

VERIFYING REDUCTIONS IN CO₂ EMISSIONS GENERATED BY THE AMERICAN ELECTRIC POWER COMPANY (AEP) IN OKLAHOMA FROM 2007 TO 2008 BASED ON 2006 LEVELS

Luke Armbruster
ENGR 326
Final Project

ABSTRACT

The American Electric Power Company's (AEP) CO₂ emissions within Oklahoma are determined from 2007 to 2008 based on 2006 levels. A fine against AEP is calculated for emitting above their allotment. The calculation of the company's emissions requires the correction for an offset in the continuous mixing ratio measurements taken by the Precision Gas System (PGS) at the U.S. Department of Energy's (DOE) Atmospheric Radiation Measurement (ARM) Climate Research Facility South Great Plains (SGP) site. PGS measurements are corrected to less frequent flask measurements taken by the National Oceanic & Atmospheric Administration. The corrected PGS measurements includes the implementation of a somewhat effective filter for high atmospheric variability. Using cubic spline and linear interpolation respectively, two separate calculations of the corrected PGS measurements are calculated. Both corrected PGS measurements reduce the mean difference between the nearest PGS to NOAA flask measurements. Also the performance of the corrections is similar based on qualitative and quantitative results. A sensitivity analysis shows shifting the diurnal bounds on the corrected PGS mixing ratio measurements by +1 hour and varying matched NOAA flask measurements by -0.2 ppm results in the greatest change from the original fines.

TABLE OF CONTENTS

| | |
|---|----|
| 1. Introduction | 1 |
| 2. Literature Review | 3 |
| 2.1. Planetary Boundary Layer Dynamics and Measurements..... | 3 |
| 2.1.1 Planetary Boundary Layer Dynamics | 3 |
| 2.1.2. PBL Measurements..... | 5 |
| 2.2. The South Great Plains Site..... | 6 |
| 2.3. Inverse Framework for Quantifying greenhouse Gas Emissions..... | 6 |
| 2.4. Case Study in Interpolating Missing Air Quality Data..... | 8 |
| 2.5. Current Pollutant Discharge regulations..... | 10 |
| 2.5.1. EU Emission Trading Scheme | 10 |
| 2.5.2 Emission Violation Case Study..... | 10 |
| 2.6 AEP Coverage Within Oklahoma..... | 11 |
| 3. Methodology | 12 |
| 4. Application | 17 |
| 5. Results and Discussion | 19 |
| 6. Conclusion | 25 |
| 7. References..... | 26 |
| 8. Appendix | 29 |
| 8.1 Inverse Modeling Framework | 29 |
| 8.1.1. Inverse Framework from Case Study | 29 |
| 8.1.2. Inverse Framework Specific to Firm's Investigation | 30 |
| 8.2 Format of In Situ Measurements Taken..... | 31 |
| 8.3 Main Assumptions | 31 |
| 8.4 Project Time and Costs..... | 32 |
| 8.5 R Script Including Developed Cubic Spline Numerical Method in Bold..... | 33 |

LIST OF FIGURES

| | |
|---|----|
| FIGURE 1: THE IDEAL DIURNAL BOUNDARY LAYER HEIGHT VARIABILITY (KAIMAL, 1994).. | 5 |
| FIGURE 2: LIVESTOCK EMISSIONS APPEAR TO BE GREATER THAN INVENTORY ESTIMATES AS SHOWN FROM THE A PRIORI SCALING FACTOR IN DARK GREY AND POSTERIOR SCALING FACTOR IN LIGHT GREY (ZHAO, 2009)..... | 8 |
| FIGURE 3: PERFORMANCE OF CUBIC SPLINE DROPS QUICKLY OVER TIME. CUBIC SPLINE IS DOTTED LINE, NEAREST NEIGHBOR IS DASHED LINE, AND LINEAR IS SOLID LINE (JUNNINEN, 2004)..... | 9 |
| FIGURE 4: MATCHED NOAA AND PGS SIGNALS REMOVED AFTER FILTERING FOR HIGH ATMOSPHERIC VARIABILITY USING THE TWO CRITERIONS..... | 20 |
| FIGURE 5: THE PERFORMANCE OF THE CUBIC SPLINE INTERPOLATION OF THE SIGNAL DIFFERENCE OR THE RESIDUAL INTERPOLATION SHOWN IN THE TOP GRAPH APPEARS SIMILAR TO LINEAR INTERPOLATION SHOWN IN BOTTOM GRAPH. THE RED LINE REPRESENTING THE CUBIC INTERPOLATION IS A LINE FIT THROUGH THE CUBIC SPLINE INTERPOLATED VALUES, AND THEREFORE THE FIRST DERIVATIVE OF CUBIC SPLINE APPEARS DISCONTINUOUS AS LINEAR INTERPOLATION. BOTH FIRST AND SECOND DERIVATIVES ARE CONTINUOUS FOR CUBIC SPLINE INTERPOLATION (CHAPRA, 2006). | 21 |
| FIGURE 6: MONTHLY MEAN OF CORRECTED HOURLY PGS MIXING RATIO MEASUREMENTS ALWAYS APPEAR HIGHER THAN NOAA'S GLOBAL MONTHLY AVERAGES..... | 23 |

LIST OF TABLES

| | |
|---|----|
| TABLE 1: FINES AGAINST AEP FOR VIOLATING 2006 MONTHLY EMISSION ALLOTMENTS ARE SIMILAR USING LINEAR OR CUBIC SPLINE TO CORRECT PGS SIGNALS. | 24 |
| TABLE 2: VARYING NOAA FLASK MEASUREMENTS HAS GREATEST IMPACT ON FINAL FINE AGAINST AEP | 24 |
| TABLE 3: THE EFFECT OF SHIFTING THE BOUNDS ON THE TIME INTERVAL EACH DAY WHEN HOURLY PGS SIGNALS ARE CORRECTED ARE SHOWN WITH THE +1 HOUR SHIFT SHOWING THE GREATEST IMPACT ON FINAL FINE AGAINST AEP. | 25 |

1. INTRODUCTION

Scientists are projecting significant changes in climate that will affect everyone. Both an increase in frequency and intensity of events such as floods and droughts are expected. Scientists believe that human beings are contributing significantly to climate change by emitting greenhouse gas emissions such as CO₂, CH₄, N₂O (Solomon Et al., 2007).

Presently a binding international agreement to reduce greenhouse emissions has not been reached. Instead, individual regional authorities are committing to reducing their emissions in their own jurisdictions. Oklahoma is one state where the law enforces a cap of anthropogenic emissions to 2006 levels. As a result, each anthropogenic source is given an emission allotment corresponding to the amount the source emitted for each month in 2006. Any anthropogenic source that exceeds their allotted emissions is penalized appropriately.

The current State of Oklahoma's monthly inventory of anthropogenic emissions by source assumes each source uses the monthly emission allotments based on 2006 levels. By law, the state's Department of Environmental Quality must verify that the inventory of anthropogenic emissions from each source are correct i.e., if in fact some sources emitted more or less than their allotment. If sources emit more than their allotment they are fined appropriately.

The Department contracted the verification of CO₂ reductions from 2007 to 2008 to a private environmental engineering firm. Tulsa Environmental Engineering Associates won the bid to verify inventory emissions estimates by month for the time period of interest. The firm decided to use an inverse modeling framework to verify reductions. The method uses continuous in situ mixing ratio measurements or signals, volume of gas to volume of dry air, in ppm of CO₂, and minimizes the difference with predicted measurements generated by the inverse modeling framework. The framework consists of inventory estimates, a biospheric model, a weather forecast model assimilated to an atmospheric transport model, and a statistical analysis technique (Lin Et al., 2004; Zhao Et al., 2009).

The state provides the firm with continuous mixing ratio measurements taken by the Precision Gas System's (PGS) infra-red gas analyzer at 60 m on a tall tower at the U.S. Department of Energy's (DOE) Atmospheric Radiation Measurement (ARM) Climate Research Facility South Great Plains (SGP) central facility over the period of interest. Generally, the continuous measurements are not affected by large systematic errors. However, during 2006-2008 the tower's measurements were affected by an offset dependent on time when the attached instrument became wet. The continuous measurements are particularly important because they provide a detailed level of variability of mixing ratios over time taken approximately every 12 minutes. The firm also had access to independent, co-located flask samples taken over the same period but less regularly at ~ 2 PM Central Standard Time (CST) each week by the U.S. National Oceanic and Atmospheric Administration's (NOAA) Earth System Research Laboratory (ESRL) Global Monitoring Division (GMD).

The firm has six objectives to meet the terms of the contract with the Department of Environmental Quality:

1. Remove unstable PGS measurements by identifying likely periods of high atmospheric variability.
2. Correct for the offset in hourly averaged PGS mixing ratio measurements using linear and cubic interpolation over one half of the differences between the matched nearest PGS to NOAA flask measurement.
3. Use remaining matched measurement differences not used for correcting the PGS measurements and evaluate the effectiveness of cubic spline and linear interpolation.
4. Implement a hypothetical emissions model that mimics the main input and output of the inverse modeling framework by assigning the rise and fall of monthly mean of corrected mixing ratio measurements to the emissions from an electric utility provider American Electric Power Company (AEP) operating in the State of Oklahoma.

5. Calculate any cumulative penalty against AEP for monthly mean emissions increasing from 2006 levels and compare severity of penalty with any case study.
6. Evaluate the sensitivity of any resulting fine to a change in the diurnal bounds of the corrected PGS signal, the accuracy of the measurements involved, and interpolation method.

2. LITERATURE REVIEW

The following section provides the necessary background information to begin adequately addressing the aforementioned environmental problem. Section 2.1 offers introductory information regarding the most important layer in meteorology affecting the regional transport of emissions and measurements taken within it. Next a brief description of the SGP site is contained in Section 2.2. In Section 2.3 the necessary background information to investigate the inverse framework is presented in a case study to quantify CH₄ emissions by source using in situ measurements. Next in Section 2.4 a case study on resolving missing air quality measurements presents interpolation methods. Section 2.5 contains a brief explanation of the world's first and most comprehensive mandatory emissions cut scheme. Also the details surrounding one of the most costly air quality emissions cases under the provisions of the Clean Air Act is described to provide a perspective on precedents regarding air pollutant emission discharge penalties in the United States. Finally, Section 2.6 presents a brief description of the coverage of AEP in Oklahoma closes the Literature Review.

2.1. PLANETARY BOUNDARY LAYER DYNAMICS AND MEASUREMENTS

2.1.1 PLANETARY BOUNDARY LAYER DYNAMICS

The planetary boundary layer (PBL) is a critical layer of the atmosphere to know of when evaluating regional emissions using the inverse analysis framework explained in Section 2.2. The PBL is contained in the troposphere, stretching from the ground to variable heights throughout the day. Two layers are a part of the boundary layer: First is the surface layer stretching from the ground to anywhere between 50 m to 100 m vertically. This layer primarily consists of winds heavily influenced by vertical temperature gradient

and surface friction with constant shearing forces. The layer above stretches from the top of the surface layer to 500 m-1000 m. Here winds are primarily influenced by same factors as surface but include the effects of planetary rotation. Above the PBL is the free atmosphere which is not affected by surface friction (Kaimal, 1994).

Diurnal variability of the PBL ranges from unstable to stable conditions. Ideally a turbulent, unstable layer, the convective boundary layer (CBL) rises after sunrise and continues to grow as solar input continues, in mid-afternoon eventually reaching about 1-2 km. An inversion cap developing from the night, called the stable boundary layer (SBL), before rises with the boundary layer, containing the PBL below it and limiting air exchange with the free atmosphere above. Air parcels heat within the PBL and create turbulent eddies, allowing the gases within the layer, including local emissions, to become well-mixed. Finally around sunset, radiative cooling destabilizes the turbulent eddies within and the CBL subside while wind shear becomes dominant, developing the SBL once again, characterized by less turbulence and mixing. Figure 1 shows this process of rising and falling of the CBL in the day with the SBL developing at night. In reality the rise and fall of the PBL is highly variable. One factor having a major impact on the formation of the PBL is the vegetation cover. The simplest of boundary layer formation occurs over a flat, uniform plain where the ideal PBL formation occurs (Kaimal, 1994).

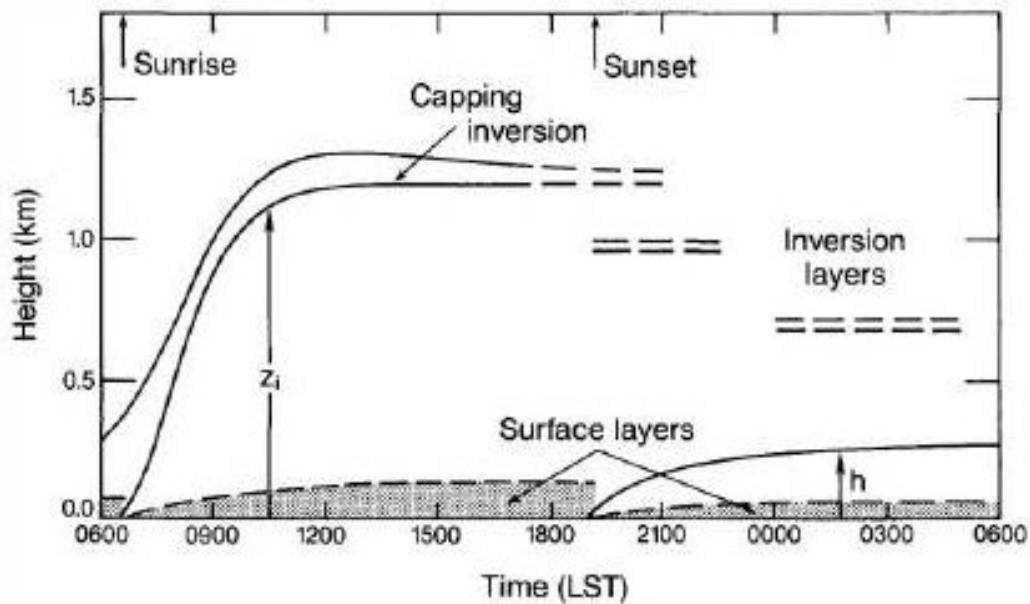


Figure 1: The ideal diurnal boundary layer height variability (Kaimal, 1994).

2.1.2. PBL MEASUREMENTS

Two categories of instruments measure properties of the PBL. Both measure properties of the PBL including wind speed and direction, fluxes of momentum and heat from various mediums including tower, mast, and aircraft. In situ measurements encompass the instruments that directly measure properties of the PBL while the remote measurements measure PBL properties indirectly through microwave, acoustic, and optical signals throughout the atmosphere. Overall in situ measurements have higher accuracy and resolution over relatively long periods of time, leaving their measurements most appropriate for quantitative studies by atmospheric researchers (Kaimal, 1994).

One such in situ instrument is the infrared gas analyzer (IRGA). This instrument is considered a slow response device because it takes minutes for the system to take a reading, though it can accurately measure long term variability of concentrations over time. IRGA measures time -averaged concentrations of CO₂ and water. To do this IRGA compares infrared absorption of a sample of air to a reference sample with a known concentration (Kaimal, 1994).

2.2. THE SOUTH GREAT PLAINS SITE

The South Great Plains (SGP) site includes around 55,000 square miles in north-central Oklahoma. SGP has maintained the site since 1992. A variety of instrumentation used to conduct large-scale independent and on-going field campaigns in the atmospheric sciences are scattered around the site. SGP was chosen for several reasons including accessibility, geographic homogeneity, and clear seasonal shifts in temperature and relative humidity. SGP consists of in situ and remote instrumentation scattered over the site. (ARM, 2012). The PGS is located at the Atmospheric Radiation Measurement Southern Great Plains Central Facility surrounded by 160 acres of wheat crops and cattle fields and about 27 miles west of Ponca City (ARM, 2010; ARM, 2012). Upwind emission sources of the PGS instrument are unknown to the firm. Perhaps land use management of wheat crops and cattle contribute to relatively elevated or reduced PGS mixing ratio measurements.

2.3. INVERSE FRAMEWORK FOR QUANTIFYING GREENHOUSE GAS EMISSIONS

A case study is presented to provide a brief explanation of verifying CH₄ emissions for a region using an inverse framework, but the same basic inverse framework may be used for CO₂ as well (Lin Et al., 2004). The study attempts to verify CH₄ emissions by source for Central California (Zhao Et al., 2009). Unlike CO₂, which is transferred between land, air and ocean over various time scales within the biosphere, CH₄ remains in the atmosphere for ~ 8.4 years before being removed by mostly tropospheric chemical oxidation (Denman et al., 2007; Solomon et al., 2007). Hence CH₄ emissions are easier to verify by source than CO₂ because the largely uncertain biospheric exchange model isn't necessary to incorporate into the framework (Lin, 2004; Zhao, 2009).

The case study is conducted over Central California Oct-Dec 2007. Over this period a four component framework is assembled. The first component is the in situ mixing ratio measurements taken on a tall tower at Walnut Grove, CA (WGC) at 91 m and 483 m every 5 min using data collection methods that reduce the sampling error to a small ~ppb range. These measurements are taken during well-mixed conditions, i.e. when the planetary boundary layer (PBL) is above the instrument, and when local and background emissions contribute to the signal (Zhao, 2009).

The second component is an atmospheric transport model assimilated with a weather forecast model to simulate the movement of air parcels, particles, from the receptor backward in time to the upstream boundary. In this study every 3 hours 100 particles are transported for 5 days until they reach the well-defined marine boundary layer, C_{BG} , allowing for the “footprint” to be calculated which shows how regional surface fluxes from various sources influence the WGC receptor (Zhao, 2009).

The third component to the framework is the surface flux emissions over the area of interest based on inventory estimates. The study uses six source inventory estimates to calculate surface flux emission over California: landfills (LF), crop agriculture (CP), livestock (LS), wetlands (WL), natural gas production and use (NG), and petroleum refining (PL). Changes in seasonal variations of emissions are ignored over the short period of the study for convenience (Zhao, 2009).

The fourth component is the Bayesian inverse analysis that minimizes the differences between the matched observed and predicted mixing ratios. The inverse analysis accounts for the total observed and predicted mixing ratio’s uncertainty, and provides improved estimates of emissions by source from inventory estimates via a posterior source scaling factor, λ_{source} with reduced total uncertainty. The assembled framework before the inverse analysis is applied is shown in Eqn. A.1 with an explanation in Section 8.1.1 (Zhao, 2009).

Figure 2 shows the results of the Bayesian inverse analysis from the aforementioned sources. The a priori scaling factors for each source are 1 with each having the same uncertainty level as shown in dark grey. The posterior scaling factors are shown in light grey with reduced levels of uncertainty. All posterior scaling factors with uncertainty lie within the uncertainty bounds of the a priori scaling factors except for livestock. Livestock is shown as having significantly greater posterior scaling factor at 1.63 ± 0.22 , than a priori. This indicates livestock emissions are $63 \pm 22\%$ greater than inventory estimates for this source (Zhao, 2009).

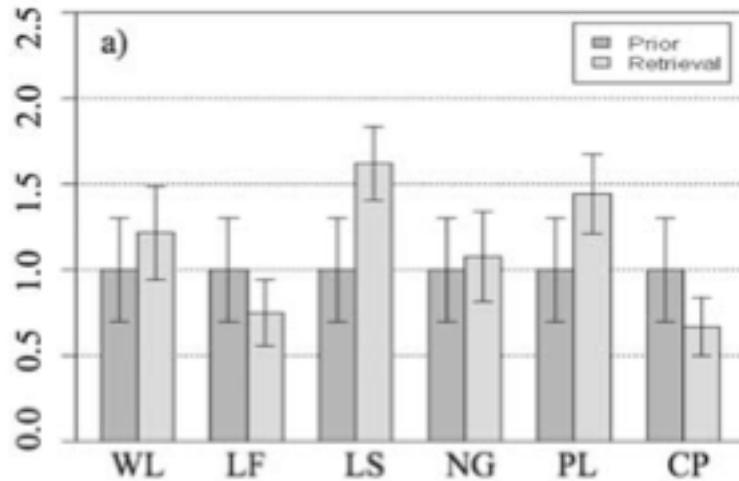


Figure 2: Livestock emissions appear to be greater than inventory estimates as shown from the a priori scaling factor in dark grey and posterior scaling factor in light grey (Zhao, 2009).

2.4. CASE STUDY IN INTERPOLATING MISSING AIR QUALITY DATA

Junninen's study (2004) uses multivariate air quality data, i.e. data varying by multiple variables over time, from two independent sets taken in 1998 extracted from the Air pollution Episodes: Modeling Tools for Improved Smog Management database. These two data sets are used to test the performance of various interpolation methods that use one independent variable to predict an independent variable over time called univariate interpolation methods and multivariate which uses many independent variables to predict an independent variable. The following summary of methods and results of his work only regard univariate methods examined, as they are most relevant to the present investigation. Included in the study are hourly concentrations of NO_x, NO₂, O₃, PM₁₀, SO₂, and CO as wells as wind speed and direction (WS and WD, respectively) temperature (T), and relative humidity (RH). The assembled matrix of hourly data measurements by type over 1998 has dimensions 10 x 8758 (Junninen Et al., 2004).

Univariate methods were evaluated by their performance to interpolate over various gap lengths in the data sets. Methods examined include linear spline (LIN), nearest neighbor, and cubic spline. The index of agreement was used to evaluate the effectiveness of univariate methods to interpolate over various gap lengths. The closer the index is to 1 the better the agreement between measured and interpolated. In figure 3, linear and cubic spline is shown to be most suitable for gaps up to 2 hours. However,

cubic spline steadily declines as gap length increases and periodically generated random error as gaps became longer than 24 hours (Junninen, 2004).

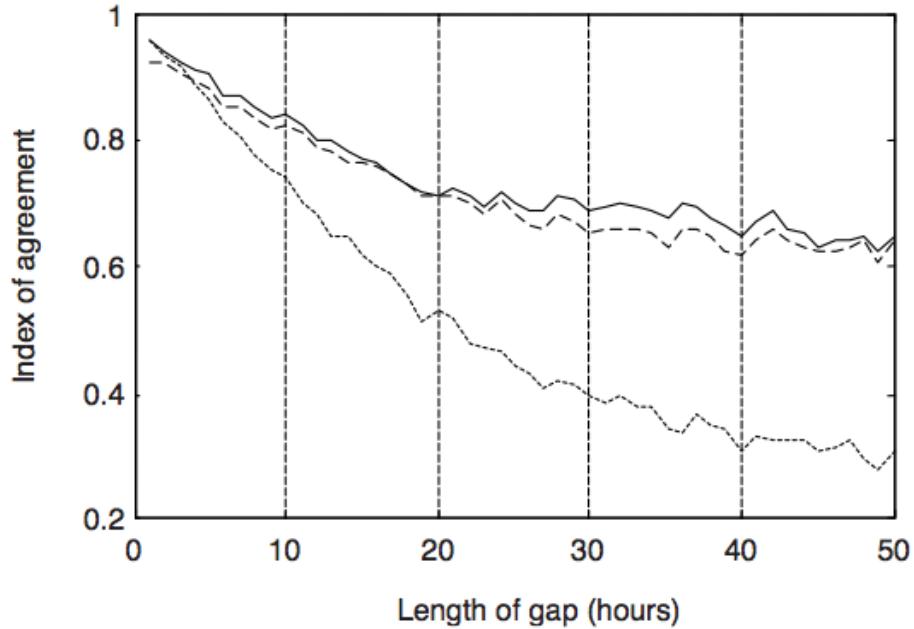


Figure 3: Performance of cubic spline drops quickly over time. Cubic spline is dotted line, nearest neighbor is dashed line, and linear is solid line (Junninen, 2004).

Baltazar-Cervantes' thesis compares 12 versions of Fourier, 5 of cubic spline and a linear spline interpolation to evaluate pseudo-gaps of 1-6 hours in length 20 samples of hourly energy use and weather data taken over the course of year at various locations throughout the United States. The versions of Fourier and cubic spline are developed providing various lengths of data before and after gaps. The coefficient of variation of the root mean square deviation error and mean bias error are used to compare the interpolation methods. CV-RMSE analysis shows linear to be best for all data types. The MBE analysis results show Fourier interpolation using 24 points before and after gaps along with 6 constants is the best for energy use data, while linear interpolation is best for weather data (Baltazar-Cervantes, 2000). The outperformance of linear with other univariate methods particularly with weather data supports Junninen's results as well.

2.5. CURRENT POLLUTANT DISCHARGE REGULATIONS

2.5.1. EU EMISSION TRADING SCHEME

Presently, the largest, most advanced climate change policy already being implemented is through the European Union's Emissions Trading Scheme (EU ETS). This mandatory cap and trade scheme which began in 2005 mandates every year governments within the EU to develop allocation plans, allowing each country to cut depending on their individual obligations. A specific set of criteria details the proper allowance, at 1 allowance = 1 metric ton of CO₂, for each establishment/business within the respective member country. A business that emits under its limit can keep or sell their extra allowances on the carbon market at a monetary amount as set by supply and demand. If an establishment emits over its allowance it can buy more from the carbon market, install energy efficient mechanisms, reduce reliance on carbon-based fuels, receive credits from investing in emission reduction projects abroad, or combine some or several options to meet the target (European Communities, 2008)

Under the EU ETS CO₂ emissions must be monitored. Every year the amount of emissions must be reported by businesses and independently verified, followed by a permit to emit and an issuance of new allowances for the next year. If a business goes over its allowances at the end of the year it is penalized by needing to cover the difference in the next year and fined 100€/Metric Tone Carbon Equivalen (MTCE) (European Communities, 2008).

In general the EU ETS rests the onus of reporting and independently verifying emissions reductions on businesses. In contrast, under an inverse modeling framework as used by the firm, businesses are not required to have emissions reported and verified for any year beyond 2006. Therefore the firm's method is less costly to businesses overall than the method used under the guidelines of the EU ETS.

2.5.2 EMISSION VIOLATION CASE STUDY

In 2007 AEP was fined an unprecedeted amount for violating U.S. pollution controls. Though the U.S. currently doesn't have similar efforts in mandatory CO₂ reductions for businesses, the U.S. Environmental Protection Agency (EPA) through the Clean Air Act (CAA) does have the power to limit the release of certain air pollutants from various

sources (EPA, 2008). The precedent set in the U.S. Et al. vs. American Electric Power provides insight into how CAA is currently being enforced, along with the penalties that can be leveled.

On Oct 9, 2007 the Plaintiffs including the United States and several states reached a settlement agreement with AEP for violating the Clean Air Act's New Source Review Provision (NSR). The total cited violations included 16 of 46 power plants located in five different states producing more than 20,000 megawatts. The case was the largest environmental enforcement settlement in history particularly in terms of injunctive relief. Also the case was the largest reduction of pollutant discharge as agreed by the owner/manager of a stationary source, a total of 813,000 tons/year (EPA, 2007).

The cost to AEP to comply with the settlement is \$4.6 billion. This compliance includes approximately a 67% and 78% reduction from the violating 16 plants in NO₂ and SO₂ from the 2006 emission amount by 2016 and 2018 respectively. Also, the settlement includes a \$15 million civil penalty, \$36 million to be directed to federal environmental mitigation projects, and \$24 million for state environmental projects (EPA, 2007). Though the total of fines against AEP are unprecedented, the total cost of the settlement is compared with any fines leveled against AEP through evaluating the corrected hourly PGS signals.

2.6 AEP COVERAGE WITHIN OKLAHOMA

The American Electric Power Company (AEP) currently serves about 5,000,000 customers across 11 states including a large number of households in the State of Oklahoma. The corporation is one of the largest generators of electricity in the United States, owning almost 38,000 megawatts of generating capacity. AEP serves about 527,000 customers in eastern and southwestern Oklahoma, about 10% of AEP's total customer base (PSO, 2012). Within the State of Oklahoma AEP was one of a handful of electric utility providers serving the approximately 1,664,378 housing units listed in Oklahoma in 2010 (U.S. Census Bureau, 2012). Much of AEP's CO₂ emissions come from the 8 power plants scattered mostly in eastern and western Oklahoma (AEP, 2007).

3. METHODOLOGY

The U.S. NOAA ESRL and U.S. DOE ARM collaborated to take their respective flask and PGS mixing ratio measurements/signals at the SGP site (ESRL, 2010; ARM, 2010). Originally the PGS was designed to measure several atmospheric properties at 2, 4, 25, and 60 m above ground. For the firm's purpose, the PGS system only includes the wet infra-red gas analyzer (IRGA) LI-6252's mixing ratio measurements taken at 60 m. The normally highly precise, accurate instrument is assumed to be functioning normally except with the additional error introduced by the time varying offset. Every 15 minutes the reference gas in the reference cell is analyzed to correct for drift. Also every 4 hours the instrument is calibrated. The accuracy of the PGS instrument is 0.05 ppm based on the blind comparison test using NOAA Climate Monitoring Diagnostic Laboratory standards (Torn, 2005). The NOAA flask mixing ratio measurements are taken using the nondispersive infrared absorption technique. The accuracy and precision of the NOAA flask measurements are ~0.2 ppm and ~0.1 ppm respectively (Conway, 2011).

The state of Oklahoma provides to the firm two data sets matching NOAA flask and PGS signals as well as the offset hourly averaged continuous PGS measurements taken over the period of interest. One of the matched data set pairs NOAA flask and PGS signals by taking the nearest PGS measurement to a NOAA flask measurement, Set 1. The second set pairs the two by taking an hourly averaged PGS signal centered around the time a NOAA flask measurement, Set 2. The former contains data in ASCII format, while the latter contains data with the same headings, but also an extra column containing hourly standard deviations in ppm (Section 8.2, Table A.1 in Appendix).

A method is decided by the firm to correct the hourly PGS CO₂ mixing ratio measurements. Equation 1 shows the firm's strategy of using the interpolated difference between half of the matched PGS and flask signal or the signal difference to correct for the offset in the PGS measurements. The firm assumes the NOAA measurements are precise, accurate enough to correct the offset in the PGS measurements. Set 1 is chosen to correct the hourly PGS signal based on the assumption less errors would be introduced from matching the PGS and NOAA flask measurements taken at different times. Half of the matched PGS and flask signals or the first half are chosen for correction by taking the

first and every other matched signal. The remaining matched signals are used to check for the performance of the correction.

$$\text{Cor_PGS}_t = \text{Uncor_PGS}_t - I(t) \quad (\text{Equation 1})$$

Where,

Cor_PGS_t = Corrected Hourly PGS signal at time t

Uncor_PGS_t = Uncorrected PGS Mixing Ratio Measurement at time t

$I(t)$ = Interpolation of flask signal subtracted from nearest PGS signal at time t

Two criterions remove matched signals from set 1 that are likely to be due to high atmospheric variability because of the turbulent dynamics of the PBL (Kaimal, 1994).

The first criterion is the matched hourly PGS and flask signals in set 2 with standard deviation greater than or equal to x_1 ppm must exclude the matched flask and PGS signals in set 1 taken within the same hour. The second criterion excludes all matched signals in set 1 with signal differences greater than x_2 number of standard deviations from the mean difference.

Another criterion removes hourly signals taken during times when the atmosphere is not likely to be well-mixed. Zhao (2009) only included time periods when the atmosphere was likely to be well-mixed to avoid measurements taken when the air surrounding the instrument was decoupled from global background gases (Section 2.2). Periods when the atmosphere are likely to be well-mixed are after solar input throughout the day contributing to a rise a in the PBL height and the mixing of background and local gases. Therefore only the PGS hourly signals between times t_1 and t_2 each day are corrected (Refer to Section 2.1.1).

From the literature, linear interpolation is recommended most for univariate atmospheric data over any other method. However Fourier and cubic spline also appear to be widely used for the type of atmospheric data acquired (Junninen Et al., 2004; Baltazar-Cervantes, 2000). Traditionally the engineering firm uses cubic spline for interpolating CO₂ mixing ratios over other methods.

Linear interpolation is sufficiently easy to implement and therefore doesn't require any further explanation. Cubic spline is more sophisticated, so an explanation of the method follows. Equation 2 shows the cubic spline interpolation, which contains 3 known PGS and flask differences and 3 unknown second derivatives of the flask and PGS differences taken at different points in time. A natural cubic spline is implemented by forcing the second derivatives at the beginning and end of the period of interest 2006 to 2008 to zero. Once the assumption of the natural cubic spline is made, a system of n-1 equations is developed to solve for n-1 unknowns (Chapra, 2006).

$$(t_i - t_{i-1})I''(t_{i-1}) + 2(t_{i+1} - t_{i-1})I''(t_i) + (t_{i+1} - t_i)I''(t_{i+1}) \quad (\text{Equation 2})$$

$$= \frac{6}{t_{i+1} - t_i} [I(t_{i+1}) - I(t_i)] + \frac{6}{t_i - t_{i-1}} [I(t_{i-1}) - I(t_i)]$$

Where,

t_i = Time at which signal difference is known

t_{i-1} = Time before t_i at which signal difference is known

t_{i+1} = Time after t_i at which signal difference is known

$I''(t_{i-1})$ = Second derivative of $I(t_{i-1})$ before t_i (ppm/min²)

$I''(t_{i+1})$ = Second derivative of $I(t_{i+1})$ after t_i (ppm/min²)

$I''(t_i)$ = Second derivative of $I(t_i)$ at t_i (ppm/min²)

$I(t_{i+1})$ = Signal difference known at time after time i (ppm)

$I(t_{i-1})$ = Signal difference known at time before time i (ppm)

$I(t_i)$ = Signal difference known at time i (ppm)

The resulting systems of equations from implementing Equation 2 over 2006 to 2008 define a tridiagonal matrix. As such there are efficient algorithms that solve specifically tridiagonal matrices. An efficient and commonly used one is called the Thomas algorithm efficiently. This algorithm uses an LU decomposition method but saves computation time and storage by dealing with the unique placement of unknowns in a tridiagonal matrix. First a decomposition method breaks the triadiagonal matrix into an upper and lower matrix. Then forward substitution is implemented using the lower triangle matrix to transform a right hand side vector. Finally back substitution is

implemented using the upper triangle matrix to solve for the system of equations (Chapra, 2006).

Once the unknown second derivatives are computed, a system of cubic equations may be calculated between all gaps with continuous first and second derivatives at the bounds, i.e. time when the PGS and flask differences are known before and after the gap. Equation 3 shows the interpolated signal between each gap or interval (Chapra, 2006).

$$I(t) = \frac{I''(t_{i-1})}{6(t_i - t_{i-1})} (t_{i-1} - t)^3 + \frac{I''(t_i)}{6(t_i - t_{i-1})} (t - t_{i-1})^3 \quad (\text{Equation 3})$$

$$+ \left[\frac{I(t_{i-1})}{t_i - t_{i-1}} - \frac{I''(t_{i-1})(t_i - t_{i-1})}{6} \right] (t_i - t)$$

$$+ \left[\frac{I(t_i)}{t_i - t_{i-1}} - \frac{I''(t_i)(t_i - t_{i-1})}{6} \right] (t - t_{i-1})$$

Where,

t = Any time between interval t_{i-1} and t_i (min.)

t_i = End of time interval (min.)

t_{i-1} = Beginning of time interval (min.)

$I(t)$ = Interpolated signal difference at t (ppm)

$I''(t_{i-1})$ = Second derivative of $I(t_{i-1})$ at beginning of interval (ppm/min²)

$I''(t_i)$ = Second derivative of $I(t_i)$ at end of interval (ppm/min²)

$I(t_i)$ = Known signal difference at end of interval (ppm)

$I(t_{i-1})$ = Known signal difference at beginning of interval (ppm)

The second objective is to use the corrected continuous PGS signals in a extremely simplified representation of the inverse framework described by Zhao (2009) to generate the results of AEP being directly, solely responsible for a rise and fall of the mean corrected PGS signals from 2006 levels. The expected assembled inverse framework before inverse analysis as formulated from the work of Zhao (2009) and Lin (2004) is shown in Equation A.2. and briefly explained in Section 8.1.2. Equation 4 represents a simplified representation of an unrealistic, but hypothetical inverse framework for a situation where AEP is solely responsible for a rise in emissions with

complete certainty. Once Equation 4 is evaluated, objective 3 may be met by penalizing any percent increase in the company's production from 2006 levels \$x/ton and summing the monthly penalty from 2007 to 2008.

$$E_{AEP,m,y} = \left[\frac{(\overline{Cor_{PGS}}_{m,y} - \overline{BG}_{m,y}) - (\overline{Cor_{PGS}}_{m,2006} - \overline{BG}_{m,2006})}{(\overline{Cor_{PGS}}_{m,2006} - \overline{BG}_{m,2006})} \right] \times 100 \quad (\text{Equation 4})$$

Where,

$E_{AEP,\text{month}}$ = AEP's percent rise or fall in emissions for month m and year

$2007 \leq y \leq 2008$

$\overline{Cor_{PGS}}_{m,y}$ = Mean of corrected PGS signal for month m and year

$2007 \leq y \leq 2008$ (ppm)

$\overline{Cor_{PGS}}_{m,2006}$ = Mean of corrected PGS signal for month m in 2006 (ppm)

$\overline{BG}_{m,y}$ = Mean of background signal for month m and year $2007 \leq y \leq 2008$ (ppm)

$\overline{BG}_{m,2006}$ = Mean of background signal for month m in 2006 (ppm)

The simplified model represented in Equation 4 is imbedded with a host of unrealistic assumptions. The purpose of the model is to mimic as simply as possible the inverse modeling framework with the mixing ratio measurements as input and amount of emissions from a source AEP as output. The assumptions of the simplified model are as follows:

1. The percent growth or decline in emissions from AEP plants for a particular month falling anywhere from 2007 to 2008 directly corresponds to a rise or fall in the mean corrected PGS CO₂ signal at SGP for the same month in 2006.
2. The contribution from background sources including all moving sources inside and outside of Oklahoma, stationary instate industry unrelated to AEP power plants, all stationary instate domestic sources, and stationary sources within regions outside of Oklahoma may be accounted for by the monthly mean background signal.
3. All AEP emissions are generated upwind of PGS instrument
4. The biospheric exchange of CO₂ between land and atmosphere may be ignored.

5. The remaining errors in the corrected PGS instrument may be ignored.

4. APPLICATION

The code in the statistical programming language R used for generating the results is shown in Appendix 8.5 (Chapra, 2006; Finney, 2012). The code includes the developed cubic spline numerical method. Linear interpolation was implemented using the built-in linear interpolation function in the R environment.

The development of the code includes assigning criteria to remove matched PGS and NOAA signals in set 1 when measurements were taken during unstable atmospheric conditions. The first criterion removes matched signals in set 1 are taken within the same hour as matched signals in set 2 and the standard deviation in set 2 is greater than or equal to $x_1=1$ ppm. X_1 was chosen as recommended by a consultant Marc Fischer, Staff Scientist at the U.S. Department of Energy Lawrence Berkeley Lab in 2010. The second criterion removes matched signals in set 1 when the signal difference is greater than $x_2=3$ times the standard deviation from the mean signal difference. $X_2=3$ was determined to be sufficiently large to remove only the few signal differences farthest from the mean signal difference.

Bounds on the time of the day to correct the hourly PGS signals are $t_1=12$ to $t_2=1$ GMT, 6 AM to 6 PM Central Standard Time (CST), based on Figure 1 and observing the CST for sunrise and sunset for the nearest city Ponca City, OK over the period of interest (U.S. Navy, 2012). The planetary boundary height ascending and descending starting at sunrise and sunset respectively is implied within the chosen time interval. The assumption is conservative because in reality the planetary boundary layer rises and falls after a few hours after sunrise and sunset. These bounds are varied by + 3 hours by 1 hour to account for annual variations of sunrise and sunset amongst other delays leading to diurnal development of the PBL.

Another portion of the sensitivity analysis includes varying the matched PGS and NOAA measurements before corrected for the offset. The matched measurements by +/-

the accuracy of each instrument, i.e. +/- 0.05 ppm and +/- 0.2 ppm for the PGS and NOAA measurements respectively (Torn, 2005; Conway, 2011)

As a means to compare the sensitivity of all potential fines to a change in the interpolation, linear and cubic spline are chosen to interpolate over the first half of the signal difference. The index of agreement with k=2 is chosen as in the work of Junninen in 2004 to compare the effectiveness of the two methods to interpolate over the other half of the signal difference or the second half of the signal difference (Equation 5) (Willmott, 1982; Junninen, 2004).

$$d = 1 - \left[\frac{\sum_{i=1}^N (P_i - O_i)^k}{\sum_{i=1}^N (|P_i - \bar{O}| + |O_i - \bar{O}|)^k} \right] \quad (\text{Equation 5})$$

Where,

d = index of agreement

P_i = Predicted value through interpolation

O_i = Observed value

\bar{O} = Mean of observed values

The background is accounted for like in the work of Zhao (2009) by a known boundary layer. The firm uses the global monthly averaged CO₂ mixing ratio calculations provided by NOAA as the background layer (Conway, 2012). The measurements used for calculating the monthly average were taken by flask at remote marine locations around the world with no major source or sink nearby. The measurements represent the global trend (ESRL, 2012). The uncertainty of the monthly averages is 0.13 ppm as determined by running the Monte Carlo technique (Conway, 2012).

AEP emissions within Oklahoma each month in 2006 is assumed to be the total emissions generated by electric utility within the State of Oklahoma in 1999 divided by the number of months in a year. The monthly amount is calculated to be 975,122 net Metric Tons of Carbon Equivalent (MTCE) (ODEQ: AQD, 2002). Taking the total and dividing by the number of months in a year assumes the emissions for AEP are constant each month in 2006 when in reality the emissions change over the year as consumer

demand changes. The assumption is by far a very conservative estimate because AEP serves only about a 10% of all residents in Oklahoma (PSO, 2012) (Section 2.6).

However the assumed amount is not greater than the total emissions that AEP produced in 2000 at 16,810,132 MTCE (AEP, 2012). Fines for AEP exceeding the allotted emissions for each month in 2007 and 2008 are set at 100€/MTCE or approximately \$133.33/MTCE using a \$1.333/€ conversion factor (Section 2.4.1).

5. RESULTS AND DISCUSSION

Qualitative evidence suggests the criteria to filter periods of high atmospheric variability appear to be somewhat effective. Ideally only signal differences noticeably different from neighboring differences would be selected for removal. Two of the greatest signal differences are removed after differences greater than three standard deviations from the mean difference were identified. The remaining signal differences removed were selected based on the hourly standard deviation of the PGS measurement having values greater than 1 ppm. The later criterion to filter atmospheric noise appears mostly effective based on Figure 4, but some problems continue to persist. However some signal differences are not removed that appear to be outliers such as the difference identified by the right arrow. Also some signal differences are removed that don't appear to be outliers such as the difference identified by the left arrow.

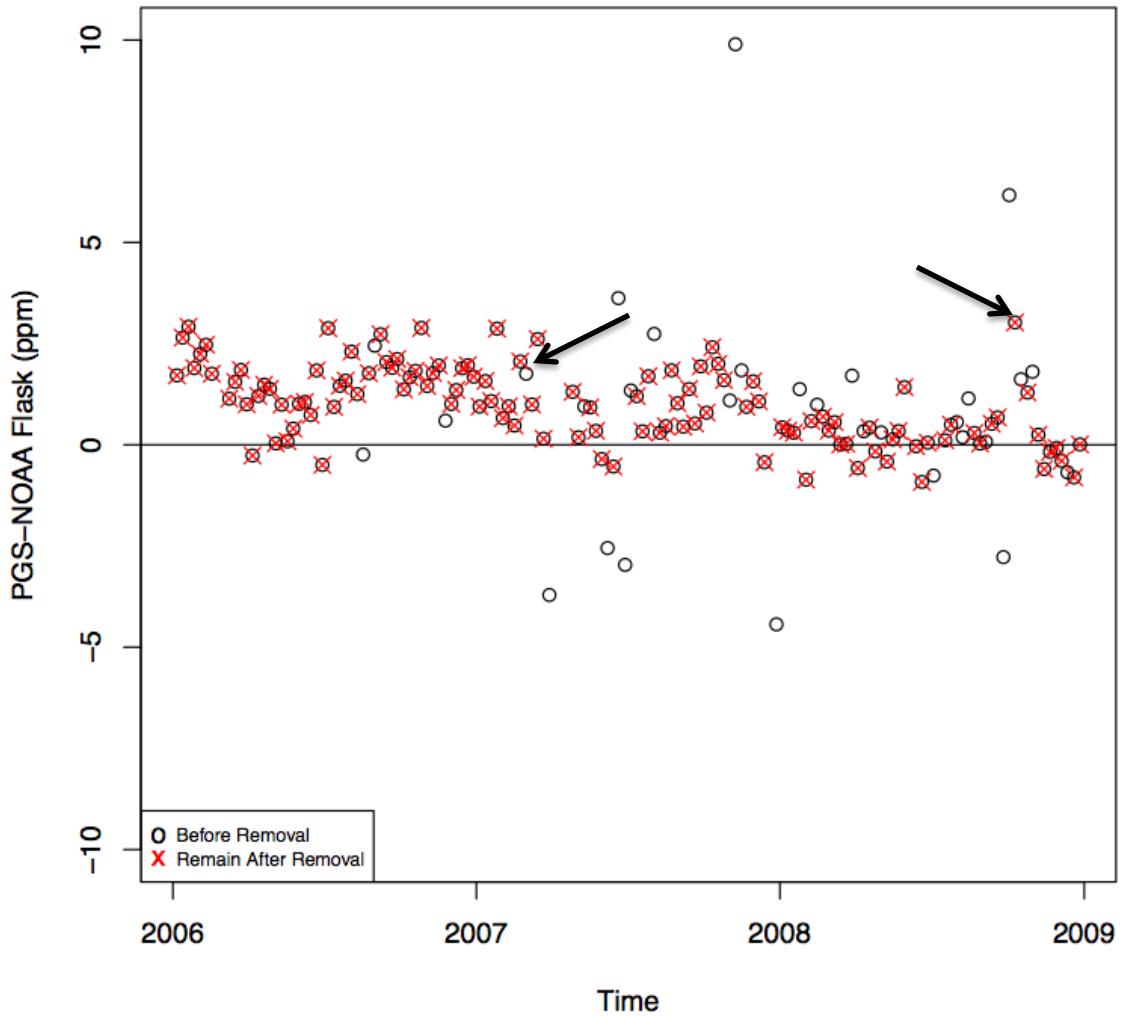


Figure 4: Matched NOAA and PGS signals removed after filtering for high atmospheric variability using the two criterions.

The index of agreement was roughly tested as a method to identify the agreement between interpolated signal difference over second half of matched PGS and NOAA signals and observed second half signal differences. The difference between the flask measurements and 0.9 of the same measurements were used to test whether an expected index of agreement near 1 was calculated. As expected the resulting index of agreement was near 1 at 0.995. Also the difference between the flask and a relatively high constant 10 returned an expected index of agreement near 0 at 0.138.

The linear and cubic spline interpolation implemented appears to be functioning as expected. Figure 5 shows the linear and cubic spline interpolation calculating the

interpolated first half signal difference exactly. As expected, the result occurs because the first half signal difference was used to develop the linear and cubic spline interpolation.

Both interpolation methods appear to perform similarly both qualitatively. The red lines in the top and bottom graph show similar interpolated signal differences, even during periods when the observed signal disagrees. The green arrow in Figure 5 identifies an instance of the occurrence. Occasionally differences appear in the performance of the interpolation methods as the blue arrow in Figure 5 identifies one instance when the interpolated signal difference of a second half point is slightly different.

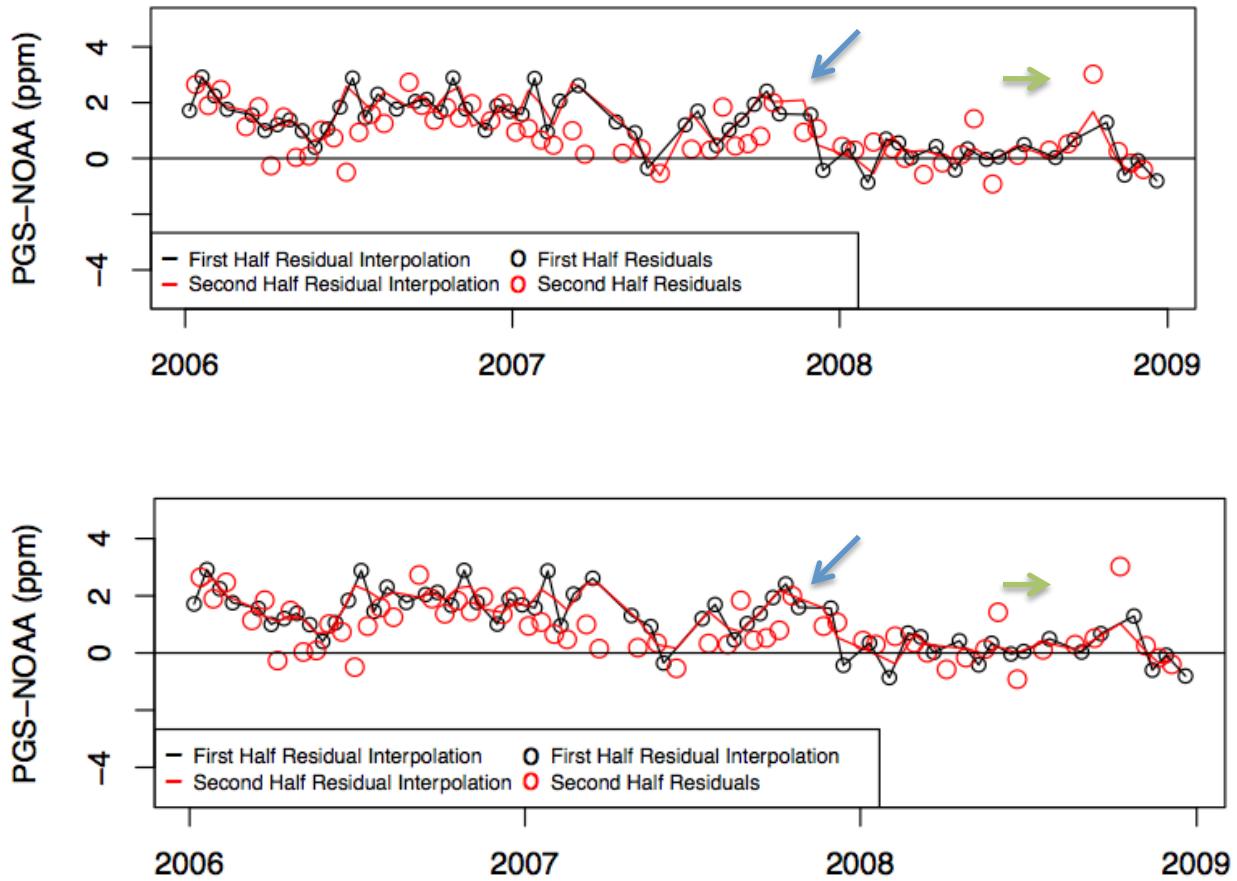


Figure 5: The performance of the cubic spline interpolation of the signal difference or the residual interpolation shown in the top graph appears similar to linear interpolation shown in bottom graph. The red line representing the cubic interpolation is a line fit through the cubic spline interpolated values, and therefore the first derivative of cubic spline appears

discontinuous as linear interpolation. Both first and second derivatives are continuous for cubic spline interpolation (Chapra, 2006).

The quantitative results show the similarity of the performance between cubic spline and linear interpolation. The index of agreement between the observed and interpolated second half of the signal differences was 0.726 using cubic spline and 0.721 using linear. The relatively high index of agreement shows the observed and predicted are more in agreement than disagreement. Also the close indices of agreement do not indicate whether cubic spline or linear are more effective at resolving the signal differences. In Junninen's study (2004) the value of the index of agreement using cubic spline over temporal gaps between known points from 0 to 50 hours diminished relative to the value using linear interpolation (Section 2.4). The temporal gaps between known signal differences in the present problem are >>50 hours, around 336 hours or about every 2 weeks. However even with large temporal gaps, cubic spline and linear interpolation appear to have similar values for the index of agreement.

The reduced mean of the second half of signal differences before after correcting the PGS signal shows using cubic spline and linear interpolation did improve PGS measurements relative to NOAA flask measurements. Before interpolating the mean of the observed signal differences was 0.840 ppm with a standard deviation of 0.900 ppm. After interpolating the mean of the second half of the signal differences became -0.343 ppm using linear and -0.354 ppm using cubic spline with respective standard deviations of 0.838 and 0.869 ppm. A small difference in standard deviation before and after correcting indicates neither cubic spline nor linear interpolation reduced the variability of the signal differences.

The corrected hourly averaged PGS signal using cubic and linear interpolation are lumped into monthly averages respectively and used with NOAA global monthly averages monthly to generate the output of the model that estimates emissions of AEP. Figure 6 shows the input into the model. Inter-annual variations and rise in the annual peak are apparent in both the monthly averaged global NOAA signals and monthly averaged corrected PGS signals from Figure 6. The corrected monthly PGS averages by cubic spline and linear interpolation are nearly identical. Also NOAA global marine

monthly averages are always below monthly averages of corrected PGS signal regardless of interpolation method used. Land management practices on the land surrounding the PGS instrument and other regional sources might be responsible for magnitude of elevation of corrected PGS measurements relative to marine NOAA (Section 2.2).

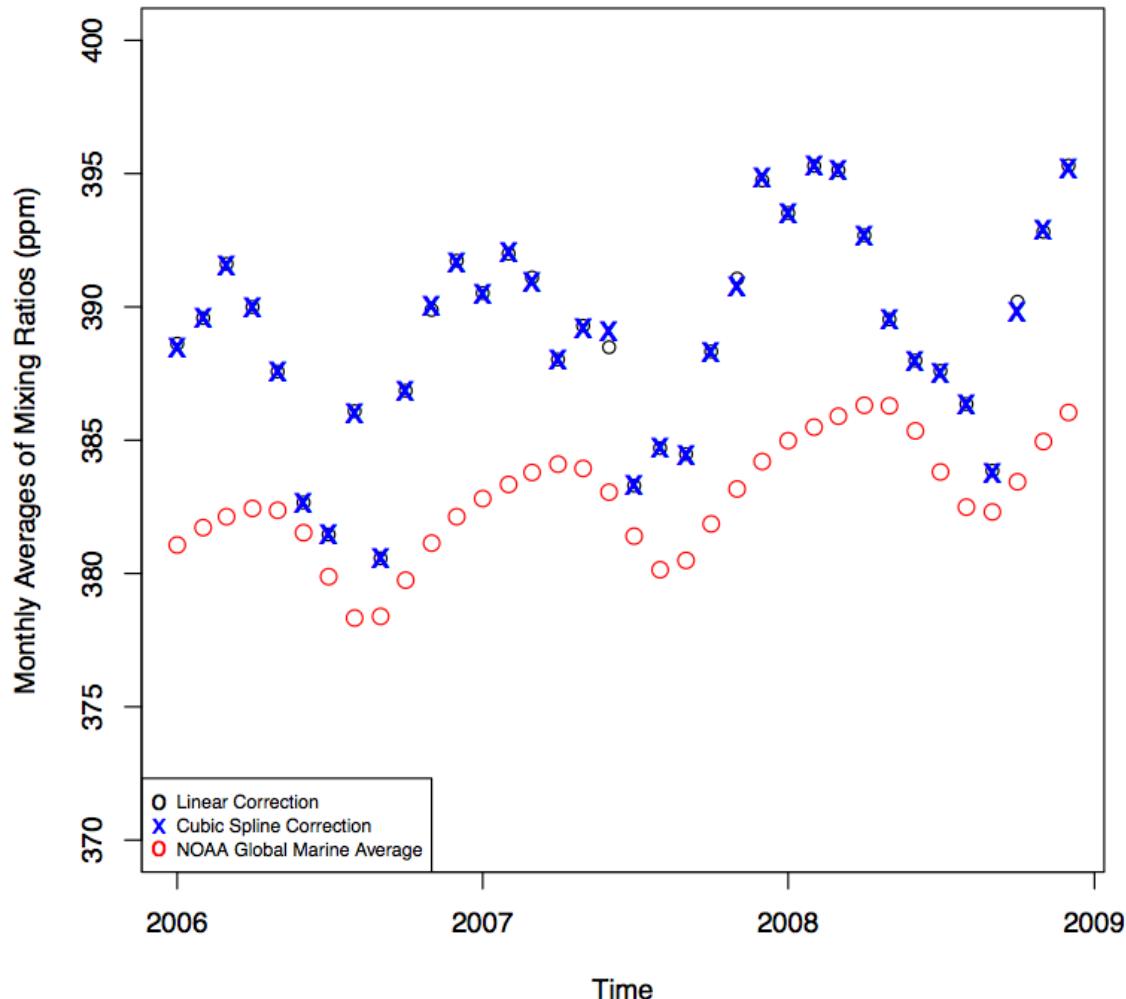


Figure 6: Monthly mean of corrected hourly PGS mixing ratio measurements always appear higher than NOAA's global monthly averages.

Fines are issued using the linear or cubic spline correction of the PGS signals. The amount of the fine has greater similarity by interpolation technique than from one interpolation in 2007 to 2008. The total cost of the fines is below the case study of cost to AEP to comply with a former settlement in 2007 at \$4.6 billion (EPA, 2007) (Section

2.5.2). However, the coverage of AEP within Oklahoma is about 10% of company's total customer base and includes 8 power plants (PSO, 2012; AEP, 2007). In contrast the \$4.6 settlement included 16 power plants over 5 states (EPA, 2007).

Table 1: Fines against AEP for violating 2006 monthly emission allotments are similar using linear or cubic spline to correct PGS signals.

| | 2007 Total (\$) | 2008 Total (\$) | Total (\$) |
|-------------------------|-----------------|-----------------|---------------|
| Linear Correction | 660,433,753 | 403,987,728 | 1,064,421,481 |
| Cubic Spline Correction | 737,367,099 | 397,987,678 | 1,135,354,777 |

PGS and NOAA measurements were varied based on the accuracy of the instruments (Beginning of Section 3). Varying NOAA flask measurements by -0.2 ppm and PGS measurements by +0.05 ppm results in greatest change in the fine amount regardless whether cubic and linear spline was used to correct the PGS signals. A -0.2 ppm reduction increases the fine the most from the original total fine amount by factors of 1.175 and 1.177 using linear and cubic spline respectively to correct PGS measurements (Table 2).

Table 2: Varying NOAA flask measurements has greatest impact on final fine against AEP.

| Varying | Value | Fraction of New Total /Original Total with Linear PGS Correction | Fraction of New Total /Original Total with Cubic PGS Correction |
|--------------------------|--------|--|---|
| Matched NOAA (ppm) | -0.200 | 1.175 | 1.177 |
| | +0.200 | 0.873 | 0.872 |
| Matched PGS Signal (ppm) | -0.050 | 0.965 | 0.964 |
| | +0.050 | 1.038 | 1.039 |

Shifting the bounds on the time interval when hourly PGS are corrected by +1 hour affects the final fine against AEP the most among the other time shifts. The magnitude of the differences for + 1 hour is ~ 9 times greater than +1 hour shift and ~18 times greater than the +3 hours shift. The reduction in the final fine for +2 and +3 hours from the +1 hour shift may be due to the inclusion of additional hourly PGS signals that

balance the relatively high positive difference between corrected monthly PGS signals and global monthly NOAA signals.

Table 3: The effect of shifting the bounds on the time interval each day when hourly PGS signals are corrected are shown with the +1 hour shift showing the greatest impact on final fine against AEP.

| Time Change | New Total /Original Total with Linear Correction | New Total/Original Total with Cubic Correction |
|-------------|--|--|
| +1 hour | 8.508 | 9.021 |
| +2 hours | 0.816 | 0.821 |
| +3 hours | 0.511 | 0.495 |

6. CONCLUSION

The following statements summarize the main findings of this report:

- The filter for high atmospheric variability appeared somewhat effective.
- Both quantitative and qualitative evidence indicates the performance of linear and cubic spline is similar for interpolating the signal difference.
- The interpolated signal differences have periods of noticeable disagreement with observed signal differences.
- Both cubic spline and linear interpolation methods reduce mean signal difference by ~0.5 ppm but the standard deviation of the mean difference remains constant at about ~0.9 ppm.
- The total fine against AEP after applying cubic spline and linear correction of the PGS signal is ~ \$1-1.1 billion
- Varying NOAA flask measurements by -0.2 ppm has the greatest impact on the final fine against AEP among varied matched measurements. .
- Increasing the bounds by +1 hour has greatest impact on the final fine against AEP among all time varied shifts.

7. REFERENCES

- American Electric Power Company (AEP). 2007. "2006 Annual Report to Shareholders." Accessed April 20, 2012. <
<http://www.aep.com/investors/financialfilingsandreports/annualReportsProxies.aspx>>
- American Electric Power Company (AEP). 2012. "IRRC Greenhouse Gas Profile." Accessed April 20, 2012. <<http://www.aep.com/environmental/reports/>>
- Atmospheric Radiation and Measurement Program (ARM). 2010. United States Department of Energy. Accessed 2010. <<http://www.arm.gov/instruments/pgs>>.
- Atmospheric Radiation and Measurement Program (ARM). 2012. "Southern Great Plains." United State Department of Energy. Accessed April 20, 2012. <
<http://www.arm.gov/sites/sgp>>.
- Baltazar-Cervantes. 2000. "Study of Cubic Splines and Fourier Series as Interpolation Techniques for Filling in Short Periods of Missing Building Energy Use and Weather Data." Office of Graduate Studies: Texas A&M University.
- Chapra, Steven, Et al. 2006. Numerical Methods for Engineers: 5th Edition. McGraw Hill: NY,NY. Pgs. 285-287; 495-505.
- Conway, T.J.; P.M. Lang, K.A. Masarie. 2011. "Atmospheric Carbon Dioxide Dry Air Mole Fractions." Carbon Cycle Cooperative Global Air Sampling Network. Earth System Research Laboratory (ESRL). National Oceanic & Atmospheric Administration. <
ftp://ftp.cmdl.noaa.gov/ccg/co2/flask/event/README_surface_flask_co2.html>
- Conway, Thomas; Tans, Pieter. 2012. Earth System Research Laboratory (ESRL). United States National Oceanic & Atmospheric Administration (NOAA). Accessed 2012. <ftp://ftp.cmdl.noaa.gov/ccg/co2/trends/co2_mm_gl.txt>
- Denman, K.L., G. Brasseur, A. Chidthaisong, P. Ciais, P.M. Cox, R.E. Dickinson, D. Hauglustaine, C. Heinze, E. Holland, D. Jacob, U. Lohmann, S Ramachandran, P.L. da Silva Dias, S.C. Wofsy and X. Zhang. 2007. "Couplings Between Changes in the Climate System and Biogeochemistry." In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report

of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z.

Earth System Research Laboratory (ESRL). 2010. United States National Oceanic & Atmospheric Administration (NOAA). Accessed 2010.
<<http://www.esrl.noaa.gov/gmd/dv/site/site.php?code=SGP>>

Earth System Research Laboratory (ESRL). 2012. United States National Oceanic & Atmospheric Administration (NOAA). Accessed April 20, 2012. <http://www.esrl.noaa.gov/gmd/ccgg/about/global_means.html>

Environmental Protection Agency (EPA). 2007. "EPA United State et al. vs. American Electric Power Information Sheet." U.S Environmental Protection Agency. Washington, DC.

Environmental Protection Agency (EPA). 2008. "Overview-The Clean Air Act Amendments of 1990." U.S Environmental Protection Agency. Washington, DC. Accessed: Mar. 4, 2012 <http://epa.gov/oar/caa/caaa_overview.html>

European Communities. 2008. "EU Action Against Climate Change: The EU Emissions Trading Scheme 2009 Edition" European Union.

Finney, Brad. Spring 2012. "Thomas Algorithm Lecture." Humboldt State University. Arcata, CA.

Juninen, Heikki Et al., 2004 "Methods for Imputation of Missing Values in Air Quality Data Sets." Journal: Atmospheric Environment of Elsevier.

Kaimal, J. C. Finnigan, J. J. 1994. Atmospheric Boundary Layer Flows : Their Structure and Measurement. Oxford University Press . Pgs. 1-21; 207-215.

Lin, J.C Et al., 2004. "Measuring Fluxes of Trace Gases at Regional Scales by Lagrangian Observations: Application to the CO₂ Budget and Rectification Airborne (COBRA) Study." Journal of Geophysical Research. Vol. 109, D15304, doi:10.1029/2004JD004754

Oklahoma Department of Environmental Quality (ODEQ): Air Quality Division (AQD). 2002. "Inventory of Oklahoma Greenhouse Gas Emissions and Sinks: 1990 and 1999." Accessed April 20, 2012. <<http://www.epa.gov/statelocalclimate/state/state-examples/ghg-inventory.html>>

Public Service Company of Oklahoma (PSO). 2012. "Facts, Figures & Bios." American Electric Power Company (AEP).Accessed April 20, 2012. <<https://www.psoklahoma.com/info/facts/>>

Solomon, S., Et al. 2007. "Technical Summary." In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. Pgs. 20-30; 90-91.

Torn, Margaret. 2005. "Precision Gas System Handbook." U.S. Dept of Energy Office of Biological and Environmental Research. Accessed March 4, 2012
http://www.arm.gov/publications/tech_reports/handbooks/pgs_handbook.pdf?id=78

U.S. Navy. 2012. "Complete Sun and Moon Data for One Day." United States Naval Observatory. Astronomical Applications Department. Accessed April 25, 2012. <http://aa.usno.navy.mil/data/docs/RS_OneDay.php>

U.S. Census Bureau. 2012. "State and County Quick Facts." Accessed April 20, 2012.
<<http://quickfacts.census.gov/qfd/states/40000.html>>

Willmott, C.J., 1982. "Some comments on the evaluation of model performance." Bulletin of the American Meteorological Society 63,1309–1313.

Zhao, C., Et al. 2009. "Atmospheric Inverse Estimates of Methane Emissions from Central California." Journal of Geophysical Research. Volume 114, D16302, doi:10.1029/2008JD011671.

Acknowledgement:

Thanks to Marc Fischer Staff Scientist at Lawrence Berkeley Lab in 2010 for providing the matched PGS to NOAA flask measurements, and advising on strategies to filter atmospheric variability and correcting the PGS instrument. .

8. APPENDIX

8.1 INVERSE MODELING FRAMEWORK

8.1.1. INVERSE FRAMEWORK FROM CASE STUDY

Equation A.1 shows the first three framework components assembled before Bayesian inverse analysis is applied. $\underline{C} - \underline{C}_{BG}$ is assumed to be the regional contributions to the receptor. $\underline{\underline{f}}$ is calculated from atmospheric trajectories. $\underline{F}(\lambda_{source})$ is the product of the source scaling factor and the surface flux emissions. Before Bayesian inverse analysis λ_{source} is set to 1 and is referred to as an *a priori* source scaling factor. After analysis λ_{source} is optimized to minimize difference between observed and predicted mixing ratios at WGC and referred to as posterior scaling factors. A significant rise or fall in posterior λ_{source} indicates actual emissions for the source are higher or lower, respectively, than inventory estimates (Zhao, 2009).

$$\underline{C} - \underline{C}_{BG} = \underline{\underline{f}} \underline{F}(\lambda_{source}) \quad (\text{Eqn. A.1})$$

Where,

\underline{C} = CH₄ mixing ratio measurements taken at WGC (ppb)

\underline{C}_{BG} = Background CH₄ mixing ratios upstream WGC (ppb)

$\underline{\underline{f}}$ = Footprint (ppb m² s nmol⁻¹)

\underline{F} = Surface flux emissions (nmol m⁻²s⁻¹)

All uncertainties from measured and predicted mixing ratios at WGC are incorporated into the Bayesian inverse analysis as shown by Zhao. Errors from WGC measurements are considered uncorrelated, random and negligible. The list of errors included in the analysis come from several sources: Errors come from releasing a finite number of particles through transport model, aggregating fluxes throughout California into a homogenous flux within a cell where in reality heterogeneous fluxes are present,

transporting particles over upstream areas when PBL height, and wind velocity vector predictions are faulty, estimates of C_{BG} , fluctuations in CH₄ concentration from turbulent eddies, from not considering ocean sources. All errors for each source are assumed to be independent of one another for convenience (Zhao, 2009).

8.1.2. INVERSE FRAMEWORK SPECIFIC TO FIRM'S INVESTIGATION

Equation A.2 develops the inverse framework that may be developed from the work of Zhao (2009) and Lin (2004). The model assumes the background contribution from neighboring states, the rest of the continental U.S., and the world may be fully accounted for by C_{BG} as well as all uncertainties and sources of error within F_{biomod} . The verified level of emissions for each source are derived from the optimized λ_{source} after the Bayesian inverse analysis is conducted. λ_{source} is a vector containing every source, including stationary, e.g. businesses, and moving sources, within Oklahoma that contributes to CO₂ emissions, with the assumption the contribution from every source in the state may be verified using the inverse framework.

$$\underline{C} - \underline{C}_{BG} - \underline{f} \underline{F}_{biomod} = \underline{f} \underline{F}(\lambda_{source}) \quad (A.2)$$

Where,

\underline{C} = Corrected PGS signal at SGP (ppm)

\underline{C}_{BG} = Background contribution to PGS (ppm)

\underline{f} = Footprint (ppm m² s nmol⁻¹)

\underline{F}_{biomod} = Surface flux of biosphere (nmol m⁻²s⁻¹)

$\underline{F}(\lambda_{source})$ = Surface flux of sources (nmol m⁻²s⁻¹)

8.2 FORMAT OF IN SITU MEASUREMENTS TAKEN

Table 1A: Header of each type of matched signal data sets.

| Matched Data Set type | Header Format |
|--|---|
| Nearest PGS to flask signal | [site][date(YYYY mm dd HH MM)] [flask id] [flask (ppm)] [PGS (ppm)] |
| Hourly averaged centered around flask signal | [site][date(YYYY mm dd HH MM)] [flask id] [flask (ppm)] [PGS (ppm)] [standard deviation (ppm)] |

8.3 MAIN ASSUMPTIONS

Filtering for High Atmospheric Variability

The two criterions for removing matched PGS and flask measurements are sufficient to identify periods of atmospheric instability.

Selecting Well-Mixed Periods

Measurements taken within a particular time interval define periods when atmospheric conditions are well-mixed.

Correcting PGS Signal using Matched PGS to Nearest NOAA Measurement

The NOAA mixing ratio measurements are accurate, precise enough to correct the offset in the PGS mixing ratio measurements

Emissions Model

1. The percent growth or decline in emissions from AEP plants for a particular month falling anywhere from 2007 to 2008 directly corresponds to a rise or fall in the mean corrected PGS CO₂ signal at SGP for the same month in 2006.
2. The contribution from background sources including all moving sources inside and outside of Oklahoma, stationary instate industry unrelated to AEP power plants, all stationary instate domestic sources, and stationary sources within regions outside of Oklahoma may be accounted for by the monthly mean background signal.
3. All AEP power plants emissions are upwind of PGS instrument

4. The biospheric exchange of CO₂ may be ignored.
5. The remaining instrument and random errors in the corrected PGS instrument may be ignored.

AEP Emissions

AEP's emissions within Oklahoma each month in 2006 are constant and equal to the total emissions generated by electric utility within the State of Oklahoma in 1999 divided by the number of months in a year.

8.4 PROJECT TIME AND COSTS

Table 2A: Deadlines for Project.

| Task Description | Deadline |
|--|--|
| 1. Search literature for project idea/development | Feb 6, 2012 |
| 2. Elevator pitch | Feb 9, 2012 |
| 3. Brad's approval for project | Feb 9, 2012 |
| 4. RFP assignment (Intro, Literature Review, Methodology, References) | March 9, 2012 |
| 5. Write working program | March 14, 2012 |
| 6. Sensitivity analysis | March 21, 2012 |
| 7. Graphs, tables, figures for report | March 26, 2012 |
| 8. First Half for Sheri (Intro, Literature Review, Methods, Application, References) | March 29, 2012 at 2PM |
| 9. Draft (Application, Results, Discussion, Conclusions) | April 9, 2012 |
| 10. Presentation | April 21, 2012 |
| 11. Finish project paper/revise draft from Sheri | April 23, 2012 |
| 12. Give presentation | April 26, 2012 or May 1, 2012 or May 3, 2012 |
| 13. Submit Final Project Report to Brad and Sheri | May 4, 2012 at 12 PM HGH |

Table 3A: Project Costs

| Task Description | Approx. Staff Hours | Approx. Costs (\$) |
|---|---|--------------------|
| 1. Search literature for project idea/development | Staff engr. 10 hr @\$100/hr | 1,000 |
| 2. Elevator pitch | Staff engr. 3 hrs @\$100/hr | 300 |
| 3. Brad's approval for project | Sen engr 1 hr d@200/hr Staff engr 6 hrs @\$100/hr | 800 |
| 4. RFP assignment (Intro, Literature Review, Methodology, References) | Staff engr 35 hrs @\$100/hr Tech editor 2 hr @\$180/hr | 3,860 |
| 5. Write working program | Staff engr 20 hrs @\$100/hr | 2,000 |

| | | |
|--|---|--------|
| 6. Sensitivity analysis | Staff engr 3 hs @ \$100/hr | 300 |
| 7. Graphs, tables, figures for report | Staff engr 4 hrs @ \$100/hr | 400 |
| 8. First Half for Sheri (Intro, Literature Review, Methods, Application, References) | Staff engr 3 hrs @ \$100/hr Tech editor 1 hr @ \$180/hr | 480 |
| 9. Draft (Application, Results, Discussion, Conclusions) | Staff engr 10 hrs @ \$100/hr Tech editor 2 hr @ \$180/hr | 1360 |
| 10. Presentation | Staff engr 3 hrs @ \$100/hr | 300 |
| 11. Finish project paper/revise draft from Sheri | Staff engr 10 @ \$100/hr | 1,000 |
| 12. Give presentation | Staff engr 2 hr @ \$100/hr | 200 |
| Total Tasks | | 12,000 |

8.5 R SCRIPT INCLUDING DEVELOPED CUBIC SPLINE NUMERICAL METHOD IN BOLD

```

Sys.setenv("TZ"="GMT")
now=Sys.time()
procdate=format(now,"%Y%m%d")

#Read in NOAA min at 60 m
sgp_noaa_pgs_60_min=read.table("sgp_noaa_lbnl_co2_min.60m.d")
names(sgp_noaa_pgs_60_min)=c("SITE","YEAR","MONTH","DAY","HOUR","MIN",
"FLASK-ID","FLASK","PGS")

data_tformat=function(data,col_beg,col_end){
time=data[,2:6]
names(time)=c("YEAR","MONTH","DAY","HOUR","MIN")

time$MONTH=paste("0",time$MONTH, sep="")
time$DAY=paste("0",time$DAY, sep="")
time$HOUR=paste("0",time$HOUR, sep="")
time$MIN=paste("0",time$MIN, sep="")

time$MONTH[nchar(time$MONTH)==3]=substr(time$MONTH[nchar(time$MONTH)
==3],2,3)
time$DAY[nchar(time$DAY)==3]=substr(time$DAY[nchar(time$DAY)==3],2,3)
time$HOUR[nchar(time$HOUR)==3]=substr(time$HOUR[nchar(time$HOUR)==3],2,3)
time$MIN[nchar(time$HOUR)==3]=substr(time$MIN[nchar(time$MIN)==3],2,3)

time=paste(time$YEAR,time$MONTH, time$DAY, time$HOUR, time$MIN,sep="")
new=data[, c(1,col_beg:col_end)]
data_new=cbind(time,new)
data_new[,1]=as.POSIXct(strptime(time, format="%Y%m%d%H"), tz="GMT")
data_new=cbind(data_new,"NUM_TIME"=as.numeric(data_new$time))
return(data_new)}

```

```

sgp_noaa_pgs_60_min_2=data_tformat(sgp_noaa_pgs_60_min,7,9)

length(sgp_noaa_pgs_60_min[,1])
length(sgp_noaa_pgs_60_min_2[,1])

#Remove all data previous to 2006
sgp_noaa_pgs_60_min_2=sgp_noaa_pgs_60_min_2[sgp_noaa_pgs_60_min_2$time>=as.POSIXct("200601010000",format="%Y%m%d%H%M") &
sgp_noaa_pgs_60_min_2$time<as.POSIXct("200901010000",format="%Y%m%d%H%M"),]

length(sgp_noaa_pgs_60_min_2[,1])
diff=sgp_noaa_pgs_60_min_2$PGS-sgp_noaa_pgs_60_min_2$FLASK

#Read in NOAA hour data
sgp_noaa_hour=read.table("sgp_noaa_lbnl_co2_hour.60m.d")

names(sgp_noaa_hour)=c("SITE","YEAR","MONTH","DAY","HOUR","MIN","ID","FLAG",
LASK_CONC","PGS_CONC", "PGS_STD")
sgp_noaa_hour_2=data_tformat(sgp_noaa_hour,7,10)

#identify which standard deviations in the noaa hour data set are less than 1 ppm or equal to and keep their row elements
loc=which(sgp_noaa_hour_2$PGS_STD<1)
length(sgp_noaa_hour_2$PGS_STD)
sgp_noaa_hour_2=sgp_noaa_hour_2[loc,]
time=sgp_noaa_hour_2$time
length(loc)
length(sgp_noaa_hour_2$PGS_STD)

sgp_noaa_pgs_60_min_2[1,]
sgp_noaa_hour_2[1,]
sgp_noaa_pgs_min_clean=sgp_noaa_pgs_60_min_2[1,]
count=1
#only transfer row elements of noaa min that have standard deviation less than or equal to 1ppm to new data frame
for(i in 2:(length(time)+1)){
  if(sum(time[i-1]==sgp_noaa_pgs_60_min_2$time)==1){
    place=which(time[i-1]==sgp_noaa_pgs_60_min_2$time)
    sgp_noaa_pgs_min_clean[count,]=sgp_noaa_pgs_60_min_2[place,]
    count=count+1}}
}

#remove nas from new data set
length(sgp_noaa_pgs_min_clean[,1])
sgp_noaa_pgs_min_clean=sgp_noaa_pgs_min_clean[!is.na(sgp_noaa_pgs_min_clean$ti

```

```

me),]

length(sgp_noaa_pgs_min_clean[,1])
length(sgp_noaa_pgs_60_min_2[,1])

#Calculate difference between flask and nearest pgs signal
sgp_noaa_pgs_min_clean_diff=sgp_noaa_pgs_min_clean$PGS-
sgp_noaa_pgs_min_clean$FLASK
sgp_noaa_pgs_min_clean=cbind(sgp_noaa_pgs_min_clean,DIFF=sgp_noaa_pgs_min_clean$clean_diff)

length(sgp_noaa_pgs_min_clean[,1])
limit1=mean(sgp_noaa_pgs_min_clean$DIFF)-sd(sgp_noaa_pgs_min_clean$DIFF)*3
limit2=mean(sgp_noaa_pgs_min_clean$DIFF)+sd(sgp_noaa_pgs_min_clean$DIFF)*3
sgp_noaa_pgs_min_clean=sgp_noaa_pgs_min_clean[limit1<=sgp_noaa_pgs_min_clean$DIFF & sgp_noaa_pgs_min_clean$DIFF<=limit2,]
length(sgp_noaa_pgs_min_clean[,1])

fnam=paste("all_flask_pgs_resid_atmos_removal_",procdate,".pdf",sep="")
pdf(fnam)
plot(sgp_noaa_pgs_60_min_2$time,diff, xlab="Time", ylab="PGS-NOAA Flask (ppm)",col="black",ylim=c(-10,10))
abline(0,0)
points(sgp_noaa_pgs_min_clean[,1], sgp_noaa_pgs_min_clean$DIFF, cex=1.3, col="red",pch=4)
legend("bottomleft",pch=c("o", "x"), c("Before Removal", "Remain After Removal"),col=c("black", "red"), cex=0.7, pt.cex=1)
dev.off()

#Split data set by taking every other point
count=2
num=round(length(sgp_noaa_pgs_min_clean[,1])*0.5)
cut1=cut2=sgp_noaa_pgs_min_clean[1,]
for(i in 2:(num+1)){
  cut1[i,]=sgp_noaa_pgs_min_clean[(count-1),]
  cut2[i,]=sgp_noaa_pgs_min_clean[(count),]
  count=count+2}
cut1=cut1[(2:length(cut1[,1]))]
cut2=cut2[(2:length(cut2[,1]))]
cut1=cut1[!is.na(cut1[,1]),]
cut2=cut2[!is.na(cut2[,1]),]

cut2=cut2[1:(length(cut2[,1])-1),]

length(cut1[,1])
length(cut2[,1])

```

```

length(cut1[,1])+length(cut2[,1])

mean(cut2$DIFF)
sd(cut2$DIFF)
sqrt((sum(cut2$DIFF^2))/length(cut2$DIFF))

#linear interpolation of first half of residuals to interpolate second half of residuals
lin_interp_resid_cut1=approx(cut1$NUM_TIME,cut1$DIFF, xout=cut1$NUM_TIME)
lin_interp_resid_cut2=approx(cut1$NUM_TIME,cut1$DIFF, xout=cut2$NUM_TIME)

```

#Cubic spline method Begins

#assembly of the tridiagonal matrix

#Chapra, Steven, Et al. 2006. Numerical Methods for Engineers: 5th Edition.

McGrawHill. New York, NY. Pages 495-505.

#assembly of tridiagonal matrix

```

num=dim(cut1)[[1]]-1
set=rep(0,num*num)
trid=matrix(set,ncol=num,byrow=TRUE)
trid2=matrix(set,ncol=num,byrow=TRUE)
RHS=uknown=rep(0,num)

for(i in 1:num){
  if(i==1){
    trid[1,1]=2*(cut1$NUM_TIME[3]-cut1$NUM_TIME[1])
    trid[1,2]=cut1$NUM_TIME[3]-cut1$NUM_TIME[2]
    RHS[1]=(6/(cut1$NUM_TIME[3]-cut1$NUM_TIME[2]))*(cut1$DIFF[3]-
    cut1$DIFF[2])+(6/(cut1$NUM_TIME[2]-cut1$NUM_TIME[1]))*(cut1$DIFF[1]-
    cut1$DIFF[2])
    part=0}
  if(i==num){
    trid[num,num-1]=cut1$NUM_TIME[num]-cut1$NUM_TIME[num-1]
    trid[num,num]=2*(cut1$NUM_TIME[num+1]-cut1$NUM_TIME[num-1])
    RHS[num]=(6/(cut1$NUM_TIME[num+1]-
    cut1$NUM_TIME[num]))*(cut1$DIFF[num+1]-
    cut1$DIFF[num])+(6/(cut1$NUM_TIME[num]-cut1$NUM_TIME[num-
    1]))*(cut1$DIFF[num-1]-cut1$DIFF[num])}
  if(i>1 & i<num){
    trid[i,i-1]=(cut1$NUM_TIME[i+1]-cut1$NUM_TIME[i])
    trid[i,i]=2*(cut1$NUM_TIME[i+2]-cut1$NUM_TIME[i])
    trid[i,i+1]=(cut1$NUM_TIME[i+2]-cut1$NUM_TIME[i+1])
    RHS[i]=(6/(cut1$NUM_TIME[i+2]-cut1$NUM_TIME[i+1]))*(cut1$DIFF[i+2]-
    cut1$DIFF[i+1])+(6/(cut1$NUM_TIME[i+1]-cut1$NUM_TIME[i]))*(cut1$DIFF[i]-
    cut1$DIFF[i+1])}
```

```
cut1$DIFF[i+1])} }
```

#Thomas algorithm

#Finney, Brad. Spring 2012. "Thomas Algorithm Lecture." Humboldt State University. Arcata, CA.

```
e=RHS[1]/trid[1,1]
f=trid[1,2]/trid[1,1]
for(j in 2:num){
  e[j]=(RHS[j]-trid[j,j-1]*e[j-1])/(trid[j,j]-trid[j,j-1]*f[j-1])
  if(j!=num){
    f[j]=trid[j,j+1]/(trid[j,j]-trid[j,j-1]*f[j-1])}
}

uknown[num]=e[num]
for(k in (num-1):1){
  uknown[k]=e[k]-f[k]*uknown[k+1]}
```

#Cubic Interpolation

#Chapra, Steven, Et al. 2006. Numerical Methods for Engineers: 5th Edition. McGrawHill. New York, NY. Pages 495-505.

```
cub_interp=function(cut1,cut2,uknown){
  part1=part2=part3=part4=0
  cub_interp_cut2=rep(-999,length(cut2$NUM_TIME))
  for(m in 1:(length(cut1$NUM_TIME)-1)){
    iter=which(cut2$NUM_TIME>=cut1$NUM_TIME[m] &
    cut2$NUM_TIME<=cut1$NUM_TIME[m+1])
    if(m==1){
      part2=(uknown[1]/(6*(cut1$NUM_TIME[2]-
      cut1$NUM_TIME[1])))*((cut2$NUM_TIME[iter]-cut1$NUM_TIME[1])**3)
      part3=(cut1$DIFF[1]/(cut1$NUM_TIME[2]-
      cut1$NUM_TIME[1]))*(cut1$NUM_TIME[2]-cut2$NUM_TIME[iter])
      part4=((cut1$DIFF[2]/(cut1$NUM_TIME[2]-cut1$NUM_TIME[1]))-
      ((uknown[1]*(cut1$NUM_TIME[2]-
      cut1$NUM_TIME[1])/6)*(cut2$NUM_TIME[iter]-cut1$NUM_TIME[1])))
      cub_interp_cut2[iter]=part2+part3+part4
      part1=part2=part3=part4=0}
    if(m==(length(cut1$NUM_TIME)-1)){
      part1=(uknown[num]/(6*(cut1$NUM_TIME[num+1]-
      cut1$NUM_TIME[num])))*((cut1$NUM_TIME[num+1]-cut2$NUM_TIME[iter])**3)
      part3=((cut1$DIFF[num]/(cut1$NUM_TIME[num+1]-
      cut1$NUM_TIME[num]))-((uknown[num]*(cut1$NUM_TIME[num+1]-
      cut1$NUM_TIME[num])/6))*(cut1$NUM_TIME[num+1]-
      cut2$NUM_TIME[iter]))
      part4=(cut1$DIFF[num+1]/(cut1$NUM_TIME[num+1]-
      cut1$NUM_TIME[num]))*(cut2$NUM_TIME[iter]-cut1$NUM_TIME[num])}
```

```

        cub_interp_cut2[iter]=part1+part3+part4
        part1=part2=part3=part4=0}
        if(m<(length(cut1$NUM_TIME)-1) & m>1){
            part1=(unknown[m-1]/(6*(cut1$NUM_TIME[m+1]-
cut1$NUM_TIME[m])))*((cut1$NUM_TIME[m+1]-cut2$NUM_TIME[iter])**3)
            part2=(unknown[m]/(6*(cut1$NUM_TIME[m+1]-
cut1$NUM_TIME[m])))*((cut2$NUM_TIME[iter]-cut1$NUM_TIME[m])**3)
            part3=((cut1$DIFF[m]/(cut1$NUM_TIME[m+1]-
cut1$NUM_TIME[m]))-((unknown[m-1]*(cut1$NUM_TIME[m+1]-
cut1$NUM_TIME[m]))/6))*(cut1$NUM_TIME[m+1]-cut2$NUM_TIME
[iter])
            part4=((cut1$DIFF[m+1]/(cut1$NUM_TIME[m+1]-
cut1$NUM_TIME[m]))-((unknown[m]*(cut1$NUM_TIME[m+1]-
cut1$NUM_TIME[m]))/6))*(cut2$NUM_TIME[iter]-cut1$NUM_TIME
[m])
            cub_interp_cut2[iter]=part1+part2+part3+part4
            part1=part2=part3=part4=0} }
        return(cub_interp_cut2)}

cub_interp_cut2=cub_interp(cut1,cut2,unknown)
length(cut2[,1])
cut2=cut2[cub_interp_cut2!= -999,]
length(cut2[,1])
cub_interp_cut2=cub_interp_cut2[cub_interp_cut2!= -999]
cub_interp_cut1=cub_interp(cut1,cut1,unknown)
sum(cub_interp_cut2== -999)

iagree_num=sum((cub_interp_cut2-cut2$DIFF)^2)
mean_obs_diff=mean(cut2$DIFF)
iagree_den=sum((abs(cub_interp_cut2-mean_obs_diff)+abs(cut2$DIFF-
mean_obs_diff))^2)

cub_iagree=1-((iagree_num)/(iagree_den))
cub_iagree

iagree_num=sum((lin_interp_resid_cut2$y-cut2$DIFF)^2)
mean_obs_diff=mean(cut2$DIFF)
iagree_den=sum((abs(lin_interp_resid_cut2$y-mean_obs_diff)+abs(cut2$DIFF-
mean_obs_diff))^2)

lin_iagree=1-((iagree_num)/(iagree_den))
lin_iagree

#test
test1=cut2$DIFF-0.1*cut2$DIFF
iagree_num=sum((test1-cut2$DIFF)^2)

```

```

mean_obs_diff=mean(cut2$DIFF)
iagree_den=sum((abs(test1-mean_obs_diff)+abs(cut2$DIFF-mean_obs_diff))^2)

test1_iagree=1-((iagree_num)/(iagree_den))
test1_iagree

test2=10
iagree_num=sum((test2-cut2$DIFF)^2)
mean_obs_diff=mean(cut2$DIFF)
iagree_den=sum((abs(test2-mean_obs_diff)+abs(cut2$DIFF-mean_obs_diff))^2)

test2_iagree=1-((iagree_num)/(iagree_den))
test2_iagree

#plot of residuals and cubic interpolation
fnam=paste("interp_resid_2006_2008_", procdate,".pdf", sep="")

pdf(fnam)
par(mfrow=c(2,1))
plot(cut1[,1], cut1$DIFF, ylab="PGS-NOAA (ppm)", ylim=c(-5,5))
points(cut2[,1], cut2$DIFF, cex=1.3, col="red")
lines(cut2$time, cub_interp_cut2, col="red")
lines(cut1$time, cub_interp_cut1)
abline(0,0)
legend("bottomleft", pch=c("-", "-", "o", "o"), c("First Half Residual Interpolation",
"Second Half Residual Interpolation", "First Half Residuals", "Second Half
Residuals"), col=c("black", "red", "black", "red"), cex=0.7, pt.cex=1, ncol=2)
#dev.off()

#plot of residuals and linear interpolation
plot(cut1[,1], cut1$DIFF, xlab="Time (year)", ylab="PGS-NOAA (ppm)", ylim=c(-5,5))
points(cut2[,1], cut2$DIFF, cex=1.3, col="red")
lines(cut1$time, lin_interp_resid_cut1$y)
lines(cut2$time, lin_interp_resid_cut2$y, col="red")
abline(0,0)
legend("bottomleft", pch=c("-", "-", "o", "o"), c("First Half Residual Interpolation",
"Second Half Residual Interpolation", "First Half Residual Interpolation", "Second Half
Residuals"), col=c("black", "red", "black", "red"), cex=0.7, pt.cex=1, ncol=2)
dev.off()

#Corrected PGS signal at 60 m
PGS_lin_corr_cut1=cut1$PGS-lin_interp_resid_cut1$y
PGS_cub_corr_cut1=cut1$PGS-cub_interp_cut1

PGS_lin_corr_cut2=cut2$PGS-lin_interp_resid_cut2$y

```

```

PGS_cub_corr_cut2=cut2$PGS-cub_interp_cut2

resid_after_cor_lin=PGS_lin_corr_cut2-cut2$FLASK
resid_after_cor_cub=PGS_cub_corr_cut2-cut2$FLASK

mean(cut2$DIFF)

sd(cut2$DIFF)

mean(resid_after_cor_lin)

sd(resid_after_cor_lin)

mean(resid_after_cor_cub)

sd(resid_after_cor_cub)

#plot of residuals before and after correction
fnam=paste("resid_before_&_after_cor_2006_2008_", procdate,".pdf", sep="")
pdf(fnam)
plot(cut2[,1], cut2$DIFF, xlab="Time (year)", ylab="PGS-NOAA (ppm)", ylim=c(-5,5))
lines(cut2$time, resid_after_cor_cub, col="blue")
lines(cut2$time, resid_after_cor_lin, col="red")
abline(0,0)
legend("bottomleft",pch=c('o', '-','-' ), c("Before Correction", "After Cubic Spline
Correction", "After Linear Correction"),col=c("black","blue", "red"), cex=0.7, pt.cex=1)
dev.off()

rm(sgp_noaa_hour,sgp_noaa_pgs_60_min,sgp_noaa_pgs_min_clean,iagree_num,iagree_
den)

data_tformat=function(data,col_beg,col_end){
time=data[,2:5]
names(time)=c("YEAR","MONTH","DAY","HOUR")

time$MONTH=paste("0",time$MONTH, sep="")
time$DAY=paste("0",time$DAY, sep="")
time$HOUR=paste("0",time$HOUR, sep="")

time$MONTH[nchar(time$MONTH)==3]=substr(time$MONTH[nchar(time$MONTH)
==3],2,3)
time$DAY[nchar(time$DAY)==3]=substr(time$DAY[nchar(time$DAY)==3],2,3)
time$HOUR[nchar(time$HOUR)==3]=substr(time$HOUR[nchar(time$HOUR)==3],2,3)
}

```

```

time_new=paste(time$YEAR,time$MONTH, time$DAY, time$HOUR,sep="")
new=data[, c(1,col_beg:col_end)]
data_new=cbind(time_new,new)
data_new[,1]=as.POSIXct(strptime(time_new, format="%Y%m%d%H"), tz="GMT")
data_new=cbind(data_new,"NUM_TIME"=as.numeric(data_new$time),
"YEAR"=time$YEAR,"MONTH"=time$MONTH,"HOUR"=time$HOUR)
return(data_new)}

pgs_2006=read.table("sgp_2006.d")
pgs_2006=pgs_2006[pgs_2006[,6]!=-999.990 | pgs_2006[,7]!=-999.990,]
pgs_2006=data_tformat(pgs_2006,6,8)

pgs_2007=read.table("sgp_2007.d")
pgs_2007=pgs_2007[pgs_2007[,6]!=-999.990 | pgs_2007[,7]!=-999.990,]
pgs_2007=data_tformat(pgs_2007,6,8)

pgs_2008=read.table("sgp_2008.d")
pgs_2008=pgs_2008[pgs_2008[,6]!=-999.990 | pgs_2008[,7]!=-999.990,]
pgs_2008=data_tformat(pgs_2008,6,8)

pgs_06_08=rbind(pgs_2006, pgs_2007, pgs_2008)

names(pgs_06_08)=c("TIME","SITE","PGS_CONC", "STD",
"Samples","NUM_TIME","Year","Month","Hour")

rm(pgs_2006, pgs_2007, pgs_2008)
pgs_06_08$Hour=as.numeric(pgs_06_08$Hour)

sum(is.na(pgs_06_08[pgs_06_08$Hour>=12,]))
length(pgs_06_08[,1])
#Removing times when PBL is not likely to be mixed, 6AM to 6PM CST or 12 to 24
GMT
pgs_06_08=pgs_06_08[pgs_06_08$Hour>=12,]
length(pgs_06_08[,1])

set=range(cut1[,1])

pgs_06_08=pgs_06_08[pgs_06_08[,1]>set[1] & pgs_06_08[,1]<set[2],]

pgs_06_08$STD=as.numeric(pgs_06_08$STD)
pgs_06_08=pgs_06_08[pgs_06_08$STD<1,]

lin_interp_resid_06_08=approx(cut1$time,cut1$DIFF, xout=as.numeric(pgs_06_08[,1]))

```

```

cub_interp_resid_06_08=cub_interp(cut1,pgs_06_08,unknown)

lin_PGS_cor_06_08=pgs_06_08$PGS_CONC-lin_interp_resid_06_08$y
cub_PGS_cor_06_08=pgs_06_08$PGS_CONC-cub_interp_resid_06_08

length(pgs_06_08$TIME)
length(lin_PGS_cor_06_08)
length(cub_PGS_cor_06_08)

lst=as.numeric(paste(pgs_06_08$Year,pgs_06_08$Month,sep=""))
cat=unique(lst)
PGS_mave=data.frame(year_month=cat,uncor_PGS=rep(-
999,length(cat)),lin_PGS_cor=rep(-999,length(cat)),cub_PGS_cor=rep(-999,length(cat)))
for(i in 1:length(cat)){
  stor=which(cat[i]==lst)
  PGS_mave$uncor_PGS[i]=mean(pgs_06_08$PGS_CONC[stor])
  PGS_mave$lin_PGS_cor[i]=mean(lin_PGS_cor_06_08[stor])
  PGS_mave$cub_PGS_cor[i]=mean(cub_PGS_cor_06_08[stor])}

global_co2=read.table("co2_global.txt")
names(global_co2)=c("year","month","decimal","average","trend")
global_co2
PGS_mave
global_co2=global_co2[313:348,]

global_co2$month=paste("0",global_co2$month, sep="")
global_co2$month[nchar(global_co2$month)==3]=substr(global_co2$month[nchar(global_co2$month)==3],2,3)

time_new=paste(global_co2$year,global_co2$month,"01",sep="")
time_new=as.POSIXct(strptime(time_new, format="%Y%m%d"), tz="GMT")

#NOAA monthly averaged marine co2 concentration
fnam=paste("global_co2_conc_",procdate,".pdf",sep="")
pdf(fnam)
plot(time_new,global_co2$average, xlab="Time", ylab="NOAA Monthly Averaged
Global Marine CO2 Mixing Ratio (ppm)",ylim=c(370,400))
dev.off()

time_new2=paste(PGS_mave$year_month,"01",sep="")
time_new2=as.POSIXct(strptime(time_new2, format="%Y%m%d"), tz="GMT")

#mave of linear and cubic correction and NOAA background
fnam=paste("mave_NOAA_&_Cor_2006_2008_", procdate, ".pdf", sep="")
pdf(fnam)
plot(time_new2, PGS_mave$lin_PGS_cor, xlab="Time", ylab="Monthly Averages of

```

```

Mixing Ratios (ppm)", ylim=c(370,400))
points(time_new2, PGS_mave$cub_PGS_cor, cex=1.3, col="blue", pch="x")
points(time_new, global_co2$average, cex=1.3, col="red")
legend("bottomleft", pch=c("o", "x", "o"), c("Linear Correction", "Cubic Spline
Correction", "NOAA Global Marine Average"), col=c("black", "blue", "red"), cex=0.7,
pt.cex=1)
dev.off()

uncor_PGS_glob_diff=PGS_mave$uncor_PGS-global_co2$average
lin_PGS_glob_diff=PGS_mave$lin_PGS_cor-global_co2$average
cub_PGS_glob_diff=PGS_mave$cub_PGS_cor-global_co2$average

PGS_mave=cbind(PGS_mave, uncor_glob_diff=uncor_PGS_glob_diff, lin_glob_diff=lin_
PGS_glob_diff, cub_glob_diff=cub_PGS_glob_diff)

uncor_PGS_06=PGS_mave$uncor_glob_diff[1:12]
uncor_PGS_07=PGS_mave$lin_glob_diff[13:24]
uncor_PGS_08=PGS_mave$cub_glob_diff[25:33]

lin_PGS_06=PGS_mave$lin_glob_diff[1:12]
lin_PGS_07=PGS_mave$lin_glob_diff[13:24]
lin_PGS_08=PGS_mave$lin_glob_diff[25:33]

cub_PGS_06=PGS_mave$cub_glob_diff[1:12]
cub_PGS_07=PGS_mave$cub_glob_diff[13:24]
cub_PGS_08=PGS_mave$cub_glob_diff[25:33]

per_growth_uncor_07=((uncor_PGS_07-uncor_PGS_06)/uncor_PGS_06)
per_growth_uncor_08=((uncor_PGS_08-uncor_PGS_06[1:9])/uncor_PGS_06[1:9])

per_growth_uncor_07
per_growth_uncor_08

per_growth_lin_07=((lin_PGS_07-lin_PGS_06)/lin_PGS_06)
per_growth_lin_08=((lin_PGS_08-lin_PGS_06[1:9])/lin_PGS_06[1:9])

per_growth_lin_07
per_growth_lin_08

per_growth_cub_07=((cub_PGS_07-cub_PGS_06)/cub_PGS_06)
per_growth_cub_08=((cub_PGS_08-cub_PGS_06[1:9])/cub_PGS_06[1:9])

per_growth_cub_07
per_growth_cub_08

uncor_fine_07=sum(per_growth_uncor_07[per_growth_uncor_07>=0]*(11701469/12)*1

```

```

00*1.3333)
uncor_fine_08=sum(per_growth_uncor_08[per_growth_uncor_08>=0]*(11701469/12)*1
00*1.3333)

uncor_fine_07
uncor_fine_08

lin_fine_07=sum(per_growth_lin_07[per_growth_lin_07>=0]*(11701469/12)*100*1.333
3)
lin_fine_08=sum(per_growth_lin_08[per_growth_lin_08>=0]*(11701469/12)*100*1.333
3)

lin_fine_07
lin_fine_08

cub_fine_07=sum(per_growth_cub_07[per_growth_cub_07>=0]*(11701469/12)*100*1.
3333)
cub_fine_08=sum(per_growth_cub_08[per_growth_cub_08>=0]*(11701469/12)*100*1.
3333)

cub_fine_07
cub_fine_08

SGP_PGS_CORR_HOURLY=data.frame(t_GMT=pgs_06_08$TIME,uncorr_ppm=pgs_
06_08$PGS_CONC,lin_corr_ppm=lin_PGS_cor_06_08,cub_corr_ppm=cub_PGS_cor_0
6_08,std_ppm=pgs_06_08$STD,samples=pgs_06_08$Samples)

#plot of corrected signal using cubic interpolation and signal before correction
fnam=paste("final_cor_sig_2006_2008_", procdate,".pdf", sep="")

pdf(fnam)
par(mfrow=c(2,1))
plot(SGP_PGS_CORR_HOURLY$t_GMT, SGP_PGS_CORR_HOURLY$uncorr_ppm,
ylab="PGS-NOAA (ppm)",pch=".")  

points(SGP_PGS_CORR_HOURLY$t_GMT,
SGP_PGS_CORR_HOURLY$cub_corr_ppm, cex=1.3, col="red",pch=".")  

legend("bottomleft",pch=c("o", "o"), c("Before Correction", "After
Correction"),col=c("black", "red"), cex=0.7, pt.cex=1, ncol=2)

#plot of corrected signal using linear interpolation and signal before correction

plot(SGP_PGS_CORR_HOURLY$t_GMT, SGP_PGS_CORR_HOURLY$uncorr_ppm,
xlab="Time (year)", ylab="PGS-NOAA (ppm)",pch=".")  

points(SGP_PGS_CORR_HOURLY$t_GMT,
SGP_PGS_CORR_HOURLY$lin_corr_ppm, cex=1.3, col="red",pch=".")
```

```
legend("bottomleft",pch=c("o", "o"), c("Before Correction", "After  
Correction"),col=c("black","red"), cex=0.7, pt.cex=1,ncol=2)  
dev.off()  
  
fnam=paste("SGP.60m.PGS.NOAAcorr.2006_2008.",procdate,".csv",sep="")  
write.table(SGP_PGS_CORR_HOURLY, file=fnam, sep=",",row.names=FALSE)  
  
fnam=paste("SGP.60m.PGS.NOAAcorr.2006_2008.",procdate,".Rdata",sep="")  
save(SGP_PGS_CORR_HOURLY, file=fnam)
```