Study Area

The study area for this analysis is the state of Texas, which is the top beef producer in the US. Fifteen percent of the United States' beef inventory lives in Texas (NASS 2022), and facilities there hold up to 2,333 cows per acre (TCEQ 2022b). These facilities produce substantial amounts of waste that emit harmful airborne pollutants (Chase and Grubinger 2014; Sierra Club 2021; Yuan et al. 2017). A map of the 2019 US Census population totals for each Texas county and the Texas CAFO locations is in Figure 1. The Texas CAFOs points come from TCEQ permits, which have the CAFO locations, the number of cows per CAFO, and the acreage of each CAFO in Texas. Of the 221 CAFO points in Texas, 204, or 92%, are in counties with less than 100,000 residents. The largest cluster of CAFOs is located in the Texas Panhandle, near Amarillo and Lubbock. There are few CAFOs near large cities like Dallas, Austin, San Antonio, and Houston. Since beef production in CAFOs produces many airborne pollutants, most of the beef production in Texas occurs in rural areas with a low population.

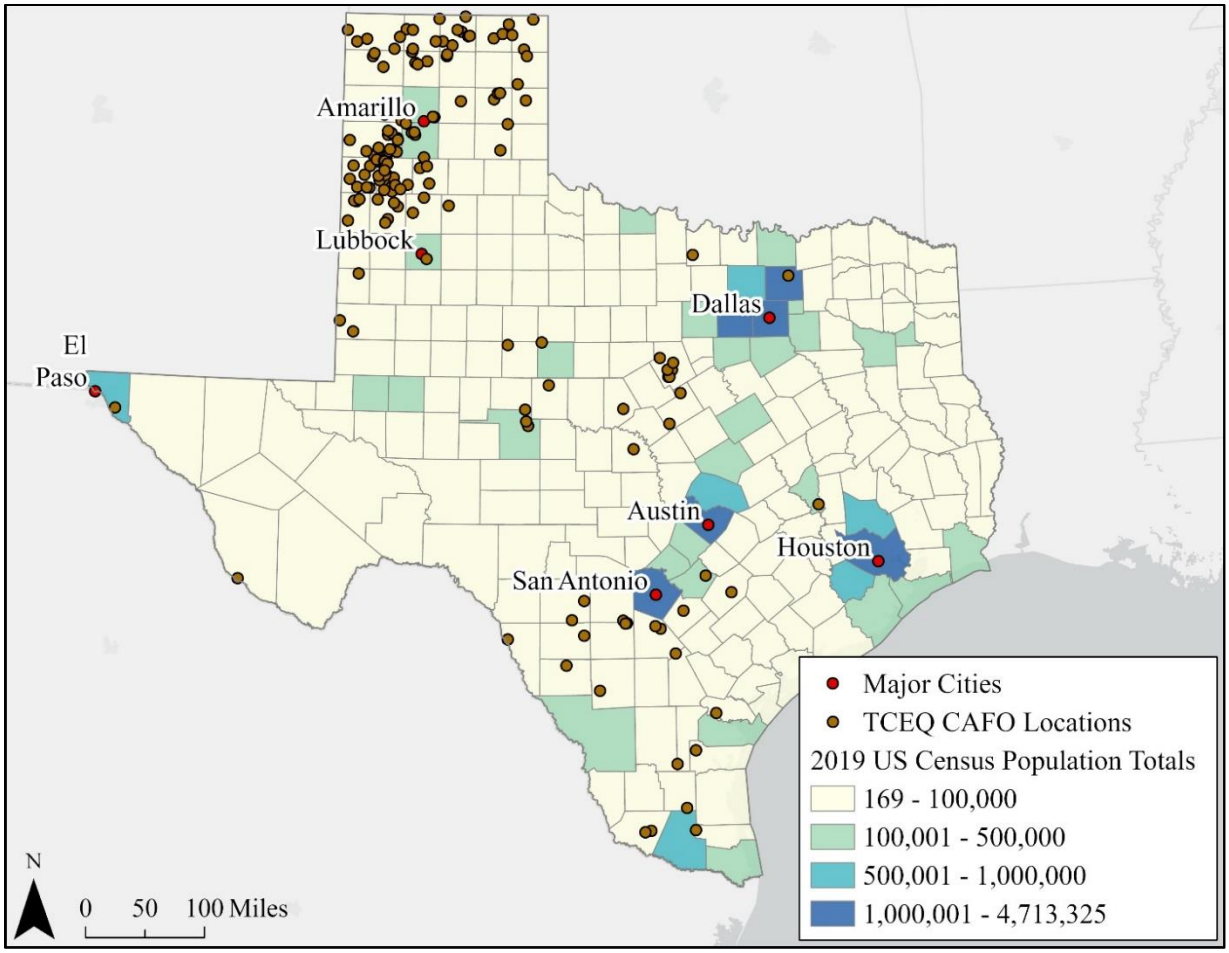

Figure 1. Texas population by county and TCEQ CAFO locations

The CAFOs in Texas are divided into three clusters: Panhandle, central, and southern coastal. Figure 2 shows the three CAFO clusters along with the Texas natural regions from the Texas Parks and Wildlife Department (TPWD). The first cluster of CAFOs is the Panhandle cluster, which is located mostly in the Panhandle Plains natural region. The Panhandle Plains region is the northernmost region in Texas, and it is also known as the Panhandle. The Panhandle is flat and grassy with very few trees. A deep canyon divides the Panhandle region; the western portion is flatter and drier, while the eastern portion is lower in elevation, has more rainfall, and has more brush. The soils in the Panhandle are coarse sands, clays, and shales. The average rainfall in this region is 15-28 inches per year, and the elevation is from 1,300-3,900 feet above

sea level (TPWD n.d.d). The CAFOs in the Panhandle cluster are located in the western portion of the Panhandle Plains region.

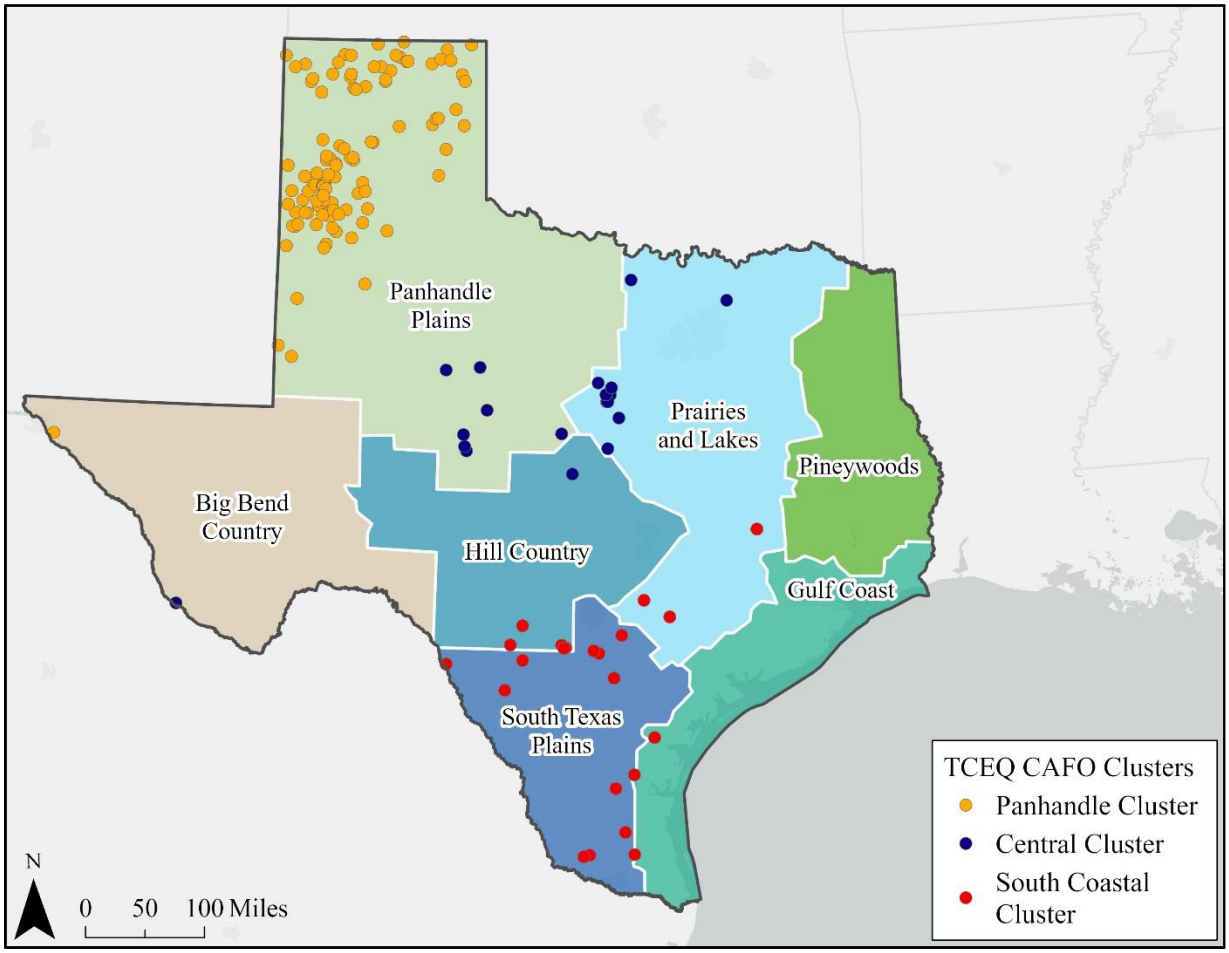

Figure 2. Texas natural regions and TCEQ CAFO clusters

The next CAFO cluster is the Central cluster, which is located partially in the eastern portion of the Panhandle Plains region. The CAFOs in this cluster are also located in the Prairies and Lakes and Hill Country natural regions. The Prairies and Lakes region is in north-central and central Texas. In this region, there is a transition from the plains of west Texas to the more forested area in eastern Texas. There are patches of woodland in the Prairies and Lakes region, while the rest is grassland prairie. The region is hilly, the soil is rich and fertile, and there are

many lakes. The elevation in the Prairies and Lakes region is 300-800 feet above sea level, and the average rainfall is 26-40 inches per year (TPWD n.d.e).

In central Texas, the Hill Country region is filled with hilly grasslands. There are springs, underground lakes, and steep canyons in this area. The Hill Country region has many oak and juniper trees, and the soil is coarse. The average rainfall in the Hill Country region is 15-34 inches annually, and elevation ranges from 500-2,250 feet above sea level (TPWD n.d.c).

The Panhandle and Central CAFO clusters also have one CAFO each in the Big Bend Country natural region. TPWD describes the Big Bend Country region as wide-open spaces with rugged plateaus with short grass. There are also desert areas with mountains that are dry and hot during the day and cool at night. The name Big Bend comes from the big bend in the Rio Grande River that runs along the southern portion of the Big Bend Country region. There are many different soil types that support diverse habitats and vegetation. This region averages 8-20 inches of rainfall per year. The elevation in the Big Bend Country region ranges from 2,500-8,749 feet above sea level (TPWD n.d.a).

Finally, the last CAFO cluster is the South Coastal cluster. These CAFOs are primarily located in the South Texas Plains natural region, though there are a few CAFOs in the Hill Country, Prairies and Lakes, and Gulf Coast natural regions. The South Texas Plains stretch from the Hill Country region to the subtropical regions of the Lower Rio Grande Valley. The South Texas Plains region is mostly dry and covered with grasses and thorny brush. However, there are patches of subtropical woodlands in the Rio Grande Valley that are humid and marshy. There are a few lakes in the region, the soil is alkaline to slightly acidic clays, and the average rainfall is 16-32 inches per year. The elevation in the South Texas Plains region is less than 1,000 feet above sea level (TPWD n.d.f).

The Gulf Coast region is located along the southern coast of Texas and contains marshes, barrier islands, estuaries, and bays. Towards the west of the region, there are prairies and grasslands. The Gulf Coast region is flat and filled with many streams and rivers that drain to the Gulf of Mexico. The soils in this region are acidic sands, sandy loams, and clay soil. The average rainfall in the Gulf Coast region is 40-60 inches per year, and the elevation of the Gulf Coast region is very close to sea level (TPWD n.d.b).

The overall beef cow inventory changed over time, but Texas has always been a large contributor to the United States' cow inventory. As of January 1, 2022, Texas had fifteen percent of the country's total beef cattle inventory. Nebraska had the second largest inventory, with only seven percent of the country's inventory (NASS 2022). Figure 3 shows a graph of the beef cattle inventory in Texas from 1920 to 2022. This graph shows how the number of cows in Texas has changed, with the peak in 1975. This data came from the National Agricultural Statistics Service (NASS).

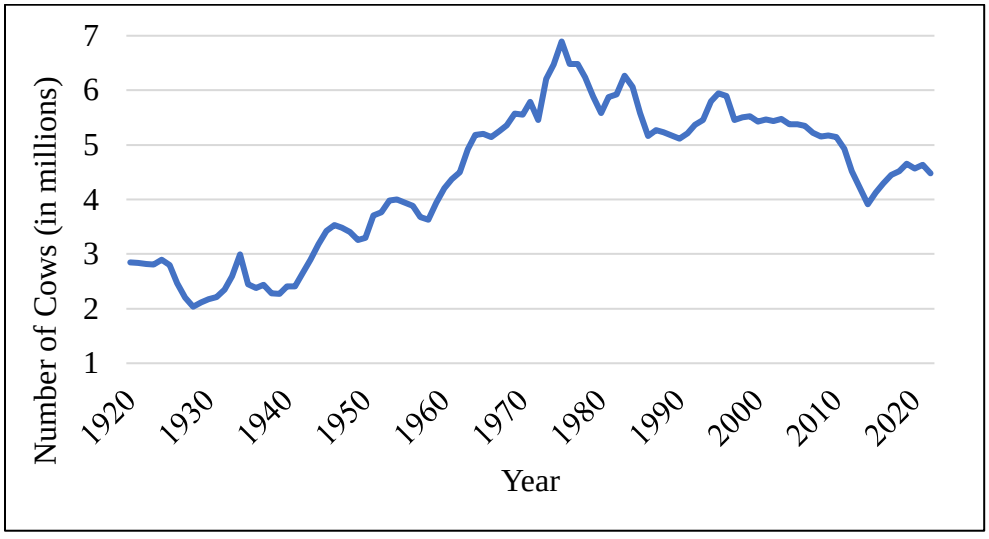

Figure 3. NASS Texas beef cattle inventory from 1920-2022

The trend of beef inventory in Texas closely mirrors the trend in beef inventory in the US. A graph of the beef cattle inventory in the US from 1920 to 2022 is in Figure 4. This graph

shows how the number of cows in the nation has changed, with the peak also in 1975. This data was also downloaded from NASS. The decrease in cows after 1975 is likely due to the shorter lifespan of cows from cow-fattening practices used in CAFOs. CAFOs became widespread for beef cows in the 1970s and 1980s.

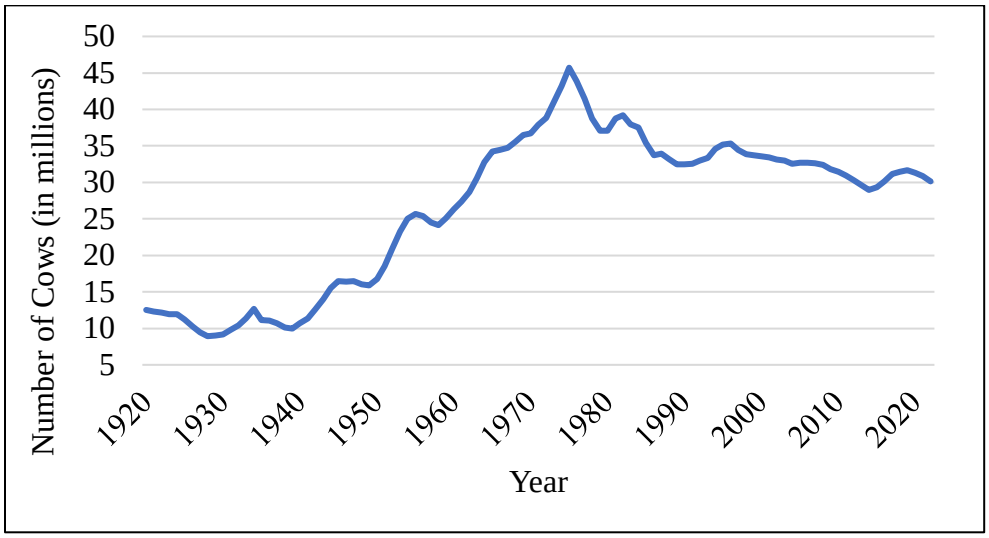

Figure 4. NASS US beef cattle inventory from 1920-2022

1.6 Project Overview

This study estimated the PM 2.5 concentrations released from CAFOs in Texas in 2016 using a Gaussian plume dispersion model. The inputs into the model include CAFO locations and meteorological variables: global horizontal irradiance, cloud fraction, wind speed, and wind direction. An emission factor was used to estimate the PM 2.5 emitted from each CAFO, and a Python script was written to execute the Gaussian plume equation. The model's results were compared to measured 2016 PM 2.5 concentrations using linear regression.

This thesis contains five parts that describe the spatial analysis of beef production. Chapter Two includes research on air pollutant dispersion modeling, the Gaussian plume dispersion model, the assessment of dispersion models, and the measurement of air pollution in

general. Chapter Three analyzes the datasets and methods used in the Gaussian plume dispersion model that estimated each CAFO's PM 2.5 contaminant distribution. Chapter Four compares the resulting PM 2.5 exposure surface to measured PM 2.5 concentrations using linear regression. Finally, Chapter Five discusses the findings, explains the limitations, and proposes further research and mitigation strategies for a more sustainable and healthy beef industry.

Chapter 2 Related Literature

Beef production in Texas takes place primarily in concentrated animal feeding operations. These facilities emit substantial amounts of airborne pollutants into the atmosphere, and the emissions are largely unregulated. Measuring the airborne pollutants from CAFOs is difficult due to the landscape having many different land cover types, the particle emissions coming from many sources across a large polygon extent rather than a single point source, and the particle emission not being a constant, steady flow. However, plume dispersion models try to quantify the pollutant concentration (Bonifacio et al. 2012). This chapter describes measuring air pollution in general, examines research on the dispersion modeling of air pollutants, describes the Gaussian plume equation, and explores ways to assess the dispersion models. The research guides this thesis, which used the Gaussian plume equation to estimate the concentration of PM 2.5 that CAFOs emit in Texas.

2.1 Measuring Air Pollution

Geographic information systems (GIS) are useful for measuring air pollution because they capture, store, analyze, and display data about locations on the Earth and can represent data from multiple sources. GIS can connect spatial and non-spatial data and handle massive databases. The capability of GIS makes them immensely useful for measuring air pollution. Air pollution models take emission information, meteorological data, and complex equations for how particles disperse in the air to estimate the particle concentrations. Modeling these complex relationships reduces the expense of continuous monitoring over vast areas and over long periods of time (Namrata and Wagh 2022; Teggi 2018).

GIS are a powerful tool for modeling air pollution dispersion, but challenges associated with measuring air pollution remain. The first problem is with the resolution of the data used in GIS atmospheric dispersion models. If the data are at a coarse resolution, the model will generalize the phenomenon being measured. Atmospheric dispersion is complex, so without fine-resolution data, the quality of the results is limited. The spatiotemporal resolution of model inputs directly affects the model results (Yin et al. 2015). However, with good-quality input data, this problem can be reduced.

The second problem is that there are multiple emission sources that are often uncontrolled in their dispersion processes for many study settings. It is difficult to measure a single pollutant source because there are many pollutant sources occurring in most study areas (Bai et al. 2022; Rittner et al. 2020). This problem is difficult to avoid; attempting to model pollution from all sources in a study area is challenging (Bunton et al. 2007). In addition, objects like buildings and trees can disrupt pollution dispersion. A study area with objects disrupting pollutant dispersion can result in the air pollution levels differing greatly across the study area in a way that the model might not be able to predict. Atmospheric models work best in flat areas with few objects to disrupt atmospheric dispersion (Namrata and Wagh 2022).

Finally, the third problem with air pollutant measurement is that models generate structural and parametric uncertainties. Structural uncertainties come from the uncertainty of whether the model correctly reflects reality. This is because particle dispersion is complex and could depend greatly on the landscape being measured. Parametric uncertainties come from unpredictable input values like wind speed and direction and the number of emission factors. This unpredictability can be from limited data availability or quality (Holmes et al. 2009; Namrata and Wagh 2022). However, as discussed above, atmospheric dispersion models can be

very accurate if the proper assumptions are made. The Gaussian plume equation assumes that the wind direction is constant in the x direction, it is a non-reacting pollutant, there is a total reflection from the ground (Pfluger, Dacunto, and Hendricks 2014), wind speeds are over two meters per second, the pollutant travel distances is greater than one hundred meters (Namrata and Wagh 2022), and the terrain is flat (Rittner et al. 2020). With these conditions, the Gaussian plume equation performs well.

2.2 Dispersion Modeling of Air Pollutants

Researchers often use plume equation models to measure airborne pollutant dispersion. Plumes are a way to describe how airborne pollutants travel from a source; the pollutants begin concentrated at the source, then they spread out as they travel. Researchers have created multiple models to quantify the pollutant concentration as it disperses. Dispersion models can be steady-state or dynamic. Steady-state models assume that the environmental conditions and source strengths do not change during each simulation of the model (Hadlocon et al. 2015). These include models such as the Gaussian plume equation and the American Meteorological Society/Environmental Protection Agency Regulatory Model (AERMOD), discussed below. Dynamic models are more complex and simulate pollutant dispersion with variable environmental conditions, like wind speed and direction, during each simulation. These include models like Lagrangian or Eulerian (Hadlocon et al. 2015).

2.2.1 Gaussian Dispersion Model

The Gaussian is the most widely used air emission dispersion model. Under steady-state conditions, the model uses vertical and horizontal Gaussian distribution equations that measure pollutant concentration. This model is accurate and widely used in research. It functions by modeling the dispersion of airborne particles from a point source using source emission rate,

wind speed, plume geometry, and plume stability parameters. One limitation of the Gaussian model is that it does not account for the time taken for the pollutants to migrate to the measurement location or the deposition and chemical reactions of the pollutant while it travels (Bai et al. 2022). Furthermore, the model does not work well for wind speeds under two meters per second or for pollutant travel distances less than one hundred meters (Namrata and Wagh 2022; Raman 2009).

Many studies use the Gaussian model to estimate pollution dispersion from point source locations in different fields of research and study locations. Bai et al. (2022) estimate the dispersal pattern of aerosols from chicken and dairy farms in a remote area of the Guangdong Province in southern China. Specifically, Bai et al. (2022) examine airborne antibiotic-resistant genes and bacteria. They focus on livestock farms since the livestock industry heavily uses antibiotics.

Bai et al. (2022) use a Gaussian plume model to assess the airborne distribution of the aerosols. The variables used as inputs into the model include wind speed, source emission strength, Pasquill dispersion parameters, and height of the pollution source. In addition to modeling the aerosol concentrations, they measured the airborne bacteria at a human respiration level of 1.5 meters at multiple downwind sites. Their results implied that antibiotic resistance genes in the ambient air were carried by airborne bacteria that originated at the animal houses. They found that the Gaussian plume model predicted airborne antibiotic-resistant genes and bacteria concentrations 2-3 orders of magnitude lower than the measured samples. They suppose the difference in concentration is because the model ignored the ability of airborne bacteria to grow and multiply. Also, the model did not consider other sources of airborne bacteria in outdoor air-release microbes like soil and plant surfaces (Bai et al. 2022). This case study provides an

example of using the Gaussian plume equation and Pasquill dispersion parameters. It also shows how to assess the model results with measured pollutant concentrations.

Another case study using a Gaussian plume model is by Lotrecchiano et al. (2020), assessing PM 2.5 pollutant dispersion during a fire event in Avellino, Italy. Lotrecchiano et al. (2020) claim that accurate modeling of the pollutant distribution during fire incidents is essential to define human exposure to pollution. They chose the Gaussian model because it is simple, reliable, fast, valid, and the most used in pollutant emission modeling. The variables used are wind speed and direction, atmospheric stability, the source's height, and the point's measured location. To measure atmospheric stability, they use Pasquill-Gifford classes, and to estimate the PM 2.5 emission rate, they use an emission factor. To verify the Gaussian model results, Lotrecchiano et al. (2020) compared the model's output to sampling sites within the predicted plume. The model predicted much higher pollutant concentrations than those measured at the sampling sites. They think the overestimates may be due to the observed wind and air stability conditions; they think the fast wind speed and unstable atmospheric conditions made the PM 2.5 disperse out of the plume area more than the model predicted (Lotrecchiano et al. 2020). This case study provides another example of using the Gaussian plume equation, Pasquill dispersion classes, and an emission factor. It shows another instance of assessing and considering the model's results with measured pollutant concentrations.

2.2.2 AERMOD

AERMOD is a Gaussian-based steady-state air dispersion model developed by the American Meteorological Society and Environmental Protection Agency Regulatory Model Improvement Committee that models plumes. Rittner et al. (2020) used AERMOD to predict PM 10, PM 2.5, and black carbon in Scania, Sweden, from 2000 to 2011. They look at the emission

sources of road traffic, shipping, aviation, railroads, industries, major energy and heat producers, small-scale heating, non-road vehicles, and emissions from Zealand, Denmark. The variables they use in the model are the height of the source and target, temperature, wind speed, wind direction, and global solar radiation. The model runs hourly at one hundred square meter grid cells. To evaluate the AERMOD model, Rittner et al. (2020) compare the predicted results to measured data in the same place and time. They find that AERMOD performed well and had good agreement with the measured data, though they claim that the lack of data in some portions of their study area limits the results of their model.

They also mention that their model results would improve if they used more extensive meteorological data since they only used data from one point. Finally, they note that the Gaussian model does not account for topography, but since their study area is relatively flat with no large mountains, the Gaussian model is suitable (Rittner et al. 2020). The case study guides this thesis by describing pollutant sources, discussing the Gaussian plume model methods, and considering how to improve their model. Considering Rittner et al.'s (2020) critique that the Gaussian model does not account for topography, the study area in this thesis is appropriate since most CAFOs in Texas are in the Texas Panhandle, which is flat.

Another case study using AERMOD was by Bonifacio et al. (2012), where they modeled PM 10 dispersion from two beef feedlots in Kansas to develop correct emission factors for beef feedlots in Kansas. They discuss how no standard techniques and assumptions exist when generating typical emission factors for air dispersion models. For example, they state that some studies used short-term measurements, and some studies only measured on days when the cow pens were wet. Measuring only when the cow pens are wet is a concern since a typical cleaning method is to spray them with water, which washes the PM material away and removes it from

the air. The EPA uses an emission factor based on the assumption that PM emitted from cattle feedlots has the same distribution as PM emitted from agricultural soils.

Bonifacio et al. (2012) seek to model PM 10 emissions from cattle feedlots in Kansas to get a correct emission factor estimate. To do this, they considered feedlot size in cow capacity and acreage, dust control methods like water sprinkler systems and cow pen cleanings, and weather conditions like wind speed, wind direction, atmospheric pressure, air temperature, relative humidity, and precipitation. For the two feedlots they looked at, the pen cleaning procedures included scraping the manure on the pen surfaces and piling it in one location multiple times yearly. The cleaning protocols include hauling the manure piles from the pens at least once yearly. They also noted the cattle feeding schedules and the amount of water used each year by the sprinkler systems.

After Bonifacio et al. (2012) analyzed the variables affecting PM 10 dispersion, they used the AERMOD model to predict PM 10 concentration. For the dispersion model analysis, they measured the PM 10 concentrations and weather parameters at 20-minute intervals at each feedlot for two years. After they collected the data, they input the meteorological data into the AERMOD model to predict PM 10 concentrations for each feedlot (Bonifacio et al. 2012). They used inverse dispersion modeling based on the idea that a relationship exists between emission rates and downwind concentrations (Hadlocon et al. 2015). The Bonifacio et al. (2012) study used the AERMOD model to find the best-fit emission rate based on the measured PM 10 data. Using this emission rate, they calculated the emission factors using Equation 2, described below (Bonifacio et al. 2012). Bonifacio et al. (2012) inform this thesis by thoroughly discussing PM emissions from feedlots and the conditions that affect pollutant dispersion. They also describe how to measure and use emission factors accurately.

2.2.3 Lagrangian Model

The Lagrangian model is like a box model with the inside of the box defined. Lagrangian models describe a region of air as a box with an initial pollutant concentration. The input into the box is wind trajectory, and the outputs estimate pollutant particles transported along the wind path. Lagrangian models work for stable and unstable conditions and flat or complex terrain (Namrata and Wagh 2022).

An animal production-related case study that uses both AERMOD and a Lagrangian model is by Cheng, Kumar, and Wang-Li (2021), who model PM 10 concentrations emitted from an egg production facility in the southeast US. Cheng, Kumar, and Wang-Li (2021) use AERMOD and the Second-Order Closure Integrated Puff (SCIPUFF) model to simulate PM 10 concentration for two years. SCIPUFF is a Lagrangian-based dispersion model that uses Gaussian puffs. They aim to measure the emission rates from the two farms, analyze the model results, and compare the two models. Each model uses source characteristics such as whether the emissions are continuous or discrete, the release height, and the emission velocity. They also use point-of-interest characteristics like location, downwind distance, and height. Finally, they use meteorological conditions like wind speed and direction, temperature, relative humidity, solar radiation, cloud cover, and stability class.

After the analysis, they found that AERMOD performed much better than SCIPUFF. However, they note that AERMOD was sensitive to wind speed and that measuring a farm as a point source is too simplistic. They suggest instead representing the farm as an area (Cheng, Kumar, and Wang-Li 2021). For the spatial analysis in this thesis, each CAFO was digitized into a polygon and divided into multiple sections. These sections are the input locations used to run the Gaussian plume equation. Cheng, Kumar, and Wang-Li's (2021) conclusion informed these methods by representing the CAFOs as areas rather than point sources.

2.2.4 Other Dispersion Models

Other dispersion models include the Eulerian and computational fluid dynamics (CFD) models, which are the most complex plume dispersion models. The Eulerian model works by dividing the three-dimensional atmosphere into horizontal and vertical grid cells and defining models for each cell. CFD models create wind speed and pollutant concentration profiles in 3-dimensional environments, analyze dispersion in complex terrain, and allow for complex fluid flow analysis with mass and momentum calculations. Eulerian and CFD models calculate pollutant concentration in three dimensions and are computationally intensive (Namrata and Wagh 2022).

2.3 Gaussian Plume Equation

The most used air emission dispersion model is the Gaussian plume equation. This equation models the dispersion of airborne pollutants from a point source, assuming that random, turbulent wind velocities disperse the pollutants away from the plume centerline in a Gaussian distribution of concentration in the vertical and crosswind direction. This model fits well with field and laboratory evidence. Regulatory atmospheric dispersion modeling commonly uses the Gaussian equation (Namrata and Wagh 2022). The following equation outlined in Pfluger, Dacunto, and Hendricks (2014) estimates the concentration (C) of pollutants at any point (x, y, z) with the following equation:

$$C = \frac{Q}{2\pi u \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right)\left\{\exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right) + \exp\left(\frac{(z+H)^2}{2\sigma_z^2}\right)\right\} \quad (1)$$

where Q is source emission rate in grams per second, u is wind speed in meters per second, y is crosswind distance from point of interest in meters, z is vertical height of point of interest in meters, H is effective stack height in meters, σ_y is horizontal stability parameter in meters, and

σ_z is vertical stability parameter in meters (Pfluger, Dacunto, and Hendricks 2014). Figure 5 shows an example of a Gaussian plume where the grey area represents particle dispersion in the wind direction with a Gaussian distribution of particle concentration in the horizontal and vertical directions.

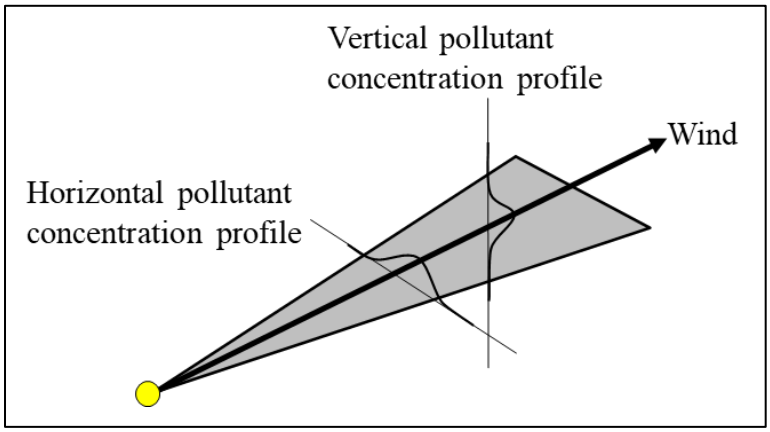

Figure 5. Diagram of Gaussian plume

Equation 1 can estimate the concentration of pollutants using the source emission rate, wind speed, and wind direction for any point. The emission rate (Q) can be measured or estimated using emission factors. Emission factors estimate the number of pollutants that the polluting source produces (Habib, Baticados, and Capareda 2022). For feedlots, the following equation, as explained by Bonifacio et al. (2012), calculates the emission factors (EF):

$$EF = \frac{Q_{yr} \times A}{10^3 \times N} \tag{2}$$

where Q_{yr} is mean annual emission flux in grams per square meter-day, A is total cow pen area in meters squared, and N is cattle in thousands. Equation 2 calculates the emission factor in kilograms per 1000 head-day. Habib, Baticados, and Capareda (2022) calculated the emission factor of a feedlot in the Texas Panhandle region by measuring PM 2.5 emissions for one week in June 2020. They had four sampling sites and multiple sampling devices at each location. They calculated the maximum mean emission factor for PM 2.5 using Equation 2 and found the

emission factor from the feedlot to be a maximum of 8.93 kg of PM 2.5 per 1,000 cows per day (Habib, Baticados, and Capareda 2022). This thesis used the emission factor to estimate the source emission rate (Q) in the Gaussian plume equation.

The equation assumes the measured wind speed (u) input is constant in the x direction. The model can calculate the crosswind distance from the point of interest (y) or it can be input as a measured value. The vertical height of the point of interest (z) is the height at which the model measures the concentration. For ground-level estimates, this height is zero. The stack height (H) is the height from which the source emits pollutants. Finally, the horizontal (σ_y) and vertical (σ_z) stability parameters are standard deviations of the pollutant concentration in the crosswind and vertical directions, respectively. Pasquill-Gifford-Turner (PGT) curves estimate these stability parameters.

Venkatram (1999) describes the developmental history of the PGT curves, beginning with Pasquill's 1961 method. Pasquill's method assumes the concentration distribution of airborne pollutants is Gaussian in the horizontal and vertical directions. Pasquill developed the curves through empirical analyses and observations of the dispersion of pollutants. In 1961, Gifford modified the dispersion equation by changing the angular spread and vertical extent in Pasquill's equation to Gaussian sigmas. Then, in 1970, Turner added the solar radiation ranges. The curves measure the stability classes, which depend on wind speed and solar radiation (Venkatram 1999). Table 1 describes the stability classes based on wind speed, day radiation intensity, and night cloudiness. There are six stability classes A – F, with class A being the most unstable and turbulent and class F being the most stable and least turbulent (Lotrecchiano 2020). The PGT stability classes calculate the horizontal and vertical stability factor variables needed in the Gaussian plume equation in this thesis.

Table 1. Atmospheric stability classes for the PGT dispersion model

Wind speed (m/s)	Day radiation intensity			Night Cloudiness	
	Strong	Medium	Slight	≥50%	<50%
< 2	A	A – B	B	F	F
2 – 3	A – B	B	C	E	F
3 – 5	B	B – C	C	D	E
5 – 6	C	C – D	D	D	D
> 6	C	D	C	D	D

Source: APSI n.d.; Raman 2009

The meteorological variable values at each CAFO determine the atmospheric stability. The strong, medium, and slight day radiation classifications are: > 700 W/m^2, 350 – 700 W/m^2, and < 350 W/m^2, respectively (APSI n.d.; Lotrecchiano 2020). Cloud fraction, the fraction of the sky covered by clouds, represents night cloudiness.

For each PGT class, PGT curves exist for the horizontal and vertical dispersion coefficients. The equations for the horizontal coefficients are in Table 2, and the vertical coefficients are in Table 3. The coefficients depend on the distance from the source to the point of interest, or the distance downwind (x), as well as the PGT class. The equations in Table 2 calculate the vertical and horizontal stability variables used in the Gaussian plume equation.

Table 2. Equations for horizontal dispersion coefficient for continuous plumes

Stability Class	σ_y (m)
A	$\sigma_y = 0.493x^{0.88}$
B	$\sigma_y = 0.337x^{0.88}$
C	$\sigma_y = 0.195x^{0.90}$
D	$\sigma_y = 0.128x^{0.90}$
E	$\sigma_y = 0.091x^{0.91}$
F	$\sigma_y = 0.067x^{0.90}$

Source: Raman 2009

Table 3. Equations for vertical dispersion coefficient for continuous plumes

Stability Class	Distance Downwind (m)	σ_z (m)
A	100 – 300 300 – 3,000	$\sigma_z = 0.087x^{0.88}$ $\log_{10}\sigma_z = -1.67 + 0.902\log_{10}x + 0.181(\log_{10}x)^2$
B	100 – 500 500 – 20,000	$\sigma_z = 0.135x^{0.95}$ $\log_{10}\sigma_z = -1.25 + 1.09\log_{10}x + 0.0018(\log_{10}x)^2$
C	100 – 100,000	$\sigma_z = 0.112x^{0.91}$
D	100 – 500 500 – 100,000	$\sigma_z = 0.093x^{0.85}$ $\log_{10}\sigma_z = -1.22 + 1.08\log_{10}x - 0.061(\log_{10}x)^2$
E	100 – 500 500 – 100,000	$\sigma_z = 0.082x^{0.82}$ $\log_{10}\sigma_z = -1.19 + 1.04\log_{10}x - 0.070(\log_{10}x)^2$
F	100 – 500 500 – 100,000	$\sigma_z = 0.057x^{0.80}$ $\log_{10}\sigma_z = -1.91 + 1.37\log_{10}x - 0.119(\log_{10}x)^2$

Source: Raman 2009

2.4 Assessment of Models

Once a dispersion model analysis is performed, there are different ways in which the model results can be assessed. These methods include linear regression and other statistical significance tests.

2.4.1 Linear Regression

Linear regression estimates a linear relationship between a dependent variable and a set of explanatory variables (Burchfield n.d.). Regression analysis is the most commonly used statistic in the social sciences to evaluate relationships between feature attributes. Regression helps better understand relationships occurring in a place, predict where something is likely to occur, or examine why things occur where they do (Esri n.d.a). Research provides examples of how linear regression compares air dispersion model results to measured data.

Mathieu, Gray, and Richmond-Bryant (2023) used linear regression to assess if diesel PM was spatially associated with COVID-19 mortality rates in 2020. The diesel PM data they used was from the 2018 AirToxScreen database, which was created by the EPA using AERMOD to estimate the air pollutant concentrations. The linear regression models they used were ordinary least squares (OLS), two global models, a spatial lag model, a spatial error model, and a geographically weighted regression. The variables used to assess the linear regression model results were the coefficient, probability, spatial autocorrelation z-score, Breusch-Pagan test for heteroscedasticity, R^2, and corrected Akaike's information criterion (AICc).

Huertas et al. (2012) estimated the total suspended PM in the air in a coal mining region in Colombia in 2008-2010 using AERMOD and another Gaussian model named ISC3 using topographic and meteorological data. They compared their results to measured values using linear regression. The criteria used to assess the linear regression model were slope of the linear regression, offset or model estimated background concentration, R^2, coefficient of correlation, probability, fractional bias (FB), geometric mean bias (MG), normalized mean square error (NMSE), root mean square errors (RMSE), and fraction of two (FAC2).

Yang, Chen, and Yu (2007) use AERMOD and ISC3 to predict effluent concentrations in Alaska and compare the results to measured data using linear regression, quantile-quantile, and residual box methods. The measures they used to assess the results were normalized correlation, FAC2, and RMSE.

This thesis used an OLS model linear regression analysis to assess the model's results. OLS is the best-known regression technique and a starting point for all spatial regression analyses. OLS provides a global model of the estimated variable using a single regression

equation (Esri n.d.a). OLS is unbiased, consistent, and the best linear unbiased estimator (Burchfield n.d.; Zhu 2022).

2.4.2 Other Statistical Significance Tests

Differences between model dependent and explanatory variables can also be compared using other statistical significance tests, as described in research. Bai et al. (2022) used Linear discriminant analysis Effect Size (LEfSe), which identifies the differences between biological samples by determining the features most likely to explain variations. LEfSe utilizes standard statistical significance tests and additional biological tests.

Lotrecchiano et al. (2020) verified their model's predictions by comparing the estimated pollutant concentrations at three monitoring stations to the measured data at those stations. They calculate the percent by which the values differ at the monitoring stations to compare the results to their measured data.

Rittner et al. (2020) similarly compared the estimated data to the data measured at monitoring stations within their study area. They calculated the mean difference between the modeled and measured concentrations, Pearson correlations, and RMSE. The Pearson correlation measures linear correlation; it is a number between negative one and one that measures the strength and direction of the relationship between two variables.

Bonifacio et al. (2012) assessed their data using statistical tests to check the data for normality and outliers. All of the datasets had non-normal distributions, so to compare the data sets for the two feedlots, they used nonparametric one-way analysis of variance (ANOVA). ANOVA is a statistical test that analyzes the difference between the means of more than two groups. Bonifacio et al. (2012) compared the estimated and measured emission fluxes and emission factors for the two feedlots using median values.

Hadlocon et al. (2015) similarly used ANOVA, t-test for paired comparisons, and Tukey-Kramer's honest significant difference for mean comparisons at monitoring station locations. They compared the model results with the measured PM concentrations, and five statistical parameters were used to evaluate the model performance, including FB, MG, NMSE, geometric variance, and FAC2. The statistical parameters measured the likelihood of the model to deviate from the actual measurement values and the bias and scatter of the results.

Cheng, Kumar, and Wang-Li (2021) also used ANOVA to assess the model performance by comparing the modeled results to the measured data. The four parameters they measured were FAC2, NMSE, FB, and MG. However, they also used a simple linear regression analysis. The linear regression measured the coefficient, p-value, and f-value for the three source types used in AERMOD.

Chapter 3 Data and Methods

This chapter outlines the data and methods used to perform a spatial analysis of the CAFO-emitted PM 2.5 concentrations in Texas. Section one describes the data with a chart, maps, and detailed summaries. Section two outlines the methods of the PM 2.5 plume equation spatial analysis using the Gaussian model and the assessment of results using linear regression with a graphic outline and detailed description.

3.1 Data and Data Preparation

The datasets used in this analysis were CAFO locations, meteorological variables, and measured PM 2.5 data. In addition to the CAFO locations, characteristic meteorological data were necessary for spatial analysis because they influence the way and the extent to which pollutants travel. The meteorological data used in this study were global horizontal irradiance (GHI), cloud fraction, wind speed, and wind direction. Table 4 contains a list of the datasets used in this study with the name, description, temporal and spatial scale, and availability of each dataset. The data collected were those required in the Gaussian plume equation and a measured PM 2.5 concentration dataset to compare the model results. The most current available PM 2.5 measured data were for 2016 values, so all the other data layers closest to 2016 were selected for use in this thesis. Additionally, all of the data layers were clipped to the Texas border. Following the table are in-depth discussions on the data processing for this study.

Table 4. Dataset description

Dataset	Description	Temporal Scale	Scale	Acquisition
TCEQ CAFO Permits	Location and other information on the permitted CAFOs in Texas	Coverage beginning in 2004	Latitude/ longitude coordinate	Obtained on June 17, 2022 through a request to TCEQ
Global Solar Atlas Global Horizontal Irradiance	Long-term average of daily total solar radiation at a horizontal plane from direct and diffuse sources	1994/1999/2007 (depending on the region) to 2021	30-arc seconds (about 1 km)	Accessed January 25, 2023 from Esri Living Atlas
NASA Earth Observations MODIS Atmosphere Science Team Cloud fraction average	Average cloud fraction	January – December 2016	1 km	Downloaded January 29, 2023 from the NASA Earth Observers site
NREL US Average Wind Speed	Average wind speed from 10m to 200m elevation	2012	2 km	Accessed on February 1, 2023 from Esri Living Atlas
NASA POWER Project Wind Direction	US wind direction at 10m	2016	.5 degrees (about 5.55 km)	Accessed on February 1, 2023 from Esri Living Atlas
SEDAC Annual PM 2.5 concentration	Annual PM 2.5 concentration	2016	1 km	Accessed January 25, 2023 from SEDAC

3.1.1 TCEQ CAFO Permits

The central dataset used in this analysis was TCEQ CAFO permits, which has the CAFO locations, the number of cows per CAFO, and the acreage of each CAFO in Texas. Robert Macias, the Open Records Coordinator at TCEQ, provided the CAFO Permit data through a Public Information Request on June 17, 2022. The CAFO locations were latitude/longitude coordinates. The TCEQ CAFO location spreadsheet was converted to spatial data using the XY Table to Point tool. The tool geocoded the latitude/longitude coordinates to a point feature class. Figure 6 shows a map of these CAFOs. Texas contains 221 CAFO locations with permit

coverage from January 17, 1991 to May 5, 2022. Within the dataset provided by TCEQ, 6,438,525 cows are within 221 CAFO points. However, TCEQ's count of cows is outdated or includes more than just beef cattle based on the graph in Figure 3 with NASS data which shows 4,475,000 beef cows in Texas as of January 1, 2022. This thesis estimated PM 2.5 concentration in 2016, so the analysis only used the CAFOs with permit dates before January 1, 2017, resulting in 165 CAFOs. The CAFOs operating in 2016 and earlier were selected using the Export Features tool.

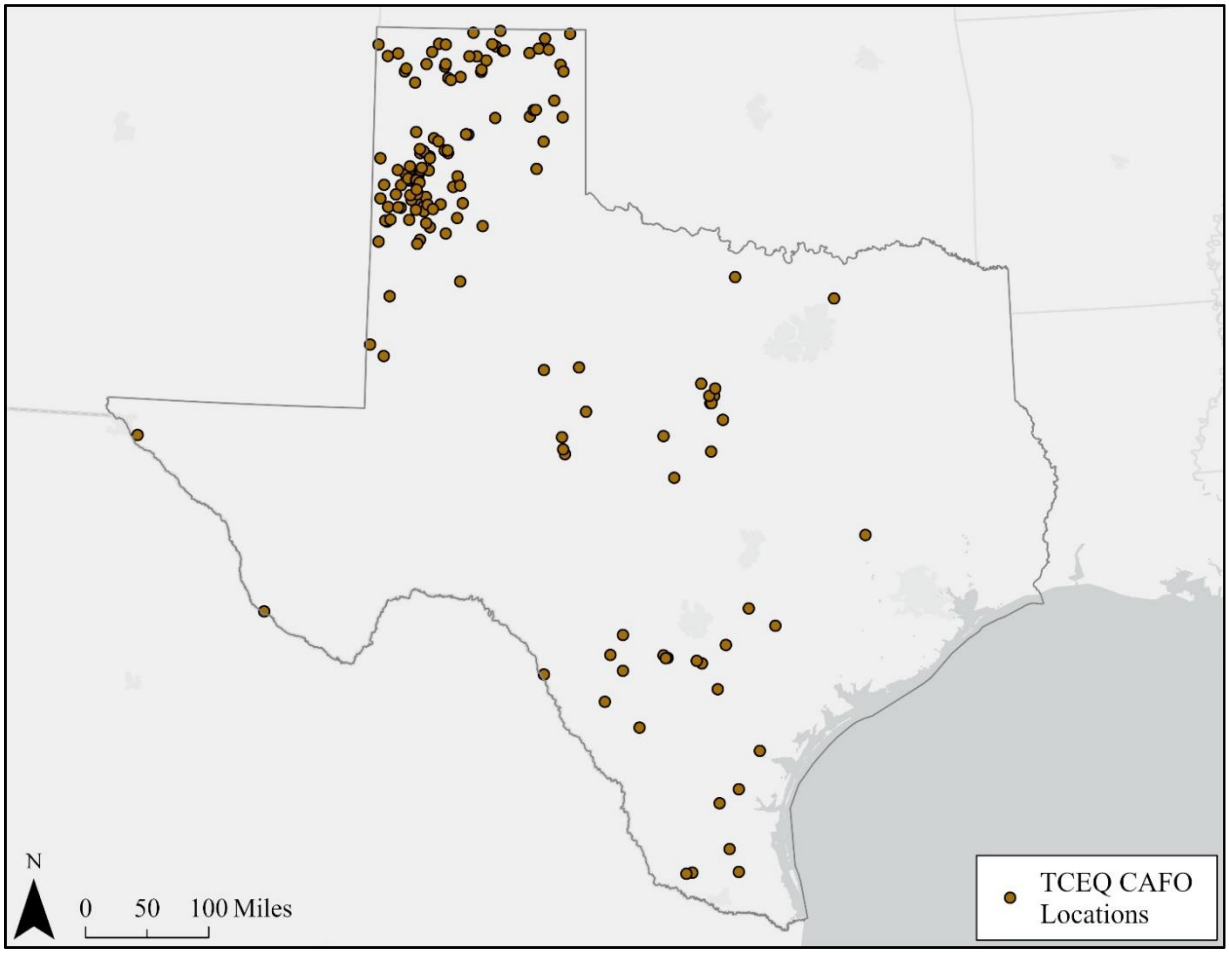

Figure 6. TCEQ CAFO locations

Since CAFOs often cover large areas, using one point source location to measure the PM 2.5 dispersion was limiting. Creating dispersion plumes from multiple points within the CAFO

extent was more representative. Since the TCEQ CAFO data were in point form, the extent of the CAFO areas needed to be determined. The extents were found by manually digitizing the CAFO boundaries using Esri's Aerial Imagery basemap. Once the boundaries were determined, each CAFO needed new source points. These points represent the new locations where the script measured the dispersion plumes.

To create the CAFO point source locations, each CAFO extent needed to be divided into one-square kilometer sections that aligned with the measured SEDAC PM 2.5 data that contained the CAFO attribute data. The Spatial Join tool was used to extract the attributes from the TCEQ CAFO points into the digitized CAFO polygons. Then, using the Create Fishnet tool, a grid covering all of Texas at one square kilometer resolution was created based on the measured SEDAC PM 2.5 data. The CAFO extent polygons were then clipped to the one-kilometer grid using the Clip tool, and the attributes from the CAFO polygons were extracted into the clipped geometries using the Spatial Join tool. Finally, the polygons were converted to points, representing the locations from which the script measured the plumes, using the Feature to Point tool. Figure 7 shows an example of the TCEQ CAFO point distribution to the CAFO extent. The red polygons show the digitized CAFO extent, the blue grid shows the one-square kilometer sections that aligned with the measured SEDAC PM 2.5 data, the red points show the new CAFO locations, and the yellow points show the TCEQ CAFO point locations.

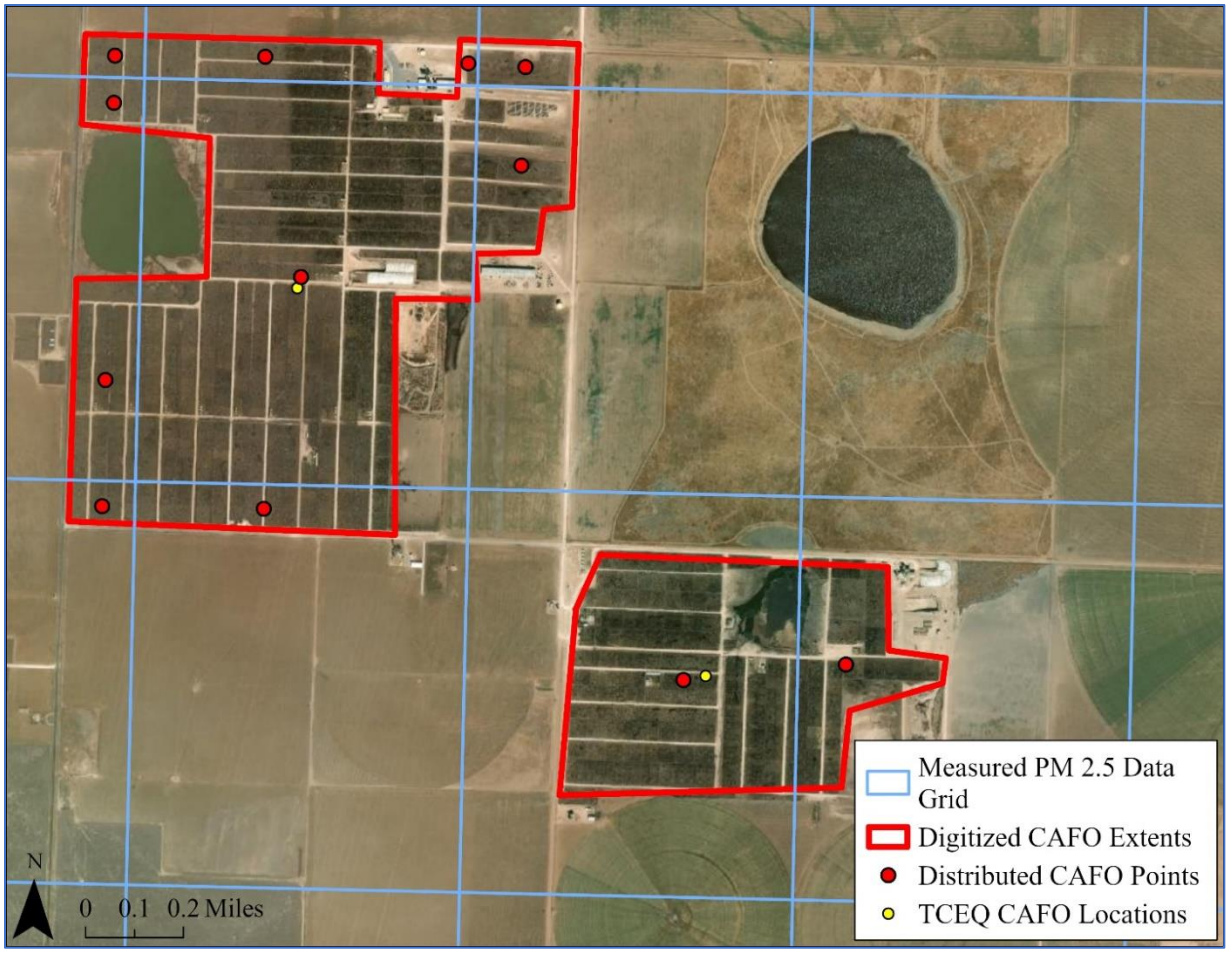

Figure 7. Example of distributed CAFO point locations

3.1.2 Global Solar Atlas Global Horizontal Irradiance

GHI plays a role in the PM 2.5 dispersion analysis by affecting atmospheric stability. GHI is the total solar radiation from direct and diffuse sources at a horizontal plane. The GHI dataset used in this thesis was from the Global Solar Atlas. The data were accessed from the Esri Living Atlas on January 25, 2023. The raster images were 30-arc second resolution with the GHI in kWh/m^2 per year. They represent the average daily totals of GHI from 1994/1999/2007 (depending on the region) to 2021. The Global Solar Atlas data only has the long-term average data, so GHI data for 2016 was not available for this analysis.

The GHI dataset required processing for the Gaussian plume equation analysis. To limit the raster's extent to Texas, the cells within the Texas border were extracted using the Clip Raster tool. Additionally, the units were converted from kWh/m^2 per year to W/m^2. A kilowatt-hour is a measure of energy, while a watt is a measure of power. The basic relationship between energy and power is that energy equals power multiplied by time. The time used to convert from energy to power for this study was the average length of daylight at the CAFO locations. The average length of daylight is necessary because it represents how much time Texas receives solar radiation.

A central point among the CAFO locations in Texas was determined using the Central Feature tool in ArcGIS Pro to calculate the average length of daylight at the CAFO locations. This point is at the latitude/longitude coordinate: 34.510555, -102.27277. This coordinate location was input into the NOAA Solar Calculator on the Global Monitoring website, which produced Sunrise and Sunset Tables for a year. Based on the tables, the average length of daylight per year is 11.953 hours.

After determining the time component, GHI was converted from energy to power. GHI was multiplied by 1,000 and then divided by (365 * 11.953) using the Raster Calculator tool to convert the units. Multiplying by 1,000 converts kilowatts to watts, dividing by 365 makes the unit per day rather than per year, and dividing by 11.953 converts the watt amount from per day to per daylight hours. Figure 8 shows a map of the resulting raster layer.

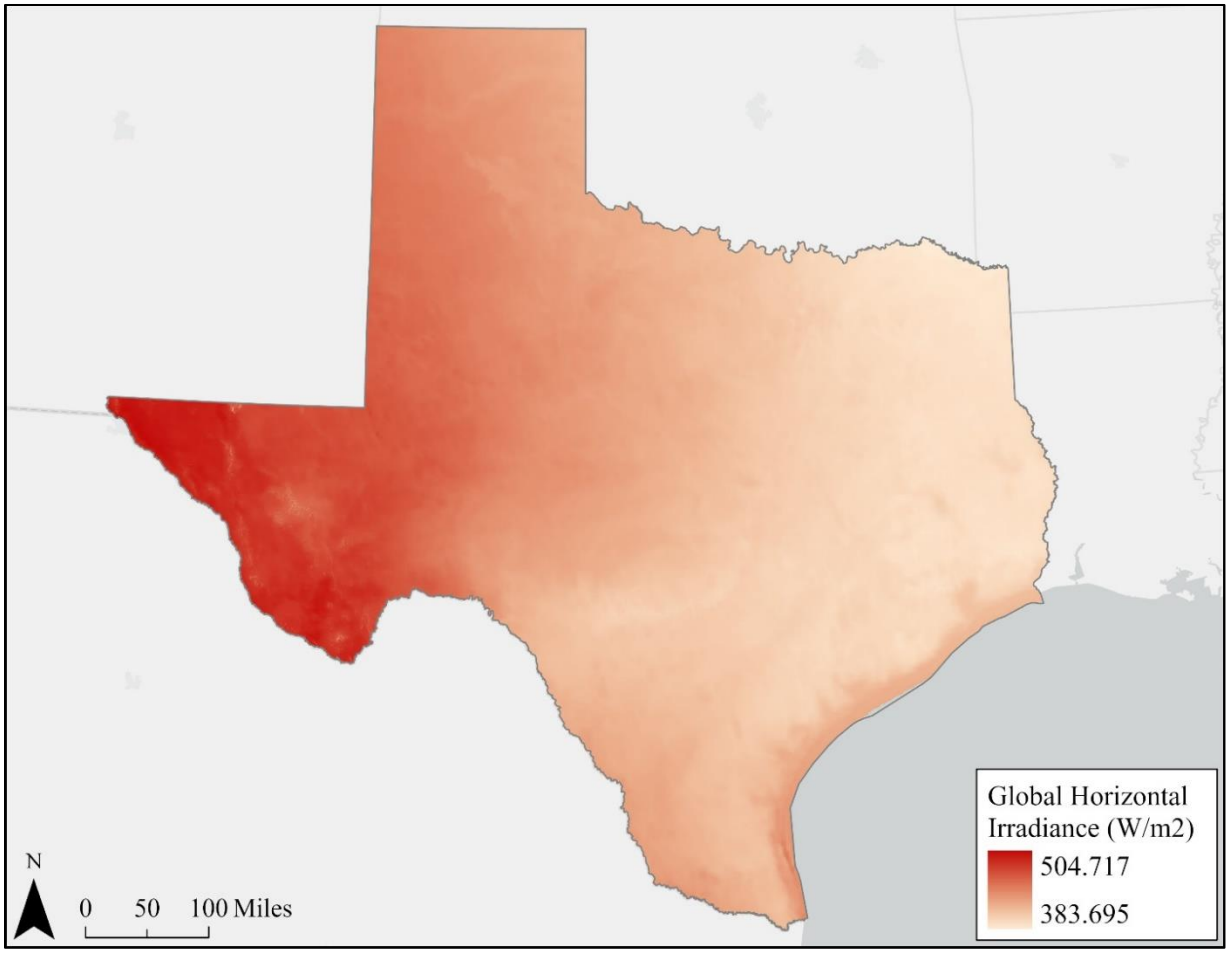

Figure 8. Global Star Atlas global horizontal irradiance in W/m^2

3.1.3 NASA Earth Observations Cloud Fraction Average

Another meteorological variable needed to determine atmospheric stability, which is a measure of the atmosphere's motion, is cloud fraction. Cloud fraction is the percent of Earth's surface that clouds cover. The cloud fraction data used in this study were accessed on January 29, 2023 through the National Aeronautics and Space Administration (NASA) Earth Observations website. The data represented the monthly average cloud fraction from January 2016 to December 2016 at one-kilometer spatial resolution. The data were downloaded as separate files for each month, with values ranging from 1-255. For this analysis, the values were reclassified to range from zero to one using the Raster Calculator tool. The reclassification was

for the values to represent a cloud fraction percentage. Once all the raster's values ranged from zero to one, the rasters were averaged and clipped to the Texas border using the Raster Calculator tool. The resulting raster represents the average cloud fraction for 2016, as seen in Figure 9.

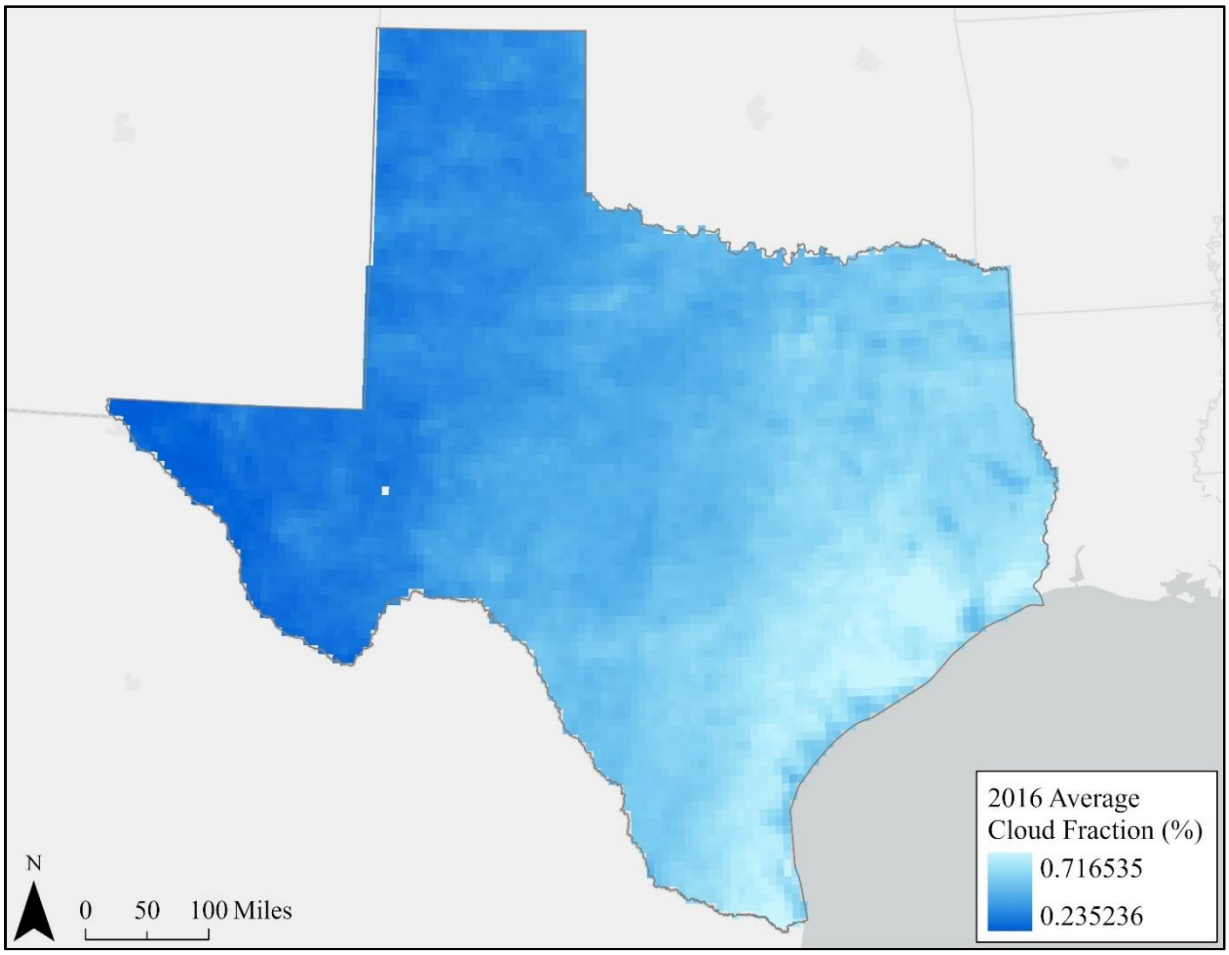

Figure 9. NASA MODIS 2016 average cloud fraction

3.1.4 NREL US Average Wind Speed

The next meteorological variable in this analysis is wind speed, which is important because it influences how far the PM 2.5 particles travel. Wind speed data were accessed on February 1, 2023 from the Esri Living Atlas and were from the Wind Integration National Dataset (WIND) Toolkit, which the National Renewable Energy Laboratory (NREL) developed.

The wind speed data was the average wind speed in 2012 in meters per second, the closest year to 2016 available in the WIND Toolkit. NREL measured the data in the US at 10 - 200 meters in height and two-kilometer grid spacing. The average wind speeds in 2012 in Texas are in Figure 10.

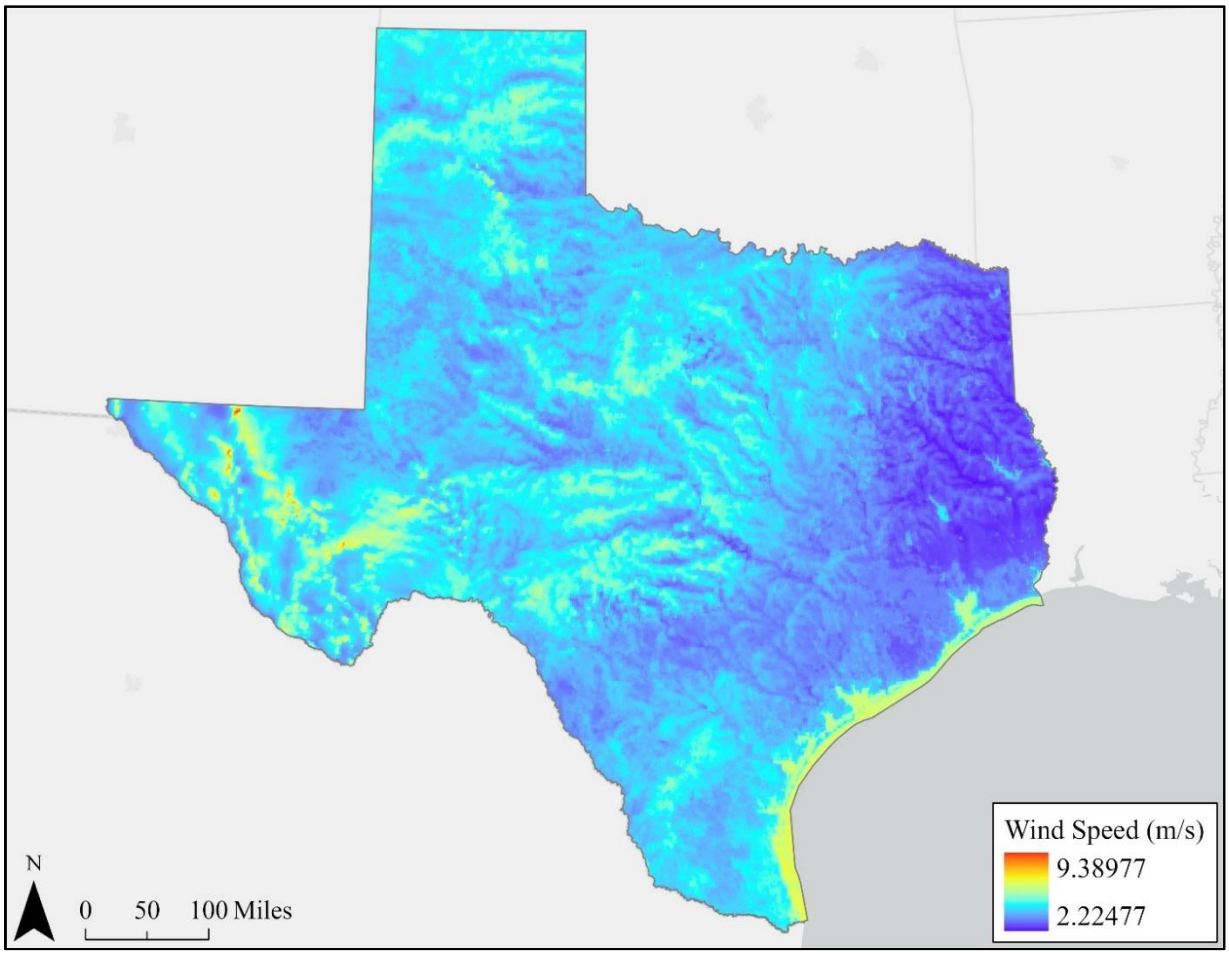

Figure 10. NREL wind speed in meters per second

3.1.5 NASA POWER Project Wind Direction

In addition to wind speed, wind direction was essential in this analysis because it determines where the pollutants disperse from CAFOs. The wind direction data were accessed from Esri's Living Atlas on February 1, 2023 and were obtained from the NASA Langley Research Center Prediction of Worldwide Energy Resource (POWER) Project funded through

the NASA Earth Science Directorate Applied Science Program. The wind direction represented an average for 2016, measured at ten-meter height and 0.5-degree grid resolution. The data were downloaded in degrees, but the wind direction values were converted to radians for this analysis by multiplying the raster by $\pi/180$ using Raster Calculator. A visual of the wind direction data is in Figure 11.

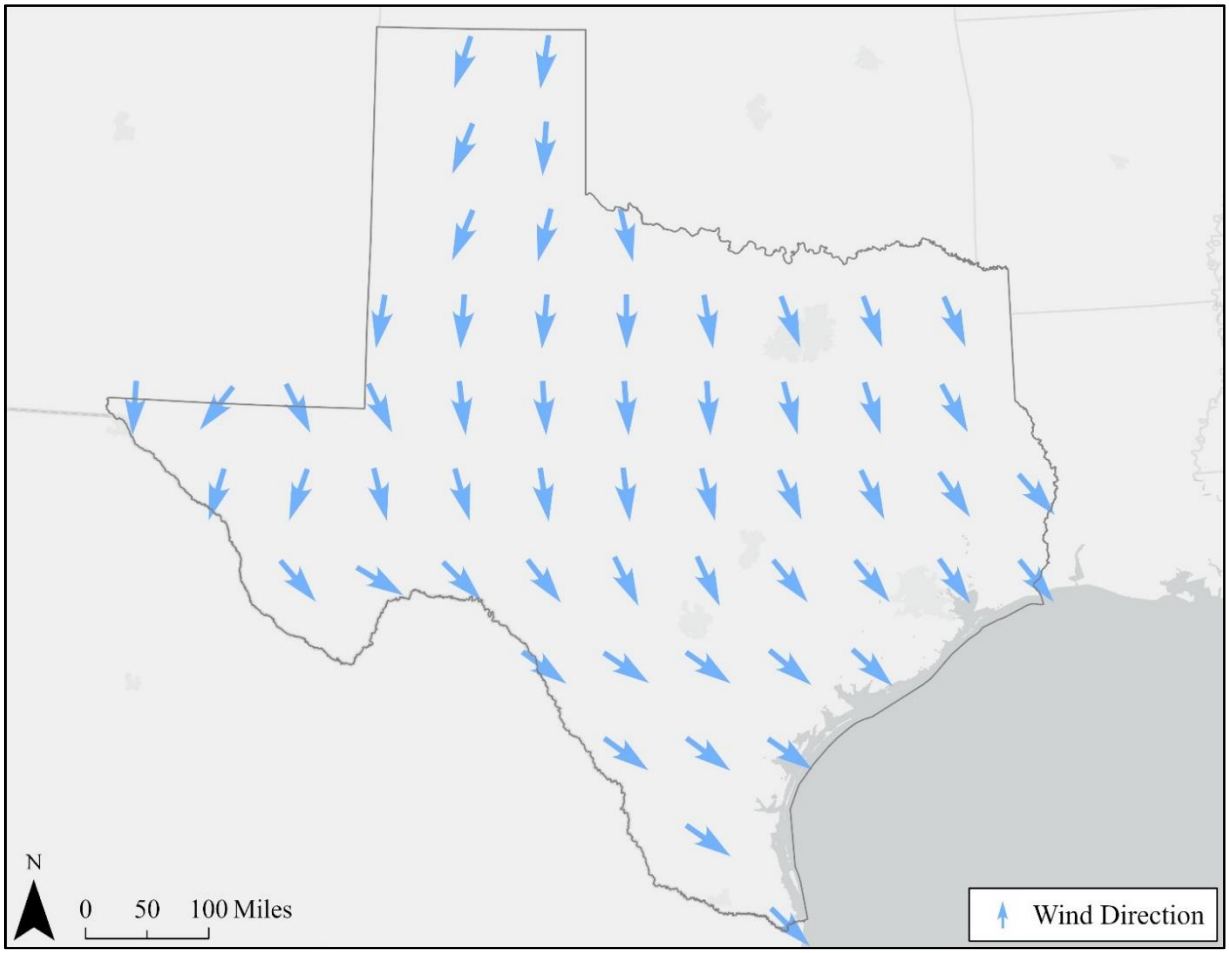

Figure 11. NASA POWER wind direction

3.1.6 SEDAC Annual PM 2.5 Concentration

The final data used in this analysis were measured PM 2.5 concentrations, used for comparison with the Gaussian plume equation model results. The measured PM 2.5 concentration data were downloaded on January 25, 2023 from the Socioeconomic Data and

Applications Center (SEDAC). The SEDAC data predicts 2016 PM 2.5 in micrograms per cubic meter ($\mu g/m^3$) in square kilometers with an extent of the contiguous US. The concentration predictions were created using data from various governmental organizations' air quality monitoring sites. These include monitors from the Air Quality System operated by the Environmental Protection Agency (EPA), the Interagency Monitoring of Protected Visual Environments, the Clean Air Status and Trends Network, and other regional or local monitoring organizations. The data used were from 2000 to 2015 (Di et al. 2021). In Texas, the monitoring site locations were sparse. There were 96 TCEQ PM 2.5 air quality monitoring sites in Texas, most of which are near cities (TCEQ 2022a).

Di et al. (2021) also incorporated one square kilometer satellite-derived aerosol optical depth data that was measured using the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument to predict PM 2.5 compositions. The satellite data were used to supplement the air quality monitoring site data. Their analysis also included chemical transport model predictions, land use data, and meteorological variables (Di et al. 2021).

To produce the output PM 2.5 surface, Di et al. (2021) trained neural network, random forest, and gradient boosting algorithms to predict PM 2.5 using the input variables. They then used a geographically weighted model to combine the PM 2.5 predictions. Di et al. (2021) used many validation methods to produce a final result with a spatial R^2 value of 0.89 and a temporal R^2 value of 0.85. These R^2 values indicated a good model performance. Di et al. also predict monthly standard deviations, which measure PM 2.5 concentration uncertainty in Texas of about 2.0-2.5 micrograms per cubic meter (Di et al. 2021). Figure 12 shows the final output of the SEDAC annual PM 2.5 concentration data and the TCEQ PM 2.5 monitoring sites (TCEQ 2022a).

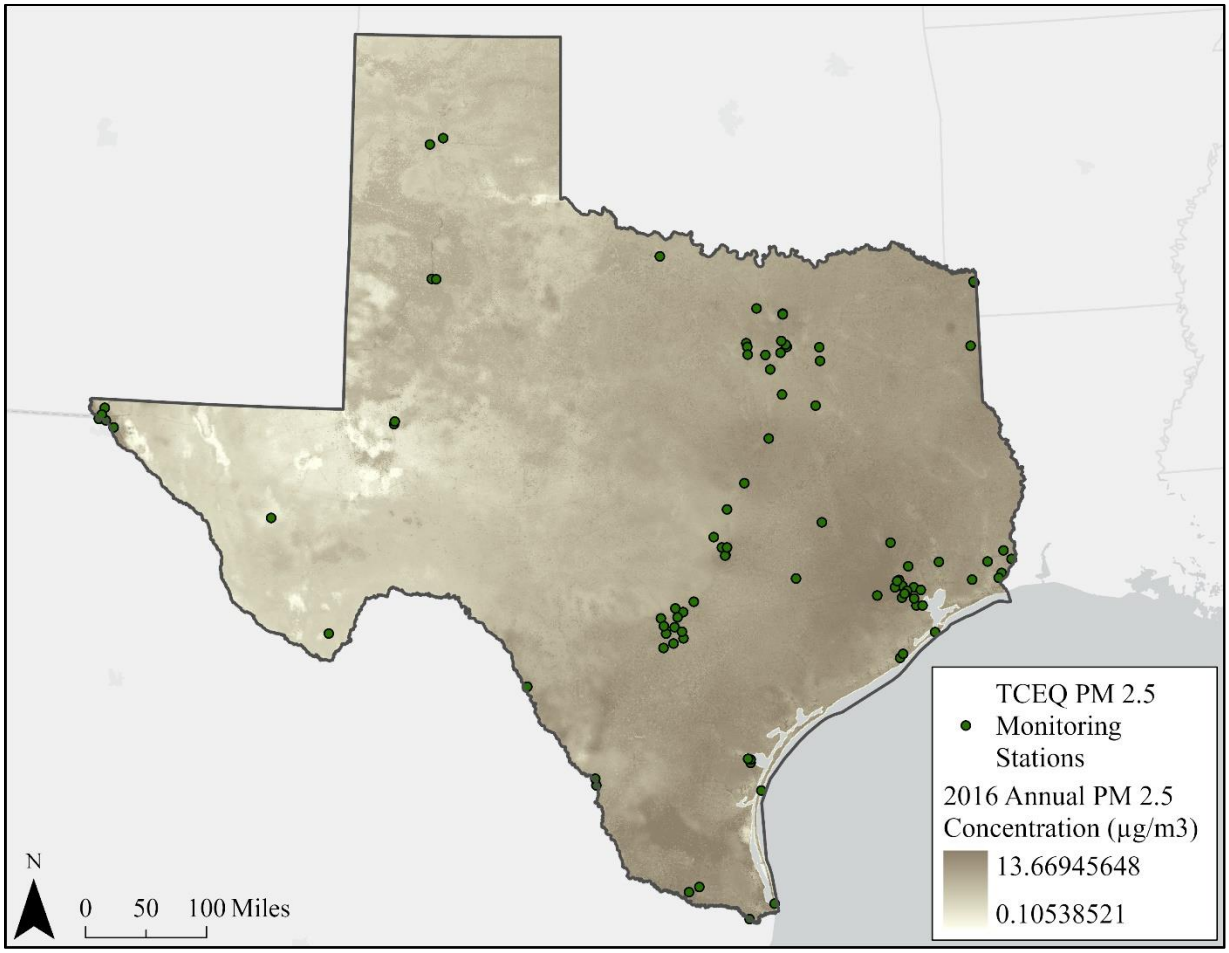

Figure 12. SEDAC 2016 annual PM 2.5 concentration in µg/m^3

Although the SEDAC measured PM 2.5 data were created using sparse air quality monitoring station locations in Texas, especially in the Panhandle and central Texas, where CAFO concentration was highest, the dataset was still reliable. Since the SEDAC data also included satellite imagery, chemical transport model predictions, land use data, and meteorological variables, the sparse air monitoring sites were supplemented, making the data trustworthy enough to use as the measured data in this project. The data reliability was demonstrated with the PM 2.5 concentration uncertainty in Texas, which was about 2.0-2.5 micrograms per cubic meter (Di et al. 2021). For the purpose of comparing the SEDAC data to the project's estimated PM 2.5 concentrations, the SEDAC data's uncertainty is acceptable.

3.2 Methods

This study performed a spatial analysis of the CAFOs in Texas and estimated the PM 2.5 concentrations they release using a Gaussian plume dispersion model. This was done using estimated PM 2.5 emission rates and meteorological variables as inputs into the Gaussian dispersion model. The model's results were compared to measured PM 2.5 data using linear regression. An outline of the methods performed in this analysis is in Figure 13, followed by a further discussion of the methods.

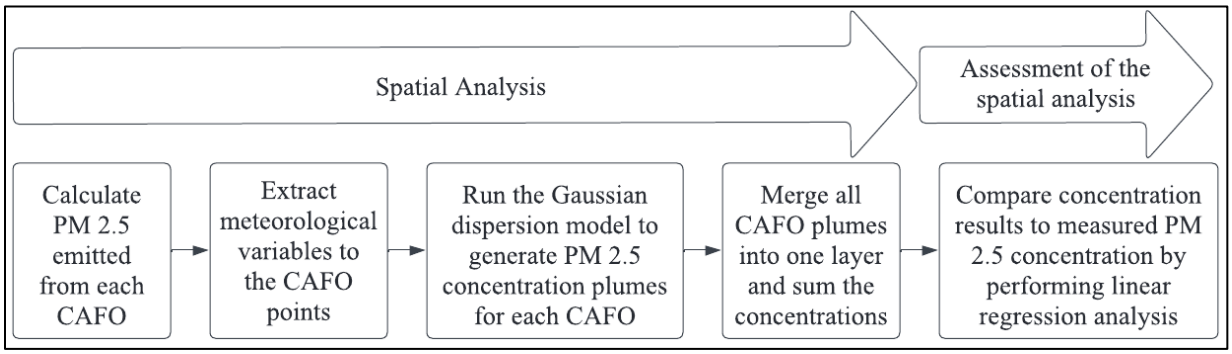

Figure 13. Methods outline

3.2.1 Spatial Analysis

The analysis in this thesis used the Gaussian plume equation model since it is the most common air emission dispersion model, and extensive research was available to guide the spatial analysis. It is a simple equation to execute, the data inputs are readily available, and the model produces accurate results. The limitations of the Gaussian model are acceptable since PM 2.5 does not undergo chemical reactions as it travels, the wind conditions near the CAFOs are strong enough, and the PM 2.5 concentrations estimated by the model are at a scale larger than 100 meters.

The spatial analysis estimated the concentration and travel extent of the PM 2.5 released from CAFOs using the Gaussian plume equation, as defined in Equation 1. To calculate the

emission rate of PM 2.5 from each CAFO, an emission factor of 8.93 kg of PM 2.5 per 1,000 cows per day was used (Habib, Baticados, and Capareda 2022). For use in the Gaussian plume equation in this analysis, the emission factor was converted to 103.356481481 micrograms per cow per second. Then, the emission factor was multiplied by the number of cows to calculate each CAFO's source emission rate in micrograms per second.

However, since the CAFO locations were spread throughout the CAFO extents to better represent the CAFOs as areas as previously discussed, a more accurate model of the source emissions was needed. A second Python script was used to adjust the emission rates to distribute the emission rate to the multiple points within each CAFO. The script used a search cursor to iterate through the new CAFO points. The script selected the other points with the same CAFO name for each point and counted the number of selected points. Then, the PM 2.5 emission rate was divided by the number of selected points. By dividing the emission rate by the number of CAFOs with the same name, it evenly distributed the total emissions to each new point within the CAFO extent. Finally, the script exported the resulting emission factors to a .csv file that was then joined to the points using the Add Join tool.

Once the new emission factors were calculated for the new CAFO points, the wind speed, wind direction, GHI, and cloud fraction were extracted from the source rasters to the new CAFO points using the Extract Values to Points tool in ArcGIS Pro. The meteorological values were extracted from the raster data layers and added to the point source CAFO data for use in the Gaussian plume equation.

The last step before running the Gaussian plume equation was to create the point locations where the equation would execute. Using the grid aligned with the measured PM 2.5 SEDAC data, the center point of each grid cell was extracted using the Feature to Point tool.

Extracting the center point aligned the resulting PM 2.5 concentration estimations with the measured data. Figure 14 shows an example of a CAFO in yellow and the point of interest points that align with the SEDAC data in purple. These purple point locations were the point of interest inputs used in the Gaussian equation and were where the equation estimated the PM 2.5 concentrations for the yellow CAFO point.

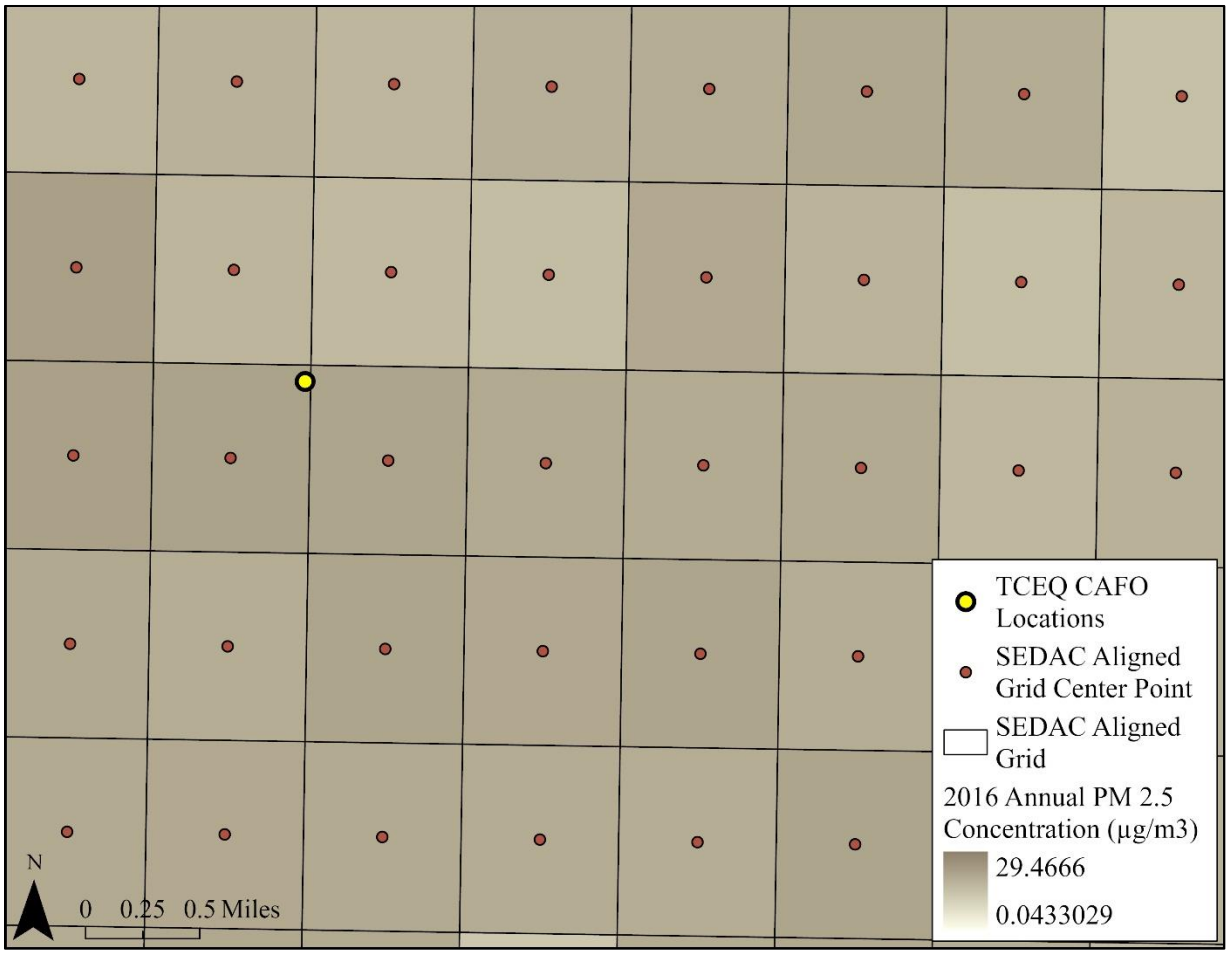

Figure 14. Example point of interest inputs into Gaussian plume equation

Once the script prepared the input datasets and created the point of interest locations, the script then executed the Gaussian distribution model. The inputs into this script were the new CAFO points, the fishnet grid polygons, and the corresponding grid points. The script iterated

through the CAFO points using a search cursor and accessed the point's shape, emission rate, wind speed, wind direction, GHI, cloud fraction, and location.

The script then iterated through the grid points and performed the Gaussian plume dispersion model calculation at each point for each CAFO point. The equation began by calculating the crosswind distance to the point of interest. The crosswind distance was calculated by first measuring the distance from the CAFO to the grid point in the x and y direction. The distance in meters between the points was calculated using the ".distanceTo" parameter of the search cursor. Then, the angle, or θ, between the CAFO and the grid point was calculated using the following equation:

$$atan\,2(\Delta y, \Delta x) \tag{3}$$

where Δy is the difference in the y direction, and Δx is the difference in the x direction between the CAFO and the grid point.

Next, the Python script calculated the angle of the wind in relation to θ. Table 5 shows the equations the script used to calculate the wind angle based on θ. The equations represent the wind direction at the CAFO by *wd*.

Table 5. Wind angle based on the angle between CAFO and grid point

θ	Wind Angle
$\theta \leq -\frac{\pi}{2}$	$(\pi - wd) + \left\lvert \frac{\pi}{2} - \lvert\theta\rvert \right\rvert$
$-\frac{\pi}{2} < \theta < 0$	$(\pi - wd) - \left(\frac{\pi}{2} - \lvert\theta\rvert\right)$
$0 \leq \theta \leq \frac{\pi}{2}$	$wd - \frac{\pi}{2} - \theta$
$\theta > \frac{\pi}{2}$	$\left(\theta - \frac{\pi}{2}\right) + wd$

Once the wind angle in relation to θ was estimated, the script calculated the crosswind distance from the grid point to the wind direction. If the wind angle was greater than $\pi/2$, the

crosswind distance was exponential. For all other angles, the crosswind distance equaled the distance between the CAFO and the grid point multiplied by *sin* (wind angle). Figure 15 shows a diagram of the variables used to calculate the crosswind distance. The yellow point represents the CAFO, and the purple point represents the point of interest. Figure 15 shows the distance between the CAFO and the point of interest, the angle between the CAFO and the grid point (θ), the wind direction (*wd*), and the crosswind distance (*y*).

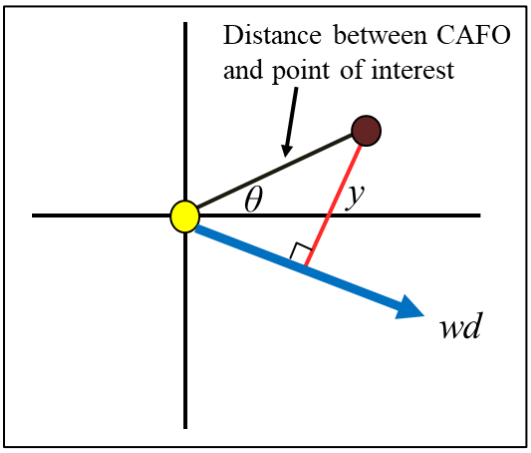

Figure 15. Wind angle calculation diagram

The next portion of the script calculated the horizontal and vertical stability parameters for the Gaussian plume equation. These parameters were determined based on the PGT stability classes. Table 1 lists the stability classes' GHI, cloud fraction, and wind speed specifications. The script first ran the concentration analysis for daytime PM 2.5 emission by looking at the GHI and wind speed at the CAFO location to determine the appropriate stability class. Then, the appropriate equations for calculating the parameters were used when the class was determined. Table 2 specifies these equations for the horizontal dispersion coefficient, and Table 3 specifies the equations for the vertical dispersion coefficient.

Similarly, after the PM 2.5 concentration was calculated for the daytime, the script calculated it for the nighttime. The script looked at the CAFO location's cloud fraction and wind

speed to determine the appropriate stability class. Nighttime cloud fraction is important because cloudy skies at night reflect infrared energy to the earth, creating a warmer atmosphere than on clear nights, which affects atmospheric stability. The cloud fraction used in this analysis was the annual average in 2016, which includes daytime and nighttime cloud coverage. The cloud fraction measurements for the stability class equations should be for nighttime cloud fraction only. However, a night-only cloud fraction dataset was unavailable, so the yearly average cloud fraction (including daytime) was used. When the stability class was determined, the corresponding equations calculated the nighttime PM 2.5 concentration.

After calculating the night and day PM 2.5 emissions, the script weighted these concentrations based on Texas's average day and night hours. As previously discussed, the average daylight hours in Texas are 11.953; therefore, the average night hours in Texas are 12.047. For weighing the concentrations, the hourly values are divided by twenty-four to get the hourly values into the fraction of the day. The day concentration was multiplied by 0.49804, and the night concentration was multiplied by 0.50195. These weighted concentrations were summed, resulting in the final PM 2.5 concentration estimate.

At this point in the script, only the vertical height of the point of interest and the effective stack height needed to be determined. For this analysis, the vertical height of the point of interest was zero because the concentrations were estimated at ground level. The effective stack height was 0.3048 meters, which estimated where the PM 2.5 emitted from, assuming it came from manure on the ground. With all the variables determined, the script then calculated the concentration of PM 2.5 at each grid point with Equation 1.

As the script iterated through each grid cell for each CAFO, the estimated PM 2.5 concentrations were output to a .csv file with the grid cell's OBJECTID and estimated PM 2.5

concentration. After the script generated the .csv file, it joined the .csv to the grid point feature class based on the OBJECTID field. The script's last step spatially joined the grid point to its grid polygon with the Spatial Join tool, resulting in an output polygon with cells containing the estimated PM 2.5 concentrations. Figure 16 shows an example plume for one CAFO; within one CAFO plume, there were about 1,610 raster cells with an estimated concentration. The yellow cells show PM 2.5 concentrations between zero and 0.5 micrograms per cubic meter, the orange cells show PM 2.5 concentrations between 0.5 and one micrograms per cubic meter, and the red cells show PM 2.5 concentrations above one microgram per cubic meter.

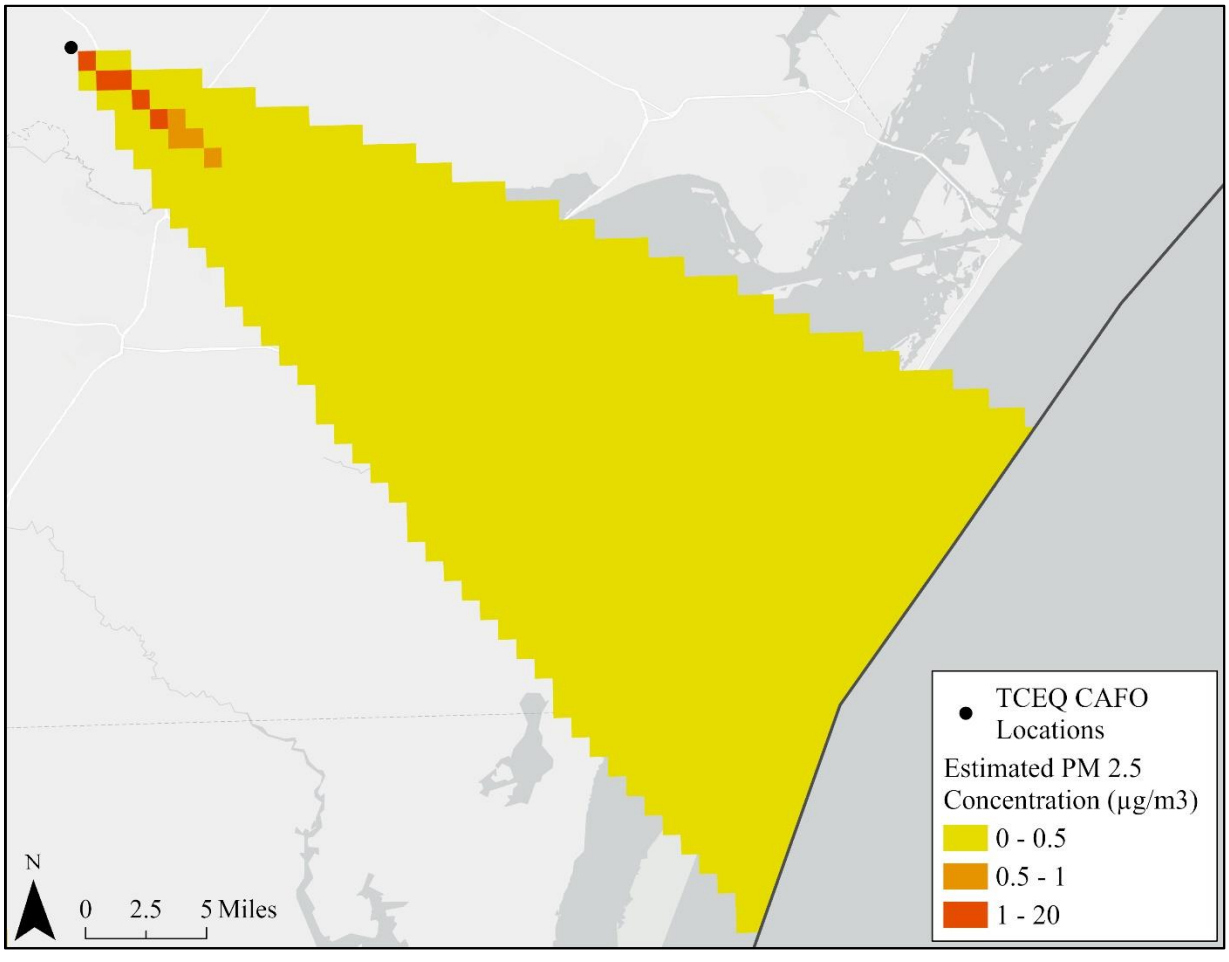

Figure 16. PM 2.5 estimation plume for a single CAFO

After the Python script finished running and each CAFO point had a plume dispersion estimation grid, all the generated plume surfaces were merged into one layer using the Merge tool. This merged layer was then dissolved to combine the overlapping grid cells using the Dissolve tool. The dissolution was performed based on the OBJECTID field. Then, the script summed the PM 2.5 concentrations to calculate each cell's estimated PM 2.5 total, resulting in the final Gaussian plume equation PM 2.5 estimation.

3.2.2 Assessment of the Spatial Analysis

This study determined a spatial estimate of the exposure to CAFO PM 2.5 in Texas. After the spatial analysis, the last step was to examine the results. The PM 2.5 dispersion analysis results were analyzed using linear regression to compare how the estimated PM 2.5 concentrations compared to the measured SEDAC 2016 PM 2.5 concentrations.

For the linear analysis measuring the spatial analysis's performance, and when using linear regression in general, the model must be checked to see if it meets the assumptions associated with linear regression. The assumptions are that the parameters are linear, there is random sampling in the observed data, there is zero conditional mean, there is no perfect collinearity, there is homoscedasticity and no autocorrelation, and the errors are normal (Zhu 2022).

Histograms of the variables and scatterplots showing the model relationships can be made to check the first assumption that the parameters are linear. OLS regression can only be used to create a linear model, so there should be a linear relationship between the dependent and explanatory variables. If the histograms are not normal distributions, they can be transformed to be more normalized. In addition, outliers can be removed to see if model performance improves (Esri n.d.b; Zhu 2022).

The histogram of the estimated PM 2.5 concentration is in the first graph in Figure 17, and the histogram of the measured SEDAC data is in Figure 18. Since linear regression assumes that the variables are normally distributed, and the estimated PM 2.5 data distribution is positively skewed, a log transformation was applied to the estimated concentrations, as shown in the second graph in Figure 17. A scatterplot of the relationship between the measured PM 2.5 data and the estimated PM 2.5 values is in Figure 19.

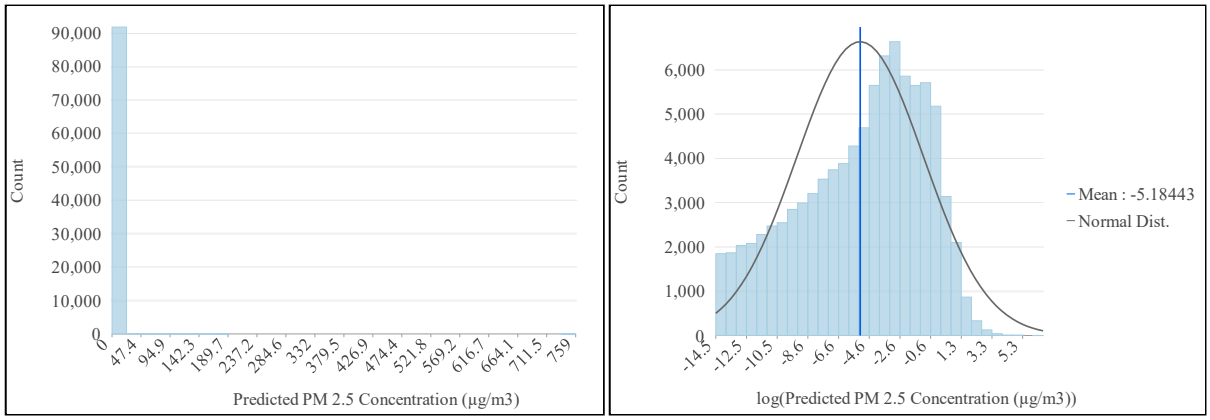

Figure 17. Estimated PM 2.5 concentration and their log transformation

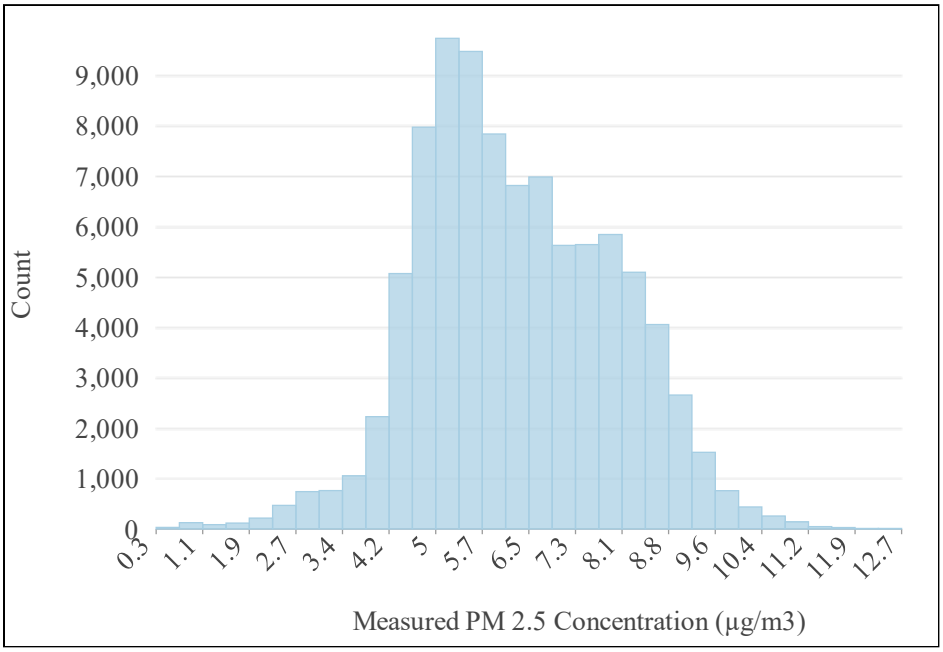

Figure 18. Measured PM 2.5 concentration

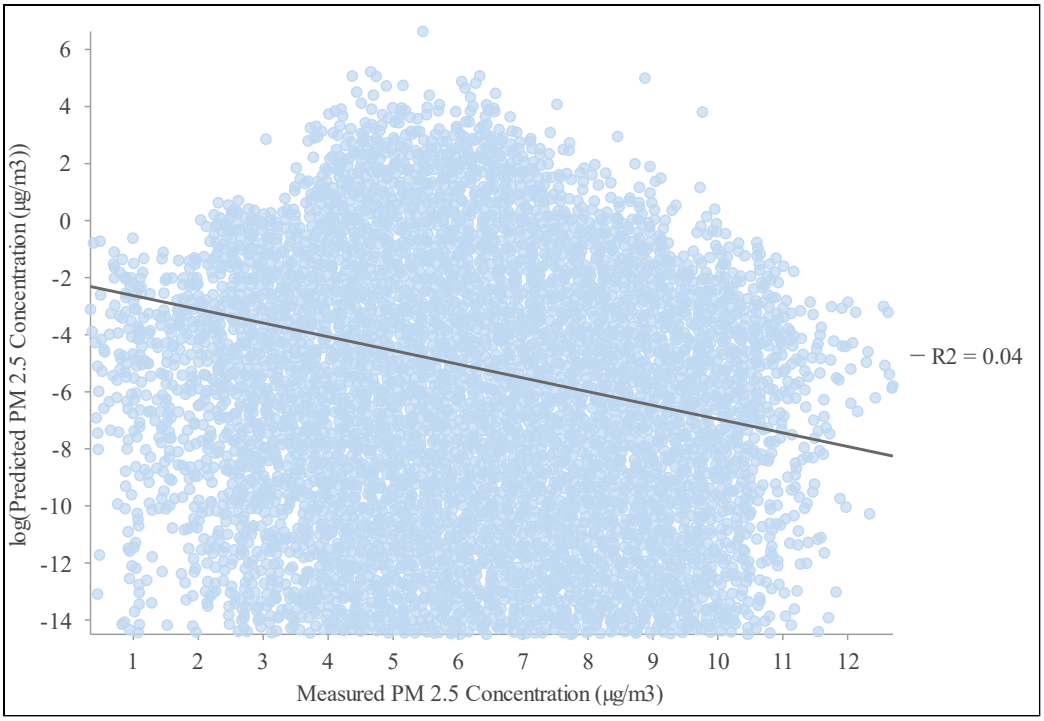

Figure 19. Measured PM 2.5 data and estimated PM 2.5 values

The next assumptions are that there is random sampling in the observed data and a zero conditional mean. The model's residuals can be plotted against the explanatory variables to check for random sampling. The residuals should not correlate when plotted against the explanatory variables. No correlation in the plots indicates randomly sampled data. The residuals can also be used to check the model for a zero conditional mean. Residuals represent the difference between the measured and estimated values. The measured values above the regression line are positive, and those below are negative. Therefore, the sum of the residuals should be zero, which means there is a zero conditional mean (Esri n.d.b; Zhu 2022). The model is biased if the residuals are normally distributed with a mean of zero. Additionally, the Jarque-Bera diagnostic can be used to check for a biased model and is calculated by the ArcGIS Pro OLS tool. The model is biased if the Jarque-Bera diagnostic is statistically significant (Esri n.d.b).

The next assumption is that there is no perfect collinearity, or explanatory variables with linear relationships. Collinearity implies that the explanatory variables are redundant, or that multiple variables tell the same story (Zhu 2022). If redundancy is present, the variables are collinear. The OLS tool calculates a variance inflation factor (VIF) for each variable, which checks for redundancy. A VIF above 7.5 suggests that variables are redundant. Since the linear regression in the project only used one explanatory variable, this check can be ignored (Esri n.d.b).

The next assumptions are that there is homoscedasticity and no autocorrelation. Homoscedasticity in the model residuals means that they have constant variance and are the same across all values of the independent variables. Autocorrelation is the clustering of the model residuals (Zhu 2022). These assumptions help determine if all the key explanatory variables have been found. If a linear regression model is performing poorly, one of the problems may be that all the explanatory variables have not been determined. If the model residuals are clustered, or the residuals are spatially autocorrelated, this is a sign that explanatory variables are missing. Spatial autocorrelation in the model residuals can be checked with the Spatial Autocorrelation tool in ArcGIS Pro. If the z-score is statistically significant, it implies that explanatory variables are missing. More research may be required to determine additional explanatory variables to assess in the linear regression (Esri n.d.b).

The final assumption is that the errors are normal, which can be determined by calculating the probability of the residuals (Zhu 2022). The robust probability of the model is calculated by the OLS tool. Robust probability assesses whether the model results are due to chance or if the model explanatory variables are statistically significant. Since spatial data often varies across the study area, the relationships between the variables are often nonstationary. OLS

calculates the Koenker (Koenker's studentized Breusch-Pagan) statistic, which measures nonstationarity, or whether the explanatory variables are consistent with the dependent variable. A statistically significant Koenker p-value means the relationships are nonstationary. When relationships are nonstationary, explanatory variable importance is assessed using robust probability (Esri n.d.b).

Once all the assumptions are met, it can be determined if the model is correctly specified. One way to check that the model is correctly specified is to determine if the explanatory variables were helping the model by performing the linear regression using the OLS tool. The OLS tool calculates a coefficient for each explanatory variable to assess whether the variable is helping the model. It calculates the coefficient by computing the probability that the coefficient is zero; if the coefficient is zero, the explanatory variable is not helping the model. If the coefficient is between zero and one and the p-value is smaller than 0.05, it indicates that the coefficient is unlikely zero and the probability is statistically significant, which means that the explanatory variable is important to the model (Esri n.d.b; Zhu 2022).

In addition to examining the value of the coefficient, the coefficient's sign can also be used to check whether the relationships are expected. Each coefficient has a positive or negative value in the report created by the OLS tool. This sign represents whether the relationship between variables is positive or negative. It is important to note the expected relationship and check that the OLS coefficients match (Esri n.d.b; Zhu 2022). In this thesis, the expected relationship was positive, meaning that if the measured PM 2.5 data increased, the estimated PM 2.5 values also increased.

The last check calculates how well the dependent variable is explained using the adjusted R^2 value. The R^2 value evaluates the model performance with a score between zero and one. This

score represents the percentage of variation in the dependent variable that the model is explaining. In addition to R^2, the AICc can be used to compare multiple models. The model with the smaller AICc value performs better for models with the same dependent variable (Esri n.d.b).

Chapter 4 Results

This chapter illustrates the spatial analysis results by discussing the results of the PM 2.5 Gaussian plume equation script and the linear regression analysis. The Gaussian plume equation is found to accurately estimate PM 2.5 from CAFOs; however, the linear regression analyses indicate that explanatory variables are missing from the model. Section one states the results of the Gaussian plume equation analysis, section two states the results of the state-wide linear regression analysis, and section three states the results of the regional linear regression analysis.

4.1 CAFO PM 2.5 Gaussian Plume Equation Analysis Results

The Gaussian plume dispersion model estimated the concentration of PM 2.5 being emitted from CAFOs in Texas and the locations affected. The resulting PM 2.5 concentration surface is in Figure 20. The PM 2.5 concentration estimations were lowest in the yellow areas, ranging from 0 – 5 $\mu g/m^3$. The orange and red areas were where the PM 2.5 concentration estimations were 5 – 11 $\mu g/m^3$ and 11 – 759 $\mu g/m^3$, respectively. The higher concentration areas were close to the CAFOs and represented only a small portion of the estimated PM 2.5 concentrations. In the PM 2.5 concentration surface, there were 92,788 grid cells. Of these cells, only 931 cells, or about one percent of the cells, estimate PM 2.5 scores greater than 5 $\mu g/m^3$. The SEDAC measured PM 2.5 data displayed with the same color scheme is in Figure 21. From visually inspecting the measured data compared to the estimated data, the estimated data appear to be estimating lower PM 2.5 values than was measured. To better analyze the results, a linear regression was performed.

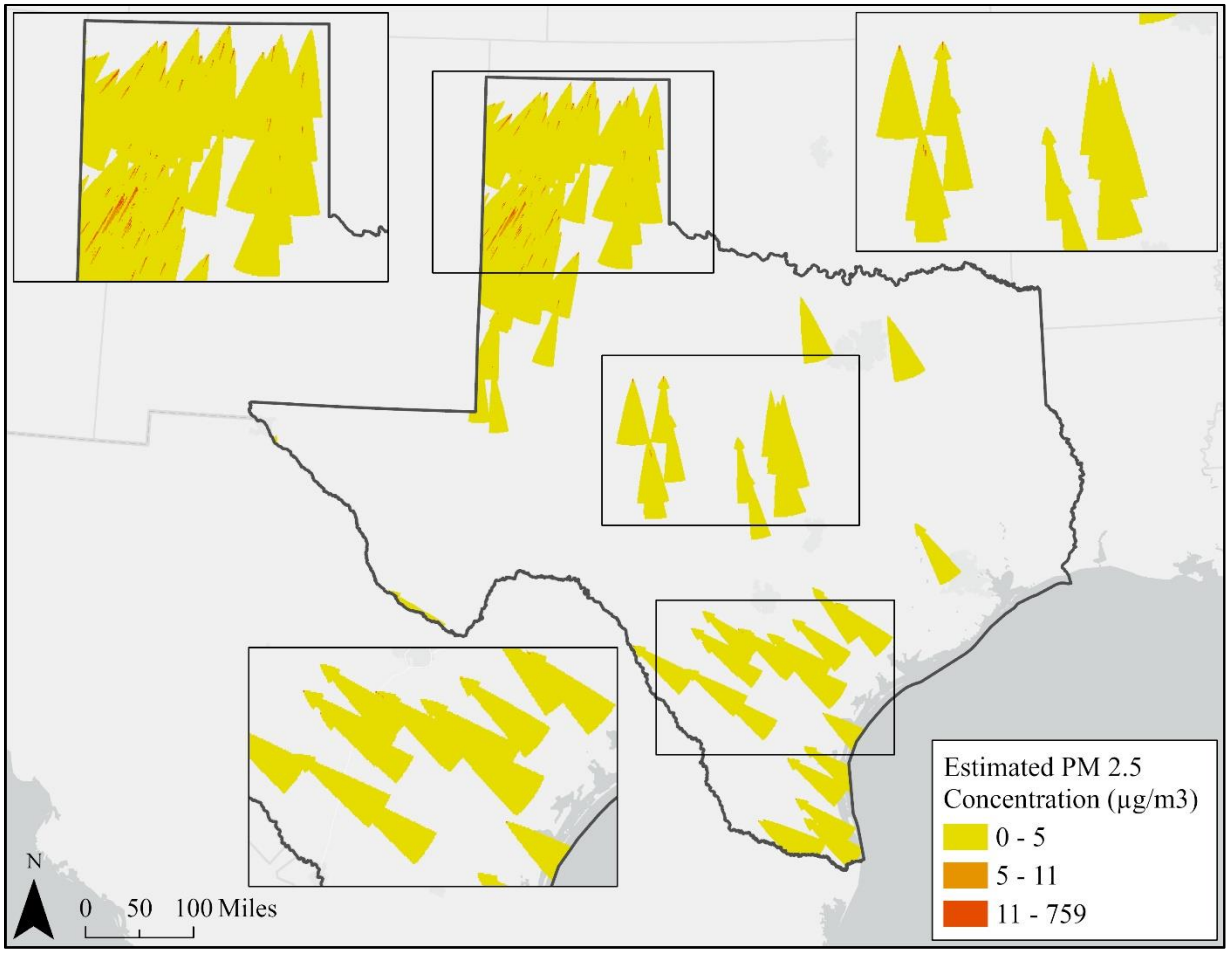

Figure 20. Estimated PM 2.5 exposure surface

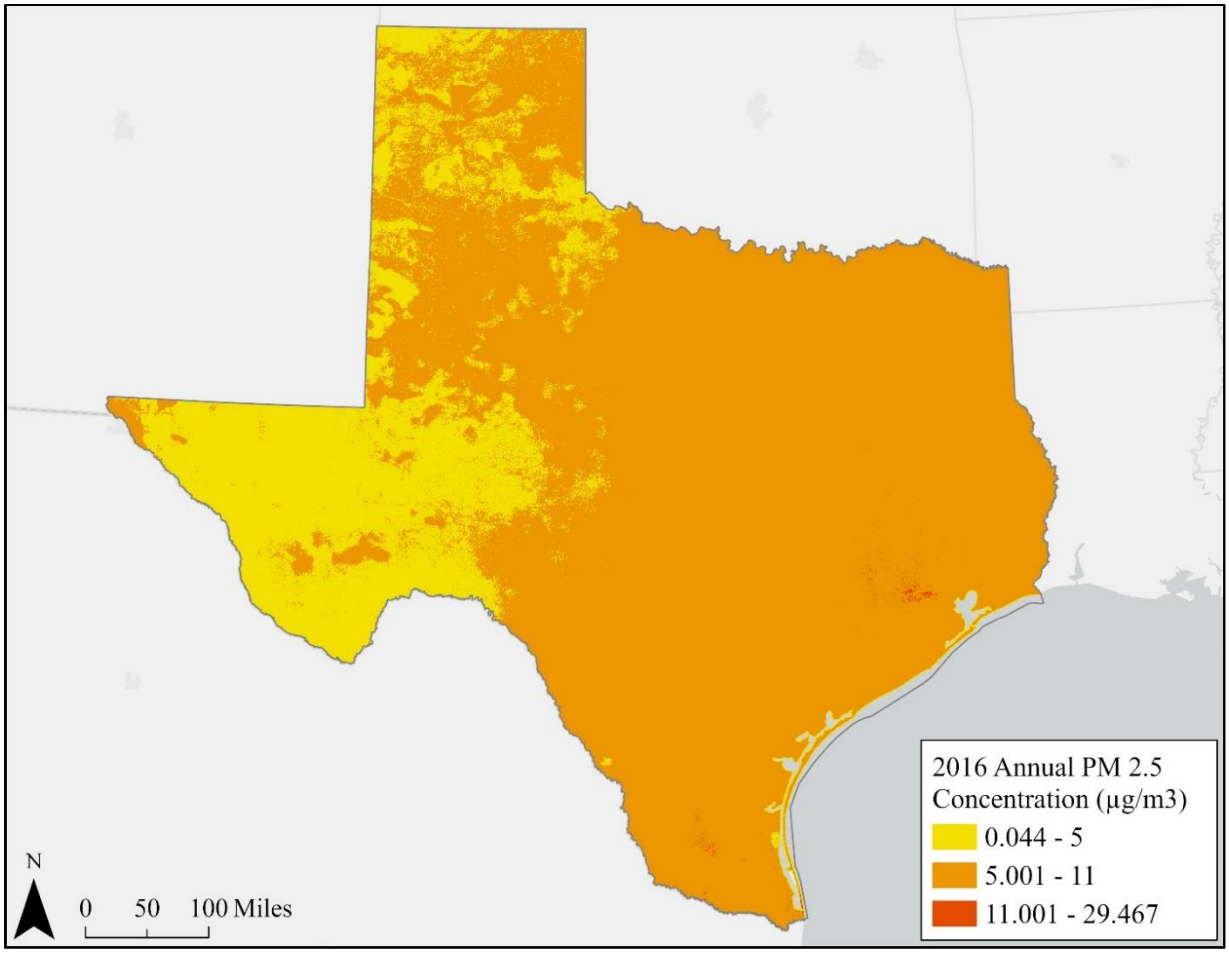

Figure 21. 2016 measured SEDAC PM 2.5 data

4.2 State-wide Linear Regression Analysis Results

The Ordinary Least Squares tool in ArcGIS Pro was used to perform the linear regression, and the checks outlined above were utilized to assess the model. The measured PM 2.5 was used as the dependent variable, and the logs of the estimated PM 2.5 data were the explanatory variable. The logs of the estimated PM 2.5 data were used since they were more normal. Table 6 shows the results of the OLS regression.

Table 6. Results from the state-wide linear regression analysis

	Linear Regression Result
Coefficient	-0.074352
Probability	0.000000
Robust Probability	0.000000
Koenker Statistic	0.000000
Jarque-Bera Diagnostic	0.000000
Spatial Autocorrelation z-score	377.542576
Adjusted R^2	0.035750
AICc Value	345493.364363
Total Raster Cells (n)	91,940

The coefficient calculated by the OLS tool was examined to evaluate whether the explanatory variable was helping the model. For the log of PM 2.5, the coefficient was -0.074352, the probability was zero, and the robust probability was zero. These probabilities were statistically significant, indicating that the explanatory variable was helping the model.

To check the model for nonstationarity, or whether the relationships vary across the study area, the Koenker statistic was used. The Koenker statistic for this model was zero, which means the relationships in the model exhibit statistically significant nonstationarity, meaning that robust probability should be used in the model.

The next check was whether the relationships, or the sign of the coefficient calculated for the explanatory variable, were as expected. In this model, the sign of the coefficient was negative, which was not expected. A negative sign represents a negative relationship. The model estimated that the measured PM 2.5 data would decrease as the estimated PM 2.5 values increased. This relationship was not expected and shows that the model has problems.

The model's bias was analyzed next by examining the distribution of the residuals. Figure 22 shows a histogram of the model residuals. The model residuals were normally distributed with a mean of zero, which suggests that the model was properly specified. However, the Jarque-Bera diagnostic was zero and statistically significant, meaning the model was biased.

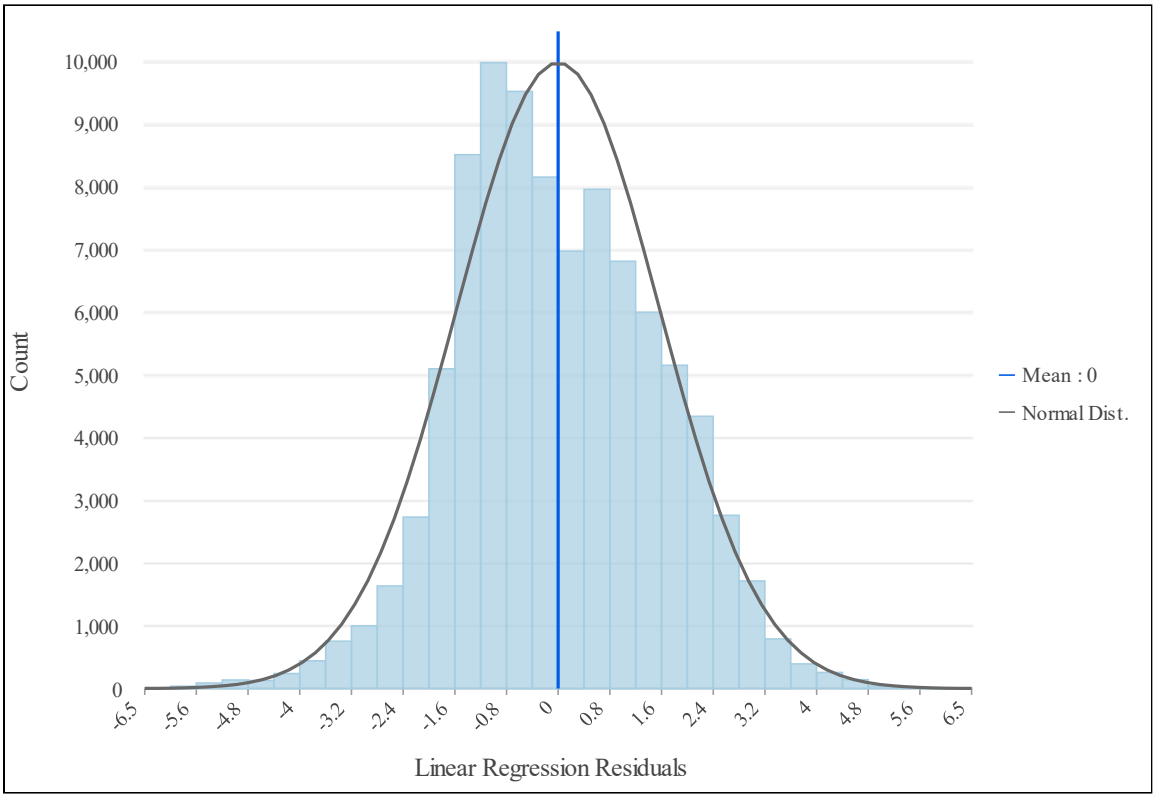

Figure 22. Linear regression residuals

The residuals can also be examined on a map to show areas where the measured values were larger or lower than the model estimated. Figure 23 shows a map of the residuals, where red areas symbolize the measured PM 2.5 concentrations larger than the model estimated. Blue areas show where the measured values were lower than estimated. The yellow color shows areas where the linear regression performed well.

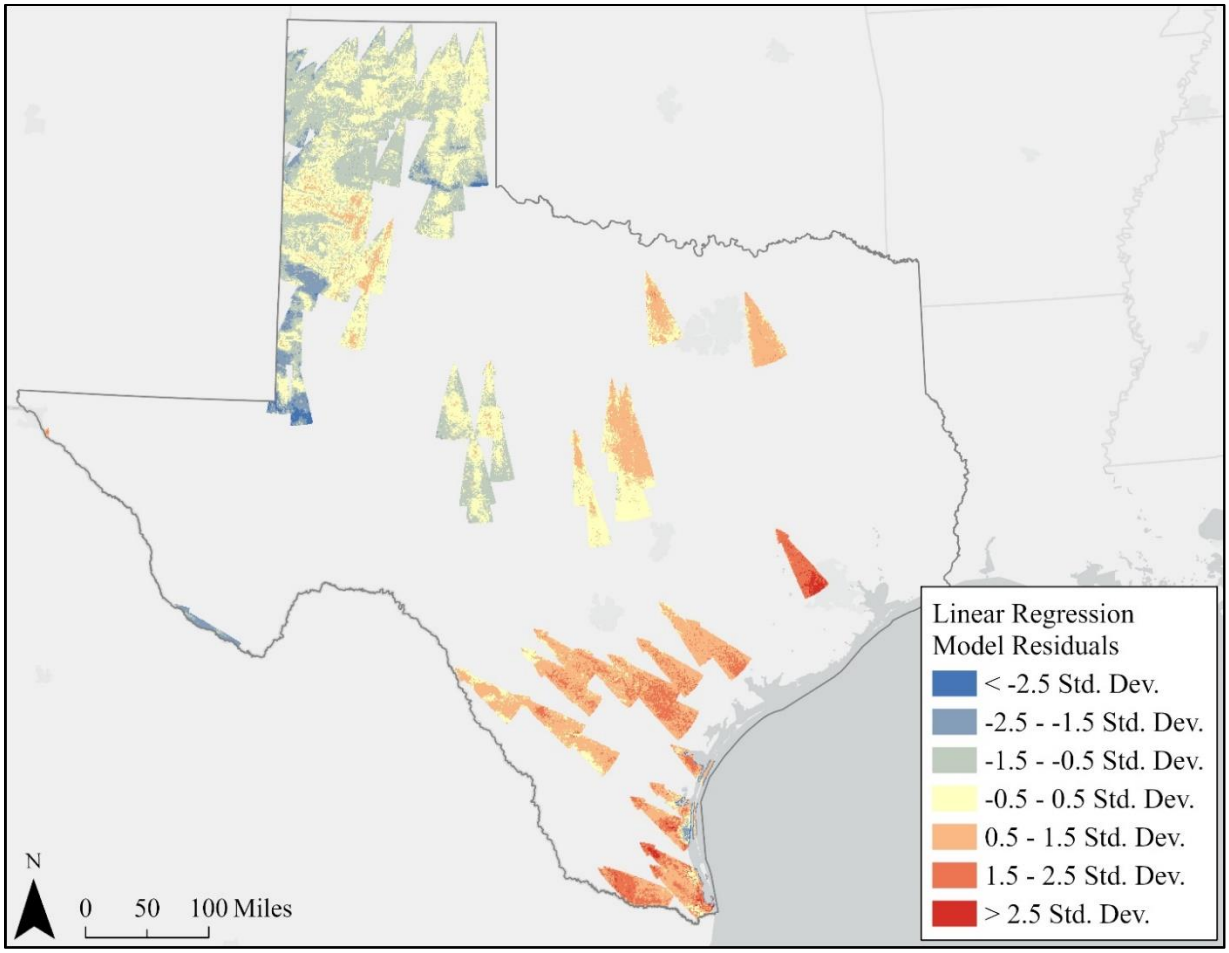

Figure 23. Spatial linear regression residuals

In addition to mapping the linear regression residuals, the residuals can be examined on a scatterplot against the explanatory variable. The residuals should not correlate in linear regression when plotted against the explanatory variables. Figure 24 shows a scatterplot of the residuals against the log of estimated PM 2.5. This scatterplot has no correlation, indicating that the data were randomly sampled.

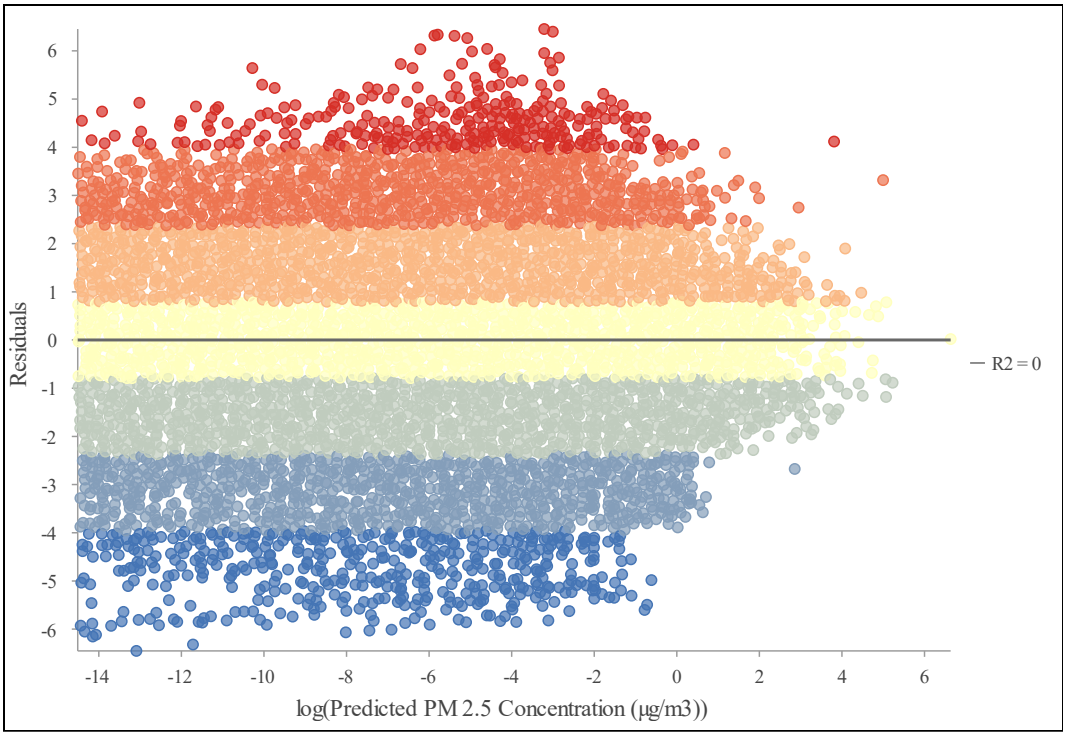

Figure 24. Linear regression residuals plotted against the log of estimated PM 2.5

The next check was to determine if all the key explanatory variables had been found. An indication that explanatory variables were missing was statistically significant spatial autocorrelation of the model residuals. Spatial autocorrelation is present if the over-estimations cluster together and the under-estimations cluster together. The Spatial Autocorrelation (Global Moran's I) tool was run in ArcGIS Pro on the linear regression residuals to assess for spatial autocorrelation in the model residuals. The z-score from the Spatial Autocorrelation tool was 377.542576. This statistically significant value indicates a less than one percent chance of the clustered residuals being the result of random chance. A statistically significant z-score suggests that key explanatory variables were missing from the linear regression.

The final check for the linear regression model was to determine how well the model explains the dependent variable by examining the adjusted R^2. Since the model did not pass all the previous checks, the R^2 value cannot be trusted. A biased model may perform well in some

areas but poorly in others if the explanatory variable explains the dependent variable better in some portions of the study area. This bias could artificially inflate the R^2 value. However, the adjusted R^2 value of the model was 0.035750, which means that the model's explanatory variable explains less than four percent of the variation in the measured PM 2.5 data dependent variable. The AICc value was 345493.364363. However, since only one linear regression model was executed, it was not helpful to consider AICc.

4.3 Regional Linear Regression Analysis Results

Since the residuals were spatially autocorrelated, linear regression was performed again for three separate regions in Texas. Three clusters of CAFOs appear in Texas: the largest cluster was in the Texas Panhandle with 123 CAFOs, the next largest was in southern Texas and along the southern coast with 23 CAFOs, and the third was in central Texas with 19 CAFOs. Figure 25 shows a map of the CAFO clusters.

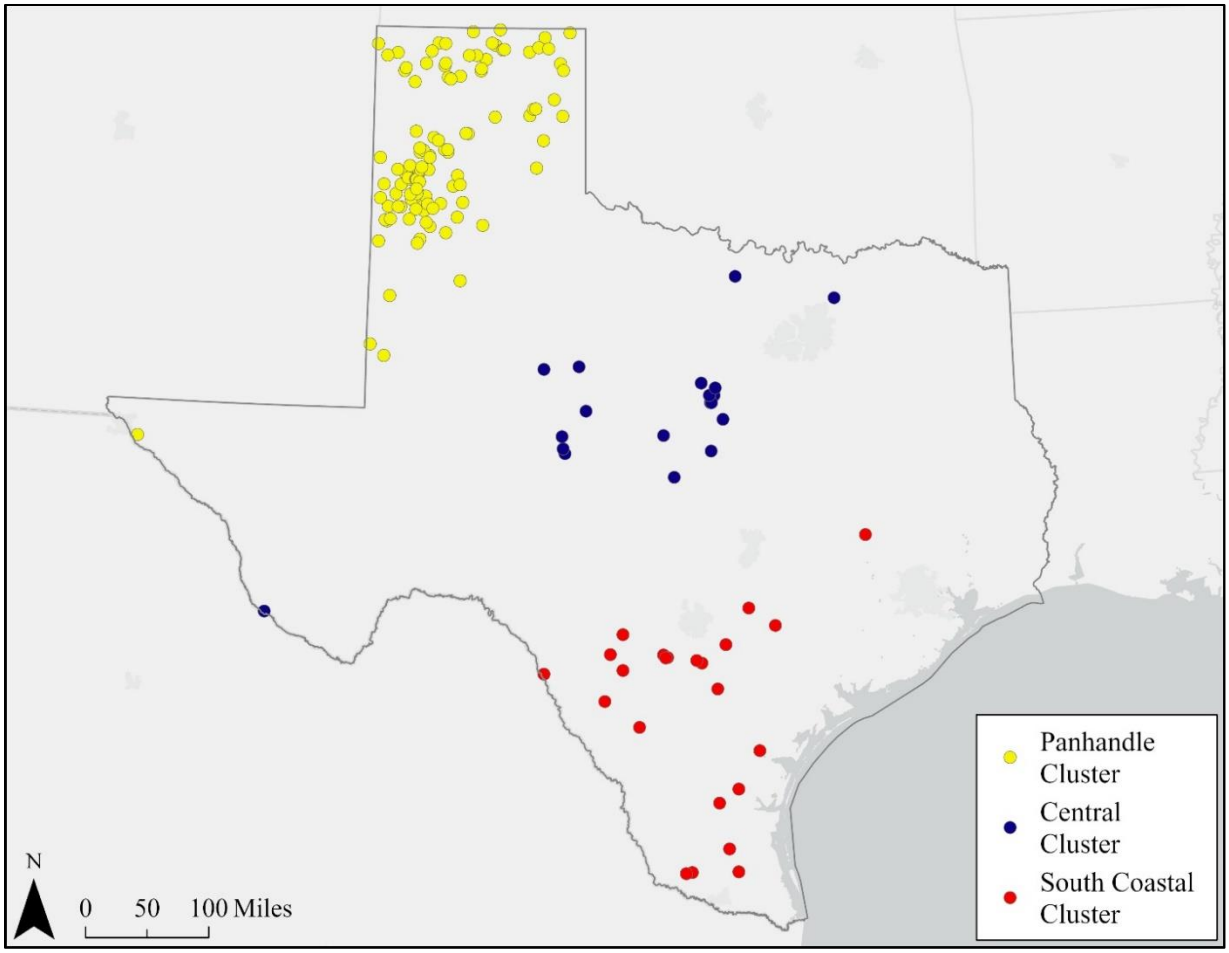

Figure 25. CAFO clusters

The linear regression was performed for each region by selecting the output grid cells from the Python script within each region and executing the Ordinary Least Squares tool in ArcGIS Pro. The measured PM 2.5 data were the dependent variable, and the logs of the estimated PM 2.5 data were the explanatory variable. The same checks outlined above were utilized to evaluate the models; Table 7 shows the output results from each region's linear regression analysis.

Table 7. Results from the regional linear regression analysis

Linear Regression Result	Overall	Panhandle Cluster	Central Cluster	South Coastal Cluster
Coefficient	-0.074352	0.008870	-0.016610	-0.012935
Probability	0.000000	0.000000	0.000000	0.000000
Robust Probability	0.000000	0.000000	0.000000	0.000000
Koenker Statistic	0.000000	0.000000	0.000010	0.657603
Jarque-Bera Diagnostic	0.000000	0.000000	0.000000	0.000000
Spatial Autocorrelation z-score	377.542576	241.878263	165.064848	148.356119
Adjusted R^2	0.035750	0.001308	0.003316	0.001467
AICc Value	345493.364363	142,387.480873	54,893.877079	74,706.820926
Total Raster Cells (n)	91,940	50,490	18,021	23,431

The coefficients calculated by the OLS tool were observed to assess whether the explanatory variable was helping the models. For the log of estimated PM 2.5, the coefficients measured were 0.008870, -0.016610, and -0.012935 for the Panhandle, central, and south coastal regions, respectively. All the regions had statistically significant probability and a robust probability of zero. These probabilities indicate that the explanatory variable was helping each model.

The Koenker statistic was used to check the models for nonstationarity. The Koenker statistic for the Panhandle region's model was zero, and the central region's model was 0.000010, meaning the relationships in these models exhibit statistically significant nonstationarity. The south coastal region's model had a Koenker statistic of 0.657603, which was not statistically significant and meant the relationship in the model did not exhibit nonstationarity.

The next check was whether the relationships exhibited in the linear regression models for each region were as expected. For the Panhandle region's model, the sign of the coefficient

was positive, which was expected. The model estimated that the measured PM 2.5 data tend to increase as the estimated PM 2.5 values increase, which was an expected relationship. For the central and south coastal regions' models, the signs of the coefficients were negative, which was not expected. The models estimated that the measured PM 2.5 data would decrease as the estimated PM 2.5 values increased. These unexpected relationships show that the models of the central and south coastal regions have problems.

The biases of the models were analyzed next by examining the distribution of the residuals. Figure 26 shows histograms of the three region's model residuals. The model residuals for each region were close to a normal distribution with a mean of zero. These residual distributions suggest that the models were appropriately specified. However, the Jarque-Bera diagnostic was zero and statistically significant for all regional models, which means the models were biased. Figure 27 shows maps of the residuals for each regional model. Red areas symbolize the measured PM 2.5 concentrations larger than the model estimated, and blue areas show where the measured values were lower than estimated. The yellow color shows areas where the linear regression performed well.

In addition to mapping the linear regression residuals, the residuals can be examined on a scatterplot against the explanatory variable. The residuals should not correlate in linear regression when plotted against the explanatory variables. Figure 28 shows scatterplots of the residuals against the log of estimated PM 2.5 for each regional model. The scatterplots had no correlation, indicating that the data were randomly sampled.

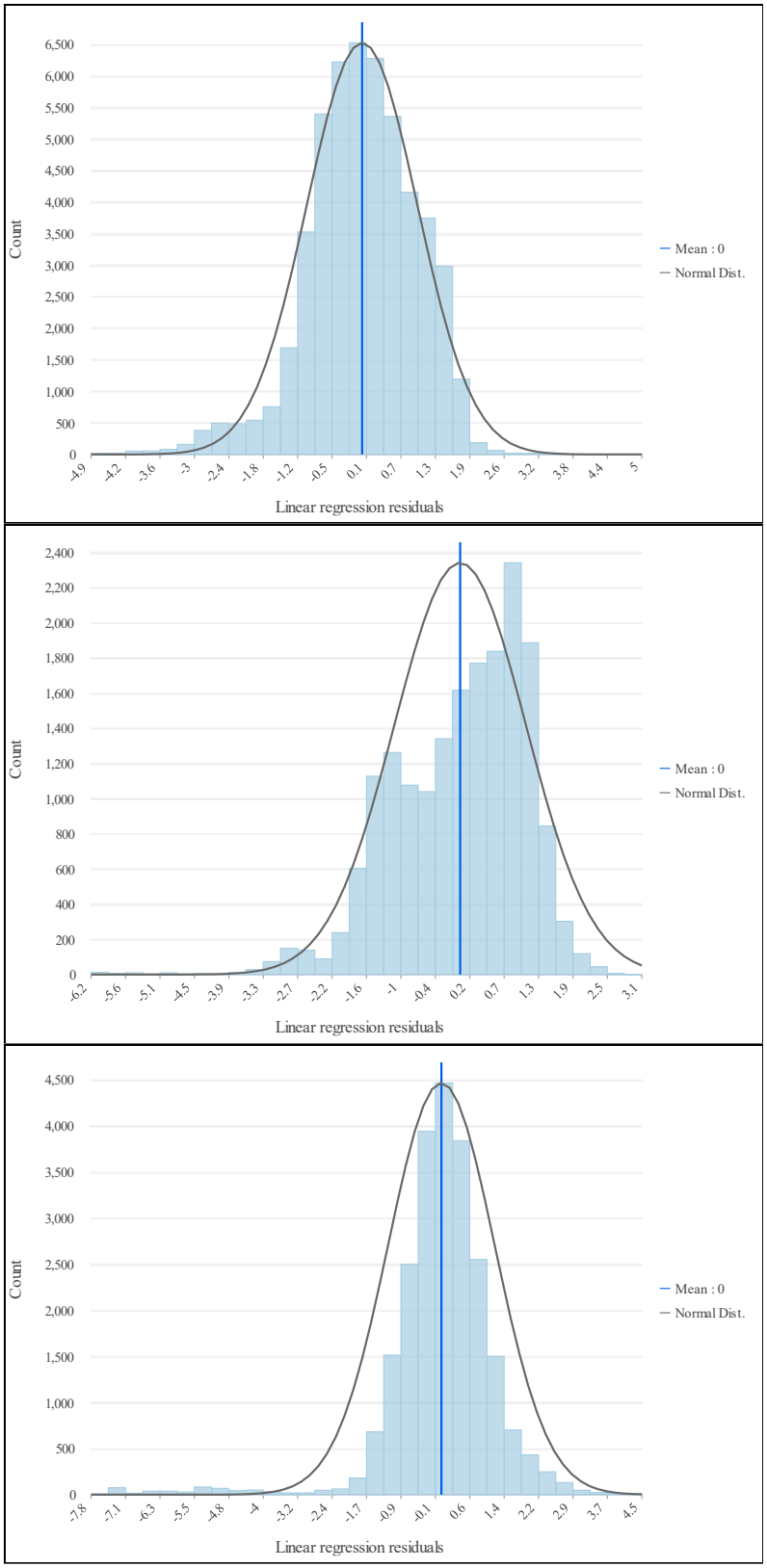

Figure 26. Regional linear regression residuals (Panhandle, central, south coastal)

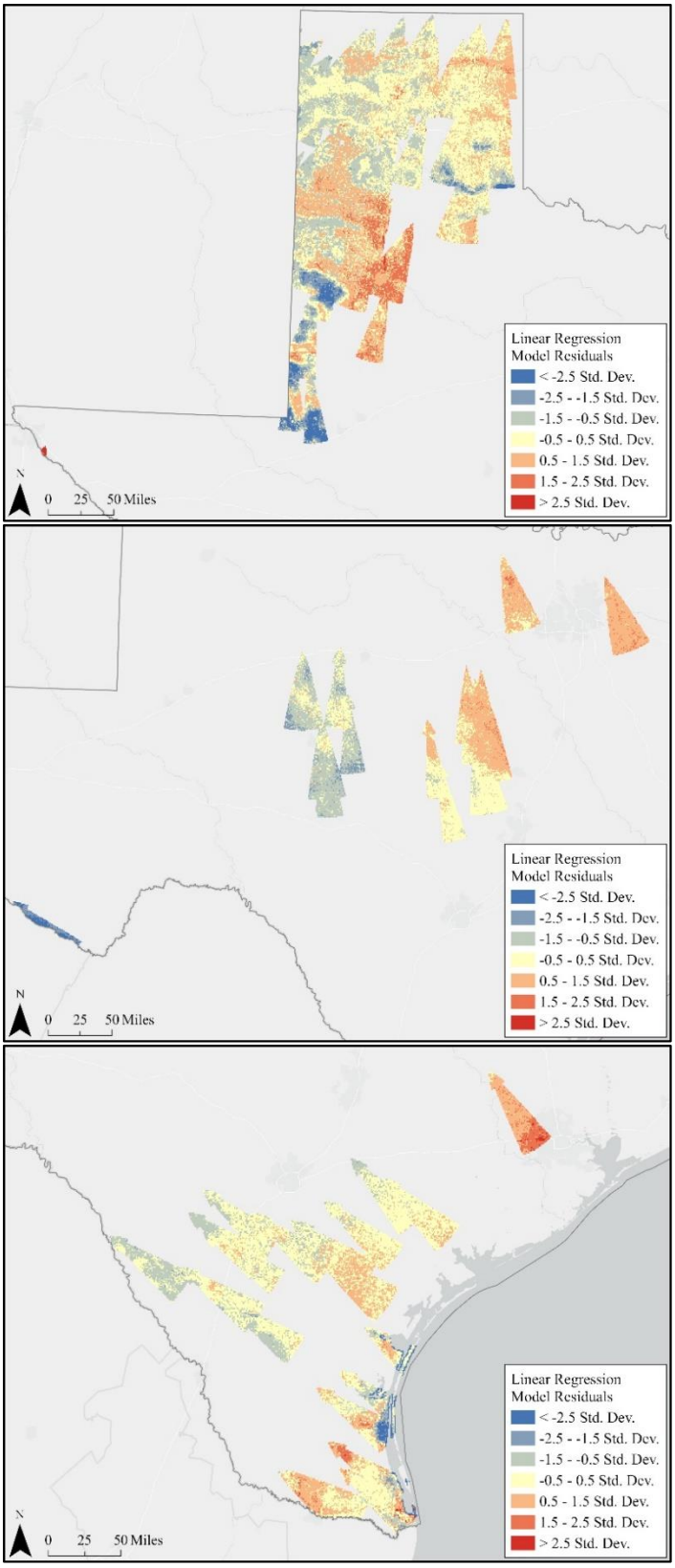

Figure 27. Spatial regional linear regression residuals (Panhandle, central, south coastal)

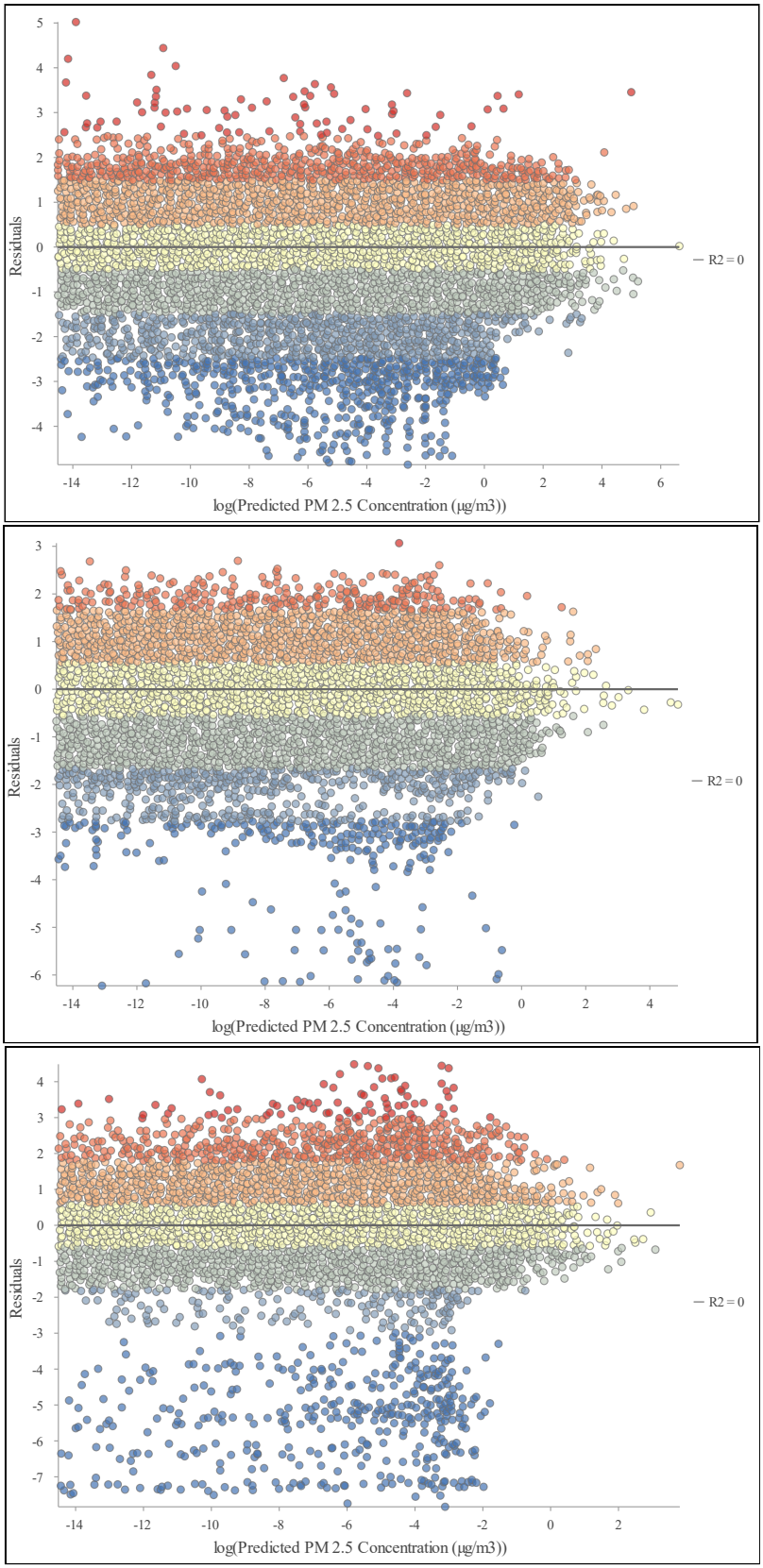

Figure 28. Regional linear regression residuals plotted against the log of estimated PM 2.5

The next check was to determine whether all the key explanatory variables had been found by looking for statistically significant spatial autocorrelation of the model's residuals. The spatial autocorrelation in the model's residuals can be assessed using the Spatial Autocorrelation (Global Moran's I) tool in ArcGIS Pro on the linear regression residuals for each regional model. The z-score from the Spatial Autocorrelation tool was 241.878263, 165.064848, and 148.356119 for the Panhandle, central, and south coastal regions, respectively. These statistically significant values indicate a less than one percent chance of the clustered residuals being the result of random chance, which suggests that key explanatory variables were missing from the linear regressions.

The final check for linear regression was to determine how well the models explain the dependent variable by examining the adjusted R^2. Since each regional model did not pass all the previous checks, the R^2 value cannot be trusted. Nonetheless, the adjusted R^2 value of the models were 0.001308, 0.003316, and 0.001467 for the Panhandle, central, and south coastal regions, respectively. These R^2 values meant the models' explanatory variable explained less than one percent of the measured PM 2.5 data dependent variable variation. The AICc values were 142,387.480873, 54,893.877079, and 74,706.820926 for the Panhandle, central, and south coastal regions, respectively. These values were all lower than the AICc of 354,470.977690 from the linear regression with the whole state, which indicates that the regional linear regression models perform better than the state-wide linear regression. However, since the models were not estimating the same dependent variable, it was not useful to consider AICc.

Chapter 5 Conclusions

This chapter discusses the findings and limitations of the project, considering the related work covered in Chapter 2. This chapter also makes recommendations for a healthier beef industry and considers opportunities for future research.

5.1 CAFO PM 2.5 Gaussian Plume Equation Analysis Discussion

The first analysis in this project was to estimate the PM 2.5 concentration emitting from CAFOs in Texas using the Gaussian plume equation. A Python script performed the analysis with the following inputs: CAFO location, estimated PM 2.5 emission rate, wind speed, wind direction, GHI, and cloud fraction. The emission rates were estimated using an emission factor of 103.356481481 micrograms of PM 2.5 per cow per second. The emission rate estimations correctly showed that CAFOs with more cows had higher emission rates. The Python script also properly applied Pasquill atmospheric stability dispersion parameter equations to calculate the atmospheric stability in the horizontal and vertical directions.

Based on research in Chapter 2 about Gaussian plume equation models, the Python script in this project estimated PM 2.5 concentration well. The output plumes estimated the highest PM 2.5 concentrations close to the CAFO locations and the concentrations dispersed in the wind direction. In the measured PM 2.5 data used in this thesis, the PM 2.5 values in Texas range from 0.105385-13.6695 $\mu g/m^3$. The estimated PM 2.5 values were appropriately within this range, with about 99% of the grid cell concentrations ranging from 0-5 $\mu g/m^3$. However, the model estimated the PM 2.5 annual estimations up to about 759 $\mu g/m^3$.

The National Ambient Air Quality Standards (NAAQS) for PM 2.5 published by the EPA for annual average standards range from 12 - 15 $\mu g/m^3$, and for 24-hour average standards are 35

$\mu g/m^3$. Close to the CAFO locations, the Gaussian plume equation estimated PM 2.5 much higher than the annual standard. The model predictions suggest that better CAFO monitoring is necessary to quantify the PM 2.5 emissions accurately, and more regulation is necessary to limit the number of pollutants emitted from CAFOs.

5.2 Linear Regression Analysis Discussion

The second analysis in this project assessed the Gaussian plume equation results using linear regression. The regression model was first performed using PM 2.5 estimated plumes for the whole state of Texas. Then, the regression was repeated three times using the plumes from three regional areas to limit the nonstationarity. The linear regression models for the state-wide and regional analyses did not perform well; however, this was expected since only CAFO-related PM 2.5 emissions were included as the independent variable. The dependent variable in the linear regression (the measured PM 2.5 data) represented PM 2.5 concentrations from all sources.

Despite the explanatory variable describing only one PM 2.5 source, the linear regression analyses were assessed using linear regression checks. For the state-wide linear regression analysis, the relationships in the model exhibited statistically significant nonstationarity, the relationships were not as expected, the model was biased, and there was statistically significant spatial autocorrelation. These results indicated that the linear regression did not use all the key explanatory variables.

The regionally divided linear regression analyses had similar results to the state-wide regression model. The Panhandle and central region linear regression models exhibited statistically significant nonstationarity, while the south coastal region did not. The relationships were not as expected for the central and south coastal regions. Though, for the Panhandle region,

the relationships were as expected. For each region, the models were biased and had statistically significant spatial autocorrelation. Like the state-wide linear regression, the results from the regional regression indicated that all the key explanatory variables were not used in the linear regression in each regional model. The regional models all had lower adjusted R^2 values than the state-wide linear regression model had.

Although the linear regression models did not pass the described checks, the models did accurately indicate that explanatory variables were missing from the models. As mentioned, this was because only CAFO-produced PM 2.5 emission estimates were used as the independent variable. Even with missing explanatory variables, the linear regression models' results, specifically the residuals' maps, help identify where the model overperformed and underperformed. The map in Figure 23, which shows the residuals from the state-wide linear regression, indicates that the estimated PM 2.5 emissions were too high in the Panhandle and too low along the southern Texas coast. The overly high estimated emissions in the Panhandle might be due to the higher concentration of CAFOs. With a high concentration of CAFOs, the emissions compound and result in high concentrations of PM 2.5.

The results of the linear regression analysis indicated that not all the key explanatory variables were used in the model. However, this was expected since the project only estimated PM 2.5 emission concentrations from CAFOs. The measured data contained PM 2.5 from all sources. Similar results were found by Bai et al. (2022), where they predicted airborne contaminant concentrations 2-3 orders of magnitude lower than the measured samples. They attribute that partially to the model not considering other sources of contamination (Bai et al. 2022). Some of the PM 2.5 sources that may be missing include road traffic, aviation, railroads, other industries, major energy and heat producers, and non-road vehicles (Rittner et al. 2020).

Additionally, Bunton et al. (2007) state that measuring dust levels downwind from CAFOs is uncommon in research due to the prevalence of confounding variables in rural areas such as unpaved roads or other dust sources (Bunton et al. 2007). Confounding variables in rural areas may explain the trend in this thesis, where the estimated PM 2.5 concentration was lower than the measured values in the most CAFO-dense area of the rural Panhandle.

The maps of the regional linear regression models' residuals show a more detailed view of the relationships between the variables. The top map of Figure 29 shows the residuals from the Panhandle region linear regression model and the TCEQ PM 2.5 air quality monitoring stations. The residuals show that the measured PM 2.5 was higher than the estimated values in the area where CAFOs were most dense. This relationship was opposite from the trend seen at a state-wide level. Four TCEQ air monitoring sites are located within the Panhandle region residuals. The northern two monitoring sites are in locations where the model performed well, and the southern two air quality monitoring sites are in locations where the measured PM 2.5 concentration data were higher than the model estimated concentrations. The bottom map of Figure 29 shows a map of the Panhandle CAFO region with aerial imagery and the TCEQ PM 2.5 air quality monitoring stations. The aerial imagery shows that the CAFOs are in a region with a very brown, dry, flat landscape. The dry soil with little vegetation easily gets blown by the wind, which produces a lot of PM 2.5 in the air. The model used in this thesis did not consider PM 2.5 created from dry soil. The TCEQ PM 2.5 air quality monitoring sites in this region are located upwind from the nearest CAFOs. These monitoring sites are not located in an optimal location to accurately measure the airborne contaminants emitted from the CAFOs.

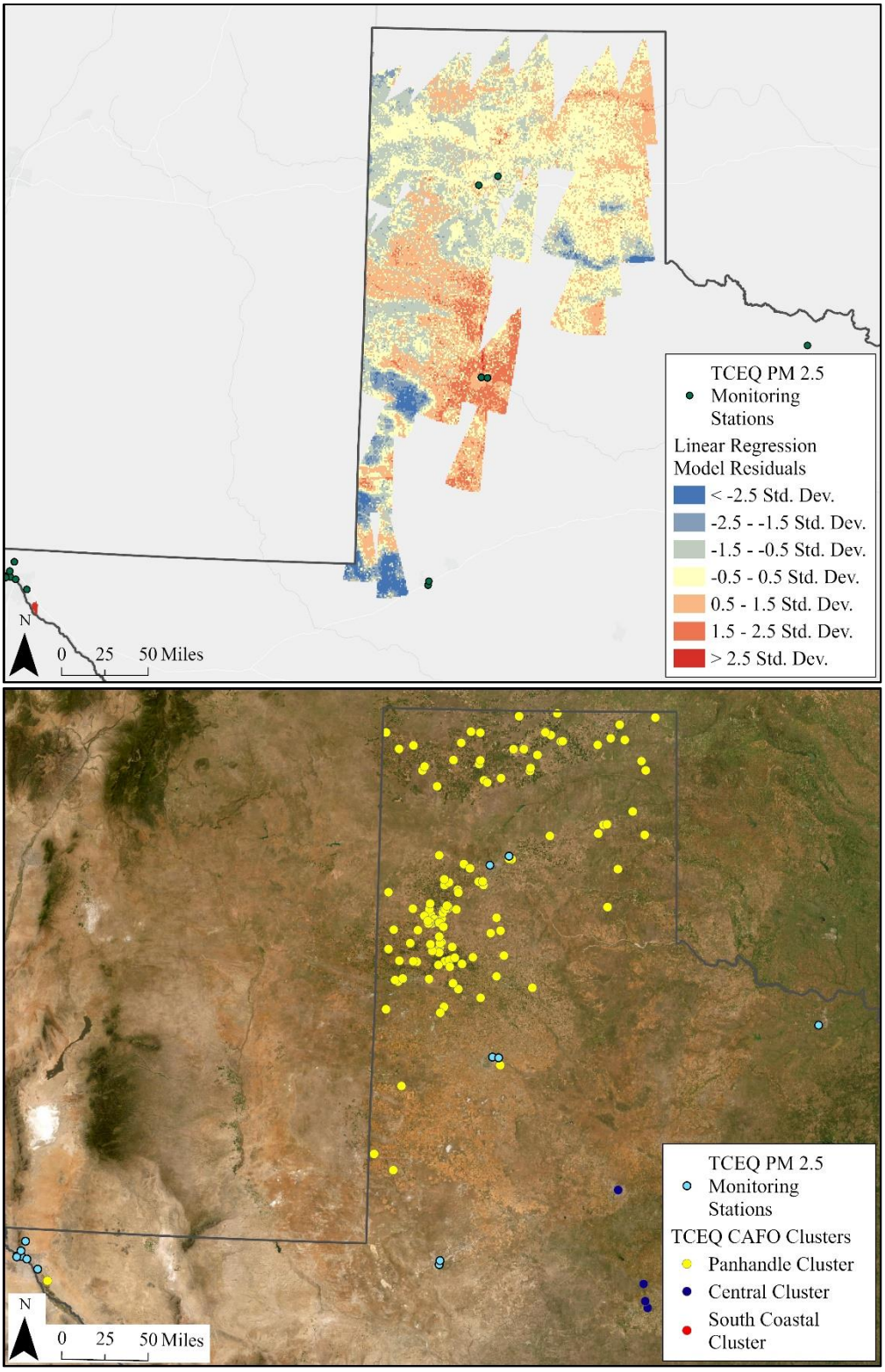

Figure 29. Panhandle CAFO region linear regression residuals and aerial imagery

The top map of Figure 30 shows the linear regression residuals of the central region model, where the measured PM 2.5 values towards the west were lower than the estimated values. The bottom map of Figure 30 shows the aerial imagery of the central CAFO region. These maps also include TCEQ PM 2.5 air quality monitoring stations. The landscape in this region is greener than in the Panhandle, which indicates that the precipitation is higher. In the western portion of this region, there is higher precipitation and more vegetation than in the Panhandle, and there are no major cities. This landscape explains why the measured PM 2.5 values are lower than the estimated values since it is likely that the PM 2.5 was washed from the air by precipitation or the PM 2.5 dispersion was disrupted by vegetation. The model did not consider precipitation or land cover. Towards the east, the measured PM 2.5 values were higher than the estimated values. The observed variation in residuals was likely since the east is near Dallas, Texas, where city-related PM 2.5 sources like cars and industry were missing from the linear regression. The bottom map of Figure 30 shows the Dallas metro region towards the northeast, where the aerial imagery is grey. Two air quality monitoring sites are within one CAFO's plume residuals near Dallas, Texas; this area's measured PM 2.5 values are higher than the model estimated. There are no close air quality monitoring sites for the rest of the central region.

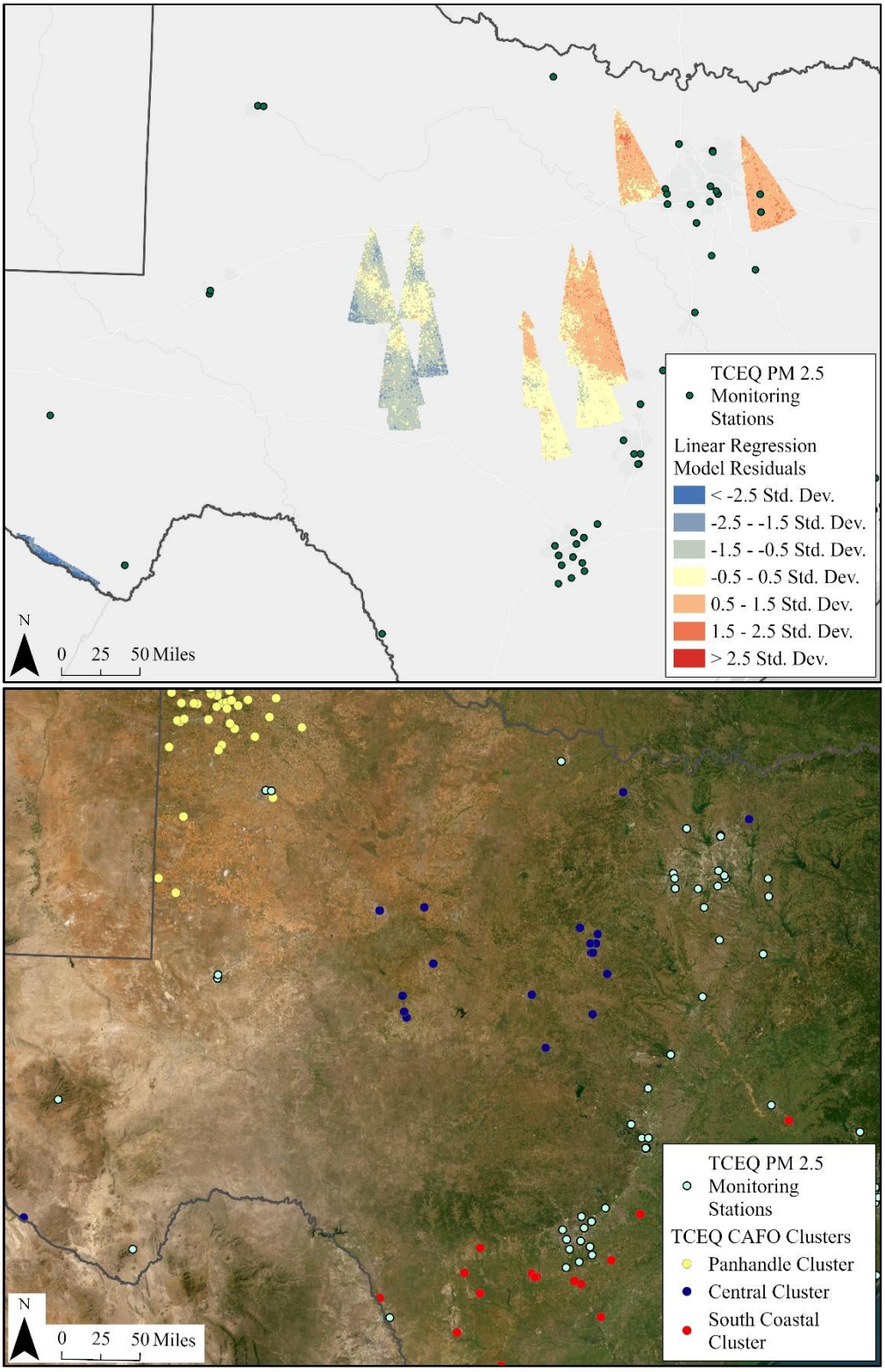

Figure 30. Central CAFO region linear regression residuals and aerial imagery

Finally, the top map of Figure 31 shows the linear regression residuals of the southern coastal region, where the estimated PM 2.5 concentrations were close to the measured values, except notably at the most northern plume. The bottom map of Figure 31 shows the aerial imagery of the southern coastal cluster. The areas that performed well in the linear regression are very green in the imagery, which indicates there is precipitation and vegetation. The PM 2.5 was likely washed from the air by precipitation or the dispersion was disrupted by vegetation. However, these areas are also close to San Antonio, Texas. San Antonio is shown in the aerial imagery in grey towards the north-central part of the map. Therefore, city-related PM 2.5 sources were missing from the model. Since this area is likely missing PM 2.5 sources, but the environmental conditions also limit the PM 2.5 that stays in the air, there is an averaging effect, and the estimated and measured PM 2.5 are similar.

The northernmost plume in the south coastal region is located near Houston, Texas, where the measured PM 2.5 concentrations were higher than the estimated values. The bottom map of Figure 31 shows Houston in grey. Near Houston, Texas, it was expected that measured PM 2.5 would be higher due to other city-related sources, like cars and industry.

The maps in Figure 31 also include TCEQ PM 2.5 air quality monitoring stations. Seven air quality monitoring stations are located within the linear regression model residuals in areas where the measured PM 2.5 values were higher than the estimated values. These air quality monitoring sites are along the coast. Most of the more inland southern CAFOs have no air quality monitoring sites nearby.

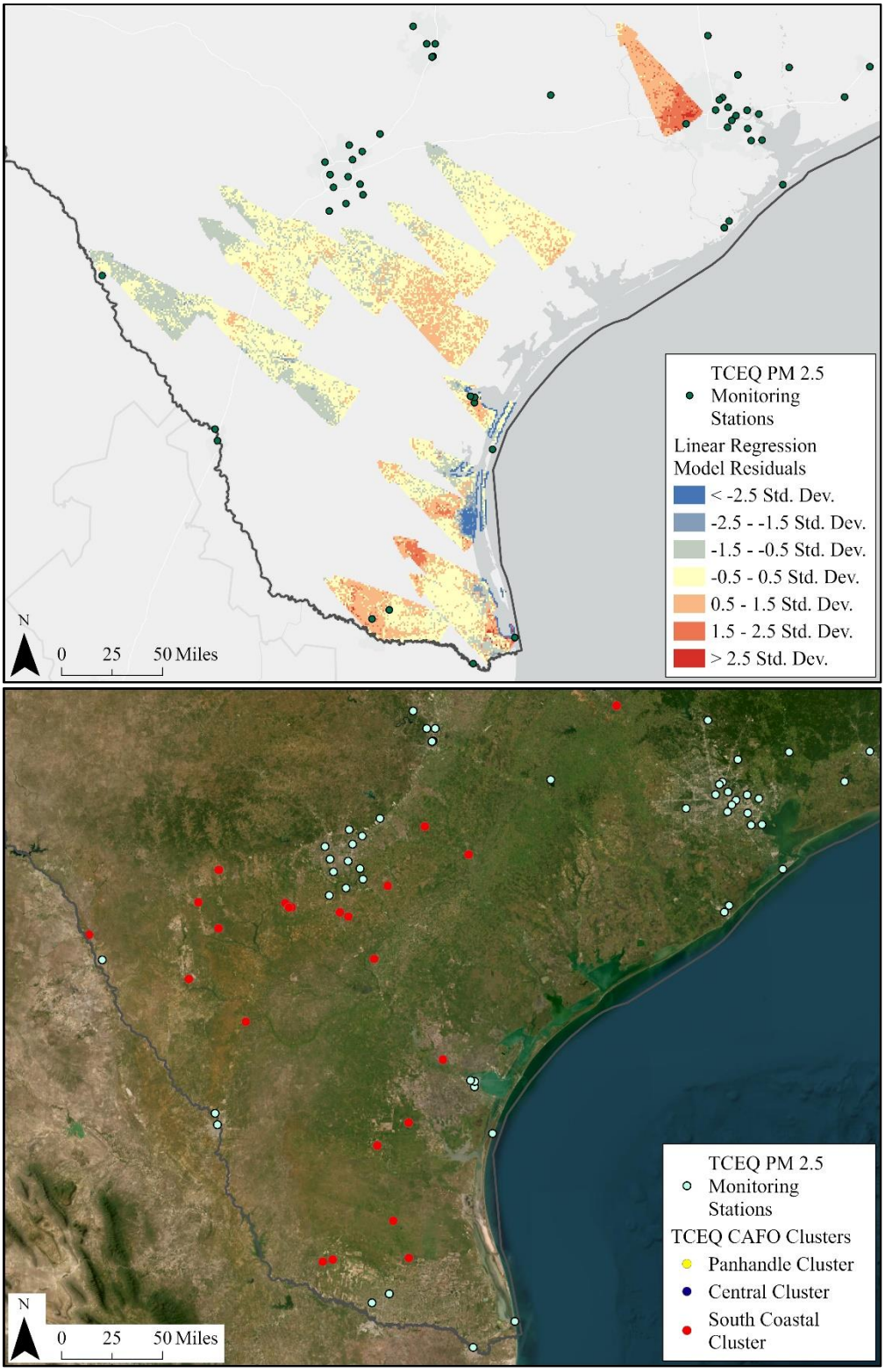

Figure 31. South coastal CAFO region linear regression residuals and aerial imagery

Besides the missing explanatory variables or the existence of confounding variables in rural areas, another possible limitation to the results may be the quality of the data input into the equation. Rittner et al. (2020) claim their model was limited by their lack of data in some portions of the study area. The areas with limited data used the closest available data point (Rittner et al. 2020). In this thesis, some input datasets were from years other than 2016. The GHI data was for a long-term average, and the wind speed was from 2012. Since the desired study year of 2016 was unavailable for GHI and wind speed, the estimated PM 2.5 concentration results may be limited. Additionally, cloud fraction measurements for the stability class equations should be for nighttime cloud fraction only, but a night-only cloud fraction dataset was unavailable, so the yearly average cloud fraction (including daytime) was used. In addition to the temporal scale accuracy, the spatial scale is important. The spatial resolution of the wind speed and direction rasters were lower quality than the rest of the one-kilometer resolution data, which may also influence the quality of the estimation results. Gaussian plume model results are sensitive to wind speed (Cheng, Kumar, and Wang-Li 2021), so accurate wind speed data is essential.

Another source of error when using plume equations is whether the input meteorological variables match the actual conditions that occurred during the study. Lotrecchiano et al. (2020) used a Gaussian plume equation to estimate PM 2.5 concentration distribution from a fire, and their model estimated higher pollutant concentrations than they measured. They suspect this was due to the more unstable and fast wind speed conditions that occurred at the time of the study than was input into the model (Lotrecchiano et al. 2020). However, for this thesis, since the data inputs were of yearly averages and the measured PM 2.5 were a yearly average, there was likely

no difference in the meteorological conditions input into the equation and those that affected the PM 2.5 concentrations in the measured data.

The last limitation that may influence the model results is the CAFO qualities that affect the amount of PM 2.5 emissions produced. Bonifacio et al. (2012) discuss qualities like dust control methods like water sprinkler systems and cow pen cleanings, manure management systems, cattle feeding schedules, and water use. These factors influence the amount of PM 2.5 emitted from each CAFO. Since PM 2.5 emanates from manure, the cleaner the pens, the less PM 2.5 emissions. Additionally, water sprinkler systems remove particles from the air and limit PM 2.5 emissions (Bonifacio et al. 2012). In this thesis, the emission factor used to estimate the emission rate of PM 2.5 from CAFOs represented the maximum mean emissions. Using the maximum mean emission rate estimated the maximum PM 2.5 concentrations. So, the PM 2.5 concentration results may be higher than the measured concentrations.

5.3 Policy Implications

The first step towards a healthier beef industry is better CAFO monitoring and understanding of how different animals and practices affect the environment. Currently, no federal agency collects accurate and exhaustive data on CAFOs (Handan-Nader and Ho 2019). As the maps in Figures 29-31 demonstrate, few air quality monitoring sites are located near the CAFO locations. Pollution from animal production varies based on the type of animals, type of feed storage, and manure handling techniques (Yuan et al. 2017). Before changes can be made to effectively mitigate the beef industry's environmental impact, gathering accurate and comprehensive information is essential. As this thesis outlined, many studies estimate the airborne pollutant concentration emitted from CAFOs and illustrate that airborne pollution occurs. This thesis performed a spatial analysis of the beef-producing facilities in Texas. It

estimated the PM 2.5 concentrations CAFOs release, which directly showed their effect on the air in neighboring communities.

However, as mentioned in the discussion of the linear analysis results, it is difficult to accurately assess model results that estimate the pollutants emitted from CAFOs due to the vast number of sources of pollutants and complex landscapes that influence particle dispersion. Governmental agencies should sample air quality data near CAFOs regularly to better monitor the pollutants CAFOs release.

Once a better understanding of the harmful effects of beef production is known, mitigation strategies for a more sustainable and healthy beef industry must occur. The existing environmental laws that are designed to limit the pollutants emitted from beef industry facilities are ineffective (Moses and Tomaselli 2017). This thesis examined CAFOs that are permitted by the TCEQ; however, the CAFO facilities that have been closed and no longer hold a TCEQ are missing from the study. The missing data are an example of how the laws are ineffective: the non-operational CAFOs may still emit pollutants, but they are not being monitored. There is also ambiguity regarding whether the state or federal government is responsible for monitoring CAFO air pollution and enforcing regulations (Sierra Club 2021). Clarification of which governmental entity is responsible for monitoring the CAFOs and enforcement of the existing policy is necessary.

Additionally, changing policy to better limit the airborne pollutants emitted by CAFOs is needed. Some policy changes include making CAFO permits more challenging to acquire, limiting animal density, requiring environmental impact studies, holding local-level public meetings, and monitoring and regulating the CAFOs more to limit pollution emissions (Donham

et al. 2007). The results of this thesis showed that CAFOs with high density are estimated to produce PM 2.5 concentration over the NAAQS annual standard.

In addition to policy change, beef production processes should be improved to decrease the pollutant byproducts created. Recommendations for the production changes that would create a healthier beef industry include lowering the air pollutants through better manure management since improperly managing cow manure is the biggest source of pollution from CAFOs. Farmers can feed cows more digestible food, move manure to off-site locations, and implement NMPs to manage manure better. NMPs optimize a CAFO's productivity while conserving nutrients and protecting the environment by specifying proper manure storage methods, manure disposal processes, and suitable irrigation practices (Clark 2020). The results showed that areas with more precipitation had lower PM 2.5 emissions; therefore, implementing cow pen water sprinkling and other manure management practices would likely decrease the airborne pollution emissions from CAFOs.

5.4 Opportunities for Future Research

This thesis provided estimates of PM 2.5 concentrations emitted from CAFOs in Texas at a state-wide level and contributed to research on beef production and its environmental and health impacts. To properly monitor and regulate beef production, it is essential to better understand how it affects the areas surrounding the production facilities. This thesis provided a state-wide perspective on CAFO airborne particle pollution emission, where most studies in the field have a smaller spatial extent. The related research discussed in this thesis included an analysis of a few animal production facilities each, while this thesis analyzed 165 CAFOs. The large-scale analysis of CAFOs was possible because the project used large geographic extent GIS data available for download online, rather than measuring the data using equipment as was

done in most of the described related research. Using accurate, downloadable, large geographic extent GIS data instead of manually measuring the values with equipment allows the study area of analysis to grow. With the downloaded GIS data, this thesis estimated the PM 2.5 concentration using a Gaussian plume equation; the methods outlined in this thesis can guide further research for other temporal and spatial extents, other pollutant types, or multiple pollutant sources.

Another area for future research is to use this analysis to understand the impacts of PM 2.5 pollution on the exposed populations. Future research could map the estimated PM 2.5 emissions with information about the populations living within the estimated plumes. Mapping the affected communities will show if there is equity in the populations affected and who the vulnerable populations are. Continued research on the pollution associated with concentrated beef production will help create a healthier beef industry.

References

Air Pollution Scientific Initiative. n.d. “Atmospheric dispersion.” Accessed January 26, 2023. https://www.apsi.tech/material/seigneur/Chapter6-AtmosphericDispersion.pdf.

Bai, H., L.Y. He, D.L. Wu, F.Z. Gao, M. Zhang, H.Y. Zou, M.S. Yao, and G.G. Ying. 2022. “Spread of airborne antibiotic resistance from animal farms to the environment: Dispersal pattern and exposure Risk.” *Environment International* 158, no. 106927 (January). https://doi.org/10.1016/j.envint.2021.106927.

Bonifacio, H.F., R.G. Maghirang, B.W. Auvermann, E.B. Razote, J.P. Murphy, and J.P. Harner III. 2012. “Particulate matter emission rates from beef cattle feedlots in Kansas—Reverse dispersion modeling.” *Journal of the Air & Waste Management Association* 62, no. 3 (February): 350-361. https://doi.org/10.1080/10473289.2011.651557.

Bunton, B., P. O’Shaughnessy, S. Fitzsimmons, J. Gering, S. Hoff, M. Lyngbye, P.S. Thorne, J. Wasson, and M. Werner. 2007. “Monitoring and modeling of emissions from concentrated animal feeding operations: Overview of methods.” *Environmental health perspectives* 115, no. 2 (February): 303-307. https://doi.org/10.1289/ehp.8838.

Burchfield, Emily. n.d. “Spatial regression.” Accessed March 28, 2023. https://www.emilyburchfield.org/courses/gsa/spatial_regression_lab.

California Air Resources Board. n.d. “Inhalable particulate matter and health (PM2.5 and PM10).” *CA.gov*. Accessed January 26, 2023. https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and-health.

Chase, L., and V. Grubinger. 2014. *Food, farms, and community: Exploring food systems*. Durham, NH: University of New Hampshire Press.

Cheng, B., A.D.S. Kumar, and L. Wang-Li. 2021. “Inverse AERMOD and SCIPUFF dispersion modeling for farm-level PM10 emission rate assessment.” *Transactions of the ASABE* 64, no. 3 (January): 801-817. https://doi.org/10.13031/trans.14311.

Chiles, R.M., and A.J. Fitzgerald. 2018. “Why is meat so important in Western history and culture? A genealogical critique of biophysical and political-economic explanations.” *Agriculture and Human Values* 35 (June): 1–17. https://doi.org/10.1007/s10460-017-9787-7.

Clark, L. 2020. “Impacts of beef concentrated animal feeding operations on environmental sustainability in the United States and practices for improvement.” Final Research Paper, Virginia Commonwealth University. https://scholarscompass.vcu.edu/cgi/viewcontent.cgi?article=1001&context=pkp.

Collins, C. 2020. “A Texas community chokes on fecal dust from cattle feedlots ... and regulators aren’t doing anything about it.” Investigate Midwest, Texas Observer, February 4, 2020. https://investigatemidwest.org/2020/02/03/a-texas-community-chokes-on-fecal-dust-from-cattle-feedlots-and-regulators-arent-doing-anything-about-it/.

Data.Texas.gov. August 16, 2021. “Texas counties map.” Texas Open Data Portal. Accessed June 26, 2022. https://data.texas.gov/dataset/Texas-Counties-Map/48ag-x9aa.

Di, Q., Y. Wei, A. Shtein, C. Hultquist, X. Xing, H. Amini, L. Shi, I. Kloog, R. Silvern, J. Kelly, M. B. Sabath, C. Choirat, P. Koutrakis, A. Lyapustin, Y. Wang, L. J. Mickley, and J. Schwartz. 2021. “Daily and annual PM2.5 concentrations for the contiguous United States, 1-km grids, v1. 2000 – 2016.” Palisades, New York: NASA Socioeconomic Data and Applications Center. Accessed January 25, 2023. https://doi.org/10.7927/0rvr-4538.

Donham, K.J., S. Wing, D. Osterberg, J.L. Flora, C. Hodne, K.M. Thu, and P.S. Thorne. 2007. “Community health and socioeconomic issues surrounding concentrated animal feeding operations.” *Environmental health perspectives* 115, no. 2 (February): 317-320. https://doi.org/10.1289/ehp.8836.

Draxl, C., B.M. Hodge, A. Clifton, and J. McCaa. 2015. “Overview and meteorological validation of the Wind Integration National Dataset Toolkit (Technical Report, NREL/TP-5000-61740).” Golden, CO: National Renewable Energy Laboratory. April 2015. Accessed January 25, 2023. https://www.nrel.gov/docs/fy15osti/61740.pdf.

———. 2015. “The Wind Integration National Dataset (WIND) Toolkit.” *Applied Energy* 151: 355366. https://doi.org/10.1016/j.apenergy.2015.03.121.

Drouillard, J.S. 2018. “Current situation and future trends for beef production in the United States of America - A review.” *Asian-Australasian Journal of Animal Sciences* 31, no. 7 (July): 1007-1016. https://doi.org/10.5713/ajas.18.0428.

Esri. n.d. “How OLS regression works.” ArcGIS Pro. Accessed March 4, 2023. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-ols-regression-works.htm.

———. n.d. “What they don’t tell you about regression analysis.” ArcGIS Desktop. Accessed February 10, 2023. https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-statistics-toolbox/what-they-don-t-tell-you-about-regression-analysis.htm.

Global Monitoring Laboratory Earth System Research Laboratories. n.d. “Global Monitoring Laboratory NOAA solar calculator.” National Oceanic and Atmospheric Administration. Accessed January 28, 2023. https://gml.noaa.gov/grad/solcalc/.

Gurian-Sherman, D. 2008. “CAFOs uncovered: The untold costs of confined animal feeding operations.” Cambridge, MA: *Union of Concerned Scientists* 1, no. 10 (April). https://www.jstor.org/stable/resrep00054.1.

Gwin, L. 2009. “Scaling-up sustainable livestock production: Innovation and challenges for grass-fed beef in the U.S.” *Journal of Sustainable Agriculture* 33, no. 2 (February): 189-209. https://doi.org/10.1080/10440040802660095.

Habib, M.R., E.J.N. Baticados, and S.C. Capareda. 2022. “Particulate matter emission factors for dairy facilities and cattle feedlots during summertime in Texas.” *International Journal of Environmental Research and Public Health* 19, no. 21 (October): 14090. https://doi.org/10.3390/ijerph192114090.

Hadlocon, L.S., L.Y. Zhao, G. Bohrer, W. Kenny, S.R. Garrity, J. Wang, B. Wyslouzil, and J. Upadhyay. 2015. “Modeling of particulate matter dispersion from a poultry facility using AERMOD.” *Journal of the Air & Waste Management Association* 65, no. 2 (January): 206-217. https://doi.org/10.1080/10962247.2014.986306.

Handan-Nader, C., and D.E. Ho. 2019. “Deep learning to map concentrated animal feeding operations.” *Nature Sustainability* 2, no. 4 (April): 298-306. https://www.nature.com/articles/s41893-019-0246-x.

Hayek, M.N., and R.D. Garrett. 2018. “Nationwide shift to grass-fed beef requires larger cattle population.” *Environmental Research Letters* 13, no. 8 (July): 84005. https://doi.org/10.1088/1748-9326/aad401.

Holmes, K.J., J.A. Graham, T. McKone, and C. Whipple. 2009. “Regulatory models and the environment: Practice, pitfalls, and prospects.” *Risk Analysis* 29 (January): 159-170. https://doi.org/10.1111/j.1539-6924.2008.01186.x.

Huertas, J.I., M.E. Huertas, S. Izquierdo, and E.D. González. 2012. “Air quality impact assessment of multiple open pit coal mines in Northern Colombia.” *Journal of Environmental Management* 93, no. 1:121-129. https://doi.org/10.1016/j.jenvman.2011.08.007.

King, J., A. Clifton, and B.M. Hodge. 2014. “Validation of power output for the WIND Toolkit (Technical Report, NREL/TP-5D00-61714).” Golden, CO: National Renewable Energy Laboratory. September 2014. Accessed February 1, 2023. https://www.nrel.gov/docs/fy14osti/61714.pdf.

Lotrecchiano, N., D. Sofia, A. Giuliano, D. Barletta, and M. Poletto. 2020. “Pollution dispersion from a fire using a Gaussian plume model.” *International Journal of Safety and Security Engineering* 10, no. 4 (August): 431-439. https://doi.org/10.18280/ijsse.100401.

Mathieu, M.E., J. Gray, and J. Richmond-Bryant. 2023. “Spatial associations of long-term exposure to diesel particulate matter with seasonal and annual mortality due to COVID-19 in the contiguous United States.” *BMC Public Health* 23, 423 (March). https://doi.org/10.1186/s12889-023-15064-5.

Mitloehner, F.M., and M.B. Schenker. 2007. “Environmental exposure and health effects from concentrated animal feeding operations.” *Epidemiology* 18, no. 3 (May): 309-311. https://doi.org/10.1097/01.ede.0000260490.46197.e0.

Moses, A., and P. Tomaselli. 2017. “Industrial animal agriculture in the United States: Concentrated animal feeding operations (CAFOs).” *International farm animal, wildlife and food safety law* (January): 185-214. https://link.springer.com/chapter/10.1007/978-3-319-18002-1_6.

Namrata, K., and N.D. Wagh. 2022. “A review on atmospheric dispersion system for air pollutants integrated with GIS in urban environment.” *Nature Environment and Pollution Technology* 21, no. 3 (February): 1553-1563. https://doi.org/10.46488/NEPT.2022.v21i04.008.

NASA Earth Observations - Moderate Resolution Imaging Spectroradiometer. 2016. “January – December 2016 cloud fraction.” National Aeronautics and Space Administration. Accessed January 29, 2023. https://neo.gsfc.nasa.gov/view.php?datasetId=MODAL2_M_CLD_FR.

NASA Earth Science Directorate Applied Science Program. 2016. “NASA Langley Research Center POWER Project wind direction at 10 meters.” National Aeronautics and Space Administration. Accessed February 1, 2023. https://power.larc.nasa.gov/.

National Agricultural Statistics Service. 2022. “NASS - Quick stats.” USDA National Agricultural Statistics Service. Accessed July 15, 2022. https://data.nal.usda.gov/dataset/nass-quick-stats.

Pfluger, A.R., P.J. Dacunto, and M. Hendricks. 2014. “A GIS-based atmospheric dispersion modeling project for introductory air pollution courses.” *American Society for Engineering Education* (June). https://doi.org/ 10.18260/1-2—19944.

Purdy, C.W., R. Nolan Clark, and D.C. Straus. 2007. “Analysis of aerosolized particulates of feedyards located in the Southern High Plains of Texas.” *Aerosol Science and Technology* 41, no. 5 (May): 497-509. https://doi.org/10.1080/02786820701225838.

Raman, A. July 2009. “Lecture 1 – Putting safety into perspective – Dispersion model 2.” *Department of Chemical Engineering University Teknology Malaysia*. Accessed January 18, 2023. https://hazopmalaysia.files.wordpress.com/2009/07/3-3_dispersion2pasquill-gifford.pdf.

Riely, A. 2011. “The grass-fed cattle-ranching niche in Texas.” *American Geographical Society of New York* 101, no. 2 (May): 261-268. https://doi.org/10.1111/j.1931-0846.2011.00090.x.

Rittner, R., S. Gustafsson, M. Spanne, and E. Malmqvist. 2020. “Particle concentrations, dispersion modelling and evaluation in Southern Sweden.” *SN Applied Sciences* 2, no. 6 (May): 1013. https://doi.org/10.1007/s42452-020-2769-1.

Schultz, A.A., P. Peppard, R.E. Gangnon, and K.M.C. Malecki. 2019. “Residential proximity to concentrated animal feeding operations and allergic and respiratory disease.” *Environment international* 130, no. 104911 (September). https://doi.org/10.1016/j.envint.2019.104911.

Sierra Club. November 15, 2021. “Why are CAFOs bad?.” Sierra Club Michigan Chapter. Accessed July 16, 2022. https://www.sierraclub.org/michigan/why-are-cafos-bad.

Sigurdarson, S.T., and J.N. Kline. 2006. “School proximity to concentrated animal feeding operations and prevalence of asthma in students.” *Chest* 129, no. 6 (June): 1486-1491. https://doi.org/10.1378/chest.129.6.1486.

Teggi, A. 2018. “GIS-based atmospheric dispersion model for pollutants emitted by complex source areas.” *Science of the Total Environment* 610-611 (January): 175–190. https://doi.org/10.1016/j.scitotenv.2017.07.196.

Texas Commission on Environmental Quality. n.d. “PM 2.5 sites GIS report.” Texas Commission on Environmental Quality. Accessed July 23, 2022. https://www17.tceq.texas.gov/tamis/index.cfm?fuseaction=report.main.

———. n.d. “TCEQ CAFO permit data.” Texas Commission on Environmental Quality. Received June 17, 2022.

Texas Parks and Wildlife. n.d. “Big Bend Country.” Texas Parks and Wildlife Department. Accessed March 5, 2022. https://tpwd.texas.gov/education/resources/texas-junior-naturalists/regions/big-bend.

———. n.d. “Gulf Coast.” Texas Parks and Wildlife Department. Accessed March 5, 2022. https://tpwd.texas.gov/education/resources/texas-junior-naturalists/regions/gulf-coast.

———. n.d. “Hill Country.” Texas Parks and Wildlife Department. Accessed March 5, 2022. https://tpwd.texas.gov/education/resources/texas-junior-naturalists/regions/hill-country.

———. n.d. “Panhandle Plains.” Texas Parks and Wildlife Department. Accessed March 5, 2022. https://tpwd.texas.gov/education/resources/texas-junior-naturalists/regions/panhandle-plains.

———. n.d. “Prairies and Lakes.” Texas Parks and Wildlife Department. Accessed March 5, 2022. https://tpwd.texas.gov/education/resources/texas-junior-naturalists/regions/prairies-and-lakes.

———. n.d. “South Texas Plains.” Texas Parks and Wildlife Department. Accessed March 5, 2022. https://tpwd.texas.gov/education/resources/texas-junior-naturalists/regions/south-texas-plains.

United States Census Bureau. March 26, 2020. “Annual estimates of the resident population for counties in Texas: April 1, 2010 to July 1, 2019.” Accessed July 1, 2022. https://www2.census.gov/programs-surveys/popest/tables/2010-2019/counties/totals/co-est2019-annres-48.xlsx.

United States Department of Agriculture Economic Research Service. September 26, 2022. "Cattle & beef sector at a glance." United States Department of Agriculture. Accessed February 25, 2023. https://www.ers.usda.gov/topics/animal-products/cattle-beef/sector-at-a-glance/.

Venkatram, A. 1999. "An examination of the Pasquill-Gifford-Turner dispersion scheme." *Atmospheric Environment* 30, no. 8 (December): 1283-1290. https://doi.org/10.1016/1352-2310(95)00367-3.

Walton, L., and K.K. Jaiven. 2020. "Regulating CAFOs for the well-being of farm animals, consumers, and the environment." *Environmental law reporter* 50, no. 6 (June): 10485. https://advance-lexis-com.libproxy1.usc.edu/api/document?collection=analytical-materials&id=urn:contentItem:6065-JYD1-JP4G-62JP-00000-00&context=1516831.

World Bank Group Energy Sector Assistance Program. June 2022. "Global Solar Atlas - Global Horizontal Irradiation." Solargis. Accessed January 25, 2022. https://earthobs3.arcgis.com/arcgis/rest/services/GlobalSolarAtlas/ImageServer.

Yang, D., G. Chen, and Y. Yu. 2007. "Inter-comparison of AERMOD and ISC3 modeling results to the Alaska tracer field experiment." *Chinese Journal of Geochemistry* 26, (May): 182-185. https://doi.org/10.1007/s11631-007-0182-8.

Yin, S., J. Zheng, Q. Lu, Z. Yuan, Z. Huang, L. Zhong, and H. Lin. 2015. "A refined 2010-based VOC emission inventory and its improvement on modeling regional ozone in the Pearl River Delta Region, China." *Science of the Total Environment* 514 (May): 426-438. https://doi.org/10.1016/j.scitotenv.2015.01.088.

Yuan, B., M.M. Coggon, A.R. Koss, C. Warneke, S. Eilerman, J. Peischl, K.C. Aikin, T.B. Ryerson, and J.A. de Gouw. 2017. "Emissions of volatile organic compounds (VOCs) from concentrated animal feeding operations (CAFOs): Chemical compositions and separation of sources." *Atmospheric Chemistry and Physics* 17, no. 8 (April): 4945-4956. https://doi.org/10.5194/acp-17-4945-2017.

Zhu, A. May 24, 2022. "Linear regression with OLS: Unbiased, consistent, BLUE, best (efficient) estimator." *Towards Data Science*. Accessed March 28, 2023. https://towardsdatascience.com/linear-regression-with-ols-unbiased-consistent-blue-best-efficient-estimator-359a859f757e.