A Method for Quantifying Historical Air Quality in Unmonitored Regions Using Statistical Relationships Developed from Regional Air Quality Model Output

David Nunes, San Joaquin Valley Air Pollution Control District, Fresno, CA

14th Annual CMAS Conference
October 5-7, 2015
Chapel Hill, NC

1. Introduction

Overview

To provide residents with air quality trends in unmonitored locations, the San Joaquin Valley Air Pollution Control District (SJVAPCD or District) estimated daily maximum 8-hr average ozone and 24-hr average PM2.5 for the past 20 years across the entire 23,700 square miles (58,600 km2) of the District. Using CMAQ model output and spatial regression, relationships were established between monitored and unmonitored locations at a "neighborhood" resolution of 4km. These daily estimates were then summarized and made available to the public in a web-based tool.

Figure 1.1: District

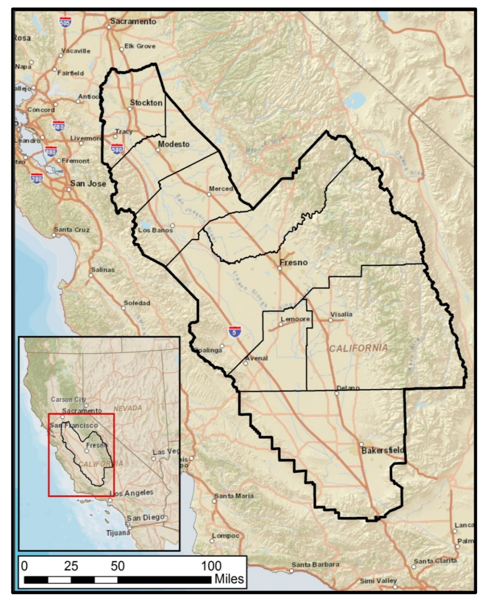
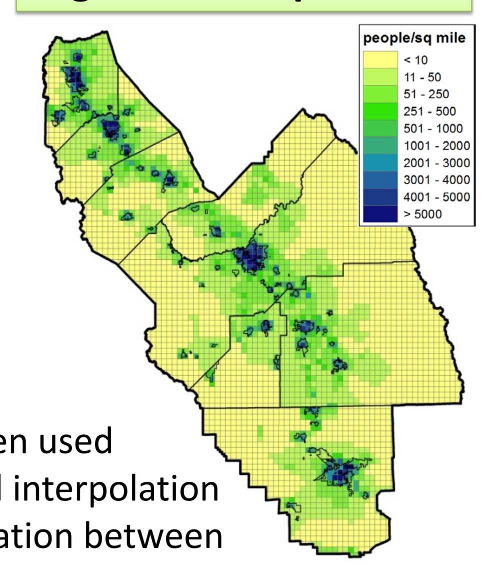


Figure 1.2: Population



Background

Monitoring networks are typically designed to measure air quality in densely populated areas or at peak locations. Often these networks are inadequate for informing all residents about air quality in their specific city or neighborhood.

Spatial interpolation and regional models are often used to estimate air quality between monitors. Spatial interpolation is quick, but may be inadequate at capturing variation between monitors. Regional models are better at capturing variation. However, they are complex, computationally expensive, need historical emissions inventories and require bias correction to match observations.

2. Methodology

Neighborhood Development

The District was divided into 3,880 "neighborhoods" using the 4x4km grid cell boundaries established by CMAQ modeling. Of these 3,880 neighborhoods, 1,827 have estimated 2010 populations to be greater than 50 people.

Site Selection and Observation Data

Daily maximum 8-hr average ozone and 24-hr average PM2.5 observations were acquired from the EPA Air Quality System (AQS). Ozone and PM2.5 monitoring sites were selected with long periods of operation and availability of future real-time observations.

Some monitoring sites were combined to create longer observation records (e.g., Fresno-First and Fresno-Garland).

Regression equations were developed to estimate missing observations by using other sites in the network as independent variables.

Statistical Equation Development and Application

Using one year of CMAQ model output (2007), equations were developed for grid cells without monitoring sites using grid cells with monitoring sites as the independent variables. R was utilized for this task using the leaps function to evaluate all possible equations for each grid cell. Bayesian Information Criteria (BIC) was used for model selection with a limit of seven variables per equation.

Using regression equations developed from the CMAQ model output, 8-hr average ozone observations from 1992 to 2014 and 24-hr average PM2.5 observations from 2002 to 2014 were used to estimate daily air quality for all neighborhoods. For years when both PM2.5 and ozone were estimated, Air Quality Index (AQI) was also calculated. To increase confidence in results, neighborhoods with adjusted R^2 less than 0.75 were excluded.

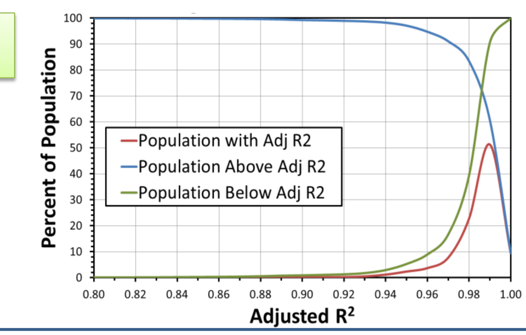
CMAQ modeling was completed by the California Air Resources Board (CARB) as part of the District's *2013 Plan for the Revoked 1-Hour Ozone Standard* (adopted September 2013) and the District's *2012 PM2.5 Plan* (adopted December 2012).

3. Evaluation

8-hr Ozone

- Regression equations developed from 23 grid cells representing 27 monitoring sites
- Daily estimates for all grid cells from 1992 to 2014
- **Figure 3.2** shows ***Madera-Pump*** and ***Santa Rosa Rancheria*** regression equations compared to CMAQ modeling and observations. These sites have long-term observations, but were not included in model development.
- For comparison, the City of ***Los Banos*** is also included in **Figure 3.2** as a city of interest near the limits of model performance.

Figure 3.3: Ozone R^2 by Population



- 97% of residents live in grid cells with R^2 of at least 0.95
- 99% of residents live in grid cells with R^2 of at least 0.90

Figure 3.1: 8-hr Ozone R^2

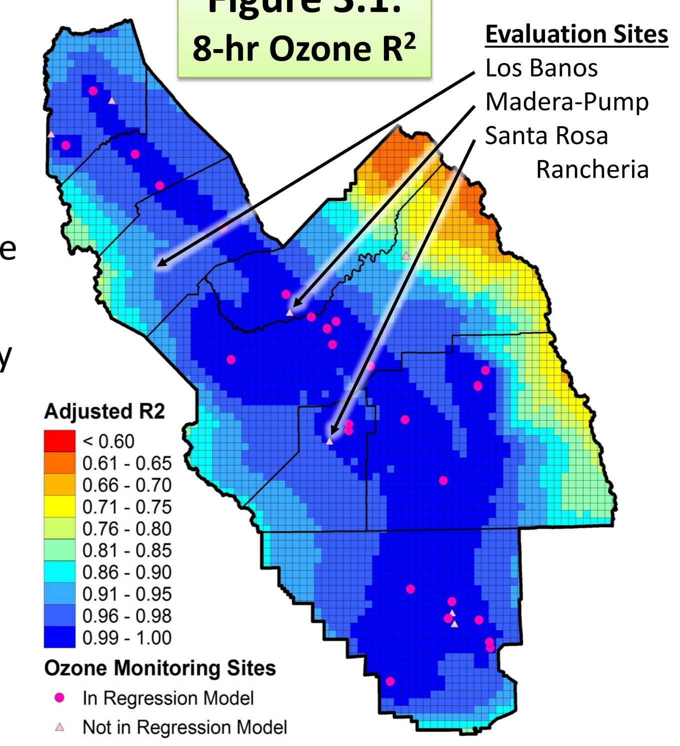
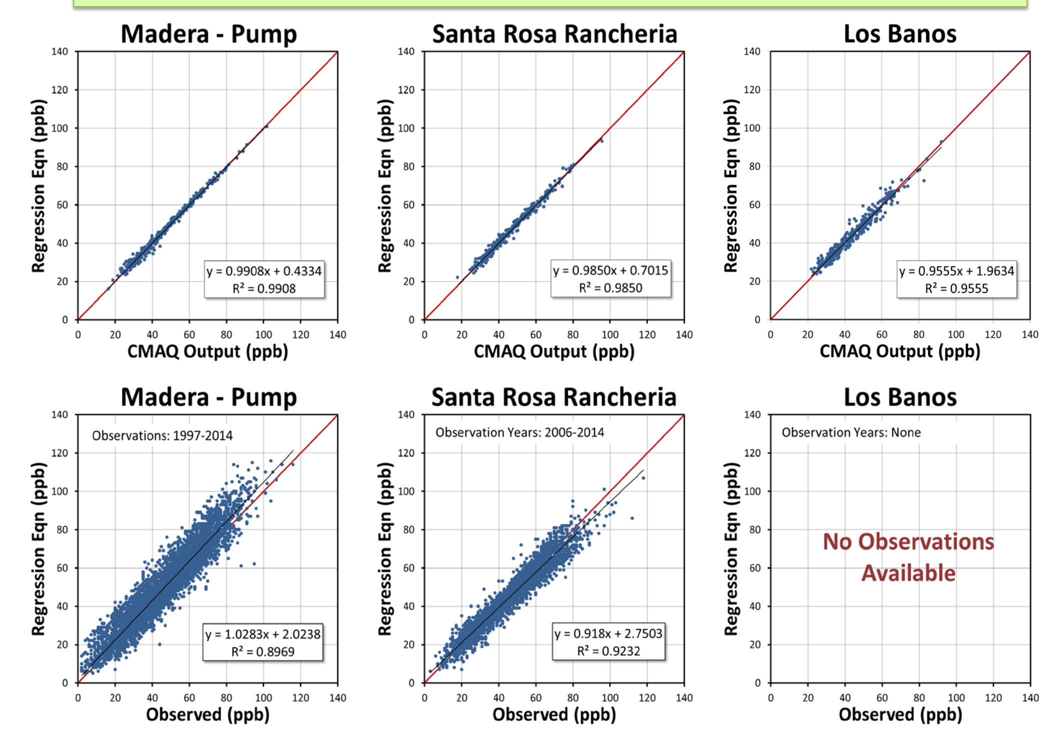


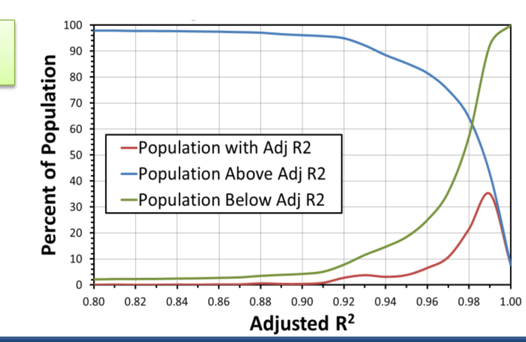
Figure 3.2: 8-hr Ozone Estimate Verification



24-hr PM2.5

- Regression equations developed from 16 grid cells representing 18 monitoring sites
- Daily estimates for all grid cells from 2002 to 2014
- **Figure 3.5** shows ***Corcoran*** and ***Fresno-Pacific*** regression equations compared to CMAQ modeling and observations. These sites have long-term observations, but were not included in model development.
- For comparison, the City of ***Los Banos*** is also included in **Figure 3.5** as a city of interest near the limits of model performance.

Figure 3.6: PM2.5 R^2 by Population



- 85% of residents live in grid cells with R^2 of at least 0.95
- 96% of residents live in grid cells with R^2 of at least 0.90

Figure 3.4: 24-hr PM2.5 R^2

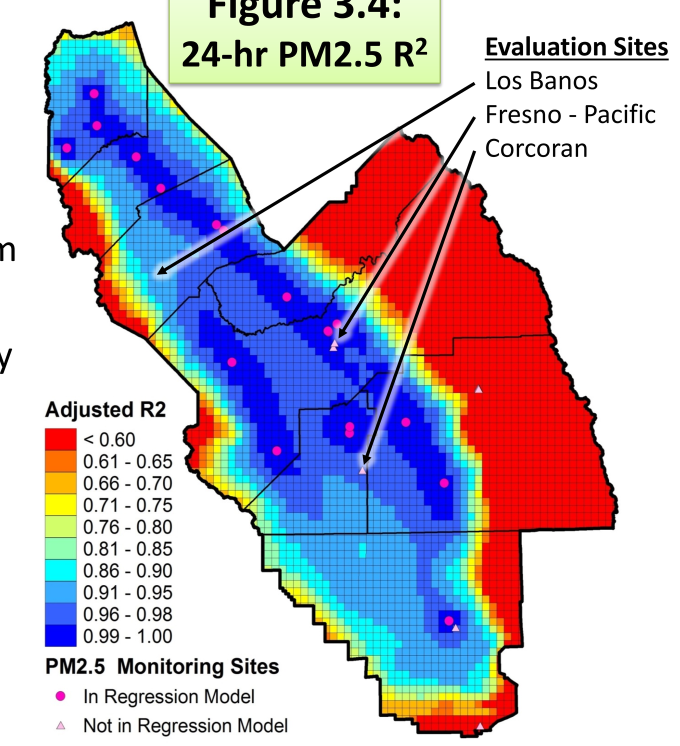
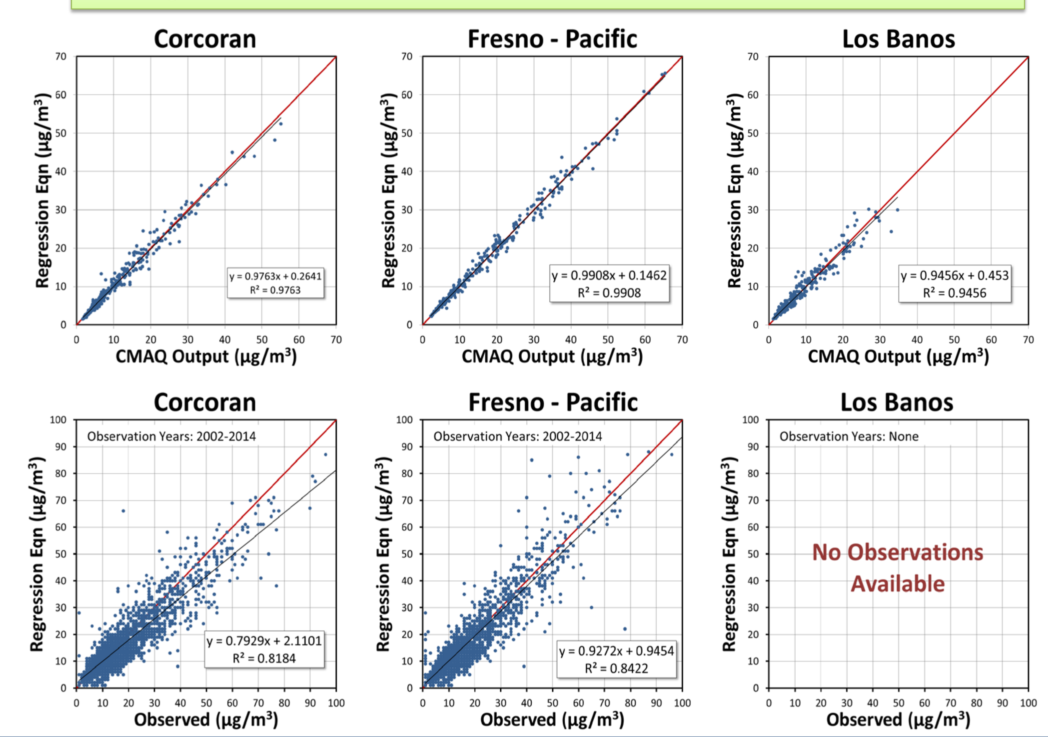


Figure 3.5: 24-hr PM2.5 Estimate Verification



4. Air Quality Trends

Figure 4.1: Days with 8-hr ozone greater than 75 ppb

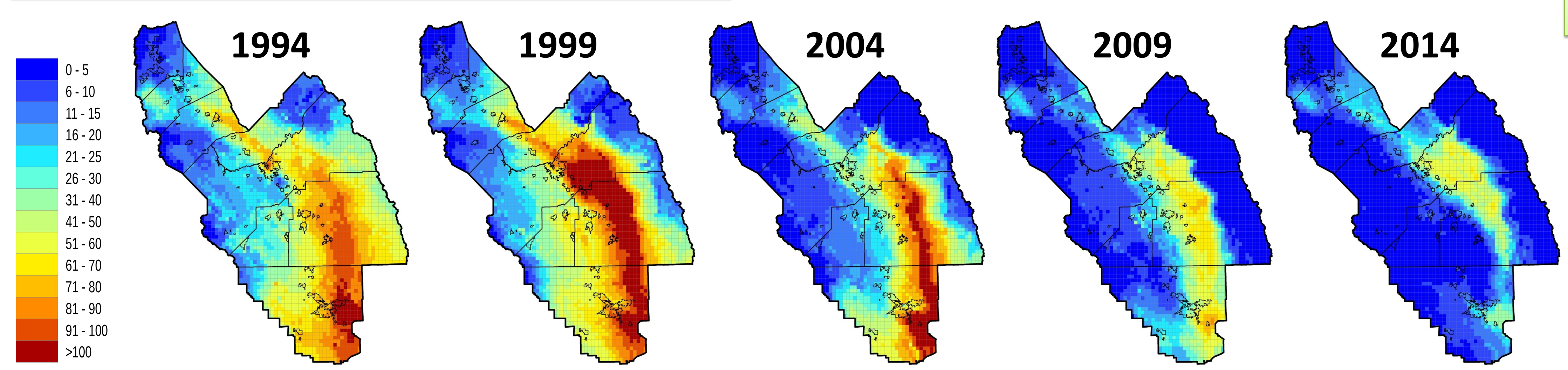


Figure 4.2: Days with 24-hr PM2.5 greater than 35 µg/m3

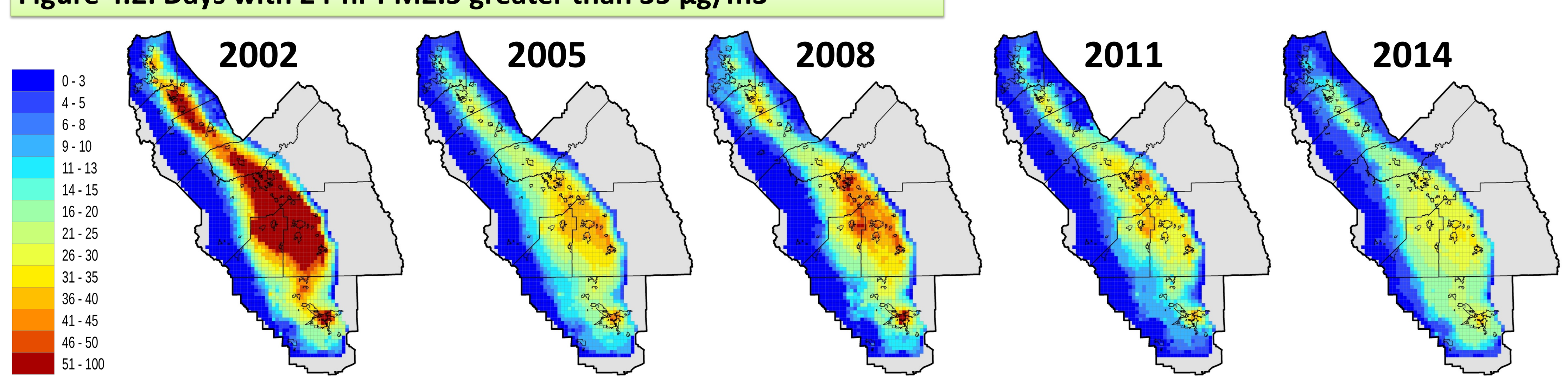
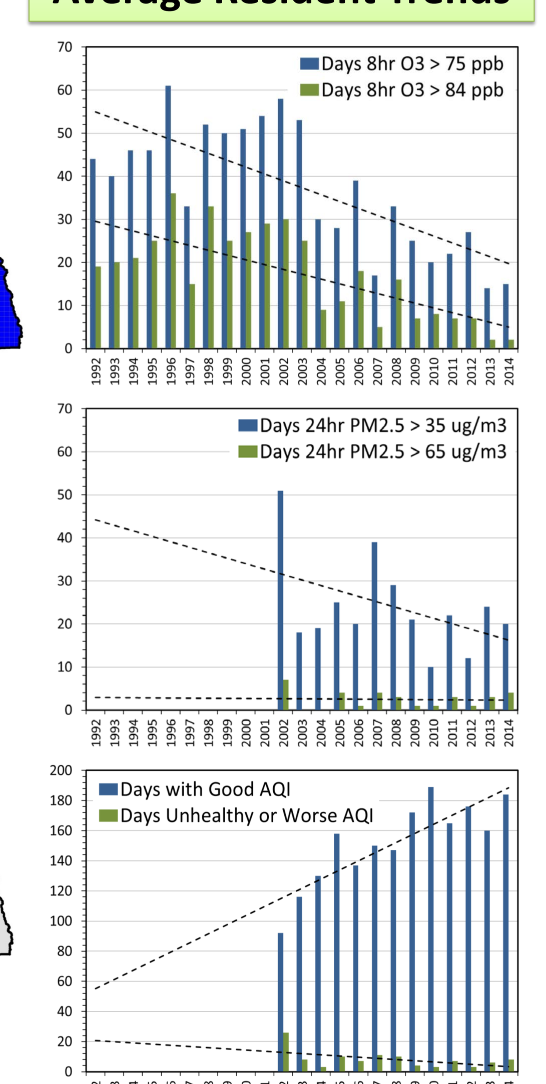


Figure 4.3: Average Resident Trends



5. Online Public Access

Daily estimates of 8-hr ozone, 24-hr PM2.5 and Air Quality Index (AQI) were summarized and placed into a web-based tool for retrieval. By entering an address, residents can compare their historical neighborhood air quality to county and district summaries.

The Web-based Archived Air Quality (WAAQ) System is available at: http://www.valleyair.org/waaqs

Figure 5.1: Public Access to Neighborhood Trends

6. Conclusion

Model Development

Regression equations do very well reproducing CMAQ model output in populated areas.

- 99% of Valley residents live in locations with R^2 greater than 0.90 for 8-hr ozone.
- 96% of Valley residents live in locations with R^2 greater than 0.90 for 24-hr PM2.5.

Reproducing Observations

Regression equations are successful at reproducing observations at sites with historical data.

- Estimates compared to observations generally have an R^2 near 0.90 for 8-hr ozone and 0.83 for 24-hr PM2.5
- Reduced performance could be due to:
  - Distance from existing monitors
  - Changing source to receptor relationships
  - Emission changes over time
  - Local impacts not modeled (e.g., fires, dust, etc.)
  - CMAQ Model performance
  - Equipment variations

Challenges

Significant challenges remain, including how to handle local emissions (e.g., fires, blowing dust) that may impact a limited number of monitors or unmonitored areas only.

7. Acknowledgements

Seyed Sadredin (APCO) and the San Joaquin Valley Air Pollution Control District Board for direction and resources.

Jeff Daniels, Seng Lee and Anton Simanov for developing the online interface.

Pete Biscay, Shawn Ferreria, Jon Klassen, Jennifer Ridgway and Stephen Shaw for insight, review and comments.

California Air Resources Board for CMAQ modeling and output.