

Contract 16RD010

Improving the CalEnviroScreen score at the US-Mexico border

Penelope J. E. Quintana

San Diego State University School of Public Health

November 18, 2019

Prepared for the California Air Resources Board and the California Environmental
Protection Agency

Disclaimer

The statements and conclusions in this Report are those of the contractor and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

Acknowledgments

The following scientists and entities contributed significantly to this report:

San Diego State University School of Public Health: Dr. Penelope Quintana, Dr. Zohir Chowdhury

San Diego State University Department of Geography: Dr. Atsushi Nara, Kristen Monteverde, graduate student, Dr. Trent Biggs, Dr. Fernando de Sales

Molina Center for Energy and the Environment, La Jolla, CA: Dr. Luisa T. Molina, Dr. Miguel Zavala, Dr. Victor Almanza

Universidad Autónoma de Baja California, Campus Tijuana: Dept of Chemistry and Chemical Engineering, Dr. Javier Emmanuel Castillo Quiñones, Lizeth Carolina Aguilar, graduate student

Universidad Autónoma de Baja California, Campus Mexicali: Dept of Geography, Judith Ley Garcia, Mónica Judith Ochoa García, graduate student, Leticia Ramírez Rubio, graduate student

Dr. James Sadd, Occidental College

Dr. Manuel Pastor, University of Southern California

Dr. Lynn Russell

Table of Contents

PAGE
List of Figures.....	v
List of Tables.....	viii
Abstract.....	ix
Executive Summary	x
Body of Report.....	1
CHAPTER 1. Mapping activities for sources in Baja California and database creation.....	1
References.....	16
CHAPTER 2. Modeling of potential areas of influence (PAIs) and potential source region analysis (PSR).....	17
References.....	40
CHAPTER 3. Border county Environmental Justice Screening Model scores after inclusion of sources in Mexico within 1.5 km of international border	42
References.....	57
CHAPTER 4. Pilot study of a method for determination of US-Mexico border burning activities though automated analysis of satellite data	58
References.....	68
CHAPTER 5. Data provided and detailed methodology in a format that can be incorporated into CES and EJSM.....	69
CHAPTER 6. Summary and Conclusions, Recommendations.....	70
Glossary of Terms, Abbreviations, and Symbols.....	72
Appendix A.....	73

List of Figures

	PAGE
Figure 1-1. Frequency distribution of remote validation location error.....	13
Figure 2-1. Methodology used for obtaining the Potential Area of Influence (PAI) and Potential Source Regions (PSR) using emissions data, meteorological observations and modeling.....	20
Figure 2-2. Map showing ambient monitoring stations in the California-Mexico border region	23
Figure 2-3. Domains used in the WRF model to capture synoptic (D1 and D2), regional (D3), and local (D4 and D5) meteorological conditions. Domains 4 and 5 (green squares) cover the San Diego-Tijuana and Calexico-Mexicali border regions, respectively.....	25
Figure 2-4. Daily number of hours with southerly wind direction in the San Diego-Tijuana (top) and Calexico-Mexicali (bottom) regions during 2014.....	26
Figure 2-5. Location of emission sources in the Tijuana (left) and Mexicali (right) regions selected for estimating their potential area of influence.....	27
Figure 2-6. Example of potential areas of influence obtained from forward trajectories for the Rosarito power plant and the “Brick kilns” sources in Tijuana during the high wind speed episode of February 28, 2014.....	29
Figure 2-7. Accumulated residence time distributions in the 100-m footprint layer during the entire meteorological episodes (28 days) for the 3 selected receptors in the San Diego–Tijuana region: San Ysidro, Chula Vista and El Cajon....	33
Figure 2-8. Zoom over the US-Mexico border region showing the transport pathways for the San Ysidro receptor.....	34
Figure 2-9. Accumulated residence time distribution in the 100 m footprint layer during the entire meteorological episodes (77 days) for the 3 selected receptors in the Calexico - Mexicali region: Calexico; El Centro and Brawley.....	36
Figure 2-10. Zoom over the US-Mexico border region showing the transport pathways for the Calexico receptor.....	37
Figure 3-1. Summary diagram showing the method for calculating the EJSM Hazard Proximity counts.....	43
Figure 3-2. Hazard Proximity counts calculated for CI polygons that are large or irregularly shaped intersected with a 1000 ft buffer to create “sliver polygons.....	44
Figure 3-3. Point location hazards used in EJSM Hazard Proximity counts in the San Ysidro and Otay Mesa region in SW San Diego County	46

Figure 3-4. Point location hazards used in EJSM Hazard Proximity counts in the Tecate region in South San Diego County.....	46
Figure 3-5. Point location hazards used in EJSM Hazard Proximity counts in the Calexico region in South Imperial County.....	47
Figure 3-6. Census tracts with EJSM Hazard Proximity Scores in the San Ysidro and Otay Mesa region in SW San Diego County.....	48
Figure 3-7. Census tract with EJSM Hazard Proximity Score in the Tecate region in Southern San Diego County.....	49
Figure 3-8. Census tracts with EJSM Hazard Proximity Scores in the Calexico region in South Imperial County.....	50
Figure 3-9. Census tracts with EJSM Hazard Proximity Scores in the Calexico region in South Imperial County.....	51
Figure 3-10. CI Polygons used to calculate EJSM Hazard Proximity Scores in the general area of the City of Calexico.....	52
Figure 4-1. MODIS imagery extent (Tile Location Code=H8V5.....	58
Figure 4-2. Example of fire pixels in the MODIS Active Fire Products.....	59
Figure 4-3. Longitudinal Time Series Fire Pixels. The number in fire pixels (orange color) represents a fire event ID.....	60
Figure 4-4. Identified fire pixels (left) and fire events (right) from February 2000 to February 2018 within the border regions.....	61
Figure 4-5. Identified monthly fire pixels from February 2000 to February 2018 within the border regions.....	61
Figure 4-6. Identified monthly fire events from February 2000 to February 2018 within the border regions.....	61
Figure 4-7. Spatial distribution of crop fields and urban areas in the US-Mexico border region.....	62
Figure 4-8. Spatial distribution of crop fields and urban areas in the Tijuana area.....	62
Figure 4-9. Spatial distribution of crop fields and urban areas in the Mexicali area.....	63
Figure 4-10. GIS overlay methods.....	63
Figure 4-11. Annual burned agricultural areas (km ²) by INEGI rural areas in the northeastern BC.....	64
Figure 4-12. Monthly burned agricultural areas (km ²).....	65
Figure 4-13. Low agricultural fire activities identified by time series clustering (Cluster 0, 1, and 2).....	65
Figure 4-14. High agricultural fire activities identified by time series clustering (Cluster 3, 4, and 5).....	66

Figure 4-15. Spatial distribution of temporal trends of agricultural fire activities.....	66
Figure 4-16. Spatial distribution of temporal trends of agricultural fire activities (Cluster 3, 4, 5).....	67

List of Tables

	PAGE
Table 1-1. Original data source for point source emission facilities.....	4
Table 1-2 All Facilities of Interest: Remote validation result.....	11
Table 1-3. Facilities of Interest s within 1.5 km of the border: Field validation result.....	12
Table 1-4. Summary statistics of remote validation location error	13
Table 1-5. Summary statistics of original data location error	14
Table 1-6. Facilities of Interest location error (n=194).....	14
Table 2-1. Meteorological data obtained for analysis of transport events and the evaluation of model performance.....	22
Table 2-2. Estimated PAI [km ²] for California side only for selected key sources in the San Diego- Tijuana and Calexico-Mexicali border regions during high and low southern wind speed episodes.....	30
Table 2-3. Receptor locations for estimating PSR.....	32
Table 3-1. Types of hazardous facilities in Mexico influencing US census tracts along the US-Mexico border.....	55
Table 3-2. Change in EJSM Regional Hazard Proximity Scores (Regional Scoring) for US Census tracts along the US-Mexico border.....	56
Table 4-1. The Pixel Classification of the MODIS Active Fire Product.....	59

Abstract

CalEnviroScreen (CES) is a screening tool developed by the California Environmental Protection Agency that scores California census tracts based on measurements of pollution burden and population vulnerability, in order to direct resources toward improving public health in disadvantaged and burdened communities. Communities near the U.S.-Mexico border face pollution burdens from sources in Baja California that are not fully incorporated into CalEnviroScreen. To improve the understanding of impacts of air pollution from Mexico on border communities and provide critical data to the California Air Resources Board, we carried out three types of analysis. The first was mapping activities for sources in Baja California, including burn events; the second was modeling the potential areas of influence (PAIs) of selected sources and potential source regions (PSRs) for California communities; and the third was to estimate the impact of sources close to the border on existing community scores in the Environmental Justice Screening Model (EJSM). A systematic methodology was developed that included satellite imagery and ground-based verification of locations of sources, and a database that contains 1174 facilities was created for use in this and future projects. Modeling results strongly suggest that areas in the US immediately adjacent to the border and farther north in San Diego County are likely to be affected by emission sources in the Mexican side, not only by within a 1.5-km buffer of the border but also from potential regions both inland and off-shore within a 50-km buffer. A pilot method was developed for measuring and mapping fire activity near the US-Mexico border, and we present mapped locations and temporal patterns of agricultural burning on both sides of the border. Data also indicated that urban burning was common in Mexicali. Air toxics emissions sources in Mexico have a clear potential to affect communities and individuals in the US-Mexico border region. Recommendations include using the tools developed in this analysis to determine border-specific approaches to improving CES that take into account Baja California sources.

Executive Summary

BACKGROUND

CalEnviroScreen (CES) is a screening tool developed by the California Environmental Protection Agency that scores California census tracts based on measurements of pollution burden and population vulnerability. Its purpose is to identify areas of the state that have the highest pollution burden with populations that are most vulnerable to pollution exposure, so that resources can be directed toward reducing these impacts and improving public health. Communities near the U.S.-Mexico border often have a high percentage of vulnerable people, including children, and are often majority low income and Latino. Communities near the US-Mexico border in both Imperial and San Diego counties face pollution burdens from the land Ports of Entry (POE) and sources in Baja California, some of which are not incorporated in CES data. The California Air Resources Board (CARB) provides critical data to CES about community burden of air pollutants. Significant gaps in knowledge remain regarding how pollution from Baja California might affect populations on the border, and in particular how this information might be used to improve CES scores for border communities. These gaps included accurate knowledge of types and locations of pollution sources in Baja California near the border that might affect populations in California situated on the border, as well as knowledge of the geographical extent of affected areas of San Diego and Imperial Valley due to sources both close (<1.5 km) and far (<50 km) from the border.

OBJECTIVES AND METHODS

To improve the understanding of impacts of pollution from Mexico on border communities in the CalEnviroScreen model, we carried out three types of analysis. The first was mapping activities for sources, including burn events; the second was modeling the potential areas of influence of selected sources and potential source regions for California communities; and the third was to estimate the impact of sources close to the border (within 1500 meters) on existing community scores in the Environmental Justice Screening Model (EJSM). For mapping activities, datasets from Baja California were obtained, including new emissions inventories. Our objective was to provide CARB with data files on air pollution emitters on the Mexican side of the California-Baja California border that have the potential to affect California border communities, in a format that can be used to augment the information on environmental exposures from California-based metrics used in environmental justice screening models (CES and EJSM). A data validation procedure was developed for geographic coordinates to identify valid locations and perform further analysis on invalid locations. The location and type of sources were verified through a series of multiple and systematic steps that included viewing satellite images, consulting public business directories, making telephone calls, and, for sites within 1500 meters of the border, visiting the sites in person. Additional data obtained, such as a change of business name, was maintained in the overall dataset created. For modeling of potential areas of influence, suitable sources, i.e. major emitters such as the POEs and brick kilns, were selected from the dataset created in the first analysis, and meteorological conditions on certain selected days were used in conjunction with forward trajectories to map the extent of pollution in California border counties. For evaluating sources in Mexico that could potentially affect border cities, backwards trajectories were modeled. For the third activity, estimating the impact of sources in Mexico on the EJSM,

the dataset generated in the first activity was used to generate new scores for border counties using existing weighting methodology.

RESULTS

A total of 1174 (92.4%) facilities were remotely verified and included in the final dataset prepared for CES, from a compiled initial database of 1271 facilities. From the 1271 facilities, 216 sites were mapped to fall within 1500 meters of the border after remote validation and underwent additional field validation. Of these, 176 were included in the final dataset, and of these 128 had the physical location corrected. For modeling of potential areas of influence (PAIs), 12 sources for the San Diego-Tijuana region and 8 sources for the Calexico-Mexicali region were selected and modeled for PAIs under high wind speed and low wind speed episodes. All sources located along the 1.5-km buffer of the US-Mexico border had relatively large PAI on the California side, even during the low wind episodes. The spatial extent of the impacts from all sources modeled depended strongly on the location of the source and meteorology. For example, the Cerro Prieto geothermal plant had a relatively large PAI of ~670 km² during the high wind speed episode, but during the low wind speed event it did not impact any region in the US side of the border. Potential source region analysis (PSR) suggests large contributions from the Tijuana urban area to the San Ysidro, California receptor site. Emissions released from sources along those pathways, including the Otay Port of Entry (POE), are likely to impact regions in the US. In addition, emissions from Tecate and farther to the east can eventually mix with those from the Tijuana urban area. The PSR results also indicate potential contributions from ocean and coastal regions, suggesting that ship emissions from Ensenada and San Diego can contribute to regional impacts. For the Calexico site, the closest receptor to the border, the PSR results suggest large and direct contributions from the Mexicali urban area to border communities. When recalculating the scores for the EJSM model, even the clustering of sources near the border did not in general affect the quintile of the EJSM score, in part due to the large size of the census tracts, which diluted local effects near the border. A pilot method for mapping fire activity near the US-Mexico border indicated locations and temporal patterns of agricultural burning and also indicated that urban burning was common in Mexicali.

CONCLUSIONS

Air toxics emissions sources in Mexico, including burning activities, have a clear potential to affect communities and individuals in the US-Mexico border region. A systematic methodology was developed that included satellite imagery to verify locations of sources, and a database was created for use in this and future projects. Recommendations include using the tools developed in this analysis to understand the impact on California from agricultural and urban burning in and near Mexicali, extending the models to examine nighttime combustion sources identified previously in Tijuana, combining potential source regions identified in this study with gridded emissions inventories to estimate the relative contributions from emission sources in those areas, and determining border-specific approaches to improving CES which take into account differences between California and Baja California sources.

Body of Report

CHAPTER 1. MAPPING ACTIVITIES FOR SOURCES IN BAJA CALIFORNIA AND DATABASE CREATION

1.1. Introduction

The first task of this project is to locate, aggregate, and characterize existing best available data on emission sources on the Mexican side of the California-Baja California border. The project focused on those sources that impact communities in California. To achieve this task, the project team conducted a series of data acquisition and evaluation processes including planning meetings, compiling existing data, and data evaluation. A list of data products is discussed in Chapter 5.

1.2. Planning Meetings (Task 1.1)

This project started with two planning meetings in April and May 2017 to inform the project investigators, collaborators and stakeholders as to objectives and tasks, data requirements, lessons learned, and suggestion for the project tasks. The first kickoff meeting was held on April 14, 2017 at San Diego State University (SDSU). Attended participants included representatives from SANDAG, ARB, OEHHA, Environmental Health Coalition, Comite Civico del Valle, SDAPCD, EPA border office, as well as investigators from US and Mexico (Dr. Castillo) and a representative from SPA (State of Baja California environmental agency) (Dr. Quintero). The second meeting in Mexico was held at the UABC Mexicali campus in May 26, 2017, in Mexicali, Baja California. At this meeting there was representation from local Mexicali city government agencies, SEMARNAT (federal, Mexico), and SPA. After the planning meetings, the project team drafted a list of emission sources to be considered in this project as follows

Large industrial facilities (non-point source)

- Distribution centers
- Railroads and Railyards
- Ports
- Refineries
- Chemical manufacture
- Power generation and distribution
- Large gasoline dispensing/storage facilities
- Recycling and solid waste treatment/disposal
- Solid waste handling, transfer and incinerators

Small area source emitters (point source)

- Mining: mining, quarrying, oil and gas development
- Meat and poultry processing/packing

- Dairy beverage and food processing
- Textile and carpet manufacture
- Wood products manufacture and processing
- Paper processing and manufacture including printing
- Chemicals manufacture, processing including paint, plastics, industrial and agricultural chemicals
- Petroleum refining and processing
- Rubber, foam and plastics manufacturing products
- Glass, ceramic and stone manufacture including concrete and plaster
- Steel and other metal manufacture and processing, smelting
- Manufacturing using steel and other metals
- Machinery manufacture
- Electronics manufacture, including computer and other office machines
- Other manufacturing: appliances, communications equipment, electronic equipment and components
- Transportation equipment manufacturing – all modes
- Warehousing and storage, freight transport
- Utilities: water, sanitary, refuse and waste disposal
- Services: automotive and truck repair and services
- Hospitals and medical laboratories
- Colleges, universities and research laboratories
- Correctional institutions
- Solid waste management and disposal
- Military facilities

1.3. Compiling Existing Data (Task 1.2)

Through consultation with government agencies and colleagues on both sides of the border, the project team compiled existing data from multiple emission sources. The US-Mexico border is defined as within 100 km of the border as specified in the La Paz agreement. The project team focused on land use, pollutant sites, and Mexican National Emissions Inventory (MNEI) concentration estimates that are as similar to the US data as possible. This project focused on location of air pollution emissions sources and certain land uses associated with high pollution emissions, following the recommendations and definitions of the CARB Air Quality and Land Use Handbook (CARB, 2005) in the border region of Baja California, within 1.5 kilometers of the border, as this distance is compatible with the ways in which both CES and EJSM characterize hazard proximity in

their respective scoring metrics. However, the locations of large- scale emissions sources and regional air quality hazards were also determined within 50 kilometers of the border.

Based on the draft list of emission sources, the Facilities of Interests (FOIs) considered in this project included industrial emitters identified in State and Federal Mexican databases, feedlots, landfills, gas stations, brick kilns, and dry-cleaning facilities.

Methodologies for data integration, remote validation and field validation of the sources were developed to produce a single dataset with FOIs from multiple data sources. Table 1-1 provides the original data source description for those point source emission facilities and describes data attributes on the dataset name, the dataset source, the year, the dataset size as it pertains to northern Baja California, if the dataset provided geographic coordinates, if the dataset provided emission or pollutant data, and the top sector identified within the dataset.

Table 1-1. Original data source for point source emission facilities

Dataset	Data Source	Year	Number of Facilities	Location	Emissions /Pollutant Data	Top Sector
MNEI (Mexico National Emissions Inventory)	Obtained by Molina Center from SEMARNAT	2013	586	Provided	Yes	Equipment & Component Manufacturing
BCEI (Baja California Emissions Inventory)	Obtained by Molina Center from SEMARNAT	2014	68	Provided	Yes	Metallurgy - Including Steel
RETC (Registry of Pollutant Release & Transfer)	Obtained by OEHHA from SEMARNAT	2011	25	Provided	Yes	Quimica & Metallurgy - Including Steel
		2012	27			Quimica
		2013	35			Metallurgy - Including Steel
	Obtained by Molina Center from SEMARNAT	2014	87			Equipment & Electronics, Electrical & Domestic Articles
		2015	102			Quimica
		2016	93			Equipment and Electronics, Electrical & Domestic Articles
Feedlots	Manual Data Collection	2018	16	Manually Collected	No	n/a
Landfills	Mexicali Atlas	2011	9	Provided	No	n/a
	Manually Collected	2018	8	Manually Collected	No	n/a
Brick Kilns	DENUE (National Statistical Directory of Economic Units)	2014	60	Provided	No	n/a
Gas Stations	PEMEX (Mexican Petroleum)	2017	472	Provided	No	n/a
Dry Cleaning	Mexicali Atlas	2011	8	Provided	No	n/a
	Manually Collected	2018	44	Manually Collected		

The following sub-sections describe each emission data source.

1.3.1. MNEI – Mexican National Emissions Inventory

Data for emission sources, pollution concentrations, and location are collected by the Mexican Federal and State Government and published as a public dataset called the Mexican National Emissions Inventory (MNEI). MNEI 2013 data was used as the primary data source for this project and was extracted from SEMARNAT (Secretariat of Environment and Natural Resources), a federal database. If FOIs within the additional datasets overlapped with MNEI dataset, they were merged into one FOI with multiple data sources (see section 1.4.3.2). The notable data associated with this MNEI dataset include: CVEIDEN identification codes, sector, business name, NAICS (North American Industry Classification System) code, address, geographic coordinates, and emissions data for pm10, pm2.5, SOX, CO, NOX, COV, CO2, CH4, NH3, and BC (see appendix for complete metadata document).

1.3.2. BCEI – Baja California Emissions Inventory

The Baja California Emissions Inventory (BCEI) is a stand-alone dataset, which is also considered as MNEI 2013+; meaning some of the data in the BCEI (2014) dataset overlaps with the current MNEI 2013 data and includes additional facilities post 2013. For clarity, we refer to this data as BCEI 2014. These data are very similar to MNEI; however, there were two notable differences. First, BCEI 2014 data included new businesses not identified in the 2013 MNEI data, thereby increasing the total number of FOIs in this study. Second, the emissions data included the same emission types as MNEI, yet also included data for NO2 emissions. The remaining data fields were identical to MNEI.

1.3.3. RETC – Registro de Emisiones y Transferencia de Contaminantes

The Registry of Pollutant Release and Transfer (Registro de Emisiones y Transferencia de Contaminantes) dataset contains data for large emitters and reports criteria pollutant data, which differs from the emissions data found in MNEI and BCEI. The years considered for this project included 2011-2016. While all years had pollutant data, 2016 was the only year pollutant data was included in the final dataset as it represented the most recent pollutant condition within the study area. RETC data were integrated into the working dataset by merging overlapping FOIs with existing MNEI and/or BCEI data or were appended to the dataset as a RETC FOI (see section 1.4.3.2 for more details). The notable data associated with RETC datasets included: Facility ID codes, NRA ID codes, business name, sector, geographic coordinates, and for 2016 only, criteria pollutant data. The 2016 criteria pollutant data required translation from the original attribute name in the RETC dataset to an English common name or chemical symbol or formula. The translation process was recorded within the metadata document (see Chapter 5).

1.3.4. Additional Data

Interest in emitters beyond industrial businesses, from datasets mentioned in the above sections, required data to be obtained via other databases and manual collection. These included: Feedlots, Landfills, Brick Kilns, Gas Stations, and Dry-cleaning facilities. Many of the additional datasets were obtained using Google services; Google Search,

Google Maps, and Google Earth. All the additional data underwent the same remote validation procedure as discussed in more detail in section 1.4 below.

1.3.4.1. Feedlots

16 feedlots were gathered using manual searches on both Google Maps and Google Earth applications. Research partners in Mexico also identified coordinates of feedlots that were not captured using Google Maps and Google Earth. Google Maps was not able to identify feedlots by simply typing ‘feedlots’, ‘ganado de engorda’ and related terms into the search bar, as many of them were not listed; therefore, Google Maps and Google Earth were primarily used as aerial imagery to manually search for feedlots, primarily within the 1.5 km buffer.

1.3.4.2. Landfills

Landfill data was obtained via extraction from the Mexicali Atlas (2011) database, contribution of landfill knowledge by research partners in Mexico, and by manual Google searches. By using Google Search and typing ‘landfills in Tijuana’, or ‘landfills nearby’ in Google Maps, FOIs were able to be identified. While a portion of the data came from the Mexicali Atlas, information on landfill activity, waste production, or practices was not provided.

1.3.4.3. Brick Kilns

Brick Kiln data was obtained via extraction from the 2014 DENU (Directorio Estadístico Nacional de Unidades Económicas, or National Statistical Directory of Economic Units in English) database. While the locations and names of the FOI were provided in the DENU dataset, information on brick production, materials, and other relevant information was not provided.

1.3.4.4. Gas Stations

Gas Station data was extracted via PEMEX (Mexican Petroleum) 2017 dataset. This dataset did not include information such as number of pumps, emissions or other detailed data, but provided location information, and name of FOI.

1.3.4.5. Dry Cleaning

Dry cleaning facilities were identified using data from the 2011 Mexicali Atlas and manual search using Google Search, Google Maps, and Google Earth. Google search was used to research business listings for dry-cleaning facilities by municipality, Google Maps was used to search and validate using Google street view, and Google Earth was used in validation. The process for validation of these data can be seen in section 1.4.

1.4. Data Validation Methodology (Task 1.3)

1.4.1. Remote Validation Procedure

The data validation procedure was developed to validate geographic coordinates resulting in either valid location or invalid locations for further analysis and evaluation for incorporation into CES. This procedure also allowed for correcting locational data by obtaining more precise geographic coordinates by utilizing 8 systematic steps to ensure locational accuracy, and consistency throughout the entire process regardless of FOI type. The steps include validating based on; 1) Latitude and Longitude plausibility 2) searching

Google Maps, 3) searching Google Earth, 4) searching address information, 5) searching business name, 6) calling business, 7) calling current business or tenant, and 8) field validation. While field validation is an important step in the process, it is not considered remote validation and will be described in detail in section 1.4.2. The first seven steps are referred to as sub-methods and are further explained below.

1.4.1.1. Sub-Method 1: Latitude & Longitude

Validation of data began with an evaluation of Latitude and Longitude coordinates using ArcGIS ArcMap 10.5.1. Provided coordinates were first visualized in GIS using an ArcGIS aerial imagery basemap. This initial visualization indicated further validation was required as some original coordinates were with zero latitude and longitude, in United States, or in rural seemingly inhabited parts of Mexico. These were later found to be incorrect once the FOI was fully validated. While these locations seemed to be inaccurate, using remote sensing technologies such as ArcGIS comes with limitations; specifically, the date of the basemap imagery. It was crucial to have locational accuracy, so each FOI required multiple ways to validate location. Thus, the remote validation procedure began with an evaluation of the geographic coordinates and concluded with phone calls.

1.4.1.2. Sub-Method 2 & 3: Google Maps & Google Earth

Both sub-methods 2 and 3 utilized online mapping services to validate FOI location. Google Maps and Google Earth Pro allowed coordinates to be visualized thereby determining if the location was valid as of imagery date, or if further remote validation was needed.

FOI coordinates were first visualized in Google Maps, then in Google Earth. Depending on the FOI type being remotely validated, various aspects of Google Maps proved insightful for validating location, such as the clickable links to businesses in the area, Google Street View, address information, street information, 3-dimensional views, and the ability to obtain precise geographic coordinates as well. Google Earth was used to look at the aerial imagery of the specified location, as well as the surrounding business and street information. Google Earth also provided detailed 3-dimensional views which were compared to the imagery shown in Google Maps. This allowed for precise identification and collection of geographic coordinates. This functionality is especially useful outside of urban regions and can serve as a surrogate for remote areas without Google Street View. The most important functionality of Google Earth used in this sub-method was the ability to obtain imagery dates. This is highly important, as validation, either remote or in the field, must consider both spatial and temporal information. While Google Earth has the ability to obtain accurate imagery dates for a specific location, Google Maps aerial imagery does not; rather, Google Maps is a mosaic of multiple aerial imagery datasets collected at various times. This does not, however, make Google Maps a less powerful tool in this framework, but rather a tool used to obtain additional information.

Once sub-methods 2 and 3 were completed, the FOI location was considered either valid or invalid. If the location was identified by using the provided coordinates or by using obtained coordinates, as a result of these two sub-methods, it was considered valid. If a

more precise location was needed or if the location could not be confirmed, sub-methods 4 and 5 were implemented.

1.4.1.3. Sub-Method 4 & 5: Google Search – Address & Business Name

A Google search was used for validating location and providing background information on the FOI being validated. Both Google search and business name search sub-methods were performed, almost simultaneously. For instance, Google searches utilized provided information in the various datasets such as FOI street names, address, and FOI (business) name. Use of these two sub-methods jointly yielded crucial information to this study such as FOI status (still operating/closed), new address (if moved locations), Facebook pages, information regarding FOI operations, and current phone number. Efforts exerted within these two sub-methods were greatly beneficial for informing research on the remaining FOIs that could not be validated using these and previous sub-methods.

1.4.1.4. Sub-Method 6 & 7: Call Business – Call Current Business or Tenant

Google and business name searches yielded phone numbers that were used in these sub-methods. If FOIs were unable to be validated using previous sub-methods, phone calls were used to obtain location. A subscription was purchased from Skype and used to call the FOIs which had yet to be validated. A Research Assistant (bilingual) called FOIs and validated on-the-fly, ensuring the location was visualized using Google Maps or Google Earth during the phone call. Google Maps was useful in this sub-method as context information, such as street names and surrounding businesses, could be discussed by the party on the phone and visualized by the Research Assistant, simultaneously. Sub-method 7 was unique in that purposeful phone calls were made to businesses which were not being validated but could perhaps provide information on the FOI in question. For instance, if a specific FOI could not be validated and contact information was not obtained, surrounding businesses whose information was readily available online would be contacted to verify if the FOI was present or not. This was beneficial to this study as some FOIs would not have been validated using previous methods only. While all data underwent the remote validation procedure, only the FOIs within the 1.5 km buffer were validated in the field.

1.4.2. Field Validation

Field validation was conducted for FOIs remotely validated within the 1.5 km buffer. Research Assistants were trained in field validation procedures that allowed for individual FOI data to be edited, updated, and validated in the field by utilizing Arc Collector technology on mobile devices such as cellular smart phones. This software allowed for Research Assistants to modify FOI location and attribute data, enter emission characteristics, take photos, and record additional comments such as the date of field validation and other relevant observations. The field data collected during ground-truthing was later synchronized with an Arc Online account for this research and evaluated. Evaluation of post field collection data involved manual assessment of the field observations to determine if any findings would prevent validity, or usage of the location as an emitter for later incorporation in CES.

1.4.3. Data Augmentation and Refinement

1.4.3.1. SIC Coding

CES requires a SIC (Standard Industrial Classification) code for the collected FOI data. For the FOI data with only NAICS codes, NAICS codes were translated into equivalent SIC codes. While some NAICS codes had directly equivalent SIC codes, many NAICS codes in the database were incomplete, or had multiple matches; thus, two methods were used in the NAICS to SIC crossover. The first method utilized the NAICS to SIC crosswalk obtained from <https://www.naics.com/product/sic-naics-cross-references/>. NAICS codes had either a direct match to a single SIC code, or a match to multiple SIC codes. When multiple SIC codes were listed as an equivalent, supplemental data such as ‘sector’ (in MNEI, BCEI, and RETC) was used to identify the closest match to the NAICS description.

The second method required more detailed investigation into the operations of the FOI in question as the NAICS code was either incomplete or missing. NAICS codes are typically 6 digits long, which will provide the most accurate and detailed information on the FOI in question. Within the datasets mentioned above, some FOIs only had a three-digit code, which only provides information on the main sector and sub-sector operations. This made the process of NAICS to SIC translation difficult. The ‘sector’ information was used for these cases as it allowed for the best possible SIC match to be made. For example, NAICS codes of 325 and 339 were matched with SIC codes of 28Chemicals and Allied Products and 30-Rubber and Miscellaneous Plastics Products where 2-digit SIC codes determine major industry groups.

When FOI data did not have NAICS code information, SIC codes were manually searched to find the best fit based on the FOI operation. For instance, Brick Kilns, Feedlots, Gas Stations and Landfills did not have NAICS codes associated with them; by manually searching related SIC codes based on their operations, SIC codes of 3255-Clay Refractories, 0211-Beef Cattle Feedlots, 5541-Gasoline Stations and 4953-Refuse Systems were assigned respectively.

1.4.3.2. RETC & BCEI Data Integration

Due to the same FOI being present on many datasets and to avoid duplication of data, a methodology was implemented to ensure accuracy in joining overlapping FOI data. MNEI 2013 was the primary dataset for this study and therefore served as a base for the additional datasets to either be merged (if overlap was observed) or appended (if no overlap was observed). Both BCEI and RETC data required careful implementation of this methodology.

BCEI data were carefully evaluated and compared to MNEI 2013 data specifically looking at the supplemental data such as Facility Name, Address, CVEIDEN, and Sector. The BCEI CVEIDEN identification codes were comparable to the MNEI CVEIDEN, which allowed for accurate evaluation. For each BCEI FOI, if overlap was observed, the FOI was joined with existing MNEI 2013 dataset. If no overlap was observed, the FOI underwent remote validation procedure (Sub-Methods 1-7) and was appended to the dataset as a new FOI.

RETC data came in the following groups, 2011-2013, 2014-2015, and 2016. Each RETC FOI had a specific identification codes called RETC FID (2011-2013) or NRA ID (2014-2016). These identification codes were compared to the MNEI and BCEI CVEIDEN identification codes. For RETC 2011-2013, Facility Name, Address, Sector, and CVEIDEN to RETC FID comparisons were used to identify overlapping or new FOIs. The next RETC dataset to be evaluated was out of chronological order and no longer had RETC FID's but rather NRA identification codes. For RETC 2016, Facility Name, Address, Sector, CVEIDEN to NRA ID (RETC) comparison, and RETC ID to NRA ID comparisons were used to identify overlapping or new FOIs. For RETC 2014-2015, Facility Name, Address, Sector, CVEIDEN to NRA ID comparison, RETC FID to NRA ID comparison, and when possible, NRA (2016) to (NRA 2014, 2015) comparisons were used to identify overlapping or new FOIs. After each stage of RETC data integration, if no overlap was observed, the FOI underwent remote validation procedure (Sub-Methods 1-7) and was appended to the dataset as a new FOI. If overlap was observed, the FOI was carefully merged into the growing dataset.

1.4.3.3. CES Selection Methods

To determine if the FOI should be considered as input to CES, a methodology was developed to systematically evaluate the results of various attribute fields within the final dataset, for each FOI. Four potential exclusions were considered for each FOI, including if the FOI was non-operating, if the SIC code was not of primary consideration in this study, if validation of the FOI was not possible, and more specifically, if the location could not be validated.

For an FOI to be considered non-operating, it had to be observed as closed during remote or field validation without replacement with an FOI of interest. For example, if a manufacturing FOI warehouse was converted to commercial office use, it was considered non-operating.

If an FOI had a SIC code which was not considered as relevant based on the NAICS to SIC crosswalk used in this study, it was not considered for inclusion in CES. This also includes the SIC code for a new business replacing an original FOI. There was, however, some manual correction of SIC codes of interest due to SIC codes which were incomplete (2 or 3 digits out of 4). This happened when provided NAICS codes were incomplete, and the main sector and/or sub-sector operations were of interest to this project. An example of this is the inclusion of SIC code 30- Rubber and Miscellaneous Plastics Products, whereas SIC code 2099-Food Preparations was not included. There were some FOIs which could not be validated by either remote and/or field validation. If the FOI was not able to be validated with confidence as to the operations, SIC, business name, etc., then the FOI was not considered for inclusion in CES.

Because this study emphasized locational accuracy, if the FOI was not able to be validated in regard to accurate location it was not considered as valid for CES. Location could be either the remotely validated or field-validated location. If no final coordinates could be obtained with confidence, the FOI in question was excluded from CES.

While these methods were developed by the researchers responsible for this task, the various attributes of the final dataset can be manipulated to remove, modify, or add new parameters for identifying potential FOIs to include in CES.

1.5. FOI Data Validation Result

1.5.1. All FOIs

Upon completion of Task 1, including data integration, 1,271 FOIs were identified from the following data sources: MNEI, BCEI, RETC, business added, feedlots, landfills, gas stations, brick kilns, & dry-cleaning facilities. Approximately 97% (1,228 of 1,271) were able to be validated remotely with 1,174 available for inclusion in CES based on the methodology described in Section 1.3 and 1.4. 97 FOIs were excluded from CES for the following reasons: the SIC code was not included in the CARB FOI dataset (n=24), the FOI was closed (n=35), the FOI could not be validated remotely (n=33), during remote validation the FOI was found to be closed (n=4), there was no access to the FOI to validate (n=1). A detailed breakdown of CES inclusion by data type can be seen in Table 1-2. For further breakdown by data source see the FOI Breakdown and Statistics document in the appendix.

Table 1-2. All FOIs: Remote Validation Result

Data Type	Total Count	CES Valid ^{*1}		CES Invalid ^{*2}	
		Count	%	Count	%
MNEI, BCEI, RETC, Business Added	654	572	87.5	82	12.5
Feedlot	16	16	100.0	0	0.0
Landfill	17	16	94.1	1	5.9
Gas Station	472	468	99.2	4	0.8
Brick Kilns	60	55	91.7	5	8.3
Dry Cleaning	52	47	90.4	5	9.6
Total	1,271	1,174	92.4	97	7.6

^{*1} Business and location were validated for inclusion in CES

^{*2} Business or location was not valid, and the FOI was not included in CES

1.5.2. FOIs within 1.5 km of the border

FOIs within the 1.5 km of the border were of great importance as these FOIs would directly inform CES. 216 FOIs locations fell within this zone based on remote validation procedures. The total of 204 FOIs underwent field validation in Spring of 2018 to validate their location and business type. 12 FOIs (out of 216 FOIs) collected from BCEI or RETC were not validated in the field since these were included after completing field validation. These 12 FOIs were validated remotely and have been flagged for inclusion in

the CES-valid dataset. 1 FOI (out of 204 FOIs) was not be able to be validated in the field due to limited access. 1 FOI's location (out of 204 FOIs) was corrected and moved outside of 1.5 km of the border after field validation. By excluding these 2 FOIs, 164 FOIs out of 202 FOIs (81.2%) have been flagged for inclusion in the CES-valid dataset. Table 1-3 shows a detailed breakdown of CES inclusion by data type, for FOIs within the 1.5 km buffer.

Table 1-3. FOIs within 1.5 km of the border: Field Validation Result

Data Type	Total Count	CES Valid*1		CES Invalid*2	
		Count	%	Count	%
MNEI, BCEI, RETC, Business Added	95	70	73.7	25	26.4
Feedlot	2	2	100.0	0	0.0
Landfill	2	1	50.0	1	50.0
Gas Station	72	68	94.4	4	5.6
Brick Kilns	4	1	25.0	4	75.0
Dry Cleaning	27	22	81.5	5	18.5
Total	202	164	81.2	38	18.8

*1 Business and location were validated for inclusion in CES final dataset

*2 Business or location was not valid, and the FOI was not included in CES final dataset

1.5.2.1. Emissions Breakdown

Of the 176 FOIs (164 FOIs validated in the field + 12 FOIs validated remotely) within the 1.5 km buffer that have been flagged for inclusion in the CES-valid dataset, 73 FOIs (41.5%) had provided emissions data. Emissions data came from one or a combination of the MNEI and BCEI datasets, while criteria pollutant data from RETC. 103 FOIs (58.5%) within the 1.5 km buffer had no emissions or criteria pollutant data.

1.5.2.2. Location Accuracy

Of the 202 FOIs validated in the field, only 52 FOIs (25.7%) matched the originally provided coordinates, demonstrating the need for remote and field validation of coordinates found in raw datasets.

Comparing remote validation to field validation, of the 202 FOIs validated in the field, 112 FOIs (55%) had no distance between remotely obtained coordinates and coordinates validated in the field. The remaining 90 FOIs had distances between the remotely obtained coordinates and coordinates validated in the field, or location errors. 4 FOIs (2%) had distances less than 10 meters, 76 FOIs (38%) had distances less than 100 meters, and 10 FOIs (5%) had distances greater than or equal to 100 meters with an average of 175.6 meters.

To measure the usefulness and performance of the remote validation procedure, a straight distance average was calculated for the 90 FOIs with location error. Table 1-4 and Figure 1-1 present the summary statistics and frequency distribution of location error respectively. The average distance (42.5m) between remotely obtained coordinates and coordinates validated in the field for these FOI is 42.5 meters. This distance indicates that while field validation is most accurate, remote validation can be very useful if exact location is not required.

Table 1-4. Summary statistics of remote validation location error.

N	Min (m)	Max (m)	Average (m)	Standard Deviation (m)
90	6.6	290.0	42.5	53.8

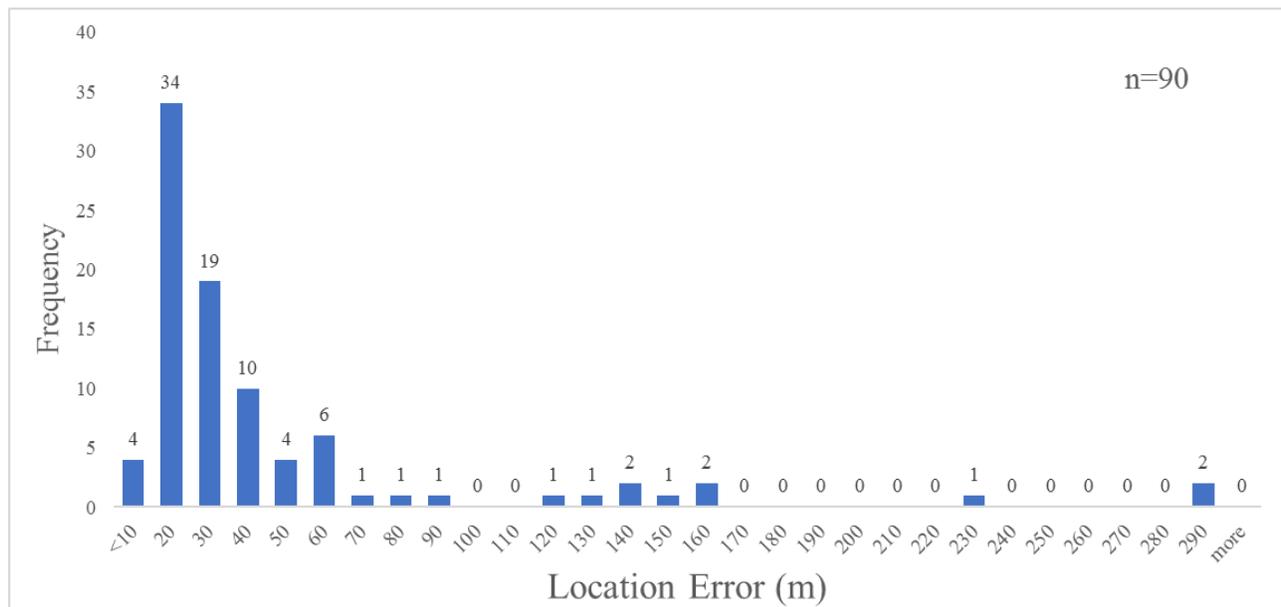


Figure 1-1. Frequency distribution of remote validation location error.

Of the 90 FOIs that were most accurately identified during field validation, 80 FOIs had less than 100 meters of distance between the coordinates obtained during remote validation and the final coordinates with an average of 26 meters. These distances were typically due to adjustments made to coordinates during field validation such as, moving datapoint in Arc Collector from the street edge to directly over the property or building. Once final coordinates were obtained, these 80 FOIs were looked over carefully and found to be considered as successfully validated using remote validation. The remaining 10 FOIs had large distances, greater than or equal to 100 meters, with an average of 175.6 meters. These were not considered successfully validated using remote validation procedures although many of them were relatively close (same block or same street).

To measure the accuracy of remotely validated locations, we combine the number of FOIs with no location error (n=112), and FOIs with small location error (n=80) form the total FOIs validated in the field (n=202). In total, 192 (95%) of 202 FOIs' locations are considered successfully validated using the proposed remote validation procedure.

1.5.2.3. Original Data Location Accuracy

Locational accuracy of original data was of great importance to this study as it provided reliable data for further analysis and demonstrated the usefulness of the developed remote validation procedure. Out of 202 FOIs that were within 1.5 km buffer from the border and undergone field validation, 194 FOIs had original coordinates provided in their data sources whereas 8 FOIs (8 Dry Cleaning Facilities) did not have originally provided coordinates. Out of 194 FOIs, coordinates with 52 FOIs (26.8%) having the same final coordinates as those originally provided and 142 FOIs (73.2%) having distance between the original and final validated coordinates.

Of the 142 FOIs that have distance between the provided coordinates (prior to validation) and final coordinates (post validation) or locational errors, 32 FOIs have less than 1 meter between the original and final coordinates with an average of 0.04 meters (Table 1-5). 69 have less than 100 meters of distance with an average of 26.1 meters. 17 FOIs have less than 1 kilometer of distance with an average of 427.2 meters. 24 have greater than 1 kilometer of distance with an average of 33,928.1 meters. This class demonstrates the need for validation of provided coordinates within the raw datasets prior to further analysis.

Table 1-5. Summary statistics of original data location error.

N	Min (m)	Max (m)	Average (m)	Standard Deviation (m)
142	0.01	115,850.12	5,798.18	19,650.76

Table 1-6. FOIs location error (n=194)

Distance between the original and final coordinates, or location error (m)	0, or same coordinates	$0 \leq d < 1$ (m)	$1 \leq d < 100$ (m)	$100 \leq d < 1,000$ (m)	$d > 1,000$ (m)
Number of FOIs	52	32	69	17	24
Average Distance (m)	0.00	0.04	26.1	427.2	33,928.1
% of FOIs	26.8%	16.5%	35.6%	8.8%	12.4%

1.6. Discussion

We encountered issues with data integration, NAICS to SIC code crossover, business turnover, and limitations. Both data integration and NAICS to SIC code crossover were dependent on the raw datasets. During integration and validation, incorrect, missing, or incomplete information introduced potential error. For example, BIO PAPPEN PRODUCTS, S.A. DE C.V. is an FOI that appeared on both the 2014 and 2016 raw RETC datasets. Strangely, it was not observed in the 2015 dataset and the 2014 ID differed from the ID in 2016. The RETC FID from 2014 differed from the RETC 2016 NRA ID; however, it was validated to the same location. Due to conflicts in the data, each RETC FOI name, sector, and address was checked against the MNEI dataset and in other RETC datasets from other years. Data integration was therefore carefully conducted so as to not duplicate FOIs while still recording occurrences of the FOI in question within the raw datasets. This was one of the major challenges in this Task.

SIC codes were manually chosen for certain FOIs as discussed in section 1.2; this also introduced potential errors. The SIC crossover result was dependent on provided data and data available through remote validation to provide as accurate a SIC match as possible. This also shows the possible errors in the raw datasets; if the raw dataset had issues, it would be amplified throughout data manipulation within Task 1 and potentially impact the final result for CES inclusion.

Business turnover also played a role in identifying FOIs for CES inclusion. Businesses were found to be closed during both remote and field validation. However, some FOIs were only identified as either closed or replaced by another business after field validation. This was one of the primary benefits of ground truthing FOIs and conducting field validation.

Lastly, one limitation to consider when conducting remote validation is the imagery date. While online remote sensing and mapping technologies such as Google Maps and Google Earth were very useful for remote validation, the imagery was not collected at the time of validation. In other words, the imagery date collected during remote validation ranged from 9/17/2004 to 2/25/2018 with a majority of FOIs having imagery dates from 11/30/2017 and 10/14/2017. This led to FOI turnover not being recorded prior to field validation. It would be difficult to remotely validate all locations without employment of multiple remote validation techniques; hence, phone calls were utilized which helped combat remote validation ambiguity. The remote validation procedure, while useful, cannot describe current conditions with 100 percent accuracy.

1.7. Summary, Conclusions, and Recommendations

Task 1 concluded with 1174 (92.4%) of the 1271 facilities of interest being successfully remotely validated for inclusion in CES and future versions of CES, as verified by subsequent steps. The developed remote validation procedures, methodologies for data integration, and developed ruleset for CES screening will be useful for future work using various emission or criteria pollutant datasets. The database and annotations provided information for future use and subsequent activities as detailed in Chapters 2 and 3 of this report.

1.8 References

California Air Resources Board (CARB). Air Quality and Land Use Handbook: A Community Health Perspective. 2005. <http://www.arb.ca.gov/ch/handbook.pdf>.

Directorio Estadístico Nacional de Unidades Económicas (DENUE). 2014. <https://www.inegi.org.mx/app/mapa/denue/>

Mexicali Atlas. 2011. <http://www.mexicali.gob.mx/transparencia/administracion/atlas/pdf/4.pdf>

Petróleos Mexicanos (PEMEX). 2017. <http://guiapemex.pemex.com/SitePages/home.aspx#!>

CHAPTER 2. MODELING OF POTENTIAL AREAS OF INFLUENCE (PAIs) AND POTENTIAL SOURCE REGION ANALYSIS (PSR)

2.1. Introduction

CalEnviroScreen is a tool used to identify communities in California with high pollution burden and with populations that are most vulnerable to pollution exposure. The analysis for estimating an impact score in CalEnviroScreen is done using census tracts and a wide array of indicators including health and socioeconomic vulnerability, exposure, sensitive populations, environmental effects from multiple sources of pollution, as well as proximity to emission sources, among others. An important feature of the CalEnviroScreen tool is that it can be continuously refined as improved and updated information becomes available.

The California-Mexico border region has communities with a high percentage of vulnerable people that are often of low income and with high composition of Latinos. In particular, communities near the border in both Imperial Valley and San Diego counties may be impacted by emission sources located in both the US side and the Mexico side of the border, some of which are not incorporated in CalEnviroScreen data. Therefore, a relevant research question for improving CalEnviroScreen in the California-Mexico region is at what extent the emissions sources on the Mexican side can potentially impact the border communities. This question is especially important not only for sources near the border, but also for high-emitters located at longer distances during south-to-north air pollution transport events.

In the current version of CalEnviroScreen, some data from Mexico have been incorporated, including traffic, PM_{2.5} and ozone levels, diesel PM, and toxic releases indicators using both direct observations and proxy data. For example, to account for additional diesel PM emissions, CARB compared the results of its gridded diesel PM calculation with estimated diesel PM measurements at the Calexico and Otay Mesa ambient monitoring stations using measured nitrogen oxides (NO_x) levels as a surrogate. However, to estimate the impacts of emission sources located in Mexico on California-Mexico border communities, a much better characterization of the local emissions is needed, as well as a better understanding of the meteorological conditions that drive the air pollution impacts. Task 1 of this project addresses the first issue by improving the databases of local emission sources in the Mexican side by gathering existing data and performing validation and ground truthing.

Task 1 of this project listed a number of emission sources on the Mexican border cities; the most prominent including traffic induced emissions from the ports of entry, maquiladoras and other industries, agricultural burning, electric generation. However, there are several emission sources that are prevalent along the Mexican border cities that are less likely to appear on the US side. These include:

- Trash burning. Vacant lots are common across Mexicali and Tijuana. When public services for garbage recollection are not available, the lots are often used to dispose trash and other items (including tires) that are prone to burning. The burning can be both intentional and non-intentional and certainly puts in high risk people living nearby.

- Dust. Dust resuspension from unpaved roads and farmlands are a large source of PM at the border region.
- Plastic burning for recovery of metals. Locals relate how people could use the vacant lots for burning of cables and electronic equipment to recover copper and other metals that are subsequently sold. Being an illegal activity, the burning often occurs at night, posing additional health and safety risks to families nearby.
- Maquiladoras are predominantly over the Mexican side of the US-Mexico border. These are manufacturing industries with mostly small to medium size combustion sources but handle large amounts of chemicals and metals for producing glass and plastic-based products. However, in addition to their direct emissions there are other sources linked to maquiladoras, including intensive traffic from heavy-duty trucks handling these products along the maquiladoras corridors.
- Steel mills. Huge piles of scrap metals are used in the steel mills processing. Impurities are separated from the molten iron, leaving behind waste that can contain iron, chromium, manganese, lead and other metals and pollutants.
- Brick kilns. There are numerous brick kilns on the Mexican border cities that still operate using artisanal production methods. These are emission sources that emit large amounts of air pollutants from intensive combustion for sustained periods of time (1-5 days) using non-conventional highly polluting materials (including tires) as fuels.
- Street food cooking. Food preparation using LPG, coal and wood in the streets is a common practice in many Mexican cities. These are even more prevalent at the border near maquiladoras and other industries as vendors set up their business nearby for easy access to consumers.

Lagrangian Particle Dispersion Models (LPDM) simulate the long-range transport and other physical processes of tracers released by a source of interest (area, line or point) (Stohl et al., 2005). These trajectory models can be useful tools for identifying upwind emission sources and to quantify their impact (Hüser et al., 2017). Trajectory models have also been used for studying short-scale dispersion of point sources, determining the flux footprints at specific measurement sites, simulating the long-range transport of biomass burning plumes, during designing of emergency and forecasting systems, and in improving emissions inventories of atmospheric species (Brioude et al., 2013 and references therein; Hüser et al., 2017). For instance, Gadhavi et al., (2015) used Lagrangian modeling to evaluate several emissions inventories of black carbon (BC) for southern India. For the State of California, Fisher and Jeong (2013) used a combination of measurements and trajectory and inverse modeling to better estimate regionally resolved methane emissions.

In Task 2 of this project we use the database obtained from Task 1 to identify the emission sources with the largest Potential Areas of Influence (PAI) impacting areas in San Diego, Calexico and adjacent US-Mexico border communities in California, using a combination of meteorological modeling simulations and trajectory analysis. We also identify Potential Source Regions (PSR) that have the potential for impacting selected receptor sites in the California side of the border. To accomplish this task, forward and backward trajectory analyses were performed using the FLEXPART-WRF model during

key selected northward transport episodes driven by outputs of the Weather Research and Forecasting (WRF) meteorological model. The following sections describe the methodology used and summarize the results obtained, which are useful for prioritizing the most important emission sources and for guiding the decision of the radius of influence to use when assigning hazard proximity scores to border communities in California.

2.2 Methodology

The methodology for estimating PAI and PSR in the California-Mexico border region uses the information obtained in Task 1 on the emission characteristics of the sources in Mexico together with trajectory modeling during selected transport events, as shown in Figure 2.1; while the estimation of PSR is based on methods described in the literature (see, e.g., Stohl et al., 2003; Seibert and Frank, 2004; Hüser et al., 2017). To our knowledge this is the first time that the proposed methodology for estimating PAI is presented. The following sections present the main methodological steps for estimating both PAI and PSR.

2.2.1 Air quality databases

A key part of the analysis consisted in identifying periods with wind patterns that could bring pollutants from Baja California, Mexico, northward into California. We examined persistently north-bound wind events using available observations. For this purpose, a series of databases containing relevant air pollution and meteorological data in the San Diego, Calexico and adjacent US-Mexico border communities were collected from diverse networks and institutions from both US and Mexico. These include data from the Air Quality System of EPA (AQS-EPA), the Air Quality and Meteorological Information System from CARB (AQMIS-CARB), the National Data Buoy Center (NDBC), NOAA's National Estuarine Research Reserve System (NERRS), National Center for Atmospheric Research (NCAR), and the Mexican National Weather Service (Servicio Meteorológico Nacional from Mexico, SMN-Mexico). Of these data sources, the primary datasets for the analysis were those of AQS-EPA and AQMIS-CARB. In addition, reanalysis data from NCAR was obtained for driving the meteorological model boundary conditions. Table 2.1 summarizes the datasets collected, the sampling periods and sampling frequencies. Figure 2.2 shows the location of the monitoring stations with data collected. The figure shows that the number and spatial extent of the monitoring stations with continuously valid data is much larger in California than in Mexico.

With exception of the SMN data, all datasets used were already processed and flagged with quality assurance and quality control (QA-QC) operations. We performed basic QA-QC to SMN data by removing spurious data and outliers. Our wind frequency analysis focused only on surface wind direction and wind speed and all data was homogenized to a common format.

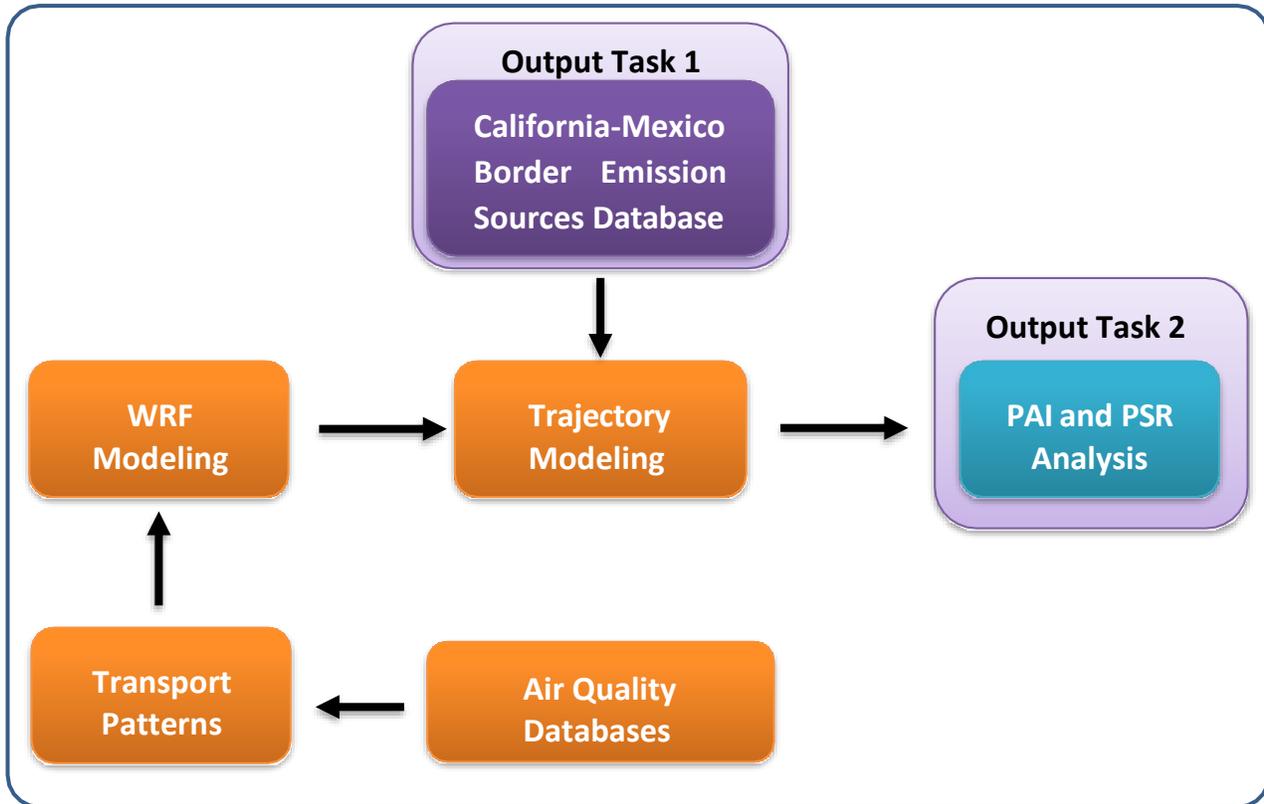


Figure 2-1. Methodology used for obtaining the Potential Area of Influence (PAI) and Potential Source Regions (PSR) using emissions data, meteorological observations and modeling.

2.2.2 Transport patterns

A prerequisite to identify the influence of major emission sources impacting sensitive areas is to understand the meteorological patterns that drive the dispersion, formation and fate of air pollutants in the region. In particular, we are interested in air pollution transport events that can potentially lead to impacts in the California side of the border. Data from monitoring stations at the San Diego-Tijuana and Calexico-Mexicali border areas were used to obtain hourly frequency analysis of wind directions during 2014. The year 2014 was selected as the modeling and analysis year based on its El Niño Southern Oscillation (ENSO) “neutral” classification and being a year with a high percentage of data completeness in the monitoring stations’ databases. The primary criterion considered in selecting the meteorological episodes was the continuous presence of southerly winds registered in locations close to the border urban areas (days with > 12 hours of southerly winds). Using this criterion, a total of 28 and 77 days with southern wind in 2014 were

identified for the San Diego-Tijuana and the Calexico-Mexicali regions, respectively. This condition allows us to focus on meteorological events that are conducive to air quality impacts in southern California urban areas. All the selected events were simulated using the WRF model.

2.2.3 WRF modeling

The WRF model (Skamarock et al., 2005) version 3.8 in a 5-domain configuration (D1-D5) with resolutions (nodes) of 45 (116 x 185), 15 (151 x 127), 3 (161 x 136), and two domains of 1 (127 x 127) km, respectively, was used to simulate the meteorology for each day of the selected episodes. The modeling domain was configured to capture the synoptic conditions that can drive the meteorological events (D1 and D2), as well as the regional (D3) and local (D4 and D5) meteorological conditions respectively (see Figure 2.3). The 1 km domains cover the main urban area of the San Diego-Tijuana region (D4), and the main urban and rural areas in the Calexico-Mexicali region (D5). The parameterizations used in this study are based on reported configurations of previous simulations of the Cal-Mex field campaign (Bei et al., 2013) and include the Purdue Lin microphysics (Lin et al., 1983; Chen and Sun, 2002), the NOAA Land Surface Model (Chen and Dudhia, 2001), the Yonsei University (YSU) scheme for the Planetary Boundary Layer (PBL) (Hong et al., 2006), the Monin-Obukov model for the surface layer (Skamarock et al., 2005), and the RRTMG schemes for the longwave and shortwave radiation respectively (Iacono et al., 2008). The Grell-Freitas convective parameterization (Grell et al., 2014) was used only in the two outermost domains (D1 and D2) and a total of 35 vertical layers. Six hourly 1-degree Global Final Analysis (FNL) were used for initial and boundary conditions. For each selected event from step 2), the model was run with a 1-way nesting configuration in 36-h simulation slots taking the first 12 hours as spin-up. To better represent the meteorological conditions, analysis nudging was included every 6 hours in the first two domains (D1 and D2) only. The model output frequency for the meteorological fields was set to 15 minutes.

The model performance was evaluated to ensure that the local meteorology was represented in the simulations with sufficient accuracy in order to simulate trajectories. The evaluation focused on the selected meteorological episodes using several statistical criteria. The data obtained from the ambient monitoring stations in the San Diego-Tijuana and Calexico-Mexicali regions were used to evaluate the wind speed and wind direction fields from the WRF model. The fields of surface radiation and PBL height were examined to ensure consistency and representativeness of the model outputs. The model had an acceptable performance in representing both wind direction and wind speed during the selected periods, indicating that the meteorological fields from the WRF model can be used to drive the trajectory modeling analyses (see appendix)

2.2.4 Trajectory Model

Trajectory models are useful tools for simulating the transport pathways of a species from their source to a specific site (Hüser et al., 2017). The FLEXPART-WRF model (Brioude et al., 2013) was used to simulate the regional and local transport processes in the San Diego-Tijuana and Calexico-Mexicali areas as well as in the surrounding regions. The parameterizations used in this study for these two sets of simulations include the calculated PBL turbulence mixing and diagnosed surface parameters such as the sensible

heat flux and friction velocity by FLEXPART and the land use from WRF. The model was run in parallel mode to improve computational performance. This model tracks the path of a tracer particle (i.e., does not undergo any chemical reaction) forward or backward in time to identify source regions of pollution or emissions and to quantify their contribution to the atmospheric composition at a specific site (Hüser et al., 2017). In the context of Lagrangian dispersion models, the so-called particles refer to infinitesimal small air parcels, and this is the terminology adopted in this study.

Table 2-1. Meteorological data obtained for analysis of transport events and the evaluation of the model performance.

Database	Data type	Sampling period	Sampling frequency	Spatial Extent
Air Quality System (AQS-EPA) Air Quality and Meteorological Information System (AQMIS-CARB)	Meteorological parameters: P, T, RH, wind	1980-2017	1 hour	Southern California
National Data Buoy Center (NDBC)	Meteorological parameters: P, T, RH, wind	1980-2017	1 hour	Southern California
NOAA's National Estuarine Research Reserve System's (NERRS)	Meteorological parameters: P, T, RH, wind	1980-2017	1 hour	Southern California
Servicio Meteorológico Nacional (SMN)	Meteorological parameters: P, T, RH, wind	2000-2016	10 minutes	Baja California
National Center for Atmospheric Research (NCAR)	Reanalysis	2014	6 hours	Global

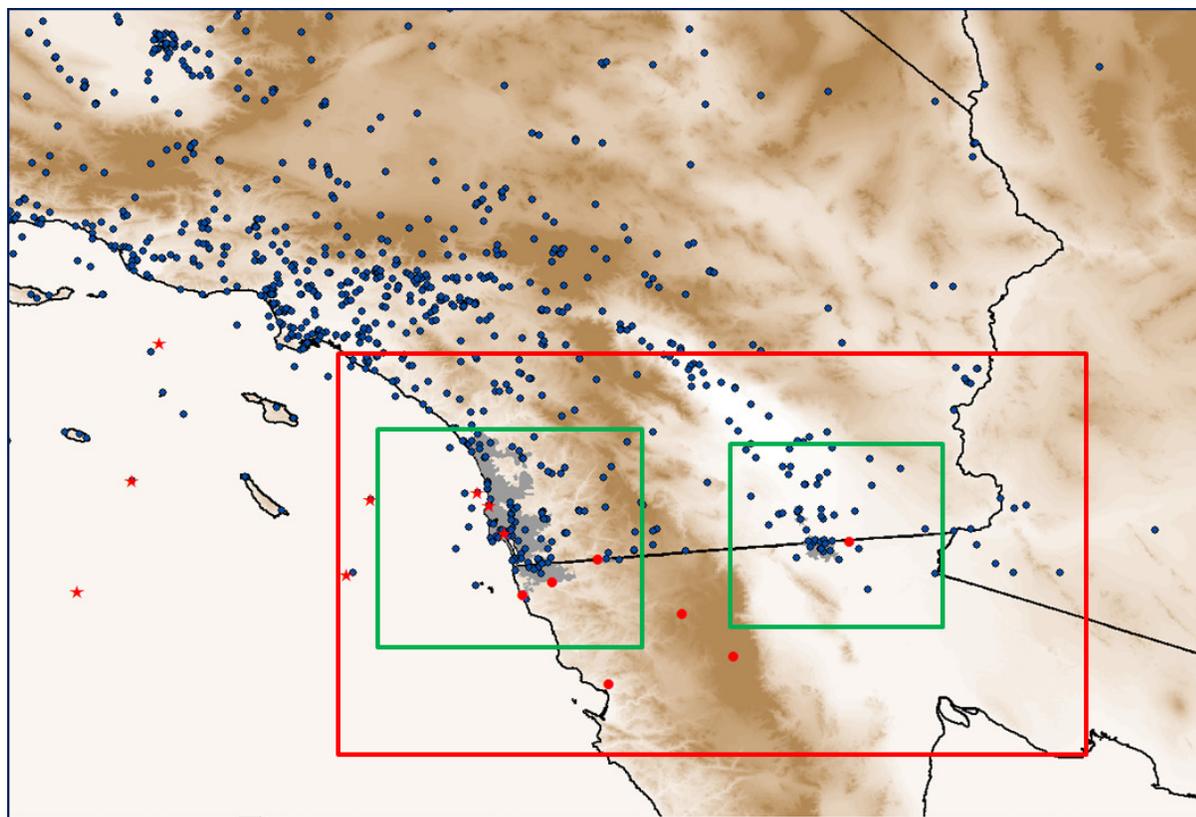


Figure 2-2. Map showing ambient monitoring stations in the California-Mexico border region. Green boxes represent the model domains 4 and 5 used for the San Diego-Tijuana and Calexico- Mexicali border regions, respectively. Blue dots represent AQS-EPA and AQMIS-CARB monitoring stations, whereas red stars and red dots represent NDBC and SMN monitoring stations, respectively. White to brown shadowing indicates low to high topography. The red box represents the output domain of backward simulations with FLEXPART-WRF.

2.2.5 PAI and PSR

Forward trajectory simulations were used to identify PAI's in the northern part of the US-Mexico border of selected key emission sources from the database developed in Task 1, whereas backward trajectory simulations were used to identify PSR's in the southern part of the US-Mexico border to identify areas with potential air quality impacts. The analysis of PAI was carried out using event- specific selected meteorological simulations, whereas PSR's were obtained considering the entire database of selected episodes of prevalent southern wind component.

2.2.5.1 Potential Area of Influence

For each of the selected key emission sources, the PAI for a given meteorological episode is estimated using the following procedure:

- a. *Forward trajectory run.* FLEXPART-WRF is run in forward mode for the selected meteorological episode using the emission characteristics for the key source. Time and three-dimensional location for each of the released particles are

continuously tracked.

- b. *Define a gridded domain.* A spatial domain and cell resolution are selected so that all the particles are included. This can be done by using the maximum and minimum east/west locations of particles and a computationally reasonable horizontal cell size. The domain is three-dimensional so the height of the column must also be selected (e.g. first 100 m above surface for considering local-to-regional transport)
- c. *Particle counting.* The number of particles per cell is counted.
- d. *Particle number fraction.* The fraction of the number of particles per cell is obtained by dividing the number of particles in each cell by the total of particles in the domain.
- e. *Accumulative fraction spatial distribution.* The accumulation is done by first sorting from high-to-low the particle number fractions per cell obtained in the previous step, and then by consecutive summation. Note that the spatial location identity of the cells remains intact in this step. The result is an accumulative spatial distribution of the particle number fractions in the domain.
- f. *Define PAI fraction criteria.* Selecting an accumulative particle number fraction to identify the cells that together account for a given percentage of the released particles. For example, an accumulative particle number fraction of 0.5 will allow identifying the first cells that together account for 50 % of the particles in the domain.
- g. *Obtain the PAI.* The PAI is obtained by counting the cells that together account for the selected accumulative fraction number and multiplying it by the cell spatial resolution.

Figure 2-3. Domains used in the WRF model to capture synoptic (D1 and D2), regional (D3), and local (D4 and D5) meteorological conditions. Domains 4 and 5 (green squares) cover the San Diego-Tijuana and Calexico-Mexicali border regions, respectively.

2.2.5.2 Potential Source Region

For each of the proposed receptors of both border regions, the PSR is estimated for the entire set of episodes using the following procedure:

- a) *Backward run.* FLEXPART-WRF is run in backward mode for all receptors in a predefined simulation period of 24 hours. All released particles from the ensemble are tracked individually as the retroplume is transported in the simulation.
- b) *Residence Time.* We obtain the time that the particles reside in every grid cell and in all vertical layers of the output grid (~1.5 km in resolution).
- c) *Total residence time.* The resulting 4D fields (space + time) of residence time are integrated in the time dimension in the footprint layer for each receptor site, to calculate the accumulated response function for all the meteorological episodes. The footprint layer is set as 100 m agl (meters above ground level).
- d) *Visualization.* Plot the resulting 2D field of total residence time in a logarithmic scale.
- e) *Identify the PSR.* Select a color map to indicate regions of enhanced values of the

response function, so that higher values of residence time suggest a potential greater contribution from that/those region(s).

2.3 Results and Discussion

2.3.1 Selected Meteorological Episodes

The analysis was focused on the monitoring stations located within modeling domains 4 and 5 that correspond to the San Diego-Tijuana and Calexico-Mexicali border regions, respectively, for the year 2014. Figure 2-4 shows the results of the frequency analysis for the selection of meteorological episodes with predominant southern ($112.5^{\circ} - 247.5^{\circ}$) wind direction. The number of hours per day that have a southern wind direction was used as the criterion for selecting the modeling episodes. Using a threshold of minimum 12 hours of wind data within each 24-hour period, 21 (totaling 28 days) and 35 (totaling 77 days) southern wind episodes in 2014 were identified for the San Diego-Tijuana and the Calexico-Mexicali regions, respectively. The figure shows that southern wind episodes are more common during the summer than during the winter in the Calexico-Mexicali region. In the San Diego-Tijuana region, southern wind episodes are more frequent during the winter and spring, and are not common during summer. Appendix A includes a description of the selected transport events and the evaluation of the model performance.

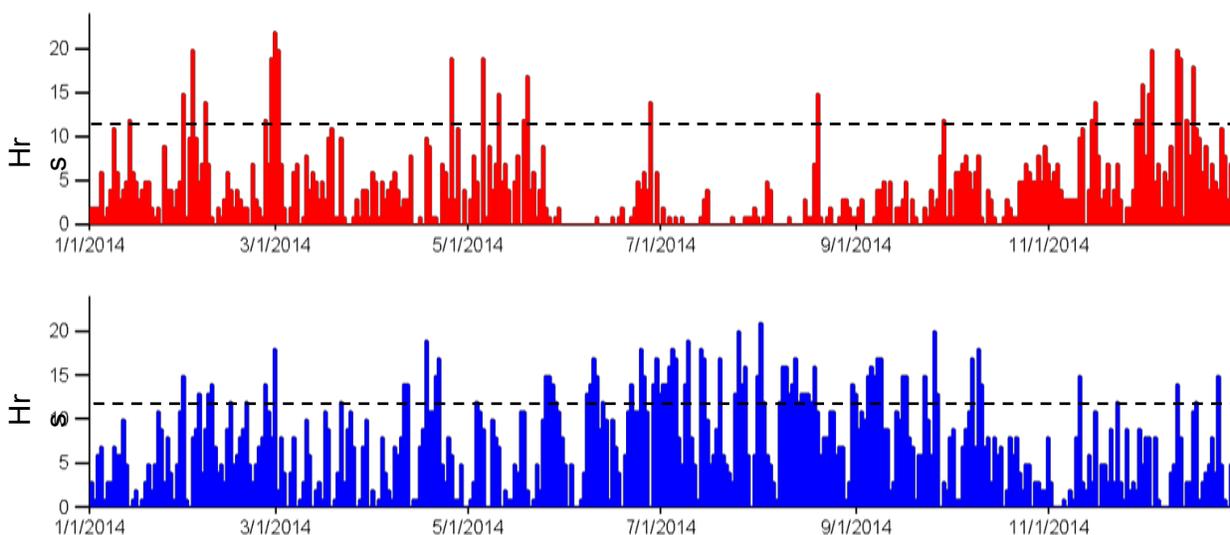


Figure 2-4. Daily number of hours with southerly wind direction in the San Diego-Tijuana (top) and Calexico-Mexicali (bottom) regions during 2014. Horizontal dashed lines represent a 12-hour threshold used for the selection of episodes.

The selected episodes were simulated using the WRF model and the FLEXPART-WRF model with a nested configuration driven by the regional (domain D3) and local (domains D4 and D5) time-averaged wind fields from WRF, to simulate the transport pathways of a generic passive tracer instead of a particular chemical species. Due to the contribution of low level winds in determining the plume transport direction (Bei et al., 2013), an

ensemble of ~ 2000 particles was released every hour for each source (forward mode) or receptor (backward mode). The output domain included 6 vertical levels up to 3000 m to represent the PBL height in the California- Mexico border based on the results of the Cal-Mex 2010 field campaign (Bei et al., 2013). Since the focus of the analysis is on short-term local dynamics, the FLEXPART-WRF model was run for each day of the selected episodes using the 15-minute output fields from WRF. For example, if one selected episode consisted of 3 consecutive days the model was run 3 times (one for each day) instead of a single 3-days continuous run, in order to reduce uncertainty as the particles disperse over the domain.

2.3.2 Potential Areas of Influence

Emission sources were selected on the basis of their emission strength reported in the database from Task 1, their location, and source type. As shown in Figure 2.5, the selected sources are located throughout the spatial extent of the Tijuana and Mexicali urban areas, and include sources near the US-Mexico border. Selected source types included power plants, major industries, maquiladoras, Ports of Entry (POE), airports, cattle feedlot and brick kilns. The following 12 sources in the San Diego-Tijuana border region were selected for estimating their PAI: “Presidente Juarez” (Rosarito) power plant, an electronics manufacturer industry, a metal mechanic industry, the San Ysidro POE, the Otay POE, a chemical processing industry, a textile industry, the airport, a maquiladora, a large cattle feedlot, a brick kiln located near the San Diego-Tijuana border (brick kiln 1), and a brick kiln located about 10-km south of the San Diego-Tijuana border (brick kiln 2). For the Calexico-Mexicali border region, 8 sources were selected: a maquiladora, the Calexico West POE, a glass manufacturer industry, the airport, a paper industry, the Cerro Prieto geothermal plant, a brick kiln located near the Calexico-Mexicali border (brick kiln 1), and another brick kiln located farther from the border (brick kiln 2). The airports and POE were treated as line sources, the Cerro Prieto geothermal plant as a small area source and the rest as point sources.



Figure 2-5. Location of emission sources in the Tijuana (left) and Mexicali (right) regions selected for estimating their potential area of influence.

The model was run in forward mode to obtain the dispersion dynamics for the set of selected sources and to understand their relation to the regional transport conditions. Forward simulations are computationally more demanding than backward simulations (Stohl et al., 2003). For this reason, this set of simulations focused on two specific days, aiming to represent a range of low and high wind speed conditions of the selected meteorological events for both the San Diego-Tijuana (February 2nd and February 28th) and the Calexico-Mexicali (October 9th and July 25th) regions, respectively. The selected low and high wind speed days for the San Diego-Tijuana basin presented 20 and 22 hours with southerly wind components, respectively. For the Calexico-Mexicali basin, the selected low and high wind speeds presented 18 and 20 hours with southerly wind components, respectively. High and low wind speed averages for these periods were 6.4 and 2.9 m/s for San Diego-Tijuana and 3.8 and 1.6 m/s for Calexico-Mexicali, respectively. We defined an emission profile (8 AM to midnight LST) to focus only on diurnal emissions. During this period, the particles were released from the surface up to 100 m agl at a constant emission rate. Night emissions (midnight to 8 AM LST) were not considered in the analysis. Daytime analysis was prioritized versus nighttime analysis for better estimating the larger (regional) spatial extent of the impacts as surface wind speeds are usually higher during daytime than during nighttime. However, measurements of black carbon during Cal-Mex 2010 campaign indicated the presence of combustion sources during nighttime, suggesting clandestine industrial activities (Shores et al., 2013). Therefore, it might be of interest for future studies to include nighttime burning emissions in the analysis of local impacts from Mexican sources.

Figure 2-6 shows an example of the spatial extent of the PAI for the power plant in Rosarito (“Presidente Juarez” thermoelectric plant) and the brick kiln source (brick kiln 2) located farther from the San Diego-Tijuana border. The figure shows the estimated PAI during the high persistent northward wind event of February 28, 2014. Similar spatial distributions of PAI for the other sources were obtained for both high and low speed wind events (not shown) for San Diego-Tijuana and Calexico-Mexicali border regions. The PAI for the Rosarito power plant, which is located close to the shore, passes over the coastal side of the Tijuana and San Diego urban areas and extends far north over large parts of northern San Diego communities towards Escondido and Riverside. The PAI of the brick kiln emissions extend over a narrower northeast path overpassing the Laguna mountain ridge and rural areas without impacting any large urban community. The results show that the potential areas of influence of both emission sources are relatively large, indicating that local emissions in the Mexican side of the border can travel large distances and contribute to regional air quality impacts.



Figure 2-6. Example of potential areas of influence obtained from forward trajectories for the Rosarito power plant and the “Brick kilns” sources in Tijuana during the high wind speed episode of February 28, 2014.

Table 2-2 summarizes the results of the PAI estimated for the selected key sources during high and low southern wind speed episodes in 2014. The PAI shown correspond to the spatial extent (in km²) of the areas impacted in the US side of the border (impacts on the Mexican side of the border were not added) for each particular source. The use of the high and low wind speed episodes allows estimating the potentially maximum and minimum PAI for each region in the US side; however, other PAI could be estimated under different meteorological conditions (for example, those conducive to impacts in the Mexican side of the border). PAI were much higher for the San Diego- Tijuana region than for the Calexico-Mexicali region because the average wind speeds during the selected episodes were much higher in the former than in the latter. PAI also showed different variability among wind episodes: during high wind speed episodes, which are often driven by large non-local persistent synoptic conditions, the magnitude of the PAI was considerably less than during low wind speed events in which wind transport is largely driven by local dynamics.

Table 2-2. Estimated PAI [km²] for California side only for selected key sources in the San Diego- Tijuana and Calexico-Mexicali border regions during high and low southern wind speed episodes.

San Diego–Tijuana			Calexico–Mexicali		
Source	High wind speed	Low wind speed	Source	High wind speed	Low wind speed
Rosarito power plant	1866	414	Paper industry	861	236
Electronics manufacturer*	1794	771	Airport	850	381
San Ysidro POE*	1748	853	Maquiladora*	755	257
Metal mechanic	1733	763	Cerro Prieto geothermal	672	0
International airport*	1723	770	Glass industry	633	99
Chemical industry	1717	469	Brick kiln 1*	530	373
Feedlot (livestock)	1682	20	Calexico west POE*	524	351
Textile Industry	1662	664	Brick kiln 2	195	98
Otay POE*	1626	645			
Brick kiln 1	1572	497			
Maquiladora*	1465	547			
Brick kiln 2	1196	234			

*These sources are within the 1.5-km buffer from the California-Mexico border.

The results indicate that the selected emission sources can have large local and regional impacts and that the spatial extent of the impacts depends strongly on the location of the source. For example, the Cerro Prieto geothermal plant had a relatively large PAI of ~670 km² during the high wind speed episode, but during the low wind speed event it did not impact any region in the US side of the border. Being farther away from the border (see Figure 2-5), all the emissions from the geothermal plant remained in the Mexican side of the border for that event. A similar result can be observed for the estimated PAI for the livestock feedlot located further south of the US-Mexico border. Nevertheless, all sources located along the 1.5-km buffer of the US-Mexico border had relatively large PAI, even during the low wind episodes.

2.3 Potential Source Regions

The model was run in backward mode to identify potential air masses arriving at selected receptor sites for both border regions. The rationale is that an ensemble of particles is released from the specified receptor location, transported backward in time and a response function is calculated (Stohl et al., 2005). In the context of Lagrangian dispersion models, this function is also known as source-receptor relationship or Potential Emission Sensitivity (PES) (Gadhavi et al., 2015). However, purposes of clarity, in this study we use the term ‘response function’, which maps the relative contribution in space and time of emission sources and is related to the residence time of the particles in the output grid cells weighted by the atmospheric density (Stohl, et al., 2005; Brioude et al., 2012; Hüser et al., 2017). Since we are interested in surface emissions, only the particles passing a vertical layer adjacent to the ground are considered in identifying the source regions. In the context of Lagrangian dispersion models, this layer is the footprint layer. Thus, the longer the time the trajectories reside in the footprint layer (high values of the response function), the more material (emissions) could be transported downwind towards the receptor site (Hüser, et al., 2017). This specially occurs near the surface since most anthropogenic emissions are released at lower heights (Stohl, 2007). The height of the footprint layer is usually set to 100 or 300 m agl (Hüser, et al., 2017, Stohl, et al., 2003). We defined the footprint layer to 100 m agl since it includes typical emission heights of mobile, area and point sources in the border region, and also because it is the minimum PBL height used in trajectory models (Hüser, et al., 2017). The ensemble of particles was released from the surface to a height up to 10 m. The output grid consisted of a rectangular domain of ~ 1.5 km of spatial resolution (see Figure 2-2).

In contrast to the simulations in forward mode, in the backward mode the model was run for all the selected meteorological episodes of both regions: 28 days from 10 different months for San Diego-Tijuana and 77 days from all 12 months for Calexico-Mexicali. Three receptor sites were selected for each region to represent border communities in California and locations relatively far from the border (see Table 2-3).

Table 2-3. Receptor locations and selected events for estimating PSR.

San Diego–Tijuana		Calexico–Mexicali	
Receptors	Events in 2014	Receptors	Events in 2014
San Ysidro; Chula Vista; El Cajón.	01/13; 01/30	Calexico; El Centro; Brawley.	01/30
	02/02; 02/06; 02/25; 02/27 - 03/01		02/04; 02/07 - 02/08; 02/14; 02/19; 02/25; 02/28
	04/25		03/21
	05/05; 05/10; 05/18 - 05/19		04/10 - 04/11; 04/17; 04/20 - 04/21
	06/27		05/03; 05/25 - 05/28
	08/19		06/07 - 06/10; 06/12; 06/21; 06/24 - 06/25
	09/28		06/28 - 07/05
	11/14 - 11/15; 11/28 - 11/29; 11/30		07/08 - 07/09; 07/13 - 07/14; 07/19
	12/02 - 12/03; 12/11 - 12/12; 12/14; 12/16; 12/30		07/24 - 07/27; 07/31 - 08/02
			08/07 - 08/18; 08/30 - 08/31
	09/04 - 09/08; 09/15 - 09/16; 09/22; 09/26		
	10/09 - 10/10		
	11/10; 11/22		
	12/11; 12/17; 12/24		

2.3.1 PSR in the San Diego–Tijuana region

Figure 2-7 shows the accumulated particle residence time (logarithmic scale) spatial distributions in the footprint layer for all of the selected meteorological episodes in the San Diego–Tijuana region. High residence times indicate more emission sensitivity and hence a higher potential contribution of a source region. The results indicate that the bulk of the retroplumes remained relatively near to the 3 receptor sites but they also spanned over a wider area encompassing the US-Mexico border and the Pacific Ocean. The spatial distribution of the response function shows the main geographical pathways over which any emissions could have been transported upwind to the receptors.

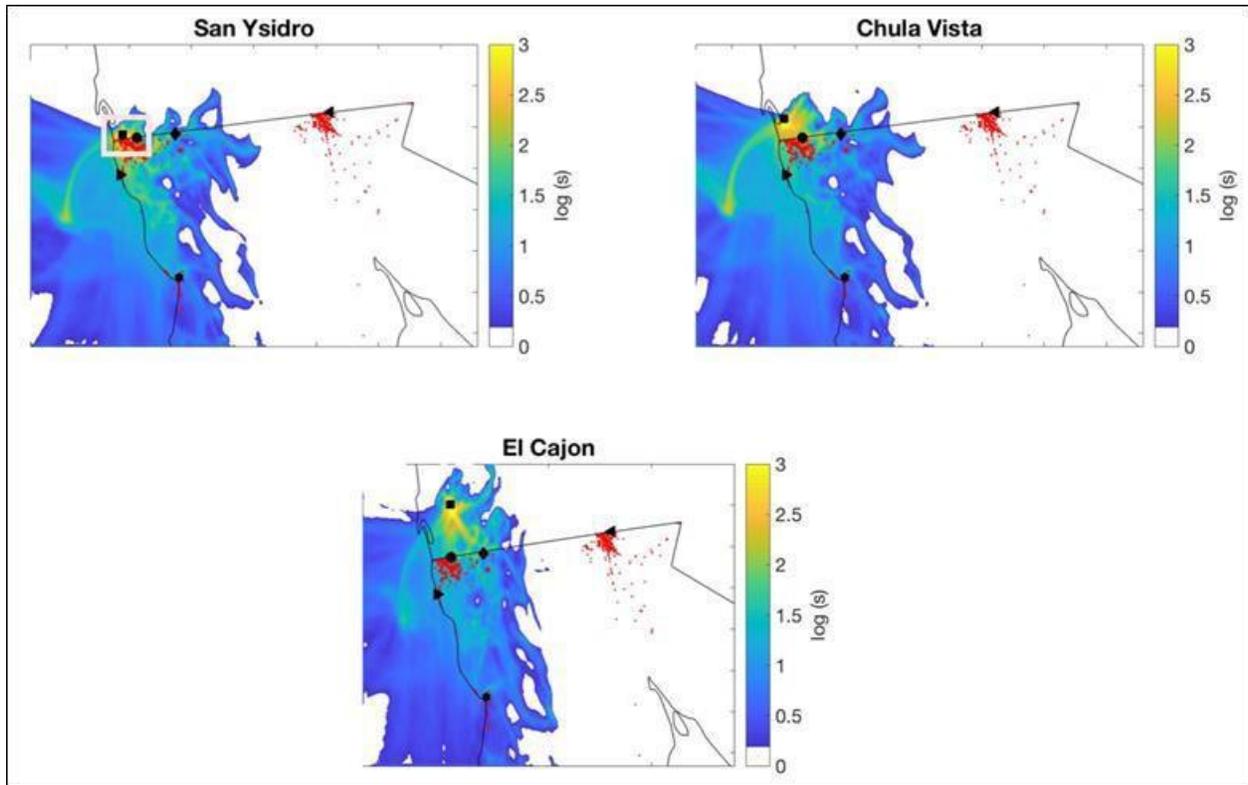


Figure 2-7. Accumulated residence time distributions in the 100-m footprint layer during the entire meteorological episodes (28 days) for the 3 selected receptors in the San Diego–Tijuana region: San Ysidro (upper left), Chula Vista (upper right) and El Cajon (bottom). The filled markers represent the location of each receptor (square); Otay POE (circle); Tecate (diamond); Rosarito (right-pointing triangle); Mexicali (left-pointing triangle) and Ensenada (star). The red dots show the location of stationary emission sources for both border regions obtained in Task 1 and the small white box on the left panel denotes the zoomed area of the US-Mexico border for Figure 2.8. The solid black line indicates the political boundaries.

The spatial distribution of residence times in the Mexican side vary for each of the US receptor locations, indicating that the local transport within the region greatly determines the impacts of emission sources. Local transport depends strongly on the interactions between regional air masses, topography and land use. The estimated PSR suggests large contributions from the Tijuana urban area (high residence time) to the San Ysidro receptor site and to locations in the US side, including near-border communities (see Figure 2.8). The transport pathways coincide with the Tijuana urban area so that emissions released from sources along those pathways (red dots), including the Otay POE, are likely to impact regions in the US side. In addition, emissions from Tecate and farther to the east, including potential emissions from the mountain regions (medium residence time), can be carried along that pathway and eventually mix with those from the Tijuana urban area. The emissions from Rosarito, including the power plant, can be

transported also along the west coast and eventually mix with those from Tijuana. Because of its emission height, the emissions from the Rosarito power plant can reach higher layers aloft and contribute also to regional and local impacts. The results also indicate potential contributions from ocean and coastal regions, suggesting that ship emissions from Ensenada and San Diego can contribute to regional impacts.

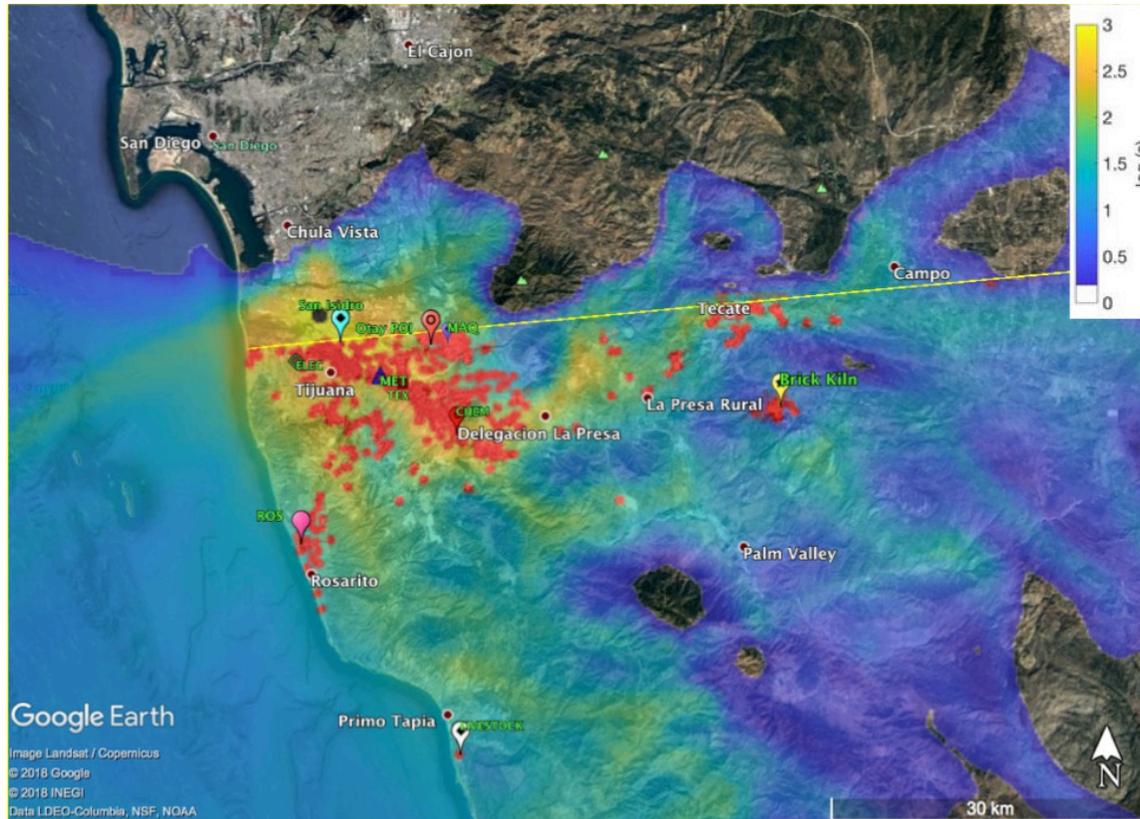


Figure 2.8. Zoom (of Figure 2.7) over the US-Mexico border region showing the transport pathways for the San Ysidro receptor. The red dots represent the emissions sources obtained in Task 1. Several industries from representative sectors are shown: Chemical (CHEM); ELEC (Electronics); MAQ (Maquiladora); MET (Metal); Otay (POE); ROS (Rosarito power plant) and TEX (textile). The logarithmic color scale from Figure 2.7 is overlaid onto the satellite plot of the region: high residence time (yellow); medium residence time (green); low residence time (blue).

The results indicate that the receptor site in Chula Vista is likely impacted mainly by local emissions in the US side, but still receives potential contributions from source regions in Tijuana, Tecate, Rosarito and the Otay POE. In addition, the transport pathways from southern areas in the ocean are also present for this receptor site so that ship emissions (but likely minimal from Ensenada) can also impact Chula Vista. For regions farther to the north, represented by the El Cajon receptor site, the transport pathways are different from the pathways for the two other receptors. The PSR results suggest that the emission source regions in the US side close to the El Cajon receptor site can have a greater contribution to the atmospheric composition than border regions near

the Mexican side. However, PSR south of the receptor site are not negligible and emissions from Tecate and Otay POE can still be transported and contribute to the local impacts. In particular, emissions from Tecate can be transported westward to regions closer to the border, and also farther towards northern areas of San Diego County. In addition, the contributions of emissions from regions over the ocean to the northern receptor site are smaller than to the other two receptors.

The results strongly suggest that areas in the US side close to the border and farther north in San Diego County are likely to be affected not only by emission sources in the Mexican side within the 1.5-km buffer of the border, but also from potential regions both inland and off-shore within a 50-km buffer. However, the regional impacts of sources located farther south from the border might be more relevant for San Diego than for Calexico, since south of Mexicali there are mostly agricultural areas, although these may be a source of agricultural burning emissions.

2.3.2 PSR in the Calexico–Mexicali region

Figure 2.9 shows the accumulated particle residence time (logarithmic scale) in the footprint layer for all of the selected meteorological episodes in the Calexico-Mexicali region. High residence times (yellow shade) indicate more emission sensitivity and hence a higher potential contribution of source regions in the area, because the longer the time the trajectories reside in the footprint layer (high values of the response function), the more material (emissions) could be transported downwind towards the receptor, especially near the surface (Hüser, et al., 2017; Stohl, 2007).

In contrast to the spatial distribution of residence time for the San Diego–Tijuana region, the associated residence time of the PSR in the Calexico-Mexicali region remained substantially wide near the 3 receptor sites and spanned greater areas, encompassing the US-Mexico border and mountain ridges to the west as the retroplume dispersed backward in time. Thus, the particle residence time suggests a more regional dynamics and more homogeneous geographical pathways to transport any emissions from Mexicali upwind to the receptor sites in Calexico.

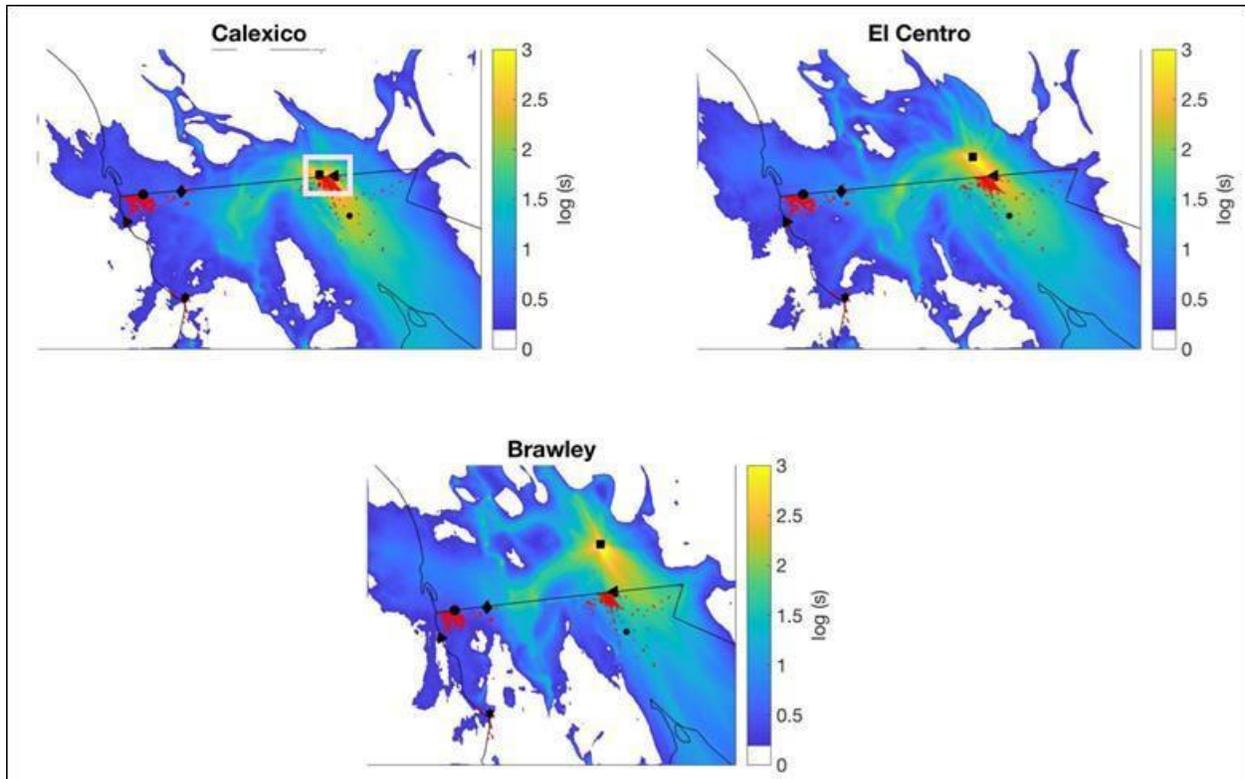


Figure 2-9. Accumulated residence time distribution in the 100 m footprint layer during the entire meteorological episodes (77 days) for the 3 selected receptors in the Calexico - Mexicali region: Calexico (upper left); El Centro (upper right) and Brawley (bottom). The filled markers represent the location of each receptor (square); Otay POE (circle); Tecate (diamond); Rosarito (right- pointing triangle); Mexicali (left-pointing triangle) and Cerro Prieto geothermal plant (circle). The red dots show the location of emission sources for both border regions obtained in Task 1 and the small white box on the left panel denotes the zoom area of the US-Mexico border for Figure 2-10. The solid black line indicates the US-Mexico border.

The results indicate that the spatial distribution of the PSR in the Mexican side vary for each of the receptor locations in the US side of the border; however, the transport within the region is more regional in extent. For the Calexico site, the closest receptor to the border, the results suggest large and direct contributions from the Mexicali urban area to border communities (high residence time, yellow shade). The transport pathways coincide with the Mexicali urban area so that emissions released from sources along those pathways (red dots), including the Mexicali POE, are likely to impact regions in the US side. The results also suggest that emissions from the Cerro Prieto geothermal plant and farther to the south can be transported along that pathway and eventually mix with those from the Mexicali urban area, further impacting the border communities. Interestingly, the results also suggest potential contributions from regions west of the mountain ridge in the southwest rural regions (see Figure 2.10).

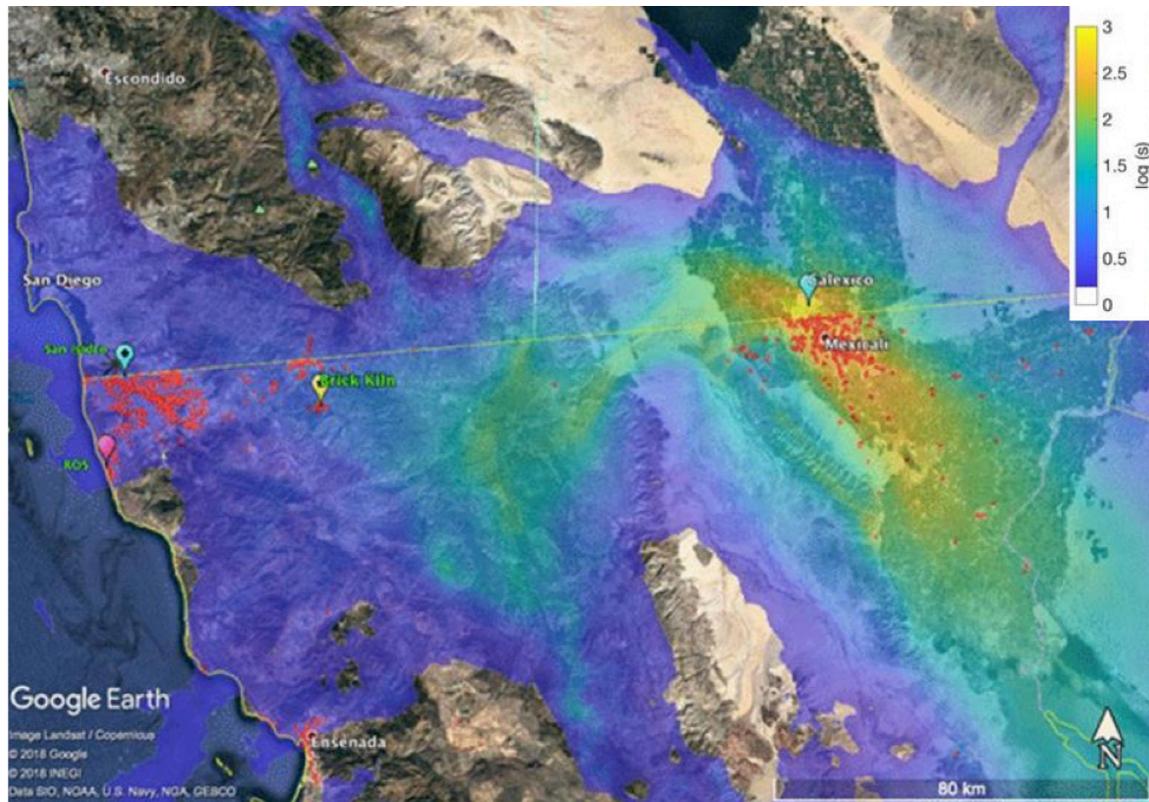


Figure 2.10. Zoom (of Figure 2.9) over the US-Mexico border region showing the transport pathways for the Calexico receptor (blue). The red dots represent the emissions sources obtained in Task 1. The markers show the location of some of industries from representative sectors: ROS (Rosarito power plant) and brick kiln (yellow). The logarithmic color scale from Figure 2.9 is overlaid onto the satellite plot of the region: high residence time (yellow); medium residence time (green); low residence time (blue).

The main characteristics of the estimated PSR are similar for the receptor sites of El Centro and Brawley. Local emissions largely contribute to the impacts at these locations but contributions from source regions in the Mexican side (including sources in the Mexicali urban area and the Cerro Prieto geothermal plant) are not negligible. Even though the WRF model tended to over- predict the magnitude of northwesterly flow (see Appendix A), the results show that emissions from rural regions to the north and northwest of these receptors (a region of high agricultural activity) as well as areas west to the Rumorosa mountain ridge can also be transported to Calexico and El Centro, therefore could have a large contribution to the air pollution impacts in the area.

2.4 Conclusions

The potential areas of influence and potential source regions impacting areas in San Diego, Calexico and adjacent California-Mexico border communities have been identified in Task 2 using a combination of meteorological and trajectory modeling and focusing on meteorological events conducive to air quality impacts in southern California communities. The year 2014 was selected for the analysis based on the “neutral” ENSO classification and data availability. Most of the selected meteorological episodes were in the winter season for the San Diego–Tijuana region, and in the mid-spring and fall

seasons for the Calexico–Mexicali region. The predominant wind directions in the Calexico-Mexicali region were from the southeast with a small influence from the northwest. In the San Diego-Tijuana region, air masses presented a wider range of wind directions spanning from northeast to south-southwest.

Potential areas of influence were obtained for a subset of key sources selected from the emission sources database compiled from Task 1 using a new methodology. The project focused on sources located at different distances from the California-Mexico border and with high emissions rates in the Tijuana and Mexicali regions, including brick kilns, international ports of entry, electric generation facilities, maquiladoras, and other industries. PAI's were estimated by calculating the area in the US side of the border containing a pre-selected amount of the initial emission released at the locations of each emission source in the Mexican side of the border.

For the determination of potential source regions, three receptor sites in the northern side of the US-Mexico border were selected in both San Diego (San Ysidro, Chula Vista and El Cajon) and Imperial Valley (Calexico, El Centro and Brawley) regions. The corresponding PSR were determined by accumulating the residence time of all backward trajectories in the footprint layer during the entire set of meteorological events with a prevailing southern wind component. Using this method, the locations with high values of residence time can be interpreted as potential source regions that are more likely to impact the selected US receptor sites.

The spatial distribution of the PAI depended strongly on the location of the emission source and the wind conditions, indicating that local transport greatly influences the extent of the impacted areas. The results of both the PAI and PSR indicate that emission sources in Tijuana and Mexicali can potentially have large local and regional impacts. In particular, border communities and surrounding areas can be impacted from sources located within the 1.5-km buffer of the US- Mexico border. It should also be noted that during selected transport episodes, these communities can also be potentially impacted from sources within 50 km and farther south from the border. For the San Diego–Tijuana region, the spatial distributions of the potential transport pathways for the identified PSRs were highly heterogeneous due to local dynamics. In contrast, the PSR in the Calexico–Mexicali region were spatially homogenous due to a larger influence of regional dynamics.

Additional research can further benefit the understanding of the impacts of local emissions in the California-Mexico border region on the local communities. Border communities are often low income and minority and this research can help to identify those that are disproportionately vulnerable to multiple sources of air pollution, contributing to improving the ranking scores of the CalEnviroScreen tool and environmental justice. The additional suggested research can be used towards investigating the impacts of different meteorological conditions and additional sources, including: 1) the identification of potential impacts from key sources that were not considered as part of Task 1, especially agricultural burning and dust resuspension emissions; 2) exploratory analysis from Task 1 on burning emissions has identified substantial burning activities in the urban area of Mexicali, and the tools developed in this study can be further applied to better understand the impacts from these sources; 3) extending the analysis to different modeling baseline years that include La Niña and El

Niño events could also help in understanding the variability of the impacts under additional meteorological conditions; 4) the impacts of shorter duration but intense southerly wind events on border communities could also be investigated; 5) measurements obtained during the Cal-Mex field campaign suggested the presence of intense nighttime combustion sources, and this analysis could be extended to include night-time hours; 6) the potential source regions identified in this study could be combined with gridded emissions inventories to estimate the relative contributions from emission sources in those areas.

2.5 Acknowledgments

The following institutions have provided data used in Task 2 of this study: Air Quality System of EPA (AQS-EPA), Air Quality and Meteorological Information System from CARB (AQMIS- CARB), National Data Buoy Center (NDBC), NOAA's National Estuarine Research Reserve System (NERRS), National Center for Atmospheric Research (NCAR), and Servicio Meteorológico Nacional of Mexico (SMN-Mexico). In addition, Mexico Ministry of the Environment and Natural Resources (SEMARNAT) has provided the Mexico National Emission Inventory and Baja California Emission Inventory used in Task 1.

2.6 References

- Bei, N. F., Li, G. H., Zavala, M., Barrera, H., Torres, R., Grutter, M., Gutierrez, W., Garcia, M., Ruiz-Suarez, L. G., Ortinez, A., Guitierrez, Y., Alvarado, C., Flores, I., Molina, L. T.: Meteorological overview and plume transport patterns during Cal-Mex 2010. *Atm. Environ.*, 70, 477-489, 2013.
- Brioude, J., Angevine, W. M., McKeen, S. A., Hsie, E.-Y.: Numerical uncertainty at mesoscale in a Lagrangian model in complex terrain, *Geosci. Model Dev.*, 5, 1127-1136, <https://doi.org/10.5194/gmd-5-1127-2012>, 2012.
- Brioude, J., Arnold, D., Stohl, A., Cassiani, M., Morton, D., Seibert, P., Angevine, W., Evan, S., Dingwell, A., Fast, J.D., Easter, R.C., Pisso, I., Burkhardt, J., Wotawa, G.: The Lagrangian particle dispersion model FLEXPART-WRF version 3.1. *Geosci. Model Dev.*, 6, 1889-1904. doi:10.5194/gmd-6-1889-2013, 2013.
- Chen, F. and Dudhia, J.: Coupling an advanced land-surface/hydrology model with the Penn State/NCAR MM5 modeling system. Part I: Model description and implementation, *Mon. Weather Rev.*, 129, 569–585, 2001.
- Chen, S. H. and Sun, W. Y.: A one-dimensional time dependent cloud model. *J. Meteor. Soc. Jpn.*, 80, 99–118, 2002.
- Draxl, C., Hodge, B.M., and Clifton A.: Overview and meteorological validation of the wind integration national dataset toolkit., National Renewable Energy Laboratory, Technical Report, USA, 2015.
- Fischer, M. and Jeong, S.: Inverse modeling to verify California’s greenhouse gas emission inventory, CARB and Cal-EPA report, USA, 2013.
- Gadhavi, H. S., Renuka, K., Ravi Kiran, V., Jayaraman, A., Stohl, A., Klimont, Z., and Beig, G.: Evaluation of black carbon emission inventories using a Lagrangian dispersion model – a case study over southern India, *Atmos. Chem. Phys.*, 15, 1447-1461, <https://doi.org/10.5194/acp-15-1447-2015>, 2015.
- Grell, G. A. and Freitas, S. R.: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos. Chem. Phys.*, 14, 5233–5250, doi:<https://doi.org/10.5194/acp-14-5233-2014>, 2014.
- Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, *Mon. Weather Rev.*, 134, 2318–2341, 2006.
- Hüser, I., Harder, H., Heil, A., Kaiser, J. W.: Assumptions about footprint layer heights influence the quantification of emission sources: a case study for Cyprus, *Atmos. Chem. Phys.*, 17, 10955- 10967, <https://doi.org/10.5194/acp-17-10955-2017>, 2017.
- Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, 113, D13103. doi: 10.1029/2008JD009944, 2008.
- Lin Y.-L., Farley, R. D., and Orville, H. D.: Bulk parameterization of the snow field in a

cloud model, *J. Appl. Meteorol.*, 22, 1065–1092, 1983.

Seibert, P. and Frank, A.: Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, *Atmos. Chem. Phys.*, 4, 51-63, <https://doi.org/10.5194/acp-4-51-2004>, 2004.

Shores, C., Klapmeyer, M. E., Quadros, M. E., Marr, L. C.: Sources and transport of black carbon at the California–Mexico border. *Atmospheric Environment*, 70, 490-499, 2013.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 2, NCAR Technical Note, NCAR/TN- 468+STR, 8 pp, 2005.

Stohl, A., Forster, C., Eckhardt, S., Spichtinger, N., Huntrieser, H., Heland, J., Schlager, H., Wilhelm, S., Arnold, F., and Cooper, O.: A backward modeling study of intercontinental pollution transport using aircraft measurements, *J. Geophys. Res.*, 108(D12), 4370, doi: 10.1029/2002JD002862, 2003.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, *Atmos. Chem. Phys.*, 5, 2461-2474, <https://doi.org/10.5194/acp-5-2461-2005>, 2005.

Stohl, A., Burkhart, J. F., Eckhardt, S., Hirdman, D., Sodemann, H.: An integrated internet-based system for analyzing the influence of emission sources and atmospheric transport on measured concentrations of trace gases and aerosols, Tech. rep., Norwegian Institute for Air Research, Kjeller, Norway, 2007.

CHAPTER 3 - BORDER COUNTY ENVIRONMENTAL JUSTICE SCREENING MODEL SCORES AFTER INCLUSION OF SOURCES IN MEXICO WITHIN 1.5 KM OF INTERNATIONAL BORDER

3.1 Introduction

The focus of Task 3 was to create geospatial data files of the facilities and emissions sources identified and mapped in the first two tasks of this project, and examine the impact of this new facilities and source information on scores for census tracts near the US-Mexico border as determined by the screening methods CalEnviroScreen (CES) and the Environmental Justice Screening Method (EJSM). Although these two screening methods are similar with results showing a very similar pattern of disparate impact and vulnerability, some key differences in the data metrics used and how scores are calculated results in differences in tract scores at the local level.

In addition, this Task considered data quality, such as facility type and emissions characteristics, for the facility and source data available for Mexico, to help identify data gaps or other data characteristics that might limit their use in the screening methods, and what might be required to make fully equivalent data for each screening method.

At the time this project was developed, and as of the start of work on the project in July 2016, the design of the data collection effort was focused on version 2 of CalEnviroScreen. OEHHA released a draft of CES Version 3 in September 2016 and a final version in January, 2017. This updated version included significant differences and improvements in characterizing pollution sources and in the scoring metrics used by the tool, with particular focus on incorporating information on cross-border effects on those southern tracts (OEHHA, 2017a).

These changes included sources of air toxics within 49 kilometers south of the California–Mexico border by incorporating them into the US EPA Risk Screening Environmental Indicators (RSEI) modelling of ambient pollution concentration. This required an independent contractor to OEHHA to use facilities from Mexico’s national emissions database RETC (Mexico Registry of Emissions and Pollutant Transfer) into the RSEI model. This CES update also considered traffic volume at border crossings and on Mexican roads close to the California–Mexico border, to update the CES Traffic Density metric, and results from new air monitors placed in the area to provide improved estimates of PM_{2.5} Diesel Particulate Matter (DPM) and Ozone.

These improvements to CES presented unanticipated changes for this project, as the approved technical plan called for work that would replicate some of the work that OEHHA had essentially already completed for CES. This has required that we be flexible and forward looking in producing data useful to the next version of CES. To respond to this challenge, we worked closely with OEHHA staff on the CES team to keep them informed on our progress and plans, and to seek their ideas and guidance on how the data collection work in this project could best inform OEHHA and provide potential data for the next release of CES. The OEHHA CES team was informed of our progress and participated in several quarterly meetings discussed the types of data collected and considered which data developed by this project could be used in CES4.

3.2 Materials and Methods

3.2.1 Environmental Justice Screening Method (EJSM)

The current version of EJSM was delivered to CARB in late 2015 and includes measures of proximity of population to sources of air toxics emissions in its exposure scoring, providing a way for facility and source locations identified in this project to be used directly. The EJSM Hazard Proximity Score is one of several metrics that contribute to the overall screening score generated by the EJSM, so changes in this value have limited potential to change the final EJSM Score. For this reason, our analysis focuses on how including these new Mexican facilities affect the Hazard Proximity Score alone

The measured distance between hazardous facilities/uses and locations to be scored are used in a calculation of the hazard proximity metric, as shown schematically in Figure 3-1, below.

DISTANCE-WEIGHTED HAZARD PROXIMITY

Each CI poly (land use and census block) receives a hazard proximity score, with the number of hazards weighted using distance (“wedding cake approach”).

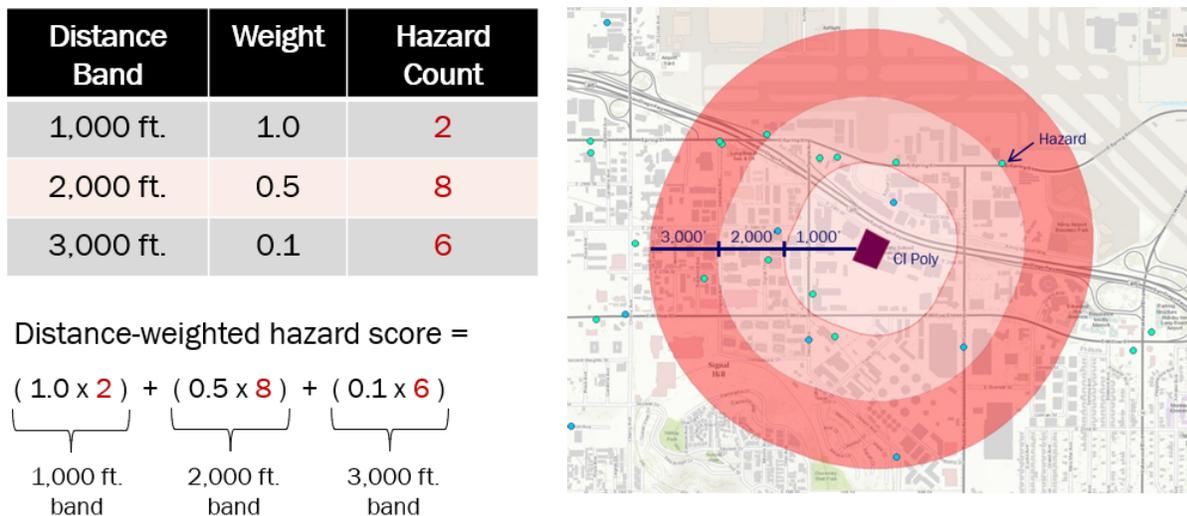


Figure 3-1. Summary diagram showing the method for calculating the EJSM Hazard Proximity counts. The distance between sensitive uses, and area and point hazards is calculated and the results aggregated into those within distance bands. These counts are distance weighted to reflect greater emphasis on those closest to the sensitive use.

Each populated spatial unit, or “CI Polygon” (CI poly – an intersect of vector geospatial data on land use, with population derived from 2010 census block-level data) is buffered

at multiple distances following CARB land use policy guidance on recommended separation distances between sensitive populations and air pollution hazards. These distances vary from 300 ft for some uses (example: gas stations and auto paint and body shops), to 1000 ft for uses with higher emissions (example, railroad uses, manufacturing facilities emitting air toxics, etc. (CARB, 2005).

In some cases, CI polygons are relatively large or have complex shapes that complicate the proximity to hazards relationship. For these cases, the large polygon is divided into smaller segments using a 1000 ft grid, and distance to hazards is calculated using the centroids of these polygon “slivers” as shown in Figure 3-2, below.

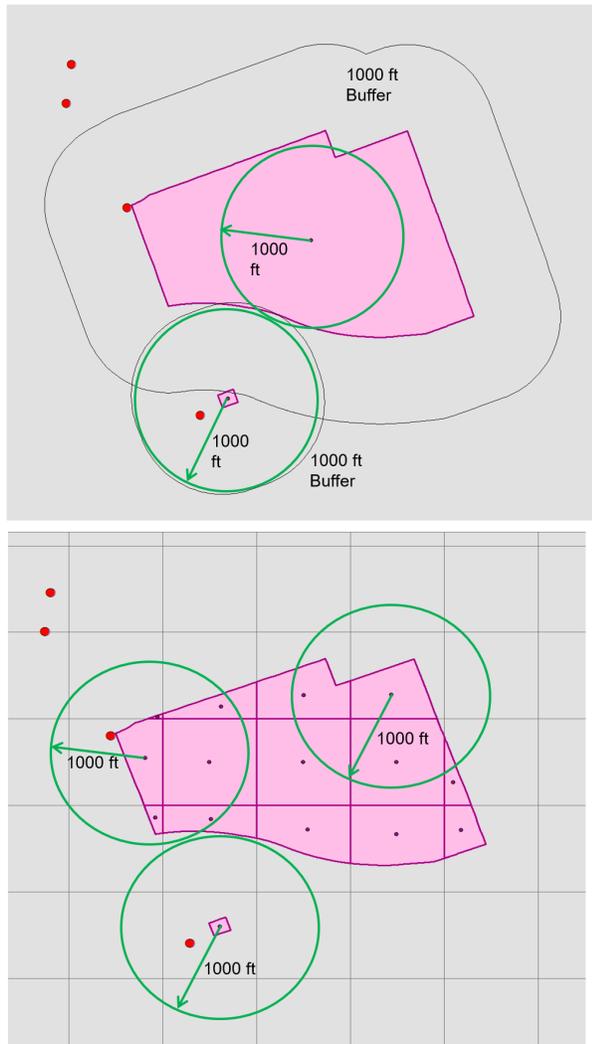


Figure 3-2. Hazard Proximity counts calculated for CI polygons that are large or irregularly shaped does not accurately reflect proximity, as the distance buffers measured from the centroid may remain entirely within the polygon itself (left). To address this problem, these CI polygons are intersected with a 1000 ft buffer to create “sliver polygons” (right).

Within these buffers, a count is made of the number of point and area hazards (air toxics emissions sources) to determine the number located proximate to the population living inside the CI poly. The count number in each buffer band is weighted according to the distance band it falls within to create a distance-weighted hazard proximity value for that CI Poly.

Once the Hazard Counts are population weighted these values are then aggregated to the census tract level using population weighting, such that portions of the tract with higher population and high hazard proximity values contribute most heavily to the resulting tract-level Hazard Proximity Score. The Scores for the region of interest, or statewide, are ranked and assigned to the appropriate quintile of the frequency distribution, resulting in tract level Proximity Hazard Scores from 1-5. For this reason, a small increase in Hazard Counts for a given tract will not result in a higher Hazard Proximity Score unless it is sufficient to place that tract in a higher quintile.

All EJSM scoring metrics for census tracts statewide, both with and without the facilities in Mexico are included in Appendix. (For details on the mechanics of this scoring metric see Morello-Frosch et al., 2016.)

Because the EJSM Hazard Proximity Score only considers hazards within a 3000 ft distance, we evaluated facilities in the “foi_1500m.csv” data set, as only facilities and sources located within that distance of the international border can increase the Hazard Counts metric which, in turn, has the potential to change an EJSM Score. Each facility or use in this dataset was then validated for use in EJSM hazard proximity scoring to determine whether it has a similar in hazard character to the types of facilities used in the EJSM in California. In most cases, the Standard Industrial Code (SIC, from a NAICS to SIC crosswalk) was used to match business type or descriptions with those of US facilities. Of the 1223 facilities and emissions sources identified by this project, 130 met these criteria and were used to update the EJSM scores (Table 3-1). These included

- Facilities from the Baja EI and RETC emissions inventories (n=62)
- Gas stations (n=68)
- Railroad land uses – used as an area source
- Assignment of highest

Most of these Mexico-based facilities and uses are clustered in areas located south of the San Ysidro, Otay Mesa, Tecate, and Calexico, as shown in the three maps, below (Figures 3-3, 3-4, 3-5). Also shown also shown for comparison purposes are facilities on the US side that were used in hazard proximity scoring, and census tract boundaries.

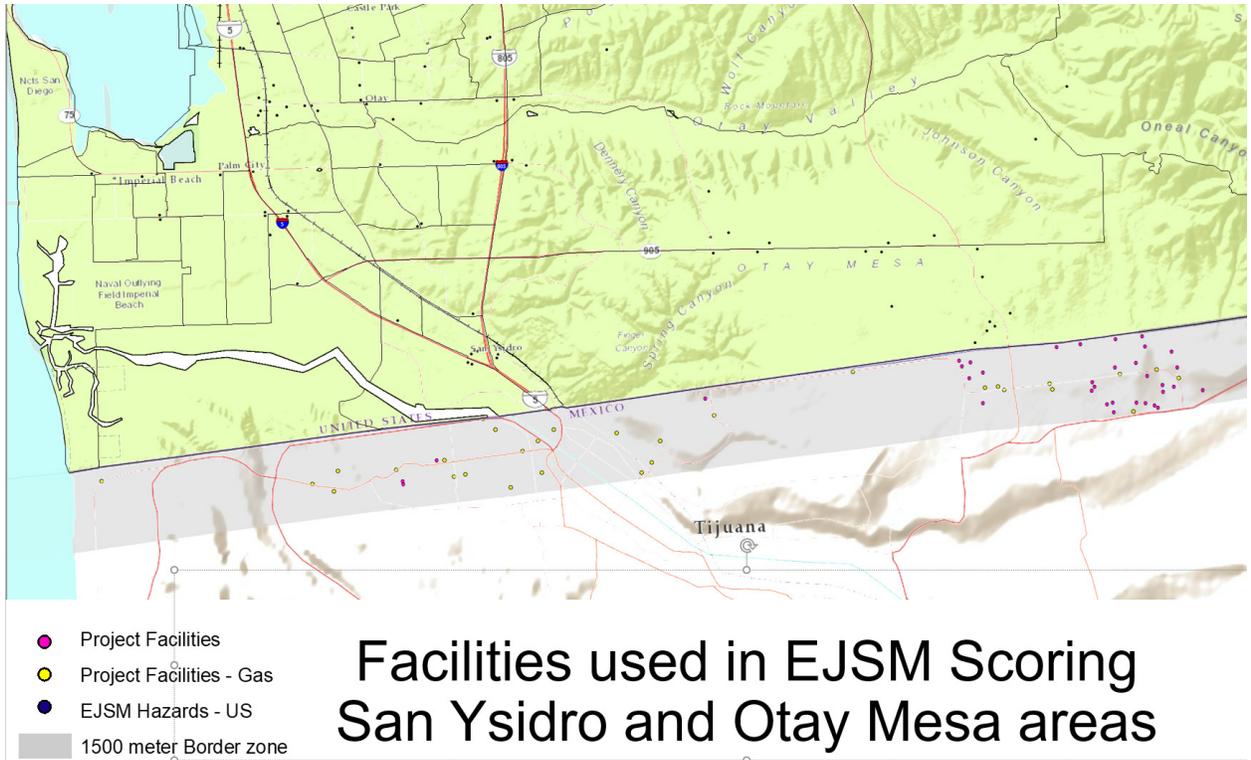


Figure 3-3. Point location hazards used in EJSM Hazard Proximity counts in the San Ysidro and Otay Mesa region in SW San Diego County.

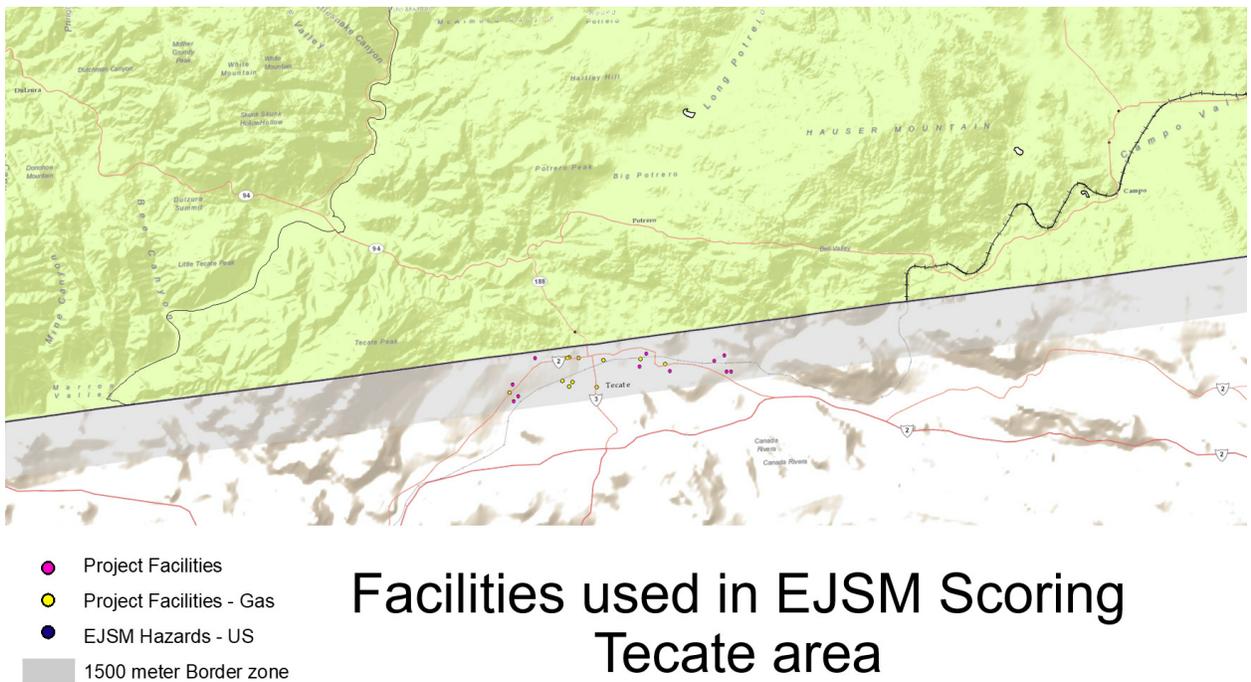


Figure 3-4. Point location hazards used in EJSM Hazard Proximity counts in the Tecate region in South San Diego County.

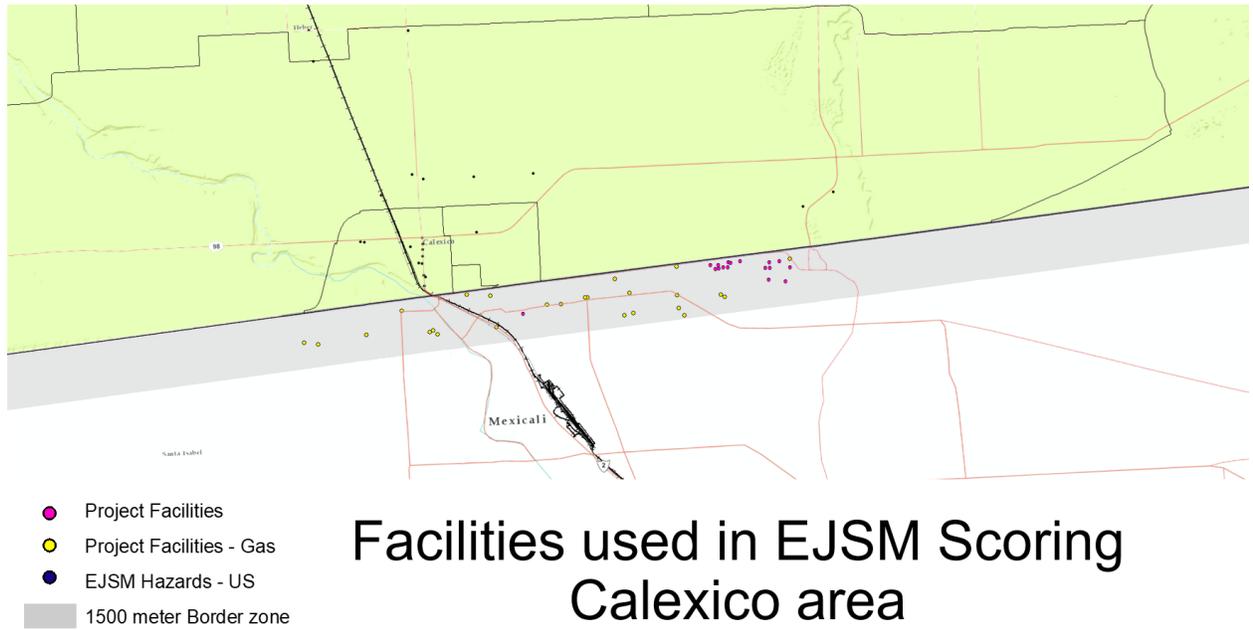


Figure 3-5. Point location hazards used in EJSM Hazard Proximity counts in the Calexico region in South Imperial County.

3.3 Results

Most US census tracts adjacent to the Border in the San Ysidro and Otay Mesa areas are relatively large compared to those in the more urbanized portions of the San Diego area, and their population centers are not located close to the hazards south of the border. Rescoring using the facilities in Mexico increased Hazard Counts for some CI polys in the region, but not enough to increase in the Hazard Proximity Score for any of the tracts in this area (Figure 3-6). This is also true when we apply greater emphasis on the EJSM Traffic Volume metric to reflect the very high vehicle emissions at the San Ysidro Port of Entry.

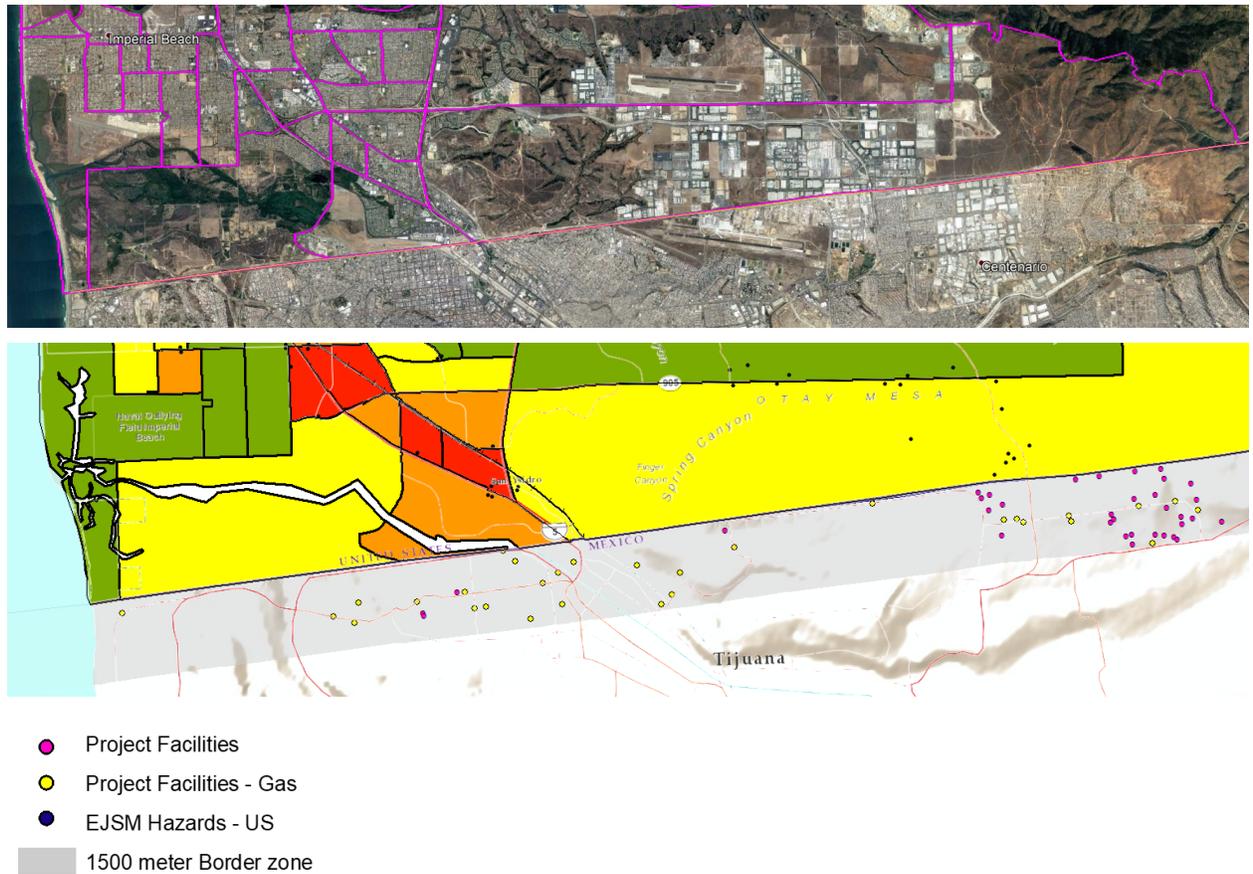


Figure 3-6. Census tracts with EJSM Hazard Proximity Scores in the San Ysidro and Otay Mesa region in SW San Diego County. Rescoring using hazards in Mexico identified during this study does not result in higher scores

Near Tecate, the one US census tract is very large and its population is very sparse, accounting for the size of this census tract. Despite numerous hazards in Tecate located near this tract boundary, the lack of population on the US side results in almost no increase in Hazard Counts, and no change in the Hazard Proximity Score. (Figure 3-7)

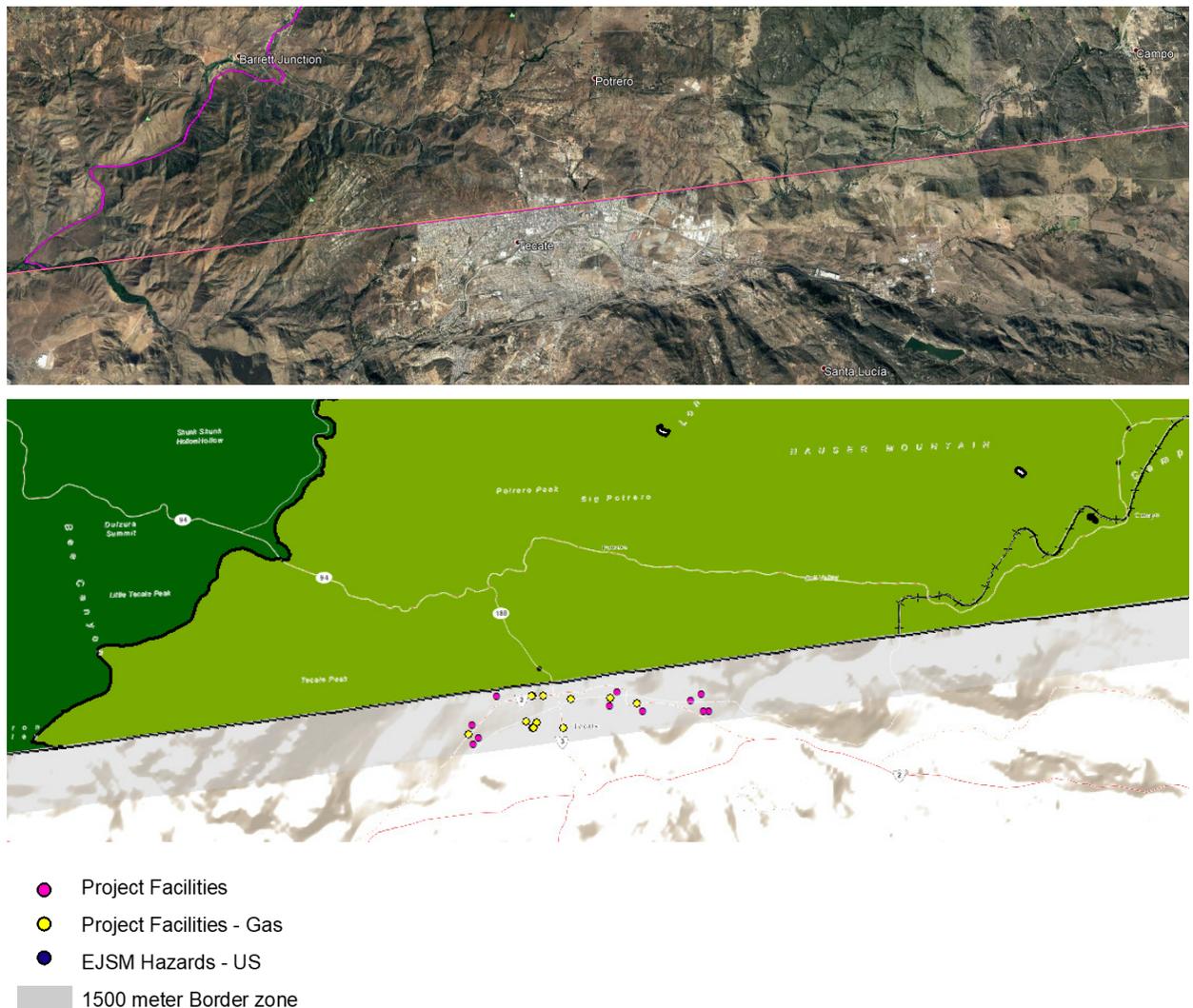


Figure 3-7. Census tract with EJSM Hazard Proximity Score in the Tecate region in Southern San Diego County. Similar to Figure 3-6, rescoring using hazards in Mexico identified during this study does not result in a higher score for this tract.

Tracts located in the city of Calexico are smaller and more numerous, as it is a regional population center. There are gas stations on the Mexico side located close to these tracts, and a cluster of other emission facilities located in the Mexican industrial center located east of Calexico. As with the previous two regions, only tracts with significant population located close to hazards have higher Hazard Counts, and one tract in this area, GEOID 06025012001, has hazard counts increased sufficiently to increase its EJSM Hazard Proximity Score by one, from 1 to 2. Note that this increase in score is lower than that of the tract adjacent to the west (GEOID 06025012100), which has a score of 3 that was unchanged by including the facilities in Mexico. Note also that this tract contains an obvious cluster of US hazards accounting for its relatively higher Score.

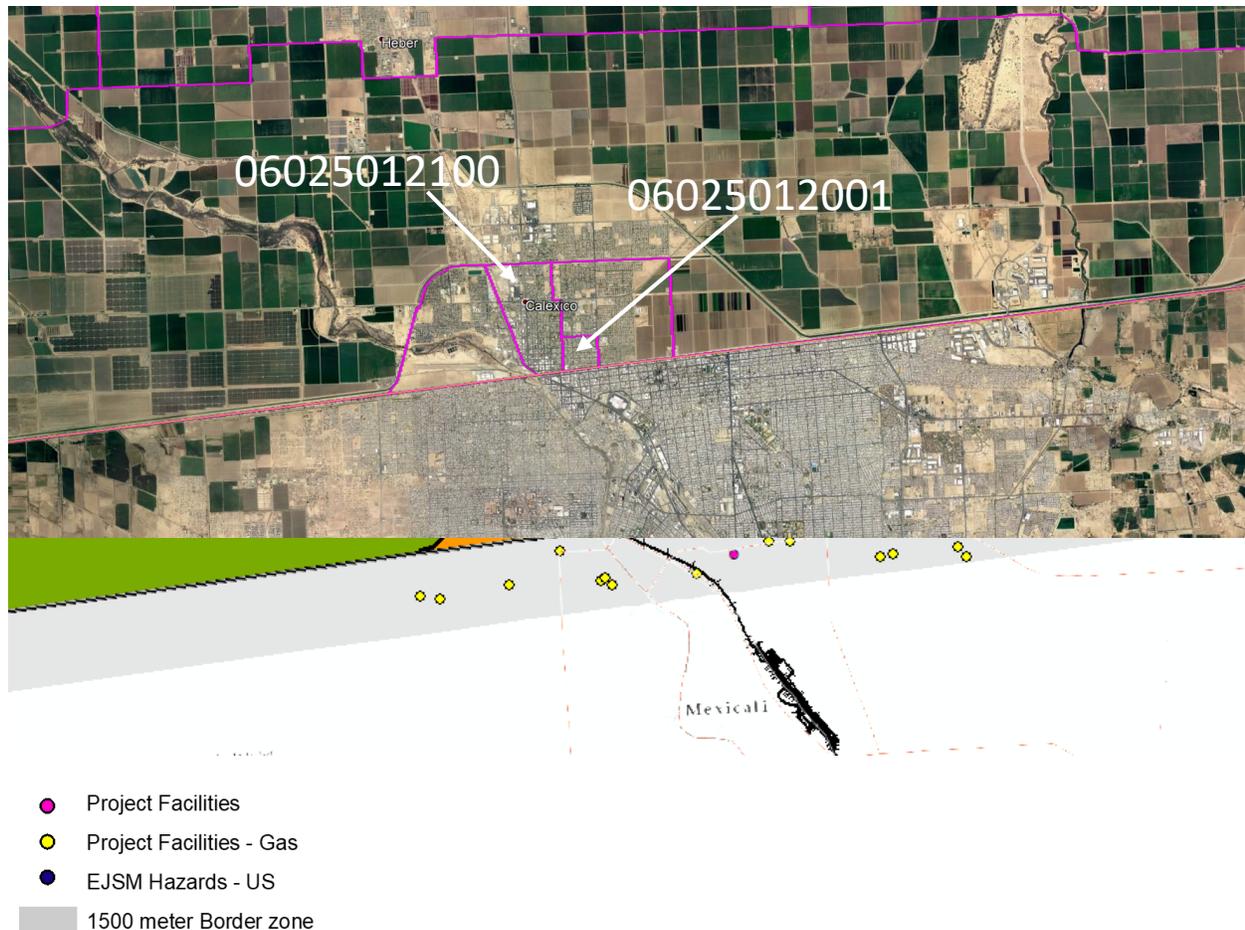


Figure 3-8. Census tracts with EJSM Hazard Proximity Scores in the Calexico region in South Imperial County. Rescoring using hazards in Mexico identified during this study increases hazard counts for tracts in tracts with relatively higher population density (within city of Calexico). This increase for tract 06025012001 is sufficient to increase its EJSM Hazard Proximity Score by one, from 1 to 2.

3.4 Discussion

A key premise in the project technical plan was that that the scoring differences would be concentrated in the census tracts in southern San Diego and Imperial Counties adjacent to the US-Mexico border. We found the effect on scores of considering additional facilities in Mexico was marginal for three reasons. First, these facilities contribute only a small increment (Hazard Counts) to final EJSM scores for a census tract, which consider many factors of cumulative impacts and community vulnerability. Second, some screening metrics and scores for tracts in this region are already within the upper end of the scoring range (see Table 3-2). Perhaps most prominent is the proximity relationship between the population distribution within a given tract, and the hazards in Mexico. EJSM Hazard Proximity Scores were recalculated for all census tracts in the Border region using the facilities in Mexico. The maps of these three areas, below, show census tracts classified by the Hazard Proximity Score on an increasing scale of 1-5. Several generalizations apply to the geography of these areas. First, census tracts in this region vary in size

significantly because of substantial differences in population and population density – these were divided into smaller segments for hazard proximity analysis, as described previously. Large tracts are mostly composed of areas with little or no residential use and are sparsely populated overall with a few denser population centers. The map below of the Calexico area illustrates four relatively small tracts that cover the city of Calexico, and one much larger tract extends over the sparsely populated, mostly agricultural land surrounding the city (Figure 3-9).

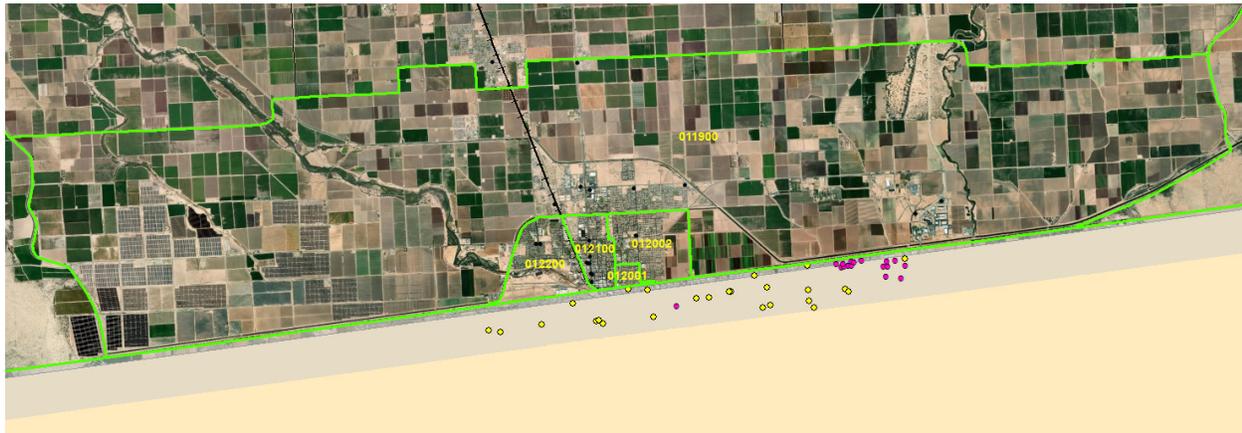


Figure 3-9. Census tracts with EJSM Hazard Proximity Scores in the Calexico region in South Imperial County. Aerial photo background shows the large tract (011900) to be primarily agricultural with very low population; the four smaller tracts that cover the City of Calexico are much more densely populated

This larger scale map image of Calexico, below Figure 3-10, shows the CI Polygons used in calculating the EJSM Hazard Proximity Score; these polygons essentially indicate where the residential population is located within the tract. Mexico-based hazards (mostly gas stations) within the 1500 meter band are located near the smaller city tracts, but only a few are very close. A large cluster of air toxics emissions sources is located very close to the border, east of the city, and close to a portion of the large tract. Note that there are no CI Polygons (residential populace) near that cluster. This geographic pattern is similar in Tecate and the San Ysidro/Otay Mesa areas, as well. The pattern also results in a relatively low impact of the Mexico facilities on the EJSM Hazard Proximity Score; buffers surrounding the CI Polygons do not intersect with many of the hazards in Mexico, so these facilities have little effect on the proximity metrics.

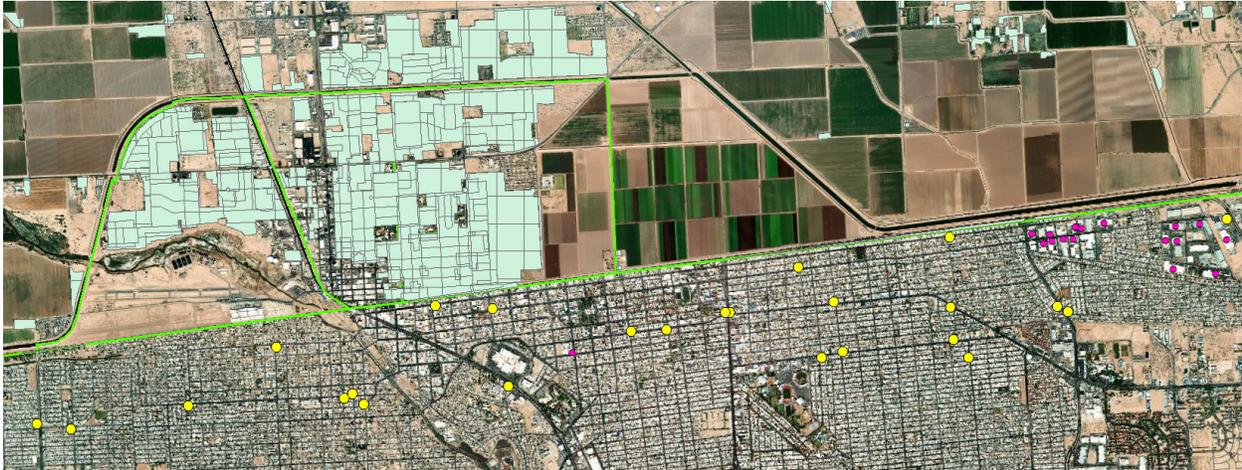


Figure 3-10. CI Polygons used to calculate EJSM Hazard Proximity Scores in the general area of the City of Calexico. Note the relatively high concentration of point hazards East of the City that are located far from the closest CI Polygons. Hazards near the City are primarily gas stations, which require a 900 ft proximity to have any effect on Hazard Proximity counts.

In order for Hazard Proximity counts to be significant for these tracts, the hazards would have to be located very close (at least within 3000 ft) to areas of high population density (CI Polygons). Smaller census tracts are parts of cities and towns with higher population density and a local economy that supports facilities that emit air toxics. Hazard counts for these tracts tend to be higher, as well.

3.4.1 CalEnviroScreen

Responding to the need for data on facilities releasing air toxics for calculating the CES toxic concentrations indicator, this project has provided three additional years of facility data from both Baja EI and RETC. These site locations have been checked and validated, and could serve as the basis for the next revision of the CES. These data would first be compared to the RETC facilities used in the current RSEI, then included in a future revision of RSEI (by Abt Associates, the private firm that manages RSEI for US EPA TRI Program).

Locations of Solid Waste Landfills/Brick Kilns/Manufacturing facilities automated in this project could be incorporated into the Environmental Effects component of CES (Solid Waste Sites and Facilities). However, additional information on these facilities would have to be determined, as the California facilities incorporated into this metric are weighted by facility-specific such as size, waste type and volume, number of employees, and violations history. It was beyond the scope of this project to investigate this information or its availability. In one conference call meeting, it was mentioned that CARB has another research contract to improve information for this source category

Railroad land use automated by this project could inform the CES Diesel PM Indicator. CES considers DPM from various emissions sources using CARB estimated concentrations from on-road and non-road sources interpolated to a 4x4km grid. EJSM

includes ambient modeled DPM concentrations in its health risk metric derived from a land use regression from the National Air Toxics Assessment that includes fate and transport model correction, with risk recalculated to reflect California Environmental Screening Levels. Air monitors for DPM are used to validate the modeled data, and show good correlation in most areas, but the correlation with monitoring at the border region is particularly low. CES3 showed marginally higher levels of DPM over CES2 for census tracts along the Border, and the degree to which these values would increase by estimating and adding emission from rail uses may be more significant.

Both CES and EJSM have struggled with the inclusion of air pollution from fires, such as wildfires and purposeful burning of agricultural waste. The MODIS satellite fire detection data is used widely to create geospatial information on fires of this type by researchers and municipal agencies nationwide. This project explored this data source experimentally, and mapped 20 years of data on 50 km on both sides of the border (see next section), but further improvements would be necessary to allow these screening tools to begin using it.

It may seem counterintuitive that hazards located close the Border with US census tracts do not increase the EJSM score to a greater degree, but this is a consequence of the specific distance weighting relationship used by the EJSM Hazard Proximity Score. In OEHHA's CES v. 3 update, the data clearly show higher ambient concentrations of some air pollutants (mostly ozone and DPM for some border tracts), and higher traffic counts for some tracts (other tracts show decreases in traffic) as compared to CES v.2 (OEHHA, 2014). The increased effect on the RSEI metric, reflecting the effect of Mexico side emissions sources, is more difficult to interpret, as CES v. 2 reported very low RSEI metrics for border tracts, most far lower than most tracts in California (OEHHA, 2017b).

3.5 Conclusions and Recommendations

Air toxics emissions sources in Mexico clearly impact communities and individuals in the US-Mexico Border region, and those located close to the border itself are of particular concern to residents of this region. However, the majority of these sources are not located sufficiently close to U.S. residential areas population in census tracts adjacent to the Border to register as increases in the EJSM Hazard Proximity Score, as census tracts are large. We found an increase of one increment of this score for one tract in the Calexico area; other border tracts' scores were unchanged.

This Score is not designed to measure potential impact, but rather to alert stakeholders and regulators of potential impacts due to the relationship between hazard/emissions proximity and residential population concentrations. This project is direct confirmation that there is significant concern over the impact of cross-border air pollution from border-area sources, and the potential for impacts is clearly shown by the changes in metrics introduced into the current version of CalEnviroScreen. The more current, and carefully checked, facility data generated by this project can clearly be used to more clearly and comprehensively assess those potential impacts and in future versions of CES, and the greater variety of emissions sources identified in this project may expand opportunities for OEHHA to add further detail to some of their pollution source metrics.

By the nature of its input data and analytical approach, as well as its intent by design, the EJSM has always provided better results identifying communities with disparate impact

and vulnerability in urban areas, where population density is high and the pollution hazards are better understood and documented. EJSM and CES are also forced by the quality and resolution of the data they use, to aggregate data and score at the census tract level. The pattern of population in many census tracts is complex, and attempting to account for this is a challenge for both screening methods. Sparsely populated tracts pose the greatest challenge, and the US-Mexico Border presents a most extreme case of this difficulty as the character and number of pollution sources in Mexico is significantly different than in border region tracts and communities. It may seem to some residents of this region that screening methods scores “undercount” their impact and vulnerability, and this may be as much of a call for improved methods of analysis in screening as it is for better characterization of pollution and sources in Mexico. The screening methods also cannot account for seasonal variability of pollutant transport, which this project’s forward/backwards trajectory studies demonstrate. They also do not presently address the role of differences in characteristics of the pollution sources themselves in Mexico vs. the US. In some cases, pollution sources may be absent in California, but present in Baja California, and in other cases the pollution sources may be present in both countries but more extreme in Baja California than in California. These differences include, for example, vehicle emissions requirements and monitoring, vapor recovery systems for dispensing gasoline and other volatile products, open burning of agricultural materials or solid waste, to name a few. Perhaps the best way to address these border-specific challenges is to build screening approaches from first principles for this region, an exercise that may identify ways to improve existing screening, as well.

Table 3-1. Types of hazardous facilities in Mexico influencing US census tracts along the US-Mexico border

Region	Census Tract	Facility Type	Number
San Ysidro Otay Mesa			
	06073010015	Computer Equipment	1
		Electrical Manufacture	8
		Metal Fabrication	1
		Metals Manufacture	2
		Plastics Manufacture	3
		Trucking	2
		Wood Products	1
		Gas Stations	15
	06073010009	Plastics Manufacture	1
		Gas Stations	7
	06073010109	Gas Stations	4
	06073010200	Gas Stations	1
Tecate			
	06073021100	Electrical Manufacture	2
		Metal Fabrication	2
		Metals Manufacture	2
		Plastics Manufacture	3
		Primary Metals	1
		Wood Products	1
		Gas Stations	12
Calexico	06025012001	Gas Stations	2
	06025012002	Electrical Manufacture	1
		Gas Stations	4
	06025012200	Gas Stations	4

	06025012100	Gas Stations	2
	06025011900	Electrical Manufacture	7
		Metals Fabrication	3
		Metals Manufacture	3
		Other Metals Foundry	1
		Plastics Manufacture	1
		Primary Metals	1

Table 3-2. Change in EJSM Regional Hazard Proximity Scores (Regional Scoring) for US Census tracts along the US-Mexico border. Other EJSM metrics and scores available in Appendix

Region	Census Tract	Population	Area (sq. mi.)	Score (no MX)	Score (w/ MX)
San Ysidro Otay Mesa	6073010009	6693	1.379	4	4
San Ysidro Otay Mesa	6073010015	2803	15.615	3	3
San Ysidro Otay Mesa	6073010109	4595	4.598	3	3
San Ysidro Otay Mesa	6073010200	6800	2.843	2	2
Tecate	6073021100	7589	448.233	2	2
Calexico	6025011900	10966	86.822	2	2
Calexico	6025012001	3831	0.221	1	2
Calexico	6025012002	9678	1.894	1	1
Calexico	6025012100	6937	1.028	4	4
Calexico	6025012200	7804	2.004	4	4

3.6 References

CARB Air Quality and Land Use Handbook: California EPA, California Air Resources Board, <https://www.arb.ca.gov/ch/handbook.pdf> , 2005.

Morello-Frosch, R, Sadd, J., and Pastor, M.P., Update and Statewide Expansion of the Environmental Justice Screening Method (EJSM): Final Report to California Air Resources Board - Contract Number # 11-336, 50 pages., (<https://ww3.arb.ca.gov/research/apr/past/11-336.pdf>) 2016

OEHHA: California Communities Environmental Health Screening Tool, version 2.0: California EPA Office of Environmental Health Hazard Assessment <https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-version-20> , 2014

OEHHA: New in CalEnviroScreen 3.0: Changes since version 2.0. California EPA Office of Environmental Health Hazard Assessment <https://oehha.ca.gov/media/downloads/calenviroscreen/document/ces3newinces3.pdf> . 2017a

OEHHA: CalEnviroScreen Data Update for the Border Region: California EPA Office of Environmental Health Hazard Assessment <https://oehha.ca.gov/media/downloads/calenviroscreen/report/ab1059borderlegreportmarch2017.pdf> , 2017b

CHAPTER 4. PILOT STUDY OF A METHOD FOR DETERMINATION OF US-MEXICO BORDER BURNING ACTIVITIES THROUGH AUTOMATED ANALYSIS OF SATELLITE DATA

4.1 Introduction

As an exploratory aim in Task 1.4, the project team developed a methodological framework to identify burning activities in the border region in the Mexico area. Specifically, this study utilized MODIS (Moderate Resolution Imaging Spectroradiometer) Active Fire products, which are time series of remote sensing imagery containing pixels labelled as fire or non-fire, and employed spatiotemporal clustering, time series clustering, and GIS overlay methods to investigate agricultural-related burning activities.

4.2 Data

The project team collected the MODIS Active Fire Product imageries from February 24, 2000 to February 17, 2018 to detect larger burning activities. The MODIS imageries, collected by two satellites Terra and Aqua daily producing 2 images per day, detects fires in 1 km pixels that are burning at the time of overpass under relatively cloud-free conditions. The collected data consists of 1,545 HDF (Hierarchical Data Format) files for the time period and each single image is approximately $10^{\circ} \times 10^{\circ}$ in size covering the study area (Figure 4-1). The MODIS Active Fire Product uses a contextual algorithm (Giglio, 2003) that exploits the strong emission of mid-infrared radiation from fires and assigns one of the 10 classes described in Table 4-1 to each pixel. In this study, we used fire pixels with “high confidence” only (Table 4-1). Figure 4-2 presents example fire pixels from the MODIS Active Fire Products observed in December 28th, 2017.

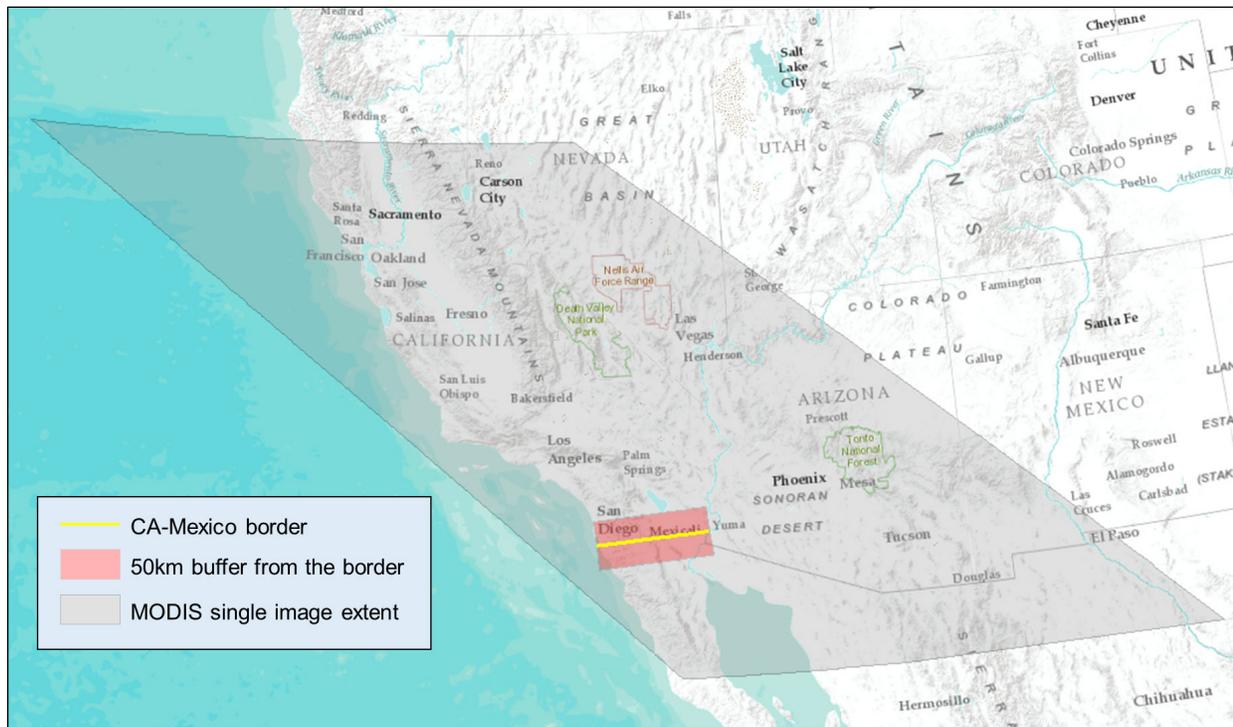


Figure 4-1. MODIS imagery extent (Tile Location Code=H8V5)

Table 4-1. The Pixel Classification of the MODIS Active Fire Product

Class	Description
0	Not processed (missing input data)
1	Not processed (obsolete)
2	Not processed (other reason)
3	Non-fire water pixel
4	Cloud
5	Non-fire land pixel
6	Unknown
7	Fire (low confidence)
8	Fire (normal confidence)
9	Fire (high confidence)

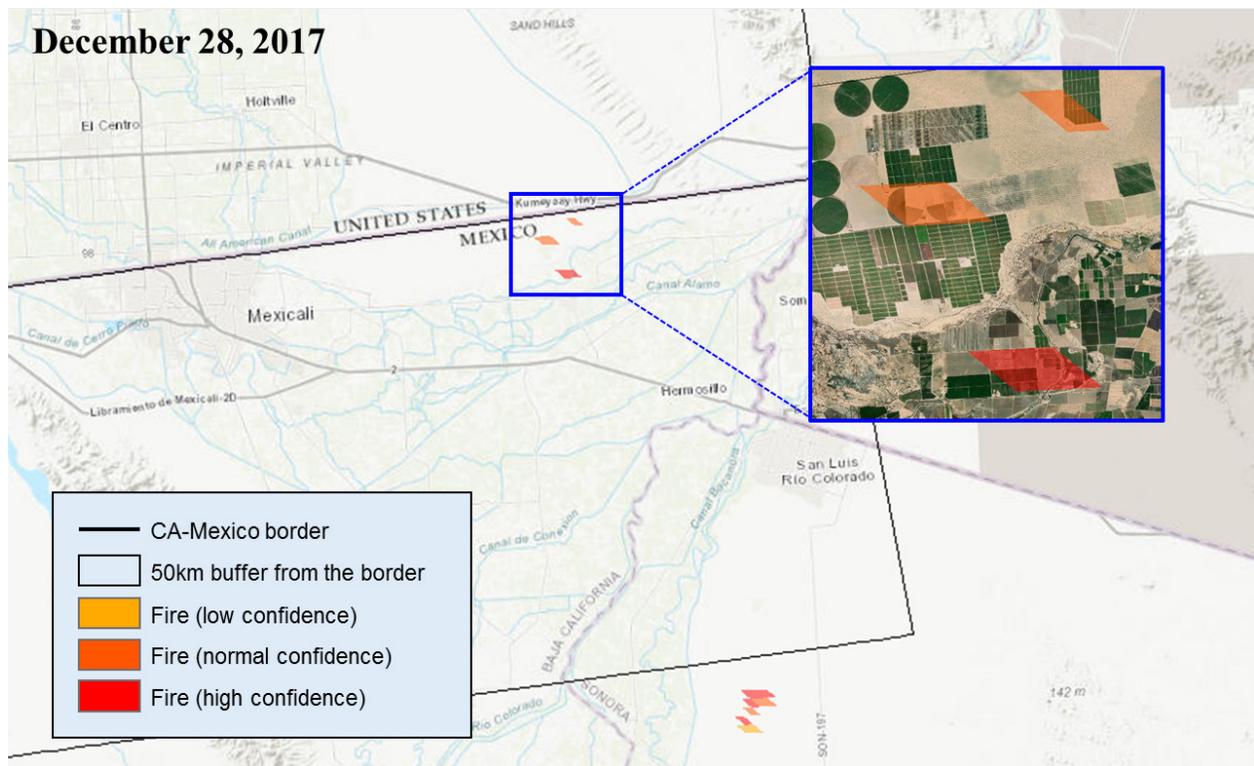


Figure 4-2. Example fire pixels in the MODIS Active Fire Products.

4.3 MODIS Data Analyses

To detect burning activities from MODIS Active Fire Products and investigate agricultural-related burning activities, this study employed spatiotemporal clustering, time series clustering, and GIS overlay methods.

4.3.1 Temporal Trends of Fire Events

A fire event can be defined as a group of fire pixels that are spatially and temporally contiguous. Identifying fire events from time series fire pixels is an important process for summarizing burning activities (e.g., fire event frequency, burned area by event) and investigate fire event spatio-temporal trends. Figure 4-3 illustrates an example of longitudinal fire pixel data, which contains three fire events based on spatiotemporal contiguity.

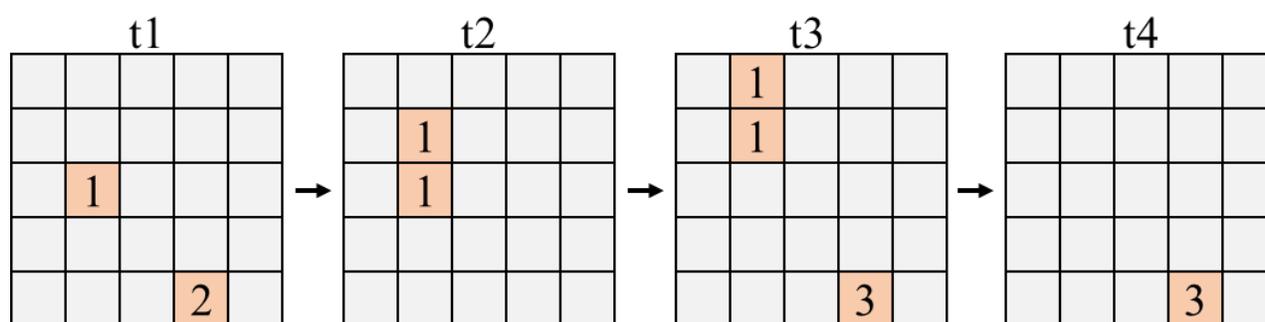


Figure 4-3. Longitudinal Time Series Fire Pixels. The number in fire pixels (orange color) represents a fire event ID.

To identify fire events from the MODIS Active Fire Products, this study employed ST-DBSCAN (Birant, 2007), which is an extension to a Density-Based Spatial Clustering of Applications with Noise, known as DBSCAN (Ester, 1996). Applied to a set of 2-dimensional points, DBSCAN discovers arbitrary-shaped spatial clusters based on two parameters, $Eps(\epsilon)$ and $MinPts$. $Eps(\epsilon)$ is a radius value to determine neighbors and $MinPts$ is the minimum number of points to form a cluster. ST-DBSCAN introduces $Eps(\epsilon)$ -space and $Eps(\epsilon)$ -time to determine neighborhoods by spatial and temporal distance values separately. In this study, $Eps(\epsilon)$ -space is set to the MODIS pixel width, 1km, and $Eps(\epsilon)$ -time is set to the MODIS temporal resolution, 1 day.

A total of 30,741 pixels (1km x 1km) were identified as fire pixels with high confidence in each side of the 50 km buffer from the U.S.-Mexico border (Figure 4-4 left). ST-DBSCAN extracted 6,194 fire events by grouping those fire pixels based on their spatial and temporal contiguity (Figure 4-4 right). To investigate general temporal trends on fire activities in the US-Mexico border region, the total count of monthly fire pixels and fire events were compared between the US side and the Mexican side of the 50 km buffers from the US-Mexico border (Figure 4-5 and 4-6). The result indicates two significant patterns, 1) high-frequent recurring fire events in the Mexico border region, and 2) smaller fragmented fire events on the eastern regions.

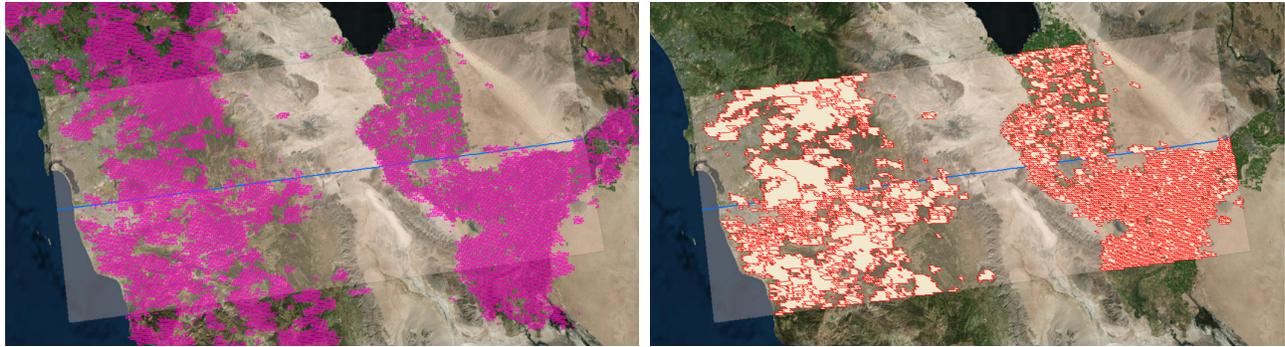


Figure 4-4. Identified fire pixels (left) and fire events (right) from February 2000 to February 2018 within the border regions.

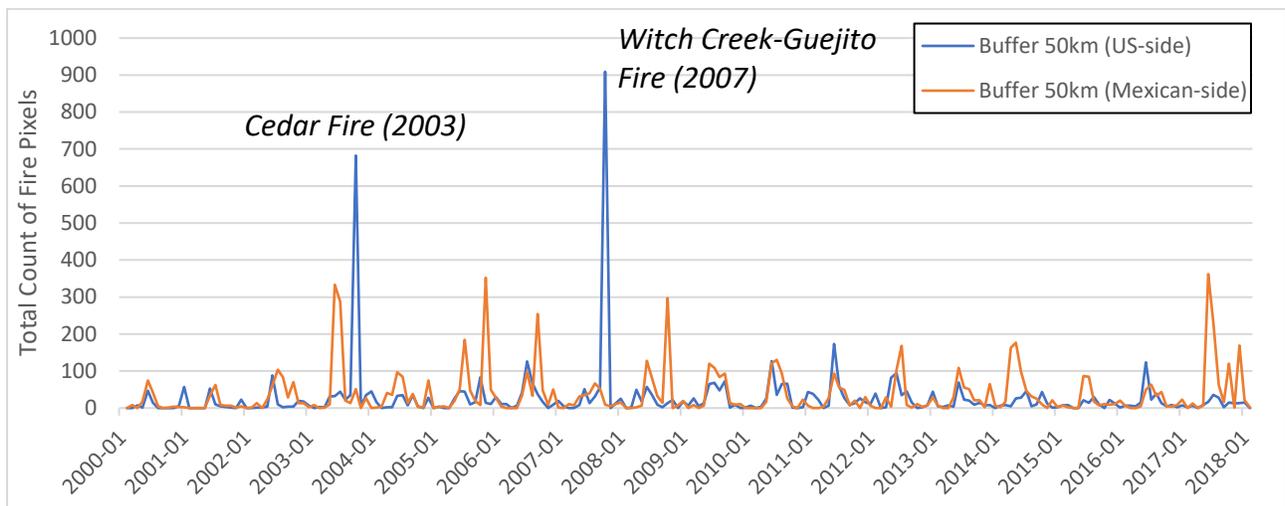


Figure 4-5. Identified monthly fire pixels from February 2000 to February 2018 within the border regions.

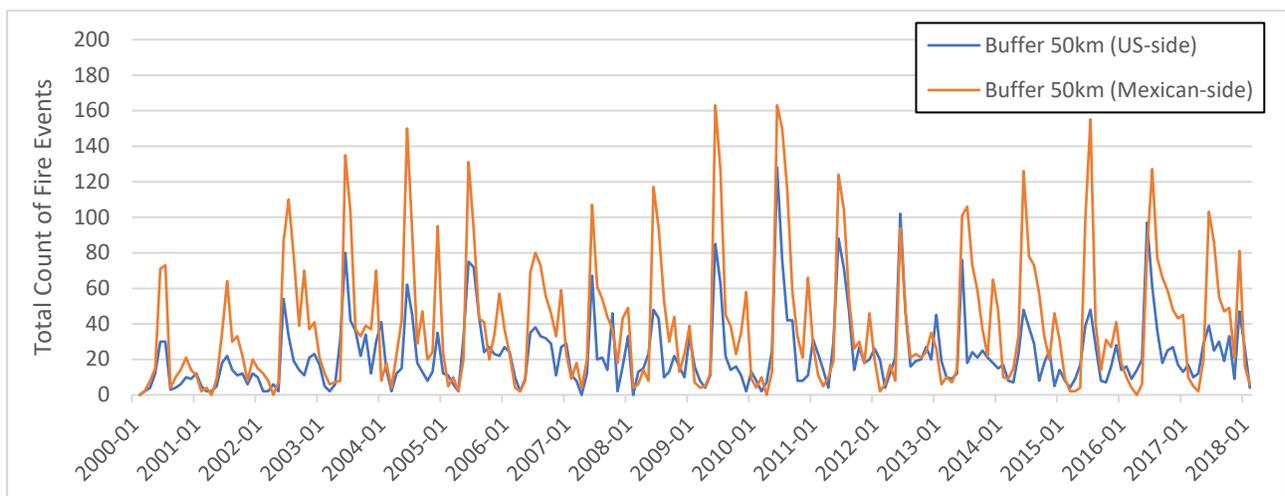


Figure 4-6. Identified monthly fire events from February 2000 to February 2018 within the border regions.

4.3.2 Fire Activities on the Agricultural Land Uses

This study further examined the relationship between those extracted fire activities from the MODIS Active Fire Products and land-uses to identify types of burning activities, specifically focusing on agricultural burns. To achieve this objective, we used GIS overlay methods to extract fire pixels on agricultural land uses derived from 2010 Landsat-based 30 meter spatial resolution Land Cover map produced by Commission for Environmental Cooperation (CEC, 2018). Figures 4-7, 4-8, and 4-9 show spatial distributions of crop fields and urban areas in the US-Mexico border region, the Tijuana area, and the Mexicali area respectively.

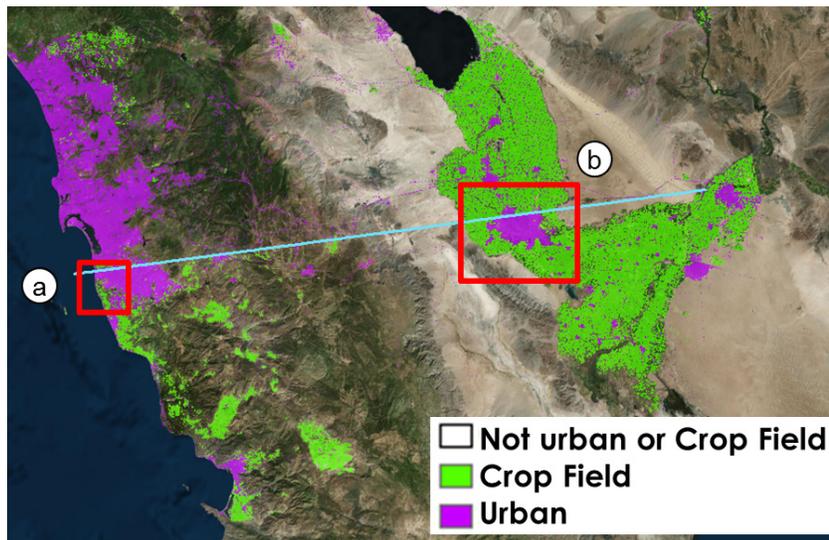


Figure 4-7. Spatial distribution of crop fields and urban areas in the US-Mexico border region.



Figure 4-8. Spatial distribution of crop fields and urban areas in the Tijuana area.

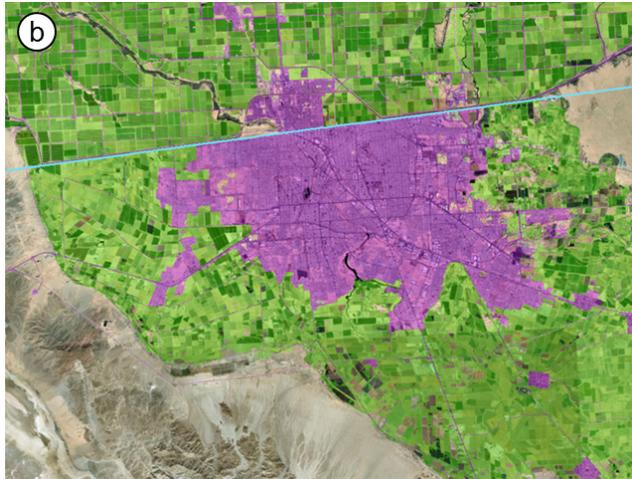


Figure 4-9. Spatial distribution of crop fields and urban areas in the Mexicali area.

We overlaid three layers, Modis fire pixels, crop fields, and rural areas defined by INEGI (Instituto Nacional de Estadística y Geografía) to calculate total areas of burned crop fields by rural areas. Figure 4-10 illustrates actual three GIS overlay steps: (1) overlaying crop fields with INEGI rural areas; (2) dissolving crop fields by INEGI rural areas; and (3) overlaying dissolved crop fields by rural areas with fire pixels to calculate areas of burned crop fields. Figure 4-11 exhibits the annual spatial distribution of fire activities on crop fields colored by the total number of fire pixels by INEGI rural areas in the northeastern Baja California. Since the pixel size corresponds to 1 km², the number burned pixels represents a rough estimate of burned agricultural areas.

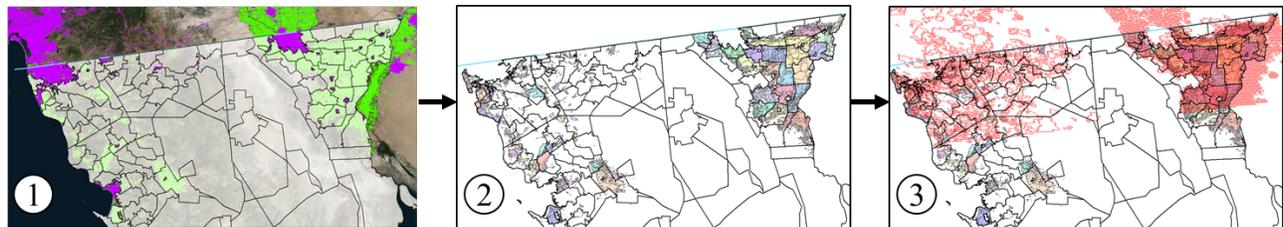


Figure 4-10. GIS overlay methods.

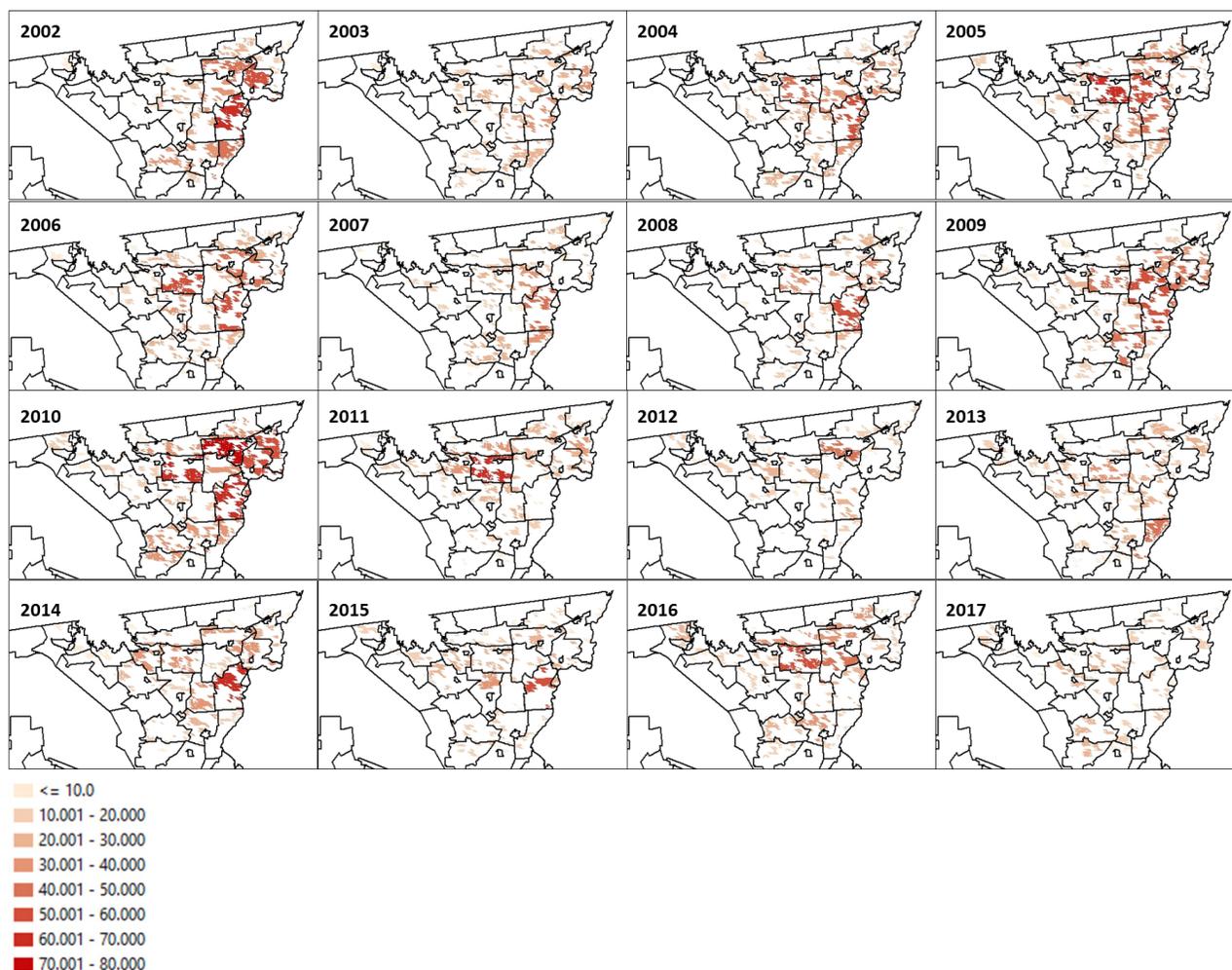


Figure 4-11. Annual burned agricultural areas (km²) by INEGI rural areas in the northeastern BC.

Figure 4-12 plots the estimate of monthly burned agricultural areas by INEGI rural areas. This time-series plot reveals strong seasonal trends on fire activities on crop fields by rural areas. As a general temporal trend, there are two peaks in each year, summer and early winter.

To further examine agricultural-related fire activities, we conducted the Euclidean k-means time series clustering analysis to group similar time series data to extract distinct temporal trends (Tavenard, 2017). The result of time series clustering unveils six unique temporal trends on agricultural fire activities (Figure 4-13, 4-14) and the identified trends as time series clusters are mapped in Figure 4-15. Time Series Cluster (TSC) 0, 1, and 2 correspond to low agricultural fire activities. TSC-2 represents the lowest fire activities in rural areas, which has relatively small areas of crop fields. Notably, all INEGI rural areas in the northwestern region of BC are assigned to TSC-2. TSC-0 and TSC-1, located in the

northeastern region of BC, shows similar seasonal fire activities with two peaks in summer and early winter. A distinct difference between these two clusters is that TSC-2 has a relative consistent time series trends on fire activities over 18 years, while TSC-1 has very small fire activities during years 2006, 2007, 2012, and 2013 and higher fire activities during years 2017 and 2018. As compared to TSCs 0, 1, and 2, TSCs 3, 4, and 5 have generally higher fire activities. INEGI rural areas with these clusters are spatially clustered and located in the central part of the crop fields in the northeastern region of BC (Figure 4-16). Further investigation is required to understand the nature of identified agricultural related fire activities and their spatial and temporal trends.

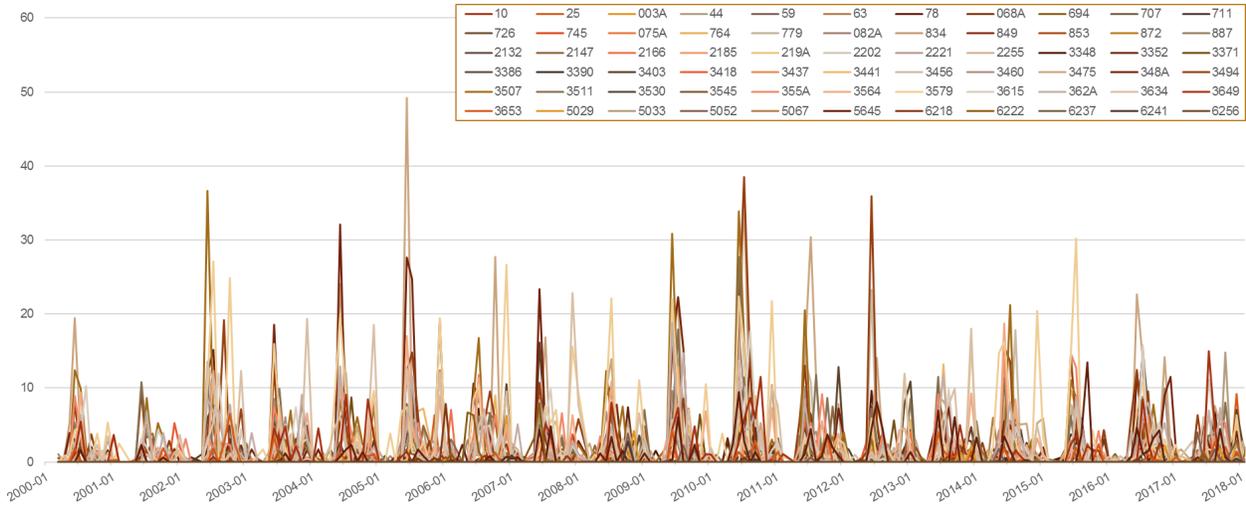


Figure 4-12. Monthly burned agricultural areas (km²)

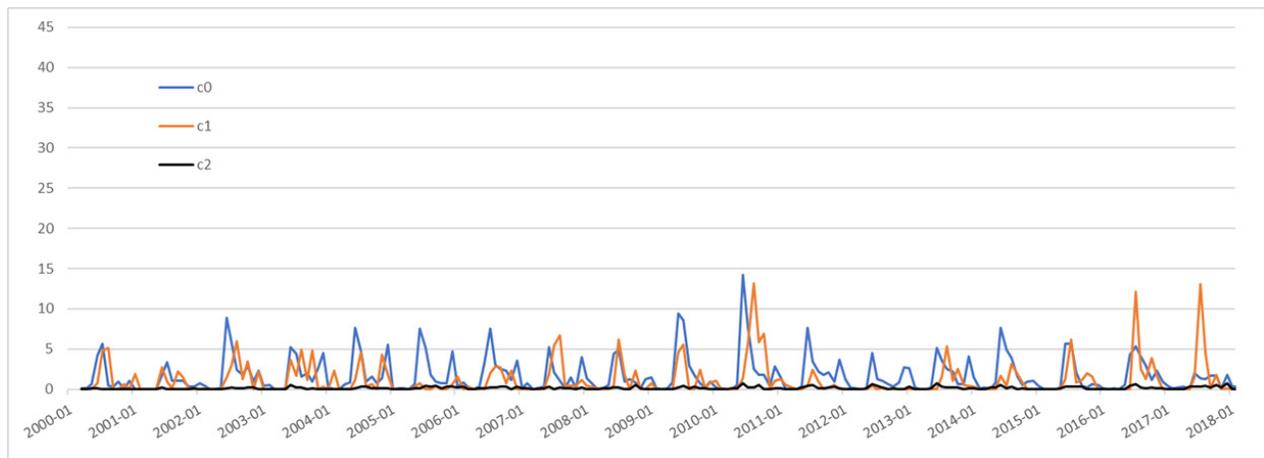


Figure 4-13. Low agricultural fire activities identified by time series clustering (Cluster 0, 1, and 2). The y-axis corresponds the frequency of fire events.

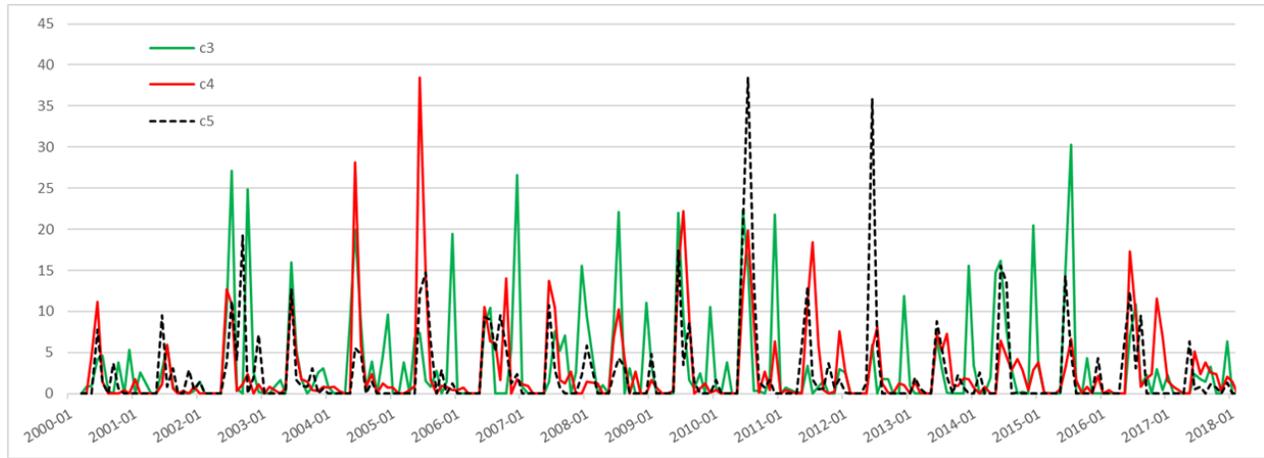


Figure 4-14. High agricultural fire activities identified by time series clustering (Cluster 3, 4, and 5). The y-axis corresponds the frequency of fire events.

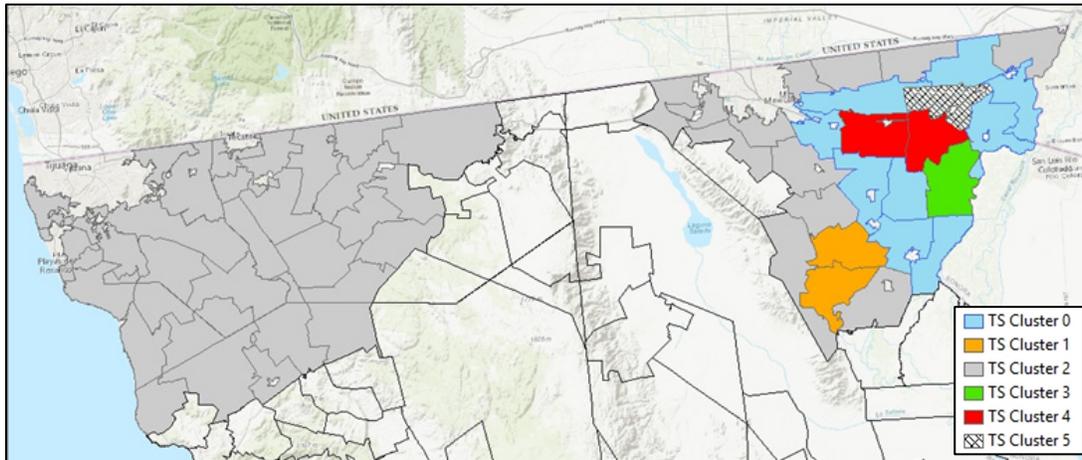


Figure 4-15. Spatial distribution of temporal trends of agricultural fire activities.

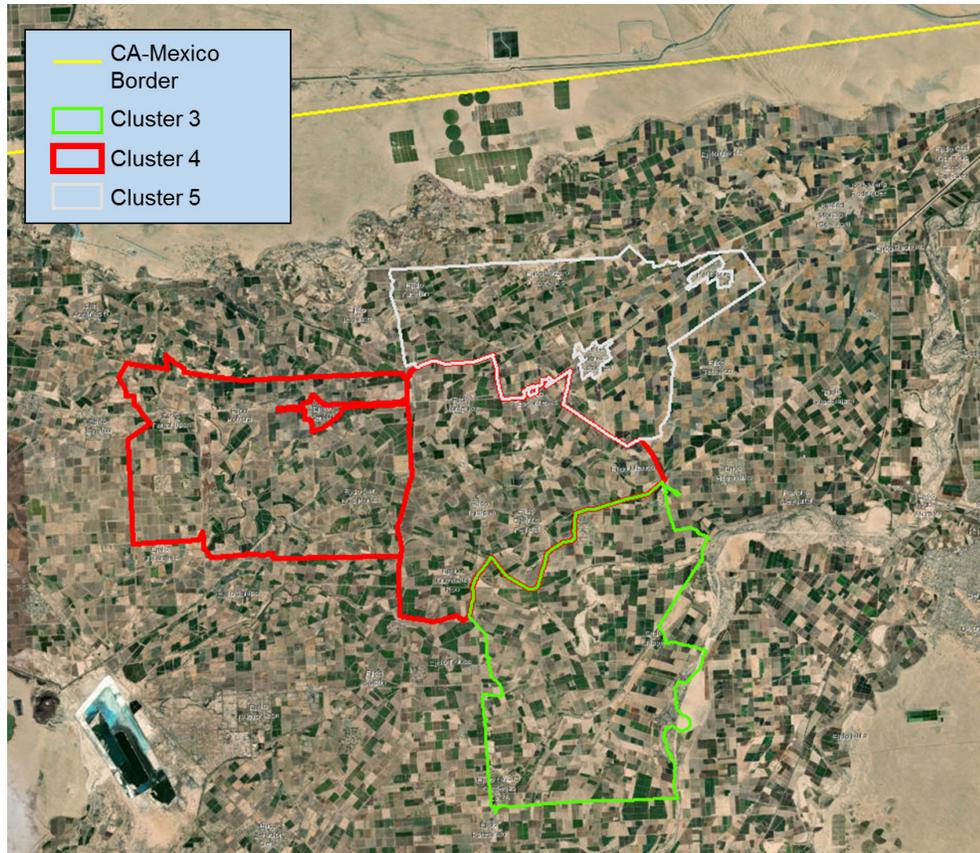


Figure 4-16. Spatial distribution of temporal patterns of agricultural fire activities (Cluster 3, 4, 5).

4.4 Summary, Conclusions, and Recommendations

The exploratory study conducted in Task 1.4 identified spatial and temporal trends of fire activities and agricultural-related fire activities in the border region in the Mexico area by using MODIS Active Fire Products and spatiotemporal analytics methodologies. While the results exhibit interesting patterns, further investigations are necessary to estimate the amount of emissions from identified fire activities in order to incorporate the result into CES. Groundtruthing is one of the key future research endeavors to validate identified fire activities. Since the MODIS spatial and temporal resolutions are 1 km by 1 km and 1 day respectively, this exploratory work may not be able to detect small fires. In addition, while this work focused on agricultural-related fire activities, there were many fire events detected in urban areas, especially in Mexicali. Further studies are needed to understand the nature of urban fires.

4.5 References

- Birant, D., & Kut, A. 2007. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering*, 60(1), 208–221.
<https://doi.org/10.1016/j.datak.2006.01.013>
- Commission for Environmental Cooperation (CEC). 2018. 2010 Land Cover Map. Retrieved from <http://www.cec.org/tools-and-resources/map-files/land-cover-2010-landsat-30m> in October, 2018.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, & U. Fayyad (Eds.), *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, 96, 226–231. AAAI Press.
- Giglio, L., Descloitres, J., Justice, C.O., Kaufman, Y. 2003. An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment*, 87:273-282
- Tavenard, R. 2017. A machine learning toolkit dedicated to time-series data: rtavenar/tslearn. Retrieved from <https://github.com/rtavenar/tslearn> in October, 2018.

CHAPTER 5. DATA AND DETAILED METHODOLOGY IN A FORMAT THAT CAN BE INCORPORATED INTO CES AND EJSM

5.1. Summary

Task 4 of this project was to produce GIS data formatted for direct input to CES and EJSM programming, as well as full documentation of methods and procedures for data collection, validation and modeling. Deliverables produced for Task 4 include a geospatial database and associated files of facilities and emission sources with documentation explaining process and decision making in data compilation, data collection, and data validation. The final metadata digitally submitted to CARB includes the following data files:

1. Geodatabase (arb_v7.gdb)
2. Facility of Interest CSV (foi.csv)
3. Facility of Interest GIS Layer Metadata (foi_metadata.xml)
4. Facility of Interest GIS Layer Metadata (foi_metadata.pdf)
5. Data Validation Procedures Document (data_validation_procedures_foi.docx)
6. Attribute Metadata Descriptions (metadata_foi_v2.docx)
7. SIC to NAICS CSV (sic_naics2017.csv)
8. ArcGIS ArcMap Document (arb_proj12.mxd)

The data validation document explains the procedural steps taken during remote validation. The remote validation framework, developed for this study, can be applied to future studies to improve locational accuracy. The metadata document provides brief descriptions of each attribute field within the final dataset. This includes details on RETC pollutant translation, data integration, data validation, and CES inclusion methods. Item 7 includes NAICS codes and their SIC codes equivalent. This list was generated by Dr. James Sadd and also includes which SIC codes should be used for CES inclusion and which identifies specific SIC's with high emissions.

The Facility of Interest GIS Layer Metadata, initially saved as stored ArcGIS metadata, was exported as a deliverable using ArcGIS 10.5.1. The export process converted the ArcGIS metadata into ISO 19139 XML (eXtensible Markup Language) format as both an .xml and .pdf.

CHAPTER 6. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary and Conclusions

Air toxics emissions sources in Mexico have a clear potential to affect communities and individuals in the US-Mexico border region. For creating databases and maps, a systematic methodology was developed and applied that included satellite imagery, Google Earth verification and groundtruthing was used to verify locations of sources, and a database was created for use in this and future projects. Modeling of potential areas of influence applied to representative sources selected from the mapped and verified locations demonstrated that emissions in Tijuana and Mexicali can affect both California cities on the border and areas farther north into San Diego and Imperial counties. Modeling of potential source regions demonstrate that the border cities of San Ysidro, CA and Calexico, CA are significantly affected by sources in Tijuana and Mexicali, respectively. Areas of San Diego and Imperial Counties are also likely to be affected not only by emission sources on the Mexican side within the 1.5-km buffer of the border, but also from potential regions both inland and off-shore within a 50-km buffer. In terms of EJSM scores, the proximity-based EJSM scores were mostly unchanged following inclusion of sources in Mexico, in part due to the large size of the census tracts which diluted local effects near the border. The exploratory study on burning events on both sides of the border in a 50-km buffer used satellite data and spatiotemporal analytics to determine that wildfires and agricultural burn events over a 20-year period could potentially be detected and mapped by this method. In addition, while this work focused on agricultural-related fire activities, there were many fire events detected in urban areas of Mexicali. Further studies are needed to verify the finding and characterize the nature of urban fires, which could be a potent source of air toxics.

6.2 Recommendations

Recommendations include using the tools developed in this analysis to understand the impact of agricultural and urban burning detected in and near Mexicali on border communities in California, beginning with verifying fire activities mapped by this method through comparison with other methods (such as mapping permitted agricultural burn events in Imperial County). Applying the potential area of influence modeling is a promising new approach for determining the geographic extent of environmental justice communities affected by specific sources. In the border region, this analysis should be expanded, for example, to examine nighttime combustion sources identified previously in Tijuana. The potential source regions identified in this study though this modeling should be combined with gridded emissions inventories to estimate the relative contributions from emission sources in those areas. Additional modeling activities could investigate the impacts of different meteorological conditions and additional sources, including the identification of potential impacts from key sources that were not considered as part of Task 1, especially agricultural burning and dust resuspension emissions. In addition, exploratory analysis presented in this report has identified substantial burning activities in the urban area of Mexicali, and the tools developed in this study can be further applied to better understand the impacts from these sources. The analysis could also be extended to model different baseline years that include La Niña and El Niño events to help understand the variability of the impacts under additional meteorological conditions. The impacts of short-duration but intense southerly wind events on border communities could also be investigated. Additional consideration should be given to determine border-specific approaches to improving CES which take into account differences between California and Baja California sources, for example, landfill emission differences and open burning differences, especially urban burning.

It was anticipated that the hazardous facilities identified and located in this study would be of value to cumulative impacts scoring. The EJSM was able to incorporate them into its method, but only to a limited degree. Both screening methods incorporate results from the US EPA Risk Screening Environmental Index (RSEI) into their scoring process. The EPA implementation of RSEI only considers facilities that report to the EPA Toxic Release Inventory (TRI); in its version 3 update, CES included a special run of the RSEI, which included some facilities located in Mexico. Now that this project has significantly broadened and validated a much larger set of facilities, the RSEI could be rerun including these facilities. This would “correct” (to the degree possible at this time) the limited impact of these facilities on RSEI air pollution estimates in San Diego and Imperial Counties, and result in better data for screening scores.

Glossary of Terms, Abbreviations, and Symbols

BCEI Baja California Emissions Inventory

CARB California Air Resources Board, California Environmental Protection Agency

CES CalEnviroScreen

EJSM Environmental Justice Screening Model

FOI Facility of Interest

MNEI Mexico National Emissions Inventory

NAICS North American Industry Classification System

NRA: Número de Registro Ambiental, or Environmental Registration Number

OEHHA Office of Environmental Health Hazard Assessment, California Environmental Protection Agency

PAI Potential Area of Influence

PES Potential Emission Sensitivity

POE Port of Entry

PSR Potential Source Region

RETC The Registry of Pollutant Release and Transfer (Registro de Emisiones y Transferencia de Contaminantes), Mexico

SIC Standard Industrial Code

SEMARNAT Secretariat of Environment and Natural Resources (Secretaría de Medio Ambiente y Recursos Naturales), Mexico

Appendix A

(Chapter 2) Transport Events and Model Evaluation

The model performance was evaluated to ensure that the local meteorology was represented in the simulations with sufficient accuracy in order to conduct the trajectory analysis. The evaluation focused on the high-resolution domains (D4 and D5) for the selected meteorological episodes with prevalent southern wind component using several statistical criteria. The data obtained from the ambient monitoring stations in the San Diego-Tijuana and Calexico-Mexicali regions were used to evaluate the wind speed and wind direction fields from the WRF model. However, additional fields, including, surface radiation and planetary boundary layer height, were examined to ensure consistency and representativeness of the model outputs. The model had an acceptable performance in representing both wind direction and wind speed during the selected periods, since the median errors of bias and the Centered Root Mean Squared Error (RMSEc) were relatively low. In addition, the overall distribution of all days comprising the selected events was adequately reproduced by the model for both regions; indicating that the meteorological fields from the WRF model can be used to drive the trajectory modeling analyses.

A1. Quality Control

The model evaluation was conducted for surface wind direction and wind speed. All observational datasets were processed and flagged with quality assurance and quality control (QA-QC) operations to remove outliers, spurious and missing data. Basic QA-QC of SMN data was performed by removing spurious data and outliers. As a final step, all data was homogenized to a common format.

The model outputs from WRF were checked for consistency and integrity to ensure representativeness of the model results. This step included visual inspection of key meteorological fields; including surface temperature, surface radiation and planetary boundary layer height. Since we generated high-temporal resolution fields (output every 15 minutes), the large database (~ 77,000 files) was continuously compressed, uncompressed and transferred among storage devices; so that additional checking of file size, model header and homogeneity of data sets on storage was also performed.

A similar procedure was conducted for the FLEXPART-WRF fields as well by checking the file structure, header and file size, even though the NETCDF format (compressed NETCDF 4) and file structure are different from those of WRF.

A2. Observations datasets

The meteorological observations used for evaluating the WRF model performance included data from the Air Quality System of EPA (AQS-EPA), the Air Quality and Meteorological Information System from CARB (AQMIS-CARB), and the Mexican National Weather Service (Servicio Meteorológico Nacional from Mexico, SMN-Mexico). After the QC of the databases, 13 stations were selected within the modeling domain covering the San Diego–Tijuana region (D4), two stations from the AQS-EPA

system (AQS-1002 and AQS-1008); nine stations from the AQMIS-CARB system (CARB-2933; CARB-3648; CARB-3690; CARB-3739; CARB-3766; CARB-5792; CARB-5795; CARB-5815) and one station from the NOAA's National Estuarine Research Reserve System's (NERRS) (NR-tjrtlmet). For the modeling domain covering the Calexico – Mexicali region (D5) we selected 5 monitoring stations, four from the AQMIS-CARB system (CARB-3135; CARB-5735; CARB-2551; CARB-5747) and one from the Mexican SMN (EM- MXCL).

A3. Statistical Metrics

The following error metrics were calculated to evaluate the model performance: Centered Root Mean Squared Error (RMSEc) for the wind velocity, the vector Root Mean Squared Error (RMSEvec) of the zonal and meridional wind components, and bias and the Pearson correlation for the wind speed and wind velocity. The RMSEc is the error after removing the bias and shows the difference around the mean. It can also be considered as the random component of the root mean squared error. The bias is the systematic component of the error (Draxl, 2015). The RMSEvec is useful to evaluate both wind speed and wind direction since it is based on the wind components. The hourly averages of wind velocity and wind direction for the model outputs were obtained using the 15 min fields of the meridional and zonal components at 10-m agl in order to have the same time interval as in the observed datasets. Additional wind roses were plotted to display whether the prevailing wind conditions were reproduced in the model.

A4. Model Evaluation

The statistical metrics were obtained for the domains with highest spatial resolution (D4 and D5). Due to the discontinuity in the temporal dimension for the days comprising the selected meteorological episodes, the error metrics were calculated in a 24-hour period for each monitoring station and further aggregated spatially and temporally using boxplots. Therefore, evaluation of the overall performance for both border regions is presented instead of a detailed performance of each station.

Figures A1 to A6 show the boxplots of bias, RMSEc, RMSEvec, and Pearson's correlation coefficient for wind speed and wind velocity at 10 m agl. In general, the WRF model tended to over-predict the wind velocity and wind direction in both geographical regions. The mean bias of wind velocity and wind direction was less than 1 m/s and less than 10 degrees, respectively. However, the model presented more variability in the bias of wind speed in the San Diego–Tijuana area. In terms of the correlation coefficient, the evaluation suggests a significant difference in the wind patterns between the two regions with an apparently lower performance in the Calexico– Mexicali region, which presented a mean correlation of about 0.2 for wind speed, compared to a mean correlation of about 0.65 in the San Diego-Tijuana region. Similar values were obtained for the correlation of wind direction.

In general, the wind direction in the Calexico-Mexicali region presented a bimodal distribution which the model reproduced (not shown). The quantile-quantile plots

(Figures A7 and A8) show that the model produced a higher incidence of directions up to 120°, and a lower incidence of directions greater than 200°. For the San Diego-Tijuana region, a similar pattern was found, but for this region the distribution of wind speed was not bimodal. In addition, the model tended to over-predict the wind velocity for values greater than 5 m/s in both regions. This is reflected in the RMSEvec, which accounts for both wind speed and direction with a median error around 2.5 to 3 m/s for each region respectively and the variability in the error was similar for both regions. A similar result is obtained with the RMSEc with a median error of about 1 m/s for both regions; nevertheless, the variability in the RMSEc was lower.

The distribution of observational wind speed data between the two regions was similar but with higher wind velocities in the San Diego-Tijuana region. In contrast, the distribution of observational wind direction was different between the regions, with more variability in the San Diego-Tijuana region, implying that the transport dynamics in these regions presented important differences (Figures A9 and A10). The air masses coming from the south to San Diego during the selected episodes were gradually transported and the shifts in direction spanned a wider range of wind direction at a slightly lower speed, and greatly influenced of the complex terrain, most likely by downslope flow from the mountain ridge to the east. In contrast, the air masses coming from the south to the Calexico area during the selected episodes were transported at a relatively higher speed but in a smaller range of wind direction, so that faster shifts in transport likely occurred in less time and probably were not reproduced in the model, even though the terrain in the Calexico area is not complex.

Despite the low correlation values, in general, the WRF model performance was reasonably good. Both the median errors of bias and RMSEc were relatively low; and the wind roses in Figures A11 and A12 show that the overall distribution during all days comprising the selected events was adequately reproduced by the model for both regions. Thus, the low correlations indicate that the timing of the events was not always captured by the model. The time series of wind velocity and wind direction for both regions were plotted (not shown) and the diurnal cycle was found to reproduce most of the stations but with a slightly shifted phase. This could be enhanced by the difference in transport dynamics between the regions. In addition, the sample size was different in these regions. There were more days in the set of selected events with a prevailing southern component (77 days) and less data from monitoring stations to compare against (5 sites). For some stations, model data at 10-m were used and compared with surface observations.

Even though the magnitude of the median RMSEc is slightly greater than the median of the wind speed bias, which could suggest that the random error could contribute more to the discrepancies between observations and model results, it is likely that the systematic error could have an important contribution as well; therefore, a change in model parameterizations is recommended for future study, in addition to a comprehensive sensitivity analysis of land surface models, planetary boundary layer schemes and number of vertical levels. Furthermore, observation nudging of wind speed and wind velocity could improve the boundary and initial conditions as well as the vertical and surface performance.

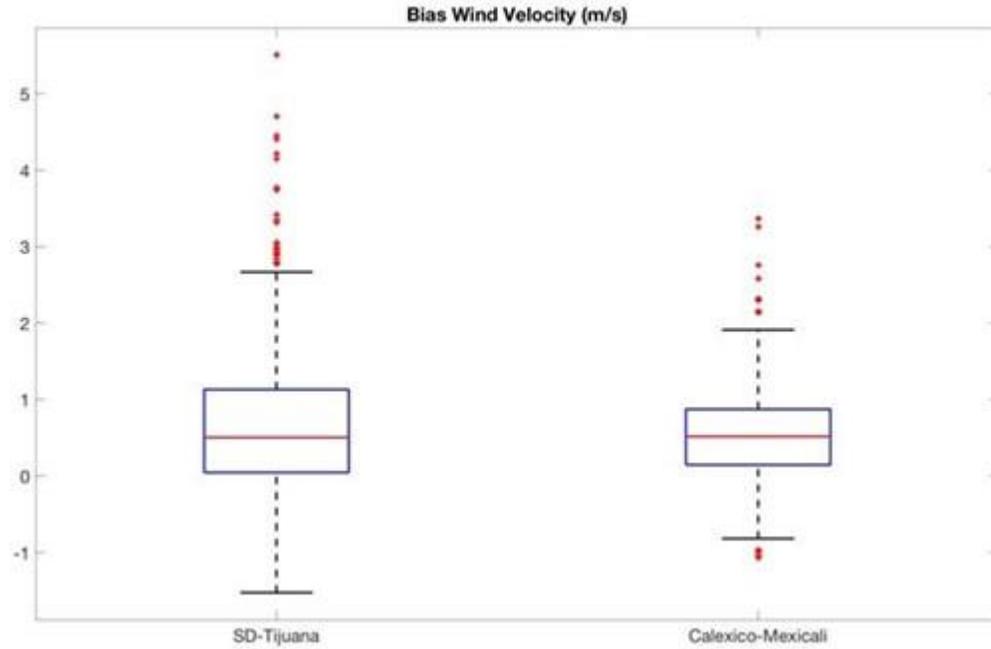


Figure A1. Boxplots of wind velocity bias (m/s) for the San Diego–Tijuana (left) and Calexico– Mexicali (right) regions during all the selected episodes. The median is shown as the red horizontal line.

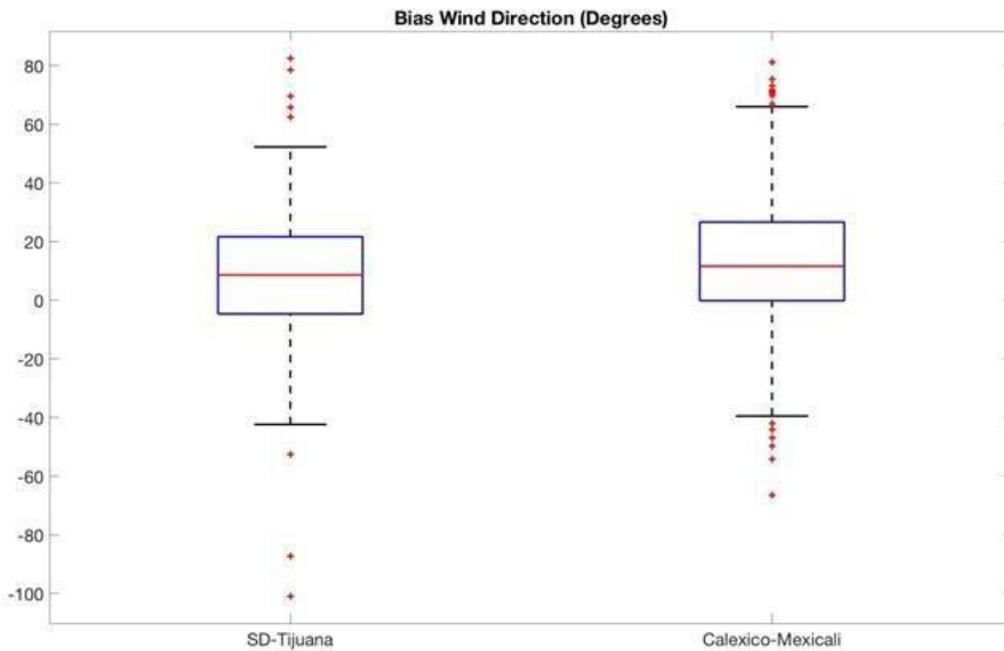


Figure A2. Boxplots of the wind direction bias (degrees) for the San Diego–Tijuana (left) and Calexico–Mexicali (right) regions during all the selected episodes. The median is shown as the red horizontal line.

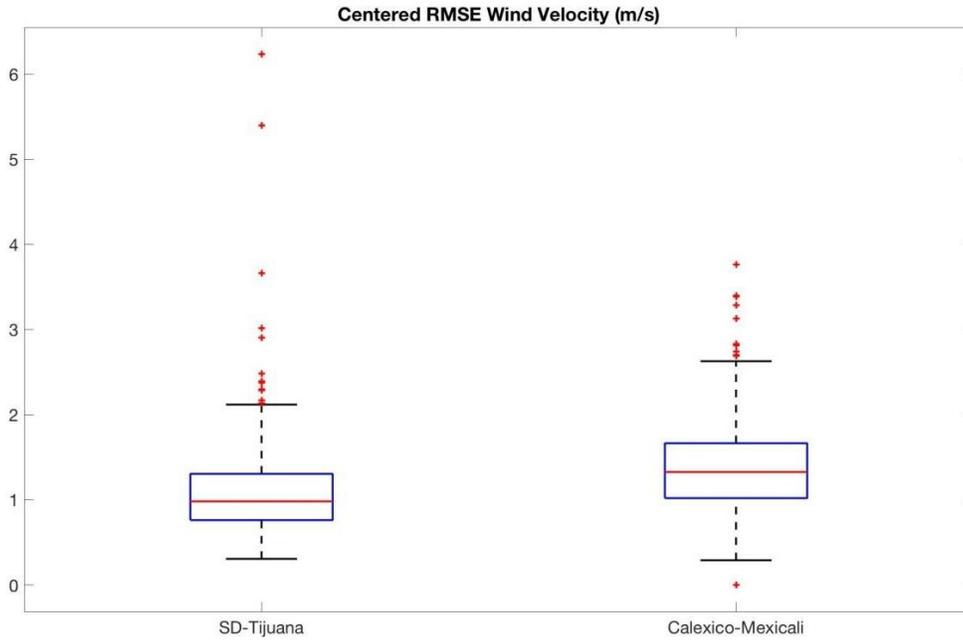


Figure A3. Boxplots of the centered Root Mean Square Error (RMSE_c) in wind velocity (m/s) for the San Diego–Tijuana (left) and Calexico–Mexicali (right) regions during all the selected episodes. The median is shown as the red horizontal line.

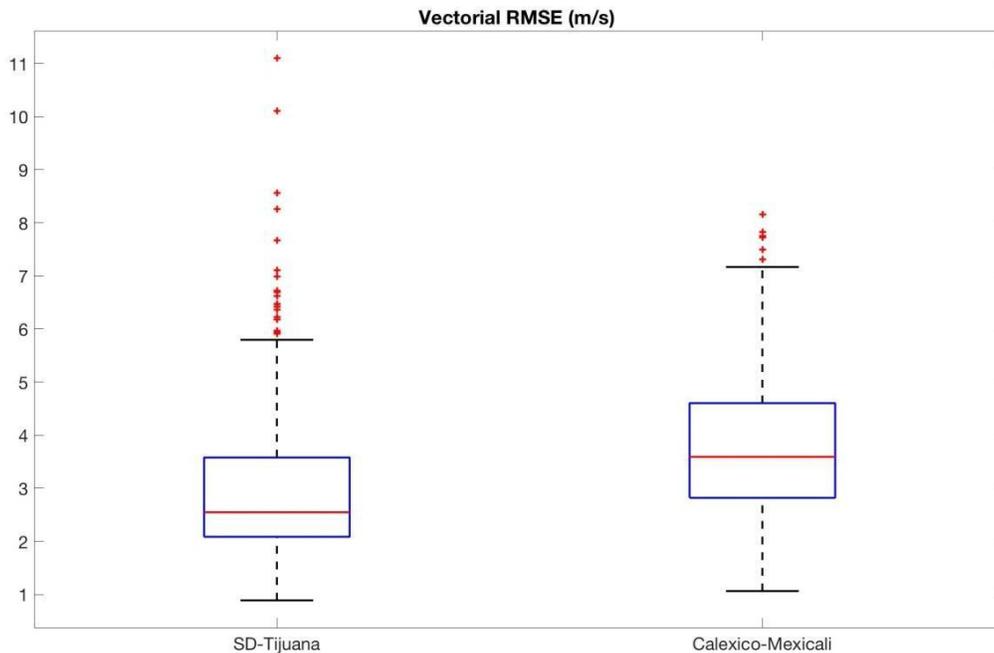


Figure A4. Boxplots of the vector Root Mean Square Error (RMSE_v) in wind velocity (m/s) for the San Diego–Tijuana (left) and the Calexico–Mexicali (right) regions during all the selected episodes. The median is shown as the red horizontal line.

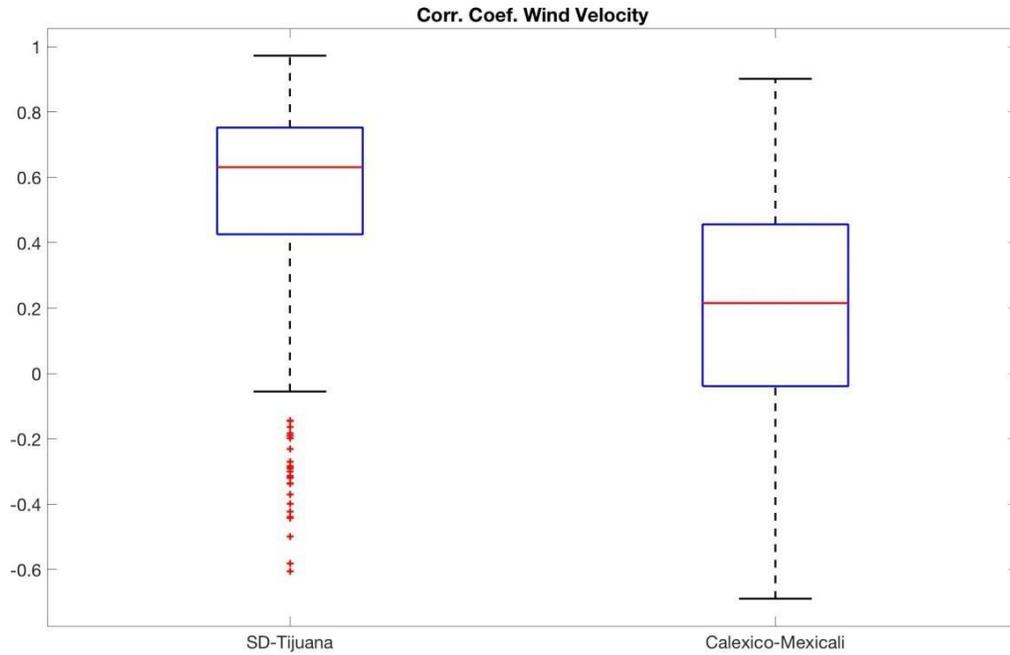


Figure A5. Boxplots of the wind velocity correlation coefficient for the San Diego–Tijuana (left) and the Calexico–Mexicali (right) regions during all the selected episodes. The median is shown as the red horizontal line.

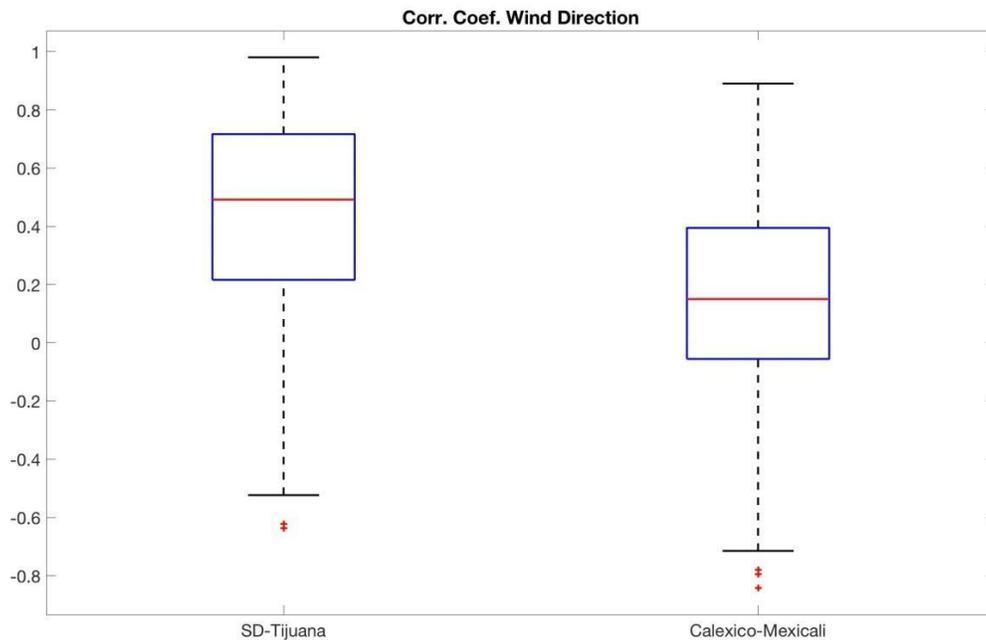


Figure A6. Boxplots of the wind direction correlation coefficient for the San Diego–Tijuana (left) and Calexico–Mexicali (right) regions during all the selected episodes. The median is shown as the red horizontal line.

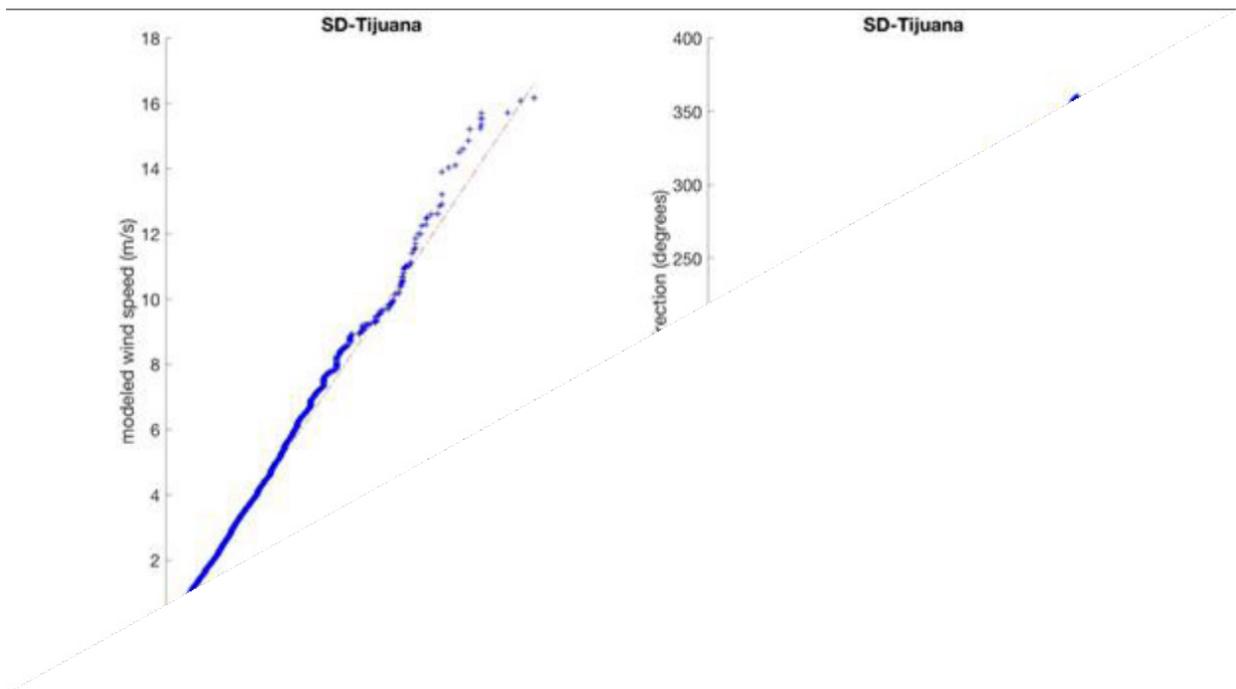


Figure A7. Quantile-quantile plots of wind speed (left) and wind direction (right) for the San Diego–Tijuana region during all the selected episodes.

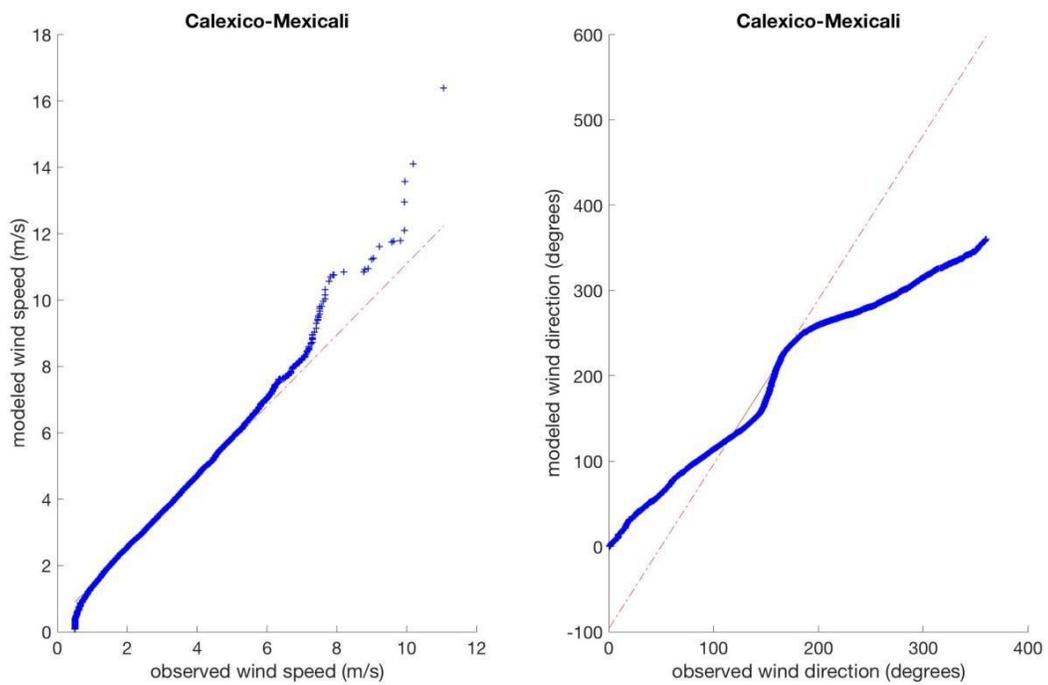


Figure A8. Quantile-quantile plots of wind speed (left) and wind direction (right) for the Calexico- Mexicali region during all the selected episodes.

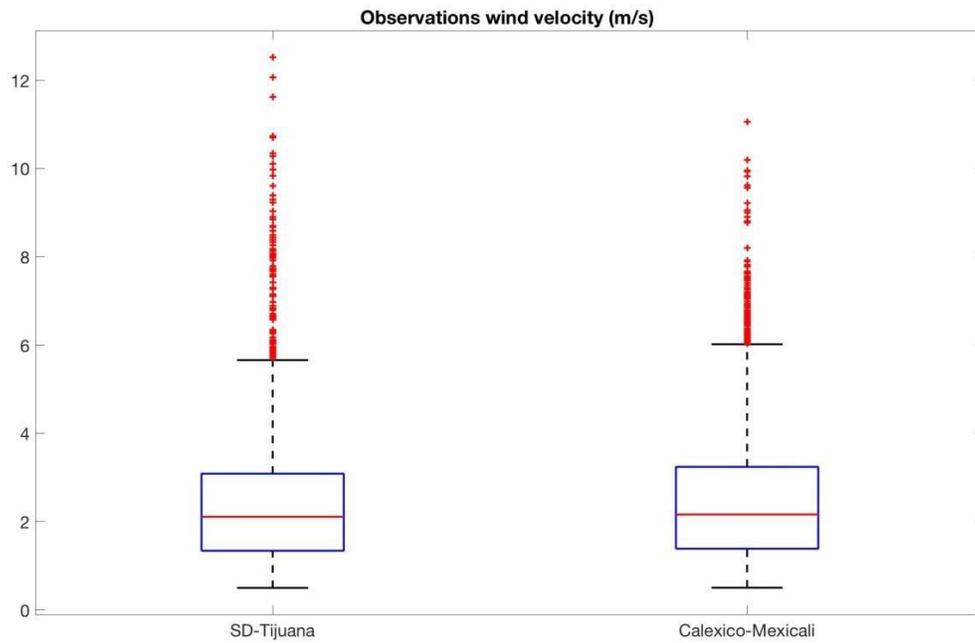


Figure A9. Boxplots of the wind velocity observations for the San Diego–Tijuana (left) and Calexico–Mexicali (right) regions during all the selected episodes. The median is shown as the red horizontal line.

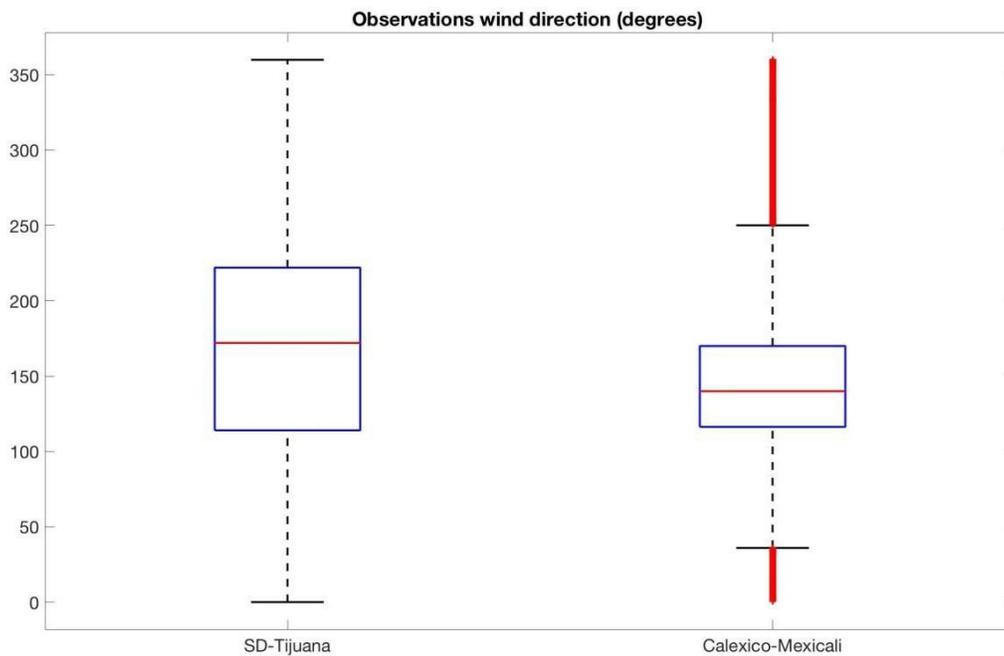


Figure A10. Boxplots of the wind direction observations for the San Diego–Tijuana (left) and Calexico–Mexicali (right) regions during all the selected episodes. The median is shown as the red horizontal line.

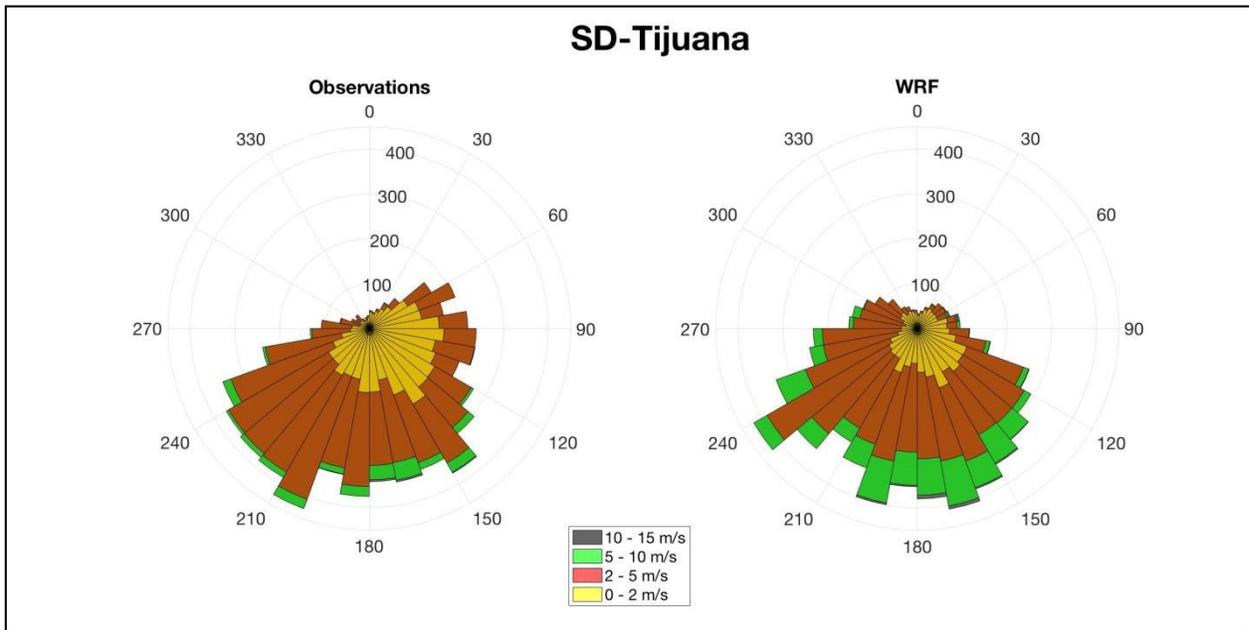


Figure A11. Wind roses for observations (left) and modeled (right) wind speeds for the San Diego–Tijuana region during all the selected episodes.

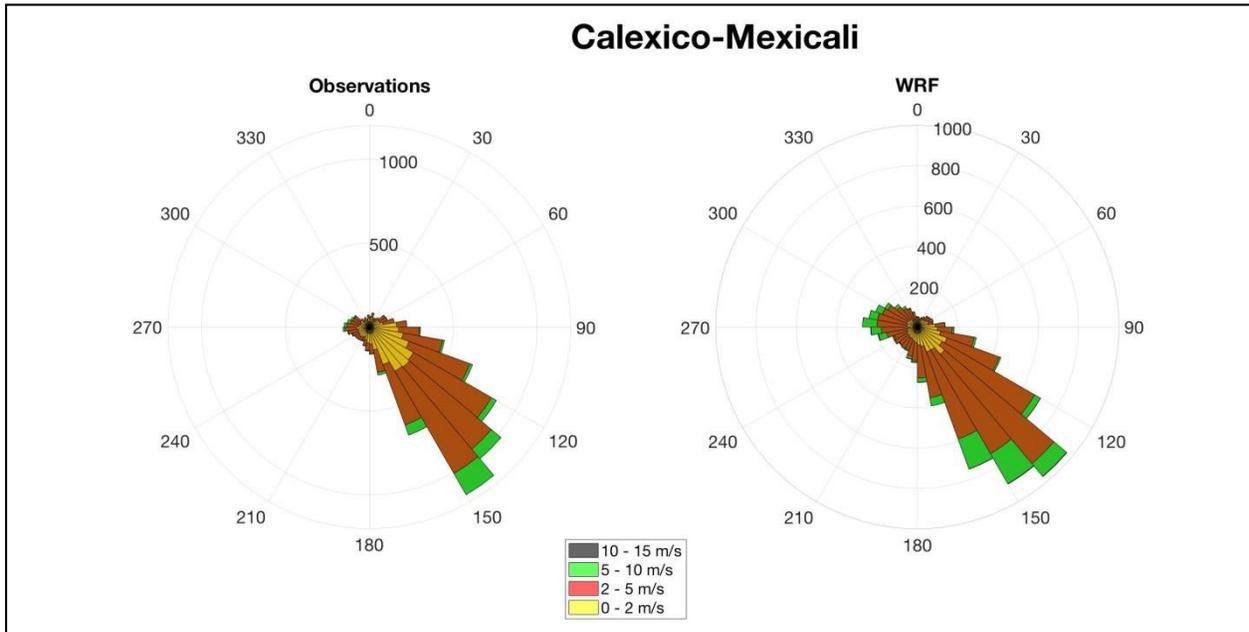


Figure A12. Wind roses for observations (left) and modeled (right) wind speeds for the Calexico–Mexicali region during all the selected episodes.