Final Report

Impact of large-scale circulation patterns on surface ozone concentrations in Houston-Galveston-Brazoria (HGB)

AQRP Project 14-010

Prepared for:

Vincent Torres
Texas Air Quality Research Program
The University of Texas at Austin

Prepared by:

Yuxuan Wang (Principal Investigator)
Department of Marine Sciences
Texas A&M University at Galveston

September 2015

QA Requirements: Audits of Data Quality: 10% Required

Acknowledgement

The preparation of this report is based on work supported by the State of Texas through the Air Quality Research Program administered by The University of Texas at Austin by means of a Grant from the Texas Commission on Environmental Quality.

Executive Summary

The Bermuda High is a key driver of large-scale circulation patterns in Southeastern Texas in summer (Davis et al., 1997). There are two mechanistic linkages between the Bermuda High (BH) and surface ozone in the Houston-Galveston-Brazoria (HGB) region: first, the western extension of the BH defines the strength of the southerly low-level jet (LLJ) that brings marine air with lower ozone background from the Gulf of Mexico (Higgins et al. 1997); second, the high pressure system allows for clear skies and high temperature conditions that are favorable for local production of ozone (O_3). This project investigates the complex effects of the BH and the related meteorological conditions on surface O_3 variations in the HGB region by analyzing the more than decade-long observational record of maximum daily 8 h average (MDA8) surface ozone and meteorology from June to September. The indicators of the BH location and strength developed/refined in this project are the longitude index of the BH western edge (BH-Lon), and two BH intensity related indices (BHI1 and BHI2). The BH indicators are proved to have significant utility in explaining the year-to-year variability in monthly mean HGB MDA8 ozone for June, July, August, and September during 1998-2013. Other indicators of large-scale meteorological conditions, including Palmer Drought Severity Index (PDSI), Arctic oscillation (AO), and HGB mean temperatures, are found to be of lesser utility than the BH indicators, but still show significant correlations with HGB ozone variability in some months. Through stepwise regression based on the Akaike Information Criterion (AIC), these meteorological predictors are employed to develop the multiple linear regression (MLR) model which reproduces more than 50% of interannual variability of the monthly mean MDA8 ozone over the HGB from June to September.

The observation-derived statistical relationships between the HGB ozone and BH are then used to develop an empirical scheme to correct for the known high bias of the GEOS-Chem chemical transport model when simulating summertime ozone along the Gulf Coast (Li et al.,

2002; Fiore et al., 2002; Reidmiller et al., 2009; Zhang et al., 2011; McDonald-Buller, 2011). A set of multiple-year simulations of HGB ozone is conducted using the GEOS-Chem model. A moderate to strong correlation is identified between the BH-Lon and the model bias for June and July, which supports the hypothesis that the model bias is caused in part by the insufficient representation of the dynamic linkage between BH and ozone inflow to HGB. The following provides a summary of major findings:

1. The MLR model we developed captures 58% - 72% of interannual variance during 1998-2013 of the monthly mean MDA8 ozone over the HGB from June to September, indicating the significant role of large-scale meteorology on ozone variability for this region. The Bermuda High variability is the most important meteorological driver of ozone variability for each month.
2. The cross-validation (CV) and hindcast analyses suggest that the MLR models have good skills in predicting interannual (1995-2013) variations of the monthly-mean MDA8 ozone over the HGB from June to September, with CV R^2 higher than 0.45 for each month.
3. The relationships between the BH-Lon and HGB ozone are valuable observational constraints that can be used to correct the high bias in simulated ozone. After the correction, the mean model bias for June and July, months with the largest model bias, shows a 70-75% decrease and the correlation coefficient between the observed and simulated ozone also improves.

Table of Contents

1. Introduction

Surface ozone, as an important air pollutant, has significant impacts on both public health and agriculture (Wang et al., 2004; Berman et al., 2012). Surface ozone is influenced not only by emissions of ozone precursors but also by circulation patterns through complex dynamical, physical, and chemical processes (e.g. Jacob and Winner 2009). Some circulation patterns will result in favorable local meteorological conditions for ozone exceedances, such as high temperatures, low wind speeds, clear skies, and stagnation (Zhu et al., 2013; Ngan et al., 2011; Jacob and Winner, 2009). Previous studies have demonstrated the associations between large-scale circulations and surface ozone concentrations over the United States (e.g. Lin et al., 2012; Shen et al., 2015; Lin et al., 2015). Surface ozone concentrations in the Western US are associated with mid-latitude cyclones which transport Asian pollutions eastward to the Pacific (Lin et al., 2012). More frequent late spring stratospheric intrusions occurring following strong La Niña winters can also elevate western US surface ozone (Lin et al., 2015). In the Midwest and Northeast U.S., polar jet frequency has been found to be a good indicator for surface ozone interannual variability (Shen et al., 2015).

During the summer in the Southeastern United States, the Bermuda High is the key large-scale circulation pattern that influences the weather and climate. The Bermuda High (BH), a quasi-permanent system located over the North Atlantic Ocean in summer (Davis et al.1997) has significant influences on the regional climate and thus surface ozone over the eastern United States (Li et al., 2012; Zhu et al., 2013; Hegarty et al., 2007; Hogrefe et al., 2004; Shen et al., 2015). In summer, the westward extension of the Bermuda High places the eastern United States under the control of high pressure, and produces high temperatures and clear skies therein, which are favorable for local production of ozone. The Bermuda High also defines the strength of the low-level jet (LLJ) over the southern Great Plains. The southerly flows at the west edge of the Bermuda High bring clean marine air from the Gulf of Mexico to the southern

Great Plains (Higgins et al. 1997). This, on the other hand, is favorable for the reduction of background ozone. Modeling results from Hogrefe et al. (2004) show that high ozone concentrations easily occur over large parts of the Northeast under the Bermuda High pressure pattern. Using observational data, Zhu and Liang (2013) found positive correlations between summer maximum daily 8 h average (MDA8) surface ozone and the intensity of the BH on the interannual time scale over the Northeast, and negative correlations over the South-Central US. Analyzing the observational data from 1980 to 2012, Shen et al. (2015) suggested that the location of the Bermuda High western edge is related to summer MDA8 ozone in the Southeast. He also stated that in the summers, when the average position of Bermuda High western edge is located west of 85.4°W, any additional westward shift of the western edge will lead to an increase of Southeast U.S. ozone by ~1 ppbv deg^{-1} in longitude.

The Houston-Galveston-Brazoria (HGB) area is usually near the western edge of the BH in summer (June, July and August). It was classified in 2012 as a "marginal" nonattainment zone for O_3 by the U.S. Environmental Protection Agency (EPA) under the 2008 standard (TCEQ, 2012). Though background ozone of the HGB area has decreased in the past decade (Nielsen-Gammon et al., 2005; Berlin et al. 2014), the HGB region still witnessed many high surface ozone episodes. Some of the high ozone episodes have been proved to be influenced by large-scale circulations and associated meteorological factors (Ngan et al., 2011; Rappenglück et al., 2008; Pakalapati et al., 2009; Haman et al., 2014). Rappenglück et al. (2008) and Ngan et al. (2011) both pointed out that ozone exceedance days over the HGB always occur under the dominance of synoptic northerly and easterly flows, since the presence of easterly flows can transport emissions from the Ship Channel into the center of the urban area, and the presence of either northerly or easterly winds can advect continental background ozone to the Houston area. Modeling results sometimes show significant differences between planetary boundary layer (PBL) heights on high ozone days and low ozone days (Haman et al., 2014). On high ozone

days, the nighttime and early morning PBL heights tend to be relatively lower (Haman et al., 2014).

The studies mentioned above mostly focused on the connections between weather patterns and high ozone concentrations on episodic cases, rather than the long-term variations. Zhu and Liang (2013) and Shen et al. (2015) have revealed a moderate-to-strong negative correlation between interannual anomalies of the BH and summer mean MDA8 ozone over the southern US. Summer means the months of June, July, and August (JJA) hereafter in this report. Until now, there has been no attempt to quantify the contribution of the BH variability on surface O_3 concentrations over HGB. Here we aim to develop a statistical model to characterize the quantitative relationship between the variability of BH and monthly MDA8 ozone over the HGB on an interannual timescale. The statistical model derived herein will be beneficial in understanding the role that meteorology plays in the interannual variability of the HGB ozone. The second aim is to assess the skill of the statistical model in hindcasting surface ozone variations over the HGB.

Furthermore, it is a known problem that the GEOS-Chem global chemical transport model (CTM), like many other global models, has a tendency to overestimate ozone at Gulf Coast sites in summer (Li et al., 2002; Fiore et al., 2002; Reidmiller et al., 2009; Zhang et al., 2011; McDonald-Buller, 2011). While inadequate marine boundary layer chemistry has been proposed as one possible explanation for this high bias in the global models, the bias can also be caused by insufficient representation of the dynamic linkage between BH and ozone inflow to the HGB from the Gulf of Mexico (Fiore et al., 2002). Thus, the third aim is to use the observation-derived relationships between BH and HGB ozone to design an empirical bias correction scheme for the GEOS-Chem global CTM to improve its simulation of the background O_3 associated with maritime inflow to the HGB.

2. Methods

2.1 Observational Ozone Data and Meteorological Data

Surface ozone concentrations over the HGB have been routinely monitored at continuous ambient monitoring stations (CAMSs) maintained by the Texas Commission on Environmental Quality (TCEQ), the City of Houston, and Harris County. In our study, observational records of MDA8 ozone during the ozone season (May 1 - September 30) from 1998 to 2013 over the HGB region were obtained from the TCEQ website. To remove the influence of trends in anthropogenic emissions over the HGB on interannual variability of ozone, the MDA8 ozone data is detrended. Since the change of ozone precursor emissions may not be linear, we detrend the raw MDA8 ozone time series by subtracting the 3-year moving average. Thus, the time series of detrended MDA8 ozone is from 1999 to 2012. We verify that the results from the project do not depend on the detrending method (c.f. Section 5). As an example, Figure 1 shows the raw (solid black line) and detrended (dashed black line) time series of monthly mean MDA8 ozone for September.

The meteorological data used in this study consists of the geopotential height at 850hPa, and sea level pressure (SLP) from the National Centers for Environmental Prediction (NCEP) Reanalysis 1 with a spatial resolution of $2.5^{o}\times2.5^{o}$ (Kalnay et al., 1996), and 2-meter temperature from the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim reanalysis with $0.5^{o}\times0.5^{o}$ spatial resolution. We also collected Palmer Drought Severity Index (PDSI) time series and time series of Arctic Oscillation (AO) from the NOAA website as potential predictors for surface ozone. Since our focus is on synoptic-scale variability, all the meteorological indices except PDSI (including AO, 2-meter temperature, and the various Bermuda High indices described in 2.3) were detrended by subtracting a best-fit linear trend from the raw time series. The time series of PDSI was not detrended because PDSI is a

normalized index by definition. As an example, Figure 1 shows the raw (solid red line) and detrended (dashed red line) time series of one meteorological index (the longitude index of the BH western edge or BH-Lon) for September.

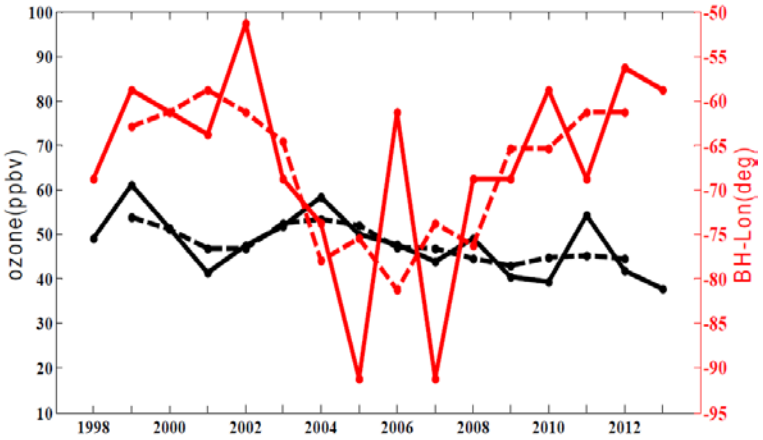

Figure 1. Time series of raw (solid) and detrended (dashed) data of the monthly-mean MDA8 ozone (black lines) and the BH-Lon (red lines) for September.

2.2 Study Area

The study area includes the region of Houston, Galveston, and Brazoria (HGB), delineated by longitude from -94.5°W to -96.0°W, and by latitude from 28.5°N to 30.5°N (black box in Figure 2). A total of 28 ozone CAMS sites are in this region. Figure 2 shows the long-term (1998-2013) mean MDA8 ozone from May 1 to September 30. Ozone concentrations for sites in Galveston and Brazoria counties are relatively lower than the sites in the Houston region, due (in part) to lower local emissions.

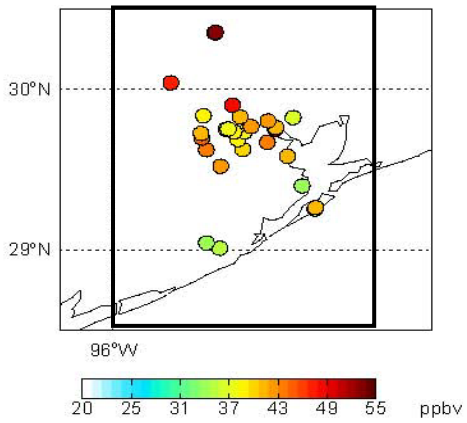

Figure 2. Map of the study area showing the locations of the sites and long-term (1998-2013) mean MDA8 ozone concentrations from May 1 to September 30.

2.3 Meteorological Indices

Several indices have been used in the literature to define the intensity and the location of the BH (Stahle and Cleaveland, 1992; Ortegren et al. 2011, Li et al. 2011; Zhu and Liang, 2013). The majority of the existing indices describing the intensity of the BH are defined on the basis of SLP differences between two locations, one near New Orleans and the other near Bermuda, with their exact locations varying among studies. In Zhu and Liang (2013) study, the BH intensity index (BHI) was defined as the regional mean SLP difference between the Gulf of Mexico (25.3°-29.3°N, 95°-90°W) and the southern Great Plains (35°-39°N, 105.5°-100°W). They found an association between their BHI and the strength of LLJ, which determines the transport of clean marine air from the Gulf of Mexico. Similar to their definition, we define our first BHI (BHI1) as the mean SLP difference between the Gulf of Mexico (25.3°-29.3°N, 92.5°-87.5°W) (box 1 in Figure 3) and the southern Great Plains (35°-39°N, 105.5°-100°W) (box 2 in Figure 3). Our second BHI (BHI2) is defined as the mean SLP difference between the Gulf of Mexico (box 1 in Figure 3) and the northeast Texas (31°-36°N, 91°-96°W) (box 3 in Figure 3). BHI1 and BHI2 are intended to indicate the meridional and zonal wind speed over the HGB region, respectively.

According to Ngan et al (2013), the possibility of exceeding the 8-h ozone standard is higher when there are easterly and northerly winds over the HGB, which corresponds to a negative BHI1 and a positive BHI2 respectively.

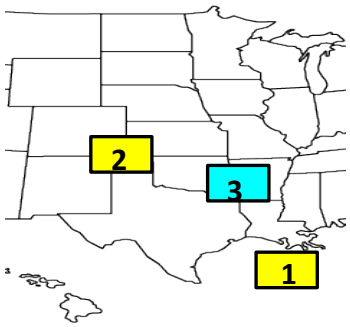

Figure 3. Regions used to define the BH intensity indices BHI1 and BHI2.

We also calculate BH-Lon to measure the westward extension of the BH, as defined by Li et al. (2011). Similar to their definition, we locate BH-Lon as the cross point of the 1560 geopotential meter (gpm) isoline and the 850hPa wind ridgeline. As shown in Figure 4, the mean seasonal variation of BH-Lon correlates well with that of surface ozone concentrations in the HGB from May to September. The trough of ozone in July is accompanied by the lowest BH-Lon of the year, which means the most westward extension of the BH occurs in July. The westward extension of BH (lower BH-Lon) is accompanied by stronger inflow of maritime air with lower ozone background into the HGB; that is why BH-Lon can explain the seasonal variations of ozone over the HGB. The whole HGB region is under the control of strong southerly winds on 850hPa from June to September (Figure 5). Since the relations between BH-Lon and the HGB ozone may differ month by month during the ozone season, we analyze the impact of the BH on the HGB ozone for individual months. In some years (e.g. 2005), the 1560-gpm isoline does not exist over the Bermuda region in May. Because of the instability of the BH in May, we only calculate BH-Lon from June to September in the later analysis. The BH-Lon for June and July is calculated using the 1560 gpm isoline as defined by Li et al. (2011). Significantly negative correlations are found between the BH-Lon and detrended HGB-mean

MDA8 ozone for June and July during 1999-2012, with the squares of the correlation coefficient (R^2) being 0.49 and 0.58, respectively. However, the BH-Lon calculated the same way for August and September does not show a significant correlation with ozone. Since the BH in August and September are much weaker than that in June and July, it may not be appropriate to use the same isoline to characterize the BH position throughout the ozone season. We tried different isolines with an interval of 4 gpm from 1560 to 1536 gpm in calculating the BH-Lon for August and September. The 1556-gpm and 1536-gpm isoline is found to be most appropriate to define the BH-Lon for August and September respectively, because the resulting BH-Lon shows the highest correlation with the detrended HGB-mean MDA8 ozone with R^2 being 0.20 for both months.

Besides the BH-based indices, we also calculate the HGB-mean PDSI, HGB-mean surface temperatures, and the Arctic Oscillation time series as candidates of potential predictors of the HGB-mean MDA8 ozone. Note that all the BH-related indices are developed on the basis of NCEP reanalysis and the HGB-mean temperatures are calculated using ERA-Interim reanalysis.

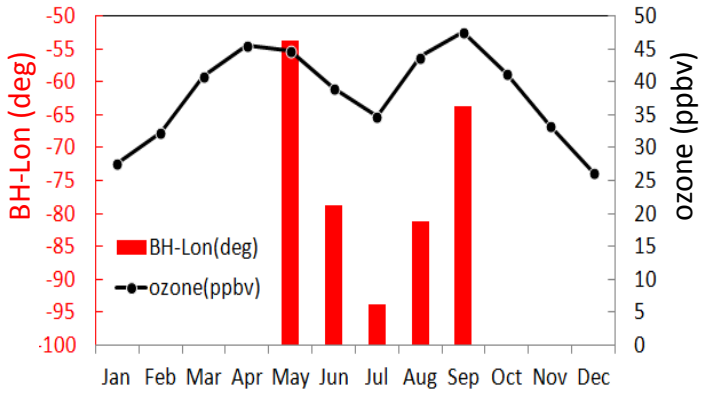

Figure 4. Seasonal variations of the BH-Lon and HGB-mean MDA8 ozone.

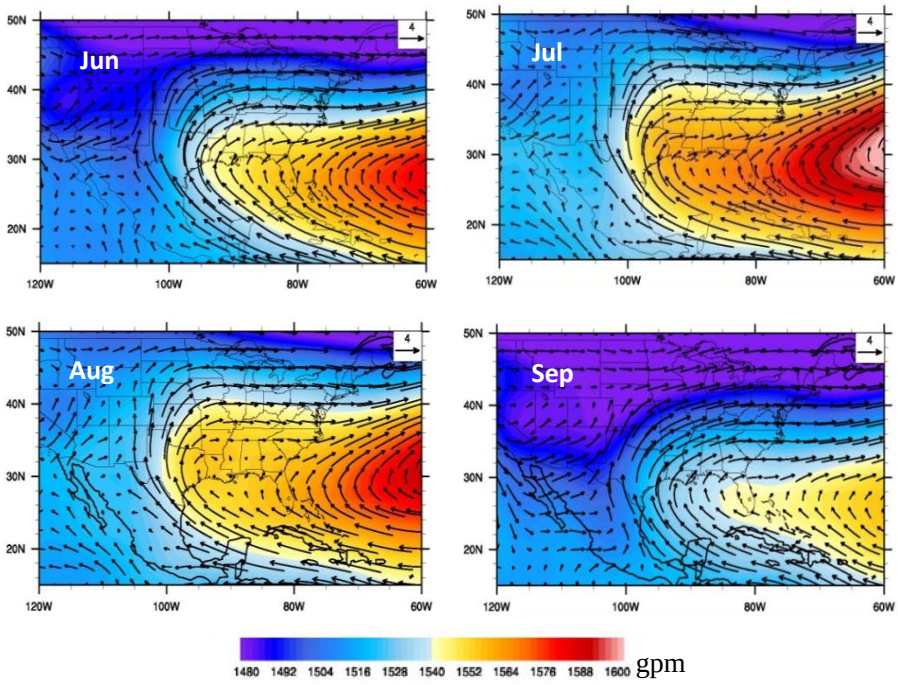

Figure 5. Distribution of the long-term mean 850hPa geopotential height and wind field in June, July, August, and September.

2.4 Statistical Method

We apply a multiple linear regression (MLR) model, which has been commonly used in air quality and climate studies (e.g. Kutner et al., 2004; Tai et al., 2010), to construct the statistical relationship between the HGB-mean monthly MDA8 ozone and the predictors selected in the previous section. For an easy comparison, we normalize all the predictors (x_k) as well as the dependent variable (y) in the later analysis. The model is of the form：

$$y=\beta_0+\sum_{k=1}^{6}\beta_k x_k \qquad (1)$$

where y is the detrended and normalized monthly mean MDA8 O_3 over the HGB region,

x_k is one of the six predictors presented above (i.e., BH-Lon, BHI1, BHI2, HGB-mean temperature, PDSI, and AO) which is detrended and normalized except for PDSI, β_k is the corresponding regression coefficient for x_k, and β_0 is the intercept. We apply a stepwise regression to add and delete terms based on the Akaike Information Criterion (AIC) statistics to obtain the best model fit (Venables and Ripley, 2003). The AIC is a measure of the relative quality of the statistical models for a given set of data. Given a collection of possible models for the data, AIC estimates the quality of each model, relative to each of the other models.

3. Quality Assurance/Quality Control Procedures

3.1 Data Quality Assurance and Quality Control

The CAMSs data of MDA8 ozone over the HGB region were downloaded in ASCII format from the TCEQ website. Even though there are some missing data, the overall data coverage is good at 99%. Since we focus on the HGB mean ozone, the missing data were simply discarded and this treatment will not affect the monthly statistics of MDA8 ozone used in the analysis.

To assess the sensitivity of the MLR model to different metrics of MDA8 ozone used in the regression, we tested the regression performance using not only the monthly mean ozone values but also the median, background ozone, and ozone enhancement (defined as measured ozone minus ozone background). The background ozone data were obtained from Mark Estes at the TCEQ. A Matlab program was used to read the ozone data and compute the statistical analyses. The best regression statistics were obtained when monthly mean MDA8 ozone was used as the dependent variable, although other ozone metrics gave consistent regression results.

The 2.5°× 2.5° meteorological reanalysis data were collected from the NOAA website and the 0.5°× 0.5° meteorological reanalysis data were collected from the ECMWF website. All the

reanalysis data were archived in NetCDF format. The time series of AO was obtained in ASCII format from the NOAA website. NCAR Command Language (NCL) program was used to read the reanalysis data.

The meteorological predictors/indicators used in the analysis were calculated using different reanalysis datasets to examine whether the analysis is sensitive to the choice of source data. The BH related indices (BH-Lon, BHI1, and BH2) derived from the coarser-resolution NCEP reanalysis were largely consistent with those derived from the finer-resolution ECMWF and the North America Regional Reanalysis (NARR), supporting our hypothesis that the BH indicators represent the large-scale circulation patterns and thus are not dependent on the resolution of reanalysis data used.

The GEOS-Chem global CTM has a standard benchmarking procedure for each major code release, using observations compiled from surface monitoring network, aircraft campaigns, and satellite retrievals around the globe. The GEOS-Chem model simulation results were thoroughly evaluated using not only surface ozone observations in the HGB but also the benchmark observations in the continental U.S. Interactive Data Language (IDL) programs were used to visualize and extract the corresponding model outputs for the comparisons with the observational data.

3.2 Audits of Data Quality

The quality of ozone and meteorological data has been audited by a member of the research team who did not compile or processed these data. At least 10% of the data have been reviewed and audited. The secondary data of meteorological observations and reanalysis has been further reviewed by comparing descriptive statistics and summary graphs generated by the project with those from the original data's website or documentation.

10% of the data used in the statistical modeling has been reviewed by a member of the

research team who did not develop the statistical model or process its input data. The statistical model outputs have been reviewed by colleagues from the University of Texas Health Science Center at Houston and Harvard University who are not involved in the project.

A member of the research team who did not conduct the GEOS-Chem simulation or develop the bias correction scheme has reviewed at least 10% of the modeling results for quality assurance purposes. A member from the GEOS-Chem model steering committee has reviewed at least 10% of the GEOS-Chem modeling results.

4. Results

4.1 Statistical Model

Using the predictors listed above, we apply a stepwise regression for the HGB-mean MDA8 ozone for each month from June to September. The regression equations for individual months are described in Table 1. Since all the predictors and the HGB ozone are normalized, the intercept β_0 equals 0 in the MLR equations. BH-Lon is the only predictor selected in the MLR equations for each month, indicating that the westward extension of BH is a key factor in influencing HGB ozone from June to September, which distinguishes BH-Lon from other meteorological predictors.

Figure 6 shows the time series of observed HGB-mean MDA8 ozone (black line) and MLR-regressed ozone (blue line) from 1999 to 2012. The squares of the correlation coefficients (R^2) for these four months are all higher than 0.55 (Table 2), which indicates that the selected predictors well capture observed interannual variability of HGB-mean MDA8 ozone. Many of the extremely high and low ozone events in June, July, and August are also captured by the MLR model. For example, HGB-mean MDA8 ozone in June 2004 is the lowest during the studied

years, and so is the regressed ozone for June 2004. However, in September, the regressed ozone shows a large inconsistency with observed ozone in the high ozone year of 2011, which indicates the potential deficiency of the MLR model to predict extremely high ozone events in September.

Table 1. The MLR model for HGB-mean MDA8 ozone concentrations in June, July, August, and September

Variables in the MLR model	June	July	August	September
Intercept (β_0)	0	0	0	0
BH-Lon	0.34	0.77	0.80	0.56
BHI1	-	-	-0.42	-
BHI2	-0.76	-	0.96	-0.50
PDSI	-	-	-1.12	-
AO	-	-	-	0.67
HGB-mean temperature	0.49	-	-	-

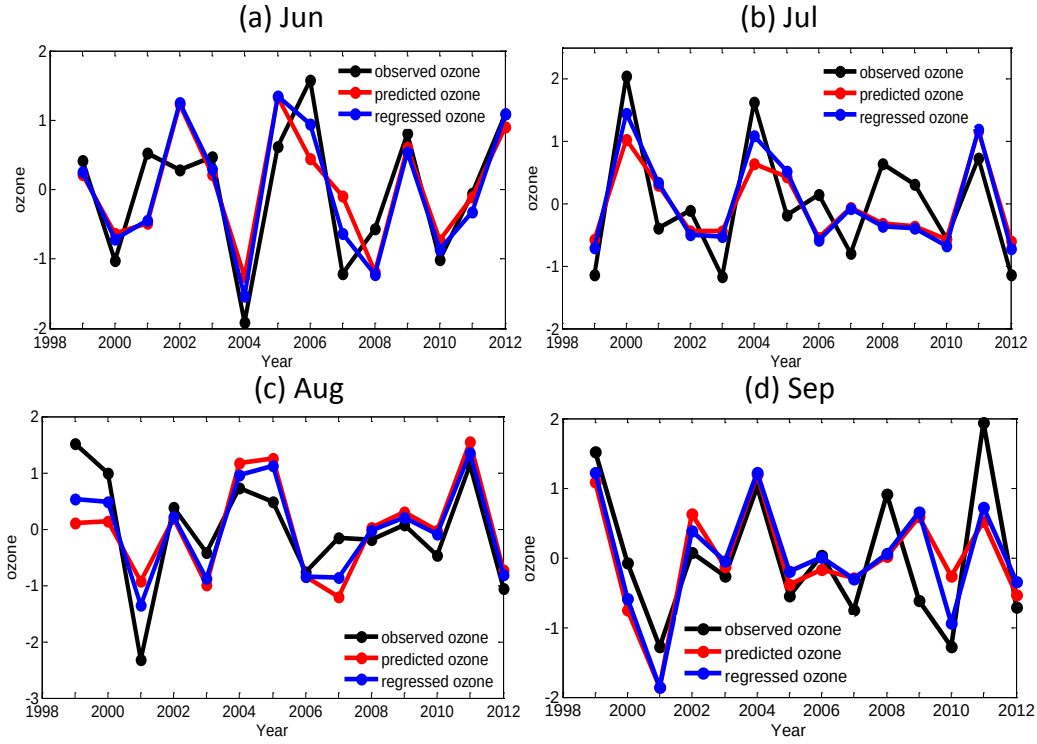

Figure 6. Time series of observed HGB-mean MDA8 ozone (black line), MLR-regressed ozone (blue line), and MLR-predicted ozone (through cross-validation) (red line) during 1999-2012.

Table 2. Squares of correlation coefficients (R^2) between observed ozone, regressed ozone, and cross-validated ozone.

	June	July	August	September
Regressed R^2	0.71	0.58	0.72	0.63
Adjusted R^2	0.62	0.55	0.59	0.52
Cross-validated R^2	0.50	0.54	0.46	0.47

According to the stepwise MLR and AIC, different predictors are selected for each month.

To test the relative importance of a single predictor for each month, we calculate the improvement of R^2 when a predictor is added in the MLR equation. Figure 7 shows the improvements of R^2 in the MLR when the predictors are added successively. In June and July, BH-Lon is the most important predictor explaining 50%-60% of the observed ozone variability, while BHI1 and BHI2 in combination play a more important role than BH-Lon in August and September. PDSI and AO are important only in August and September respectively.

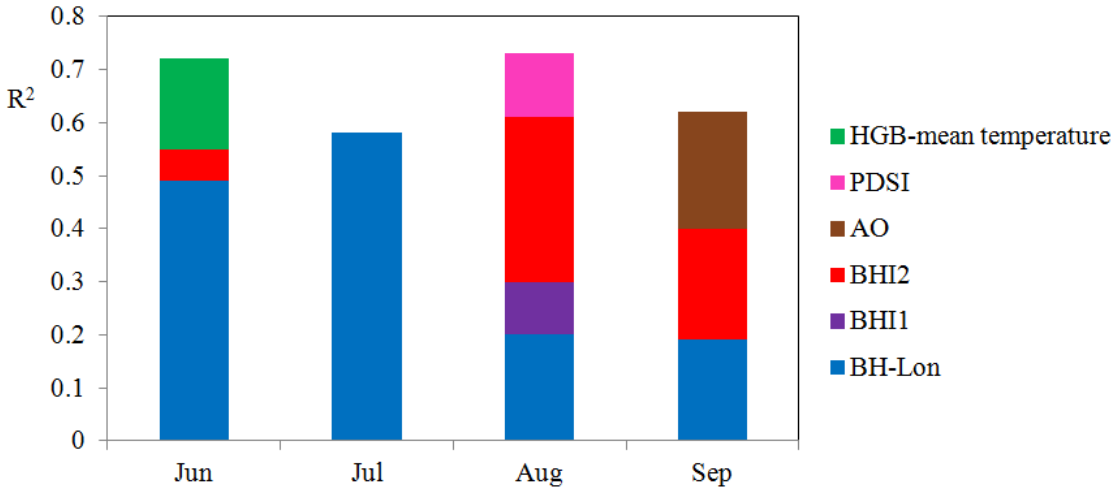

Figure 7. Improvements of R^2 in the MLR when individual predictors are added in sequence.

The MLR model developed here captures more than 50% of the interannual variance of HGB ozone from June to September, indicating the important role on ozone variability over the HGB. The definitions of BHI1 and BHI2 are both related to the regional mean SLP over the Gulf of Mexico. The BH-based indices may also have associations with the predictors that describe regional meteorological conditions (e.g. HGB-mean temperature and PDSI) on the interannual timescale. The regional meteorological predictors may be correlated with each other. Therefore, it is necessary to examine the multi-collinearity between the meteorological predictors in the MLR model for each month. To evaluate the collinearity, the variance inflation factor (VIF) is

calculated for each variable in each month. VIF is an index widely used in MLR analysis to measure how much the variance (the square of the estimate's standard deviation) of an estimated regression coefficient is increased as a result of multi-collinearity. Table 3 summarizes VIF of the predictors selected for each month. VIF for HGB-mean temperature and BHI2 in June are relatively higher (7.4 and 6.1), but still lower than 10, which is a commonly-used VIF threshold to determine collinearity (Kutner et al., 2004). For other months, VIF of all the predictors are less than 5, indicating the problem of multi-collinearity among these variables is generally unimportant.

Table 3. VIF of the predictors selected for each month

VIF	June	July	August	September
BH-Lon	2.06	1.54	3.06	1.18
BHI1	-	-	1.89	-
BHI2	6.13	-	4.88	4.76
PDSI	-	-	4.09	-
AO	-	-	-	1.47
HGB mean temperature	7.42	-	-	-

4.2 Prediction Skill of the MLR model

The MLR model developed above shows good regression performance in explaining the interannual variations of HGB-mean MDA8 ozone on the monthly scale. To evaluate the utility of the MLR model in predicting the HGB ozone, we implement a cross-validation (CV) method. First, we isolate one month at a time, perform model fitting with the remaining months, and then apply the model to predict ozone on the isolated month. The time series of CV-predicted

HGB-mean MDA8 ozone are shown as solid red lines in Figure 6 (detrended and normalized time series) and Figure 8 (actual ozone concentrations). The R^2 between observed and CV-predicted ozone is higher than 0.45 for each month (Table 2), indicating the MLR model is capable of predicting 45% or more interannual variability of monthly-mean ozone over the HGB during 1999-2012. However, some of the extreme ozone values are not very well predicted. With BH-Lon as the single predictor, the CV correlation coefficient in July is the highest among the four months, which indicates the utility of BH-Lon in predicting MDA8 ozone over the HGB peaks in the month of strongest BH influence for the region.

To further assess the prediction skill of the MLR model, it is applied to hindcast ozone for 1995-1998 and 2013, the periods outside the time series of data which were used in the model fitting. The meteorological predictors for these periods are calculated using the same method and reanalysis data as applied during the model development. The predictors are then applied in the MLR model derived above (Table 1) to hindcast the HGB-mean MDA8 ozone by month. Figure 9 compares the observed HGB-mean MDA8 ozone concentrations (black solid lines) with the hindcast values (red dashed lines) for the five-year period outside 1999-2012. Except for September 1995-1997, the MLR model correctly predicts the direction of change in ozone over the HGB for all the hindcast months during 1995-1998 and 2013. Figure 8 shows the scatter plot of observed MDA8 ozone vs the MLR hindcast value. 85% of the predicted ozone values are within ±10 ppbv of the observed values. The mean absolute error (MAE) and mean relative error (MRE) between the observed ozone and predicted ozone, as defined in Equation 2, are summarized in Table 4. The MAE and MRE are ~50% smaller for June and July when the HGB is strongly influenced by the BH than those for August and September with a weaker influence. Although only one predictor, BH-Lon, is selected in the MLR equation of July, MAE and MRE in July are the lowest among the four months, indicating the robustly strong association between BH-Lon and ozone over the HGB in July. MAE and MRE are the highest in September mainly

because the MLR equation fails to capture the extremely high ozone in 1995 and 1997, which lies above 1.2 and 2.4 times standard deviation of the 1994-2014 mean respectively. Nevertheless, high ozone concentrations observed during the other three months are well captured by the MLR model. This points to a particular deficiency of the MLR model in predicting high ozone concentrations in September.

$$MAE = \frac{1}{N}\sum_{k=1}^{N}\left|p_{simu,k} - p_{obs,k}\right|$$
$$MRE = \frac{1}{N}\sum_{k=1}^{N}\frac{\left|p_{simu,k} - p_{obs,k}\right|}{p_{obs,k}} \quad (2)$$

where $p_{simu,k}$ means simulated ozone and $p_{obs,k}$ indicates observed ozone, and N denotes the sample size.

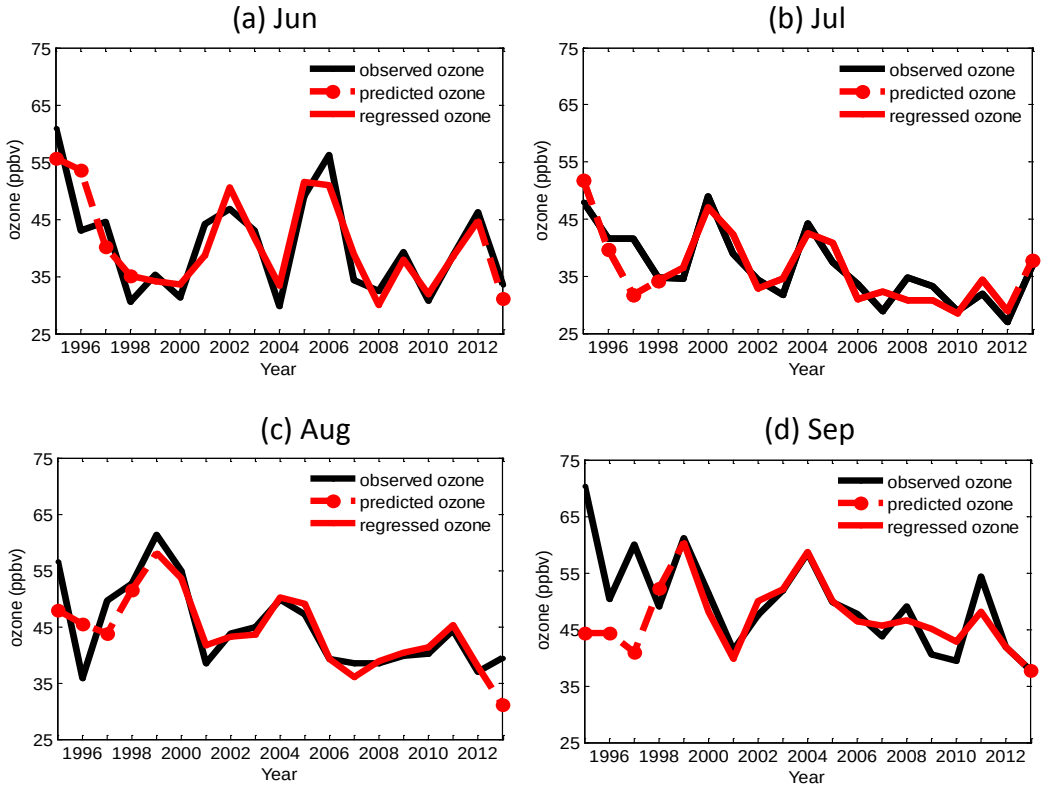

Figure 8. Time series of observed HGB-mean MDA8 ozone (black solid line), MLR-regressed ozone (red solid line), and MLR-predicted ozone (dashed line; prediction period 1995-1998 and 2013) for June, July, August and September.

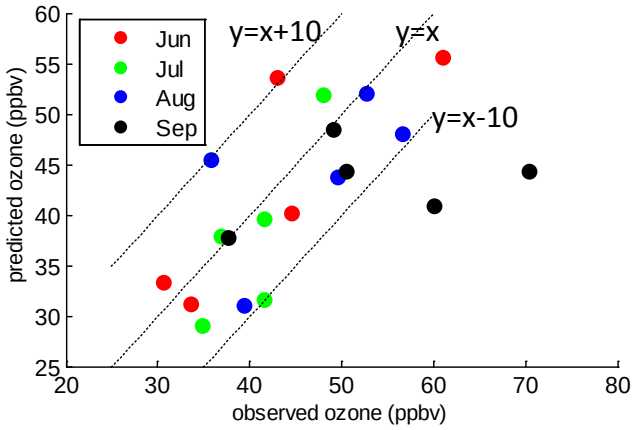

Figure 9. Scatter plot of observed HGB-mean MDA8 ozone (x-axis) vs MLR-predicted ozone during 1995-1998 and 2013 for June, July, August and September.

Table 4. Mean absolute error (MAE; ppbv) and mean relative error (MER; %) of the MLR hindcast performance. MAE and MRE are defined in Equation 2.

	June	July	August	September
MAE (ppbv)	5.10	4.46	6.57	10.34
MRE (%)	12	11	14	19

4.3 GEOS-Chem Simulation and Bias Correction Scheme

GEOS-Chem North America-nested simulations are conducted for June and July from 2004 to 2012 using the GEOS-5 assimilated meteorology and EPA National Emissions Inventory (NEI) with year-to-year changes of emissions. June and July are selected as the simulation months

because they are the months when the model has the largest positive bias in simulating surface ozone over the HGB. The model horizontal resolution is 0.5^{o} x 0.667^{o}. Figures 10 and 11 show the distribution of simulated monthly mean surface ozone over the HGB region from 2004 to 2012, overlaid with observed ozone concentrations at the CAMSs. While the model captures the observed interannual variability of surface ozone over the HGB region, the main model bias is the overestimation of surface ozone over the Galveston and Brazoria coastal region, which have lower local emissions. This overestimation is manifest in both June and July and is consistent with previous studies (Li et al., 2002; Fiore et al., 2002; Reidmiller et al., 2009; Zhang et al., 2011; McDonald-Buller, 2011). The NEI inventory used in the simulation may not correctly reflect the timing and effects of local emissions controls over the HGB or unusual emissions over this region, which partly explains the model bias.

For simplicity, mean ozone concentrations at four coastal rural sites (2 in Brazoria county: 480391004-Manvel Croix Park and 480391016-Lake Jackson; 2 in Galveston county: 481670014-Galveston airport and 481671034-Galveston 99^{th} street) are adopted to present the observed coastal ozone. Simulated coastal ozone is calculated as the mean value over the model grids containing these four sites, i.e., the black box shown in Figure 10 and 11. We note that the effects of local emissions from the coastal region are not considered in the simplified approach of diagnosing coastal ozone from both model and observations. The model captures very well the interannual variations of coastal ozone in both June and July (Figure 12). The mean bias between the simulated and observed surface ozone over the coastal region is 5.17 ppbv for June and 9.54 ppbv for July. The ozone bias is higher in July than in June partly because the bias is associated with overestimating maritime ozone inflow, which peaks in July. Since both observed and simulated coastal ozone over the HGB region can be explained by the interannual variations of BH-Lon, the bias between observed and simulated ozone is expected to be correlated with BH-Lon. We hence used BH-Lon to predict this bias.

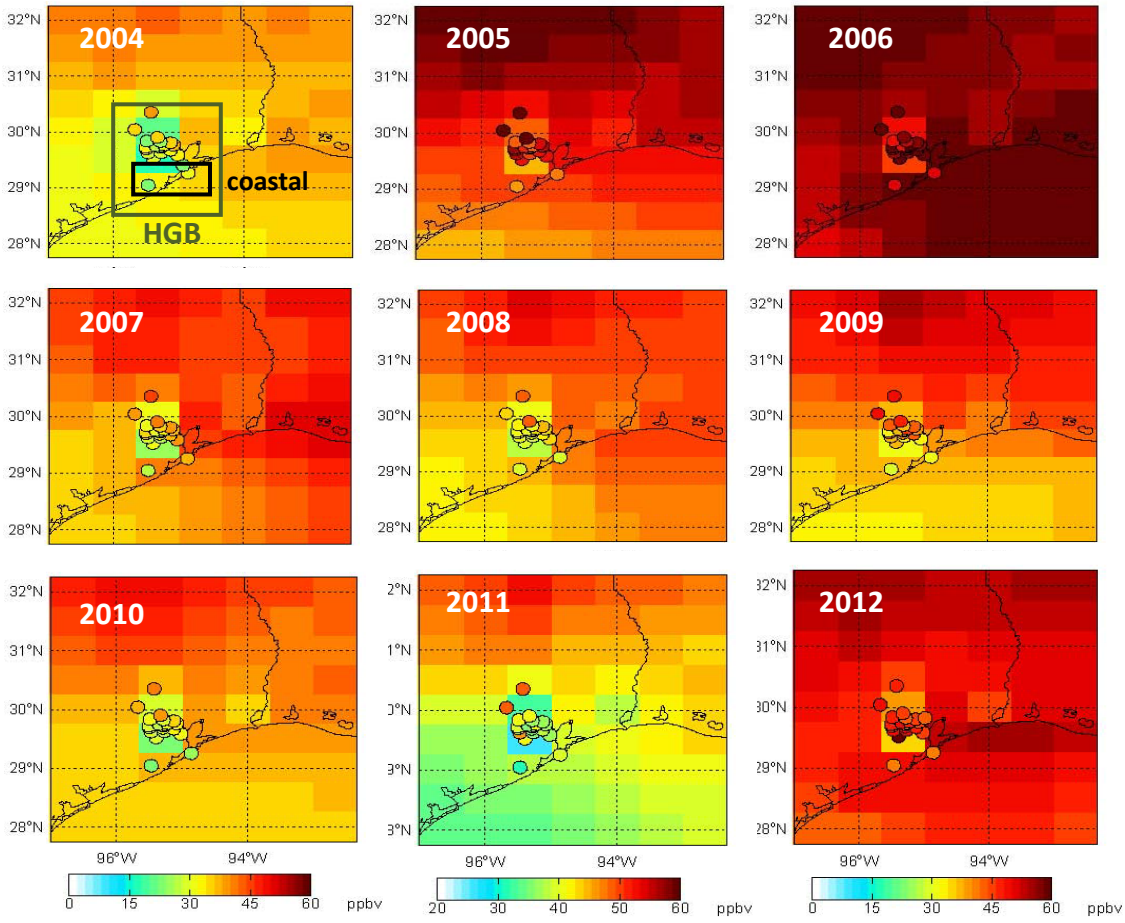

Figure 10. Observed surface ozone (filled circles) and GEOS-Chem simulated surface ozone over the HGB region in June.

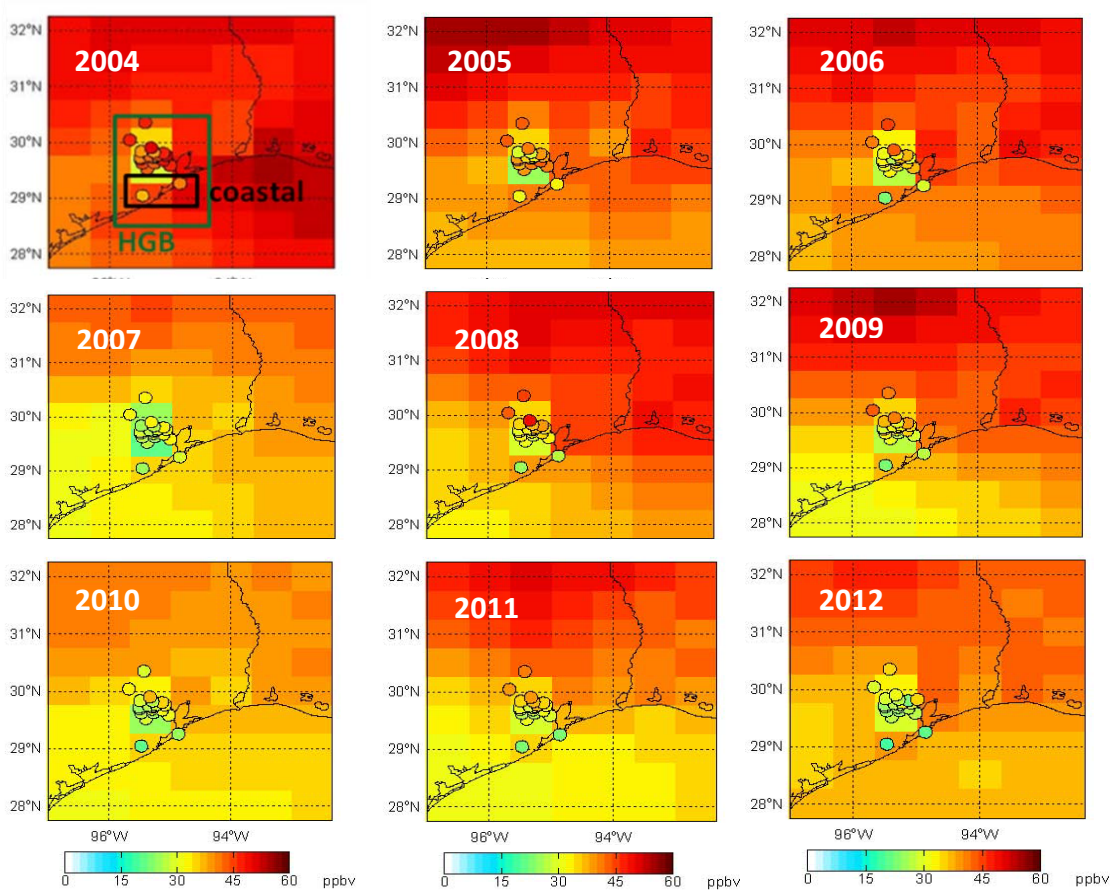

Figure 11. Observed surface ozone (filled circles) and GEOS-Chem simulated surface ozone over the HGB region in July.

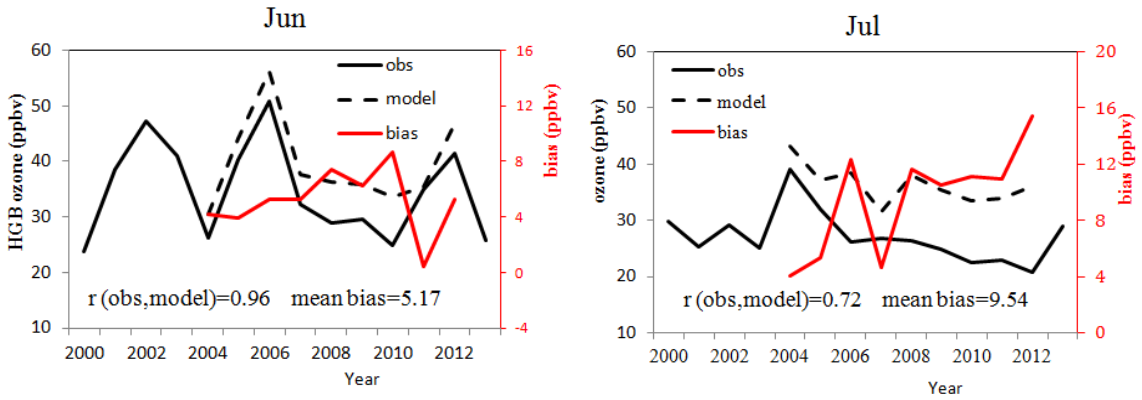

Figure 12. Time series of observed and simulated mean ozone over the coastal sites and

their differences (simulated minus observed) for June (left) and July (right).

The MLR equations for the predicted model bias (y) for June and July are as follows,

$$y_{Jun} = -0.13x + 5.23$$
$$y_{Jul} = -0.34x + 9.59 \quad (3)$$

where x indicates the detrended BH-Lon. The negative coefficients in front of the x term (BH-Lon) in June and July both indicate a higher model bias when BH-Lon locates more westward. It in turn testifies that the higher bias in July is due to the stronger maritime inflow that accompanies the westward extension of the BH from June to July. The interannual variations of the model bias is different between June and July, which supports our motivating hypothesis that different MLR equations are needed for individual months in order to capture the variance of the bias, rather than a fixed bias correction.

The time series of the model bias and MLR-predicted bias using Equation (3) are shown in Figure 13. The predicted bias basically captures the variation of the model bias. However, the discrepancy is large for 2007, 2011 and 2012 when the model bias is extremely high or low.

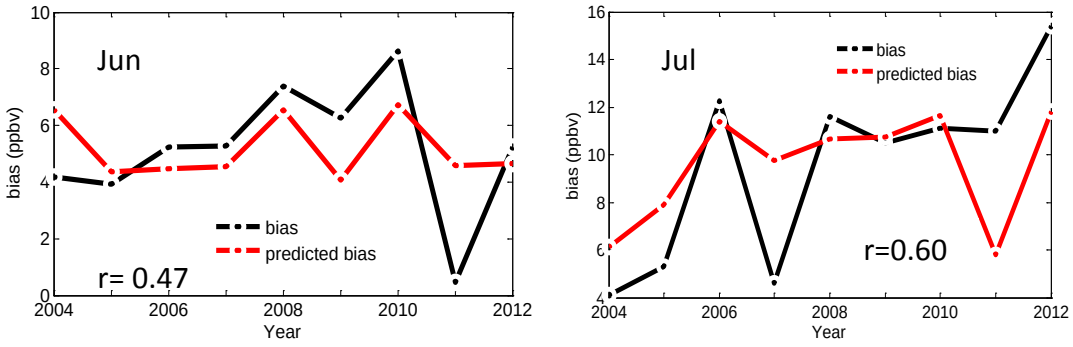

Figure 13. Time series of model bias (black line) and predicted bias (red line) for June and July.

To correct the simulation results, we then subtract the predicted bias from simulated

coastal ozone. The time series of corrected simulation is shown in Figure 14 (red line). Comparing the simulation results before correction (blue line) and after correction (red line), the correlation coefficient between observation and simulation increases from 0.72 to 0.88, and mean bias decreases from 9.54 ppbv to 2.36 ppbv in July. The correlation coefficient between the observed ozone and simulated ozone after correction is higher in June, and the mean bias in June is lower than that in July. This is probably because the stronger maritime inflow in July brings more clean background air, which in turn results in a model bias due to the overestimation of background ozone over the Gulf of Mexico.

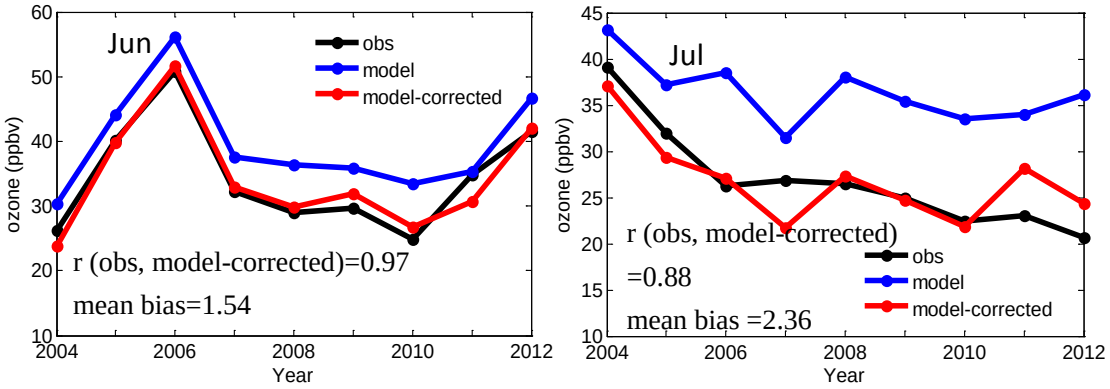

Figure 14. Time series of observation (black line), simulation results (blue line) and corrected simulation results (red line) for June and July.

5. Discussion

The MLR model developed in this project captures 58% - 72% of the interannual variance of the HGB-mean MDA8 ozone for the months of June, July, August, and September, indicating the significant role of large-scale meteorology on ozone variability over this region. To the best of our knowledge, the correlation coefficients reported in this project are significantly higher than those from previously published studies on the regression relationship between interannual variability of meteorological factors and observed ozone over the HGB region or for

the southern U.S. at large. Zhu and Liang (2013) reported a negative correlation, r, ranging between -0.5 and -0.7 between the BHI and summer-mean MDA8 ozone during 1993-2010. Shen et al. (2015) identified the polar jet, the Great Plains low level jet (GPLLJ), and the BH as major synoptic-scale patterns influencing surface O_3 variability in the eastern US in summer. They reported that the combination of those meteorological indices explains 53% of the interannual variance of summer-mean MDA8 ozone in South Central US during 1980-2010. Compared with those previous studies which averaged surface ozone over a large geographical region and onto a seasonal mean and thus smoothed out some variability, the present project tackles a more challenging question of explaining the interannual variability of monthly ozone over a smaller region (HGB) during a shorter, more recent time period (1998-2013). Yet, our MLR results reveal higher regression correlation coefficients than those from the previous studies, and the MLR model developed here also show a good prediction skill with the CV R^2 higher than 0.45 for each of the months from June to September.

In addition to the HGB mean MDA8 ozone, we also investigated the relationship between the selected meteorological predictors and other ozone metrics to assess the robustness of the relationships; the other metrics include median total ozone, mean and median background ozone, and mean and median ozone enhancement over the HGB region. The background ozone data were provided by TCEQ. Median total ozone over the HGB is calculated as the mean value of median monthly ozone of all the sites. Mean/median ozone enhancement is calculated as the difference between mean/median total ozone and mean/median background ozone. The median ozone is relatively lower than the mean, since the median value is less sensitive to extremely high ozone events (Figure 15). Monthly mean background ozone shows very similar interannual variations with monthly mean total ozone from June to September, and their correlations are highest in June (r=0.97). Table 5 summarizes the correlation coefficients between different ozone metrics and BH-Lon. Note that in calculating the correlations total

ozone and background ozone is detrended, while ozone enhancement is not detrended. In June and July, when the correlations between total ozone and BH-Lon are stronger, there are also significant correlations between background ozone and BH-Lon. In August and September, however, there are no significantly positive correlations between BH-Lon with either total ozone or background ozone at $p < 0.01$. Thus meteorological indices other than the BH-Lon have been developed for the MLR model for these months. Since the correlation coefficients between the BH-Lon and mean total ozone are higher than other metrics of ozone (Table 5), we used mean total ozone as the dependent variable y in the MLR.

The stepwise MLR show that BH-Lon is an important predictor for every month. To quantify the influence of BH-Lon on the HGB surface ozone, we calculated that as BH-Lon extend westward, mean MDA8 surface ozone in the HGB decreases at a rate of 0.73, 0.44, 0.16 and 0.23 ppbv deg^{-1} in June, July, August, and September, respectively.

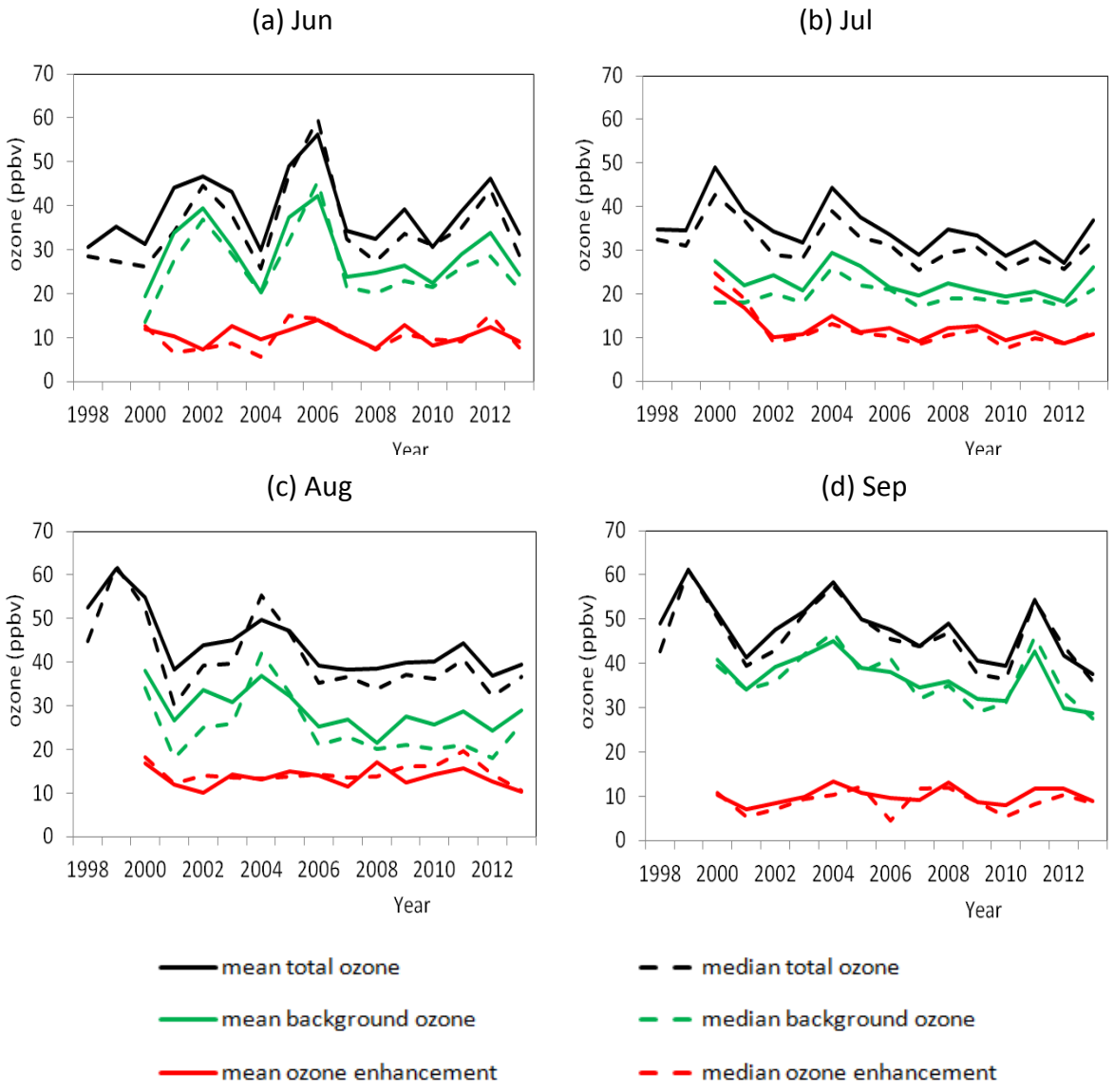

Figure 15. Time series of mean and median total ozone, mean and median background ozone and ozone enhancement.

Table 5. Correlation coefficients (R^2) between BH-Lon and different metrics of ozone (Total ozone and background ozone are detrended). Red numbers indicate significant correlations ($p<0.01$).

R^2	Jun	Jul	Aug	Sep
total ozone (mean)	0.49	0.58	0.20	0.20
total ozone (median)	0.35	0.55	0.15	0.21
background ozone (mean)	0.30	0.28	0.00	0.10
background ozone (median)	0.19	0.22	0.06	0.06
Ozone enhancement (mean)	0.29	0.44	0.40	0.12
Ozone enhancement (median)	0.23	-0.11	0.01	0.00

In the analysis presented above, raw time series of MDA8 ozone is detrended by subtracting the 3-year moving average. To test if the analysis depends on the way of detrending, we used another approach by subtracting a best-fit linear trend line from the raw time series. Stepwise regression selected the same predictors for June, July and September when the HGB ozone is detrended in two different ways. However, when the HGB ozone is detrended in the second way, BHI1 is not selected as a predictor in the MLR for August. Similar correlation coefficients are found between the regressed ozone and the observed ozone if we use the second way to detrend (Table 6). However, the second detrending method gives a lower CV R^2 for August and September.

One shortcoming of our analysis is that we examined the HGB region ozone as a whole and as such the spatial differences between the sites in the HGB region are not represented in the MLR model developed here. The meteorological predictors chosen in the MLR model are large scale in nature, such as the BH, drought, and AO. However, the spatially inhomogeneous emissions can result in large surface ozone gradients within the region. Owing to the spatial resolution of the meteorological predictors, we considered all sites in the HGB region as a whole and simply took the average of all CAMS observations as the dependent variable y in the MLR equation. Thus, the site-specific relations between ozone and meteorology were inevitably

left out.

Table 6. Correlation coefficients (R^2) between observed ozone, regressed ozone and cross-validated ozone when the ozone data are detrended by substracting a linear trend line.

	June	July	August	September
Regressed R^2	0.77	0.66	0.56	0.59
Adjusted R^2	0.72	0.63	0.45	0.49
Cross-validated R^2	0.70	0.58	0.31	0.40

6. Summary

The more than decade-long observational record of ozone and meteorology (1998 - 2013) during the months from June to September are analyzed to characterize the complex effects of the BH on surface MDA8 ozone variations in HGB. Statistical relationships are developed and tested through multiple linear regression (MLR). The indicators of the BH location and strength developed/refined in this project are the longitude index of the BH western edge (BH-Lon), and two BH intensity related indices (BHI1 and BHI2). We find that the BH-Lon alone explains 50-60% ($r = 0.7\sim 0.8$) of the year-to-year variability in June and July monthly mean ozone over HGB during 1998-2013. Such a high correlation is explained by the mechanism that the western extension of the BH determines the strength of the southerly low-level jet (LLJ) that brings marine air with lower ozone background from the Gulf of Mexico to the HGB. In August and September when the BH weakens, the correlation between BH-Lon and ozone decreases to a less significant value of ~0.2 and stepwise regression identifies the variability in BH strength (BHI1 and BHI2), PDSI, and AO as additional predictors to explain the interannual ozone

variability over the HGB for these months. The MLR model developed in this project is able to capture 58% - 72% of the interannual variance of the HGB-mean MDA8 ozone from June to September, indicating the significant role of large-scale meteorology on ozone variability for this region. Among all the predictors, the BH-Lon is the most important for each month. The MLR model developed here also show a good prediction skill with the CV R^2 higher than 0.45. To the best of our knowledge, the correlation coefficients reported in this project are significantly higher than those from previously published studies in terms of the regression relationship between interannual variations of meteorological factors and ozone over the HGB region or for the southern U.S. at large.

The statistical relationship is then applied to develop an empirical bias correction scheme to mitigate the problem of the GEOS-Chem global model in overestimating surface ozone concentrations along the Gulf coastal region in the summer. A set of multiple-year GEOS-Chem simulations is conducted. A moderate to strong correlation is identified between the BH-Lon and GEOS-Chem model bias for Jun and July, which supports the hypothesis that the model bias is caused in part by the insufficient representation of the dynamic linkage between BH and maritime ozone inflow to HGB. After the correction, the mean model bias in June and July shows a 70-75% decrease and the correlation coefficient between the observed and simulated ozone also improves.

7. Recommendation for Future Work

More in-depth analysis is needed in the future to further advance the mechanistic and quantitative understanding of the factors that drive the ozone variability over the HGB. To build upon the monthly scale analysis of the present project, the first direction of future investigation is toward the daily scale analysis. After all, ozone exceedance is counted on the daily basis, and there are large day to day changes in air pollution driving primarily by meteorology. The

monthly mean approach of the present project, while advancing the seasonal approach of many previous studies, has inevitably left out extreme events of the HGB ozone which matter most to public health and air quality management. We expect to find abnormal or extreme features of the BH and other meteorological conditions on the daily scale that may or may not contribute to the ozone exceedance. As the monthly mean analysis smoothes out the daily variability, the daily scale analysis in the future will likely reveal other important mechanisms for the influence of BH and other meteorological conditions on the HGB ozone. The second future direction is to explore finer spatial variations of ozone over the HGB. The present project averages all CAMS sites to present the mean ozone variability over the HGB region. Site specific analysis will be necessary to identify which sites are most influenced by large-scale meteorology and which ones are not. For the latter sites, the important question is what other factors are more important: local emissions, local meteorology, or regional transport. It will be helpful for air quality managers to have such knowledge to design attainment strategies. Finally, all the observation-derived understanding will provide valuable constraints for air quality models to identify their strength and deficiency in simulating the drivers of ozone variability over the HGB. If the models are proved capable of simulating those drivers, they can be employed with confidence to separately quantify the relative contribution of emission controls vs other "natural" factors on the surface ozone trend over the HGB in the past decade.

8. Acknowledgement

The preparation of this report was financed through a grant from the Texas Commission on Environmental Quality (TCEQ), administered by The University of Texas through the Air Quality Research Program. The contents, findings, opinions, and conclusions are the work of the author(s) and do not necessarily represent the findings, opinions, or conclusions of the TCEQ. We thank Mark Estes from TCEQ and Lu Shen from Harvard for their suggestions on the project.

9. References

Berlin, S.R., A.O. Langford, M. Estes, M. Dong, D.D. Parrish: Magnitude, decadal changes, and impact of regional background ozone transported into the greater Houston, Texas area, Environ. Sci. Technol., 47(24), 13985-13992, 2013.

Berman, J. D., N. Fann, J. W. Hollingsworth, K. E. Pinkerton, W. N. Rom, A.M. Szema, P. N. Breysse, R. H. White, F. C. Curriero: Health benefits from large-scale ozone reduction in the United States, Environ. Health Perspect., 120(10), 1404-1410, 2012.

Davis, R. E., B. P. Hayden , D. A. Gay , W. L. Phillips, G. V. Jones: The North Atlantic subtropical anticyclone, J. Climate, 10, 728-744, 1997.

Fiore, A.M., D.J. Jacob, I. Bey, R.M. Yantosca, B.D. Field, A.C. Fusco: Background ozone over the United States in summer: origin, trend, and contribution to pollution episodes, J. Geophys. Res., 107 (D15), 2002.

Haman, C. L., E. Couzo, J. H. Flynn, W. Vizuete, B. Heffron, and B. L. Lefer: Relationship between boundary layer heights and growth rates with ground-level ozone in Houston, Texas, J. Geophys. Res., 119 (10), 6230-6245, 2014.

Hegarty, J., H. Mao, R. Talbot: Synoptic controls on summertime surface ozone in the northeastern United States, J. Geophys. Res., 112, D14306, 2007.

Higgins, R. W., Y. Yao, E. S. Yarosh, J. E. Janowiak, K. C. Mo: Influence of the Great Plains low-level jet on summertime precipitation and moisture transport over the central United States, J. Climate, 10, 481-507, 1997.

Hogrefe,C., J. Biswas, B. Lynn, K. Civerolo, J.Y. Ku, J. Rosenthal, C. Rosenzweig, R. Goldberg, P.L. Kinney: Simulating regional-scale ozone climatologyover the eastern United States: model evaluation results, Atmos. Environ., 38, 2627-2638, 2004.

Jacob, D. J., and Winner, D. A.: Effect of climate change on air quality, Atmos. Environ., 43(1), 51-63, 2009.

Kalnay, E., et al.: The NMC/NCAR CDAS/Reanalysis Project, Bull. Am. Meteorol. Soc.,77, 437-471, 1996.

Kutner, M.H., C.J. Nachtsheim, J. Neter, W. Li: Applied Linear Statistical Models. McGraw-Hill/Irwin, New York, NY, USA, 2004.

Li, W., L. Li, R. Fu, Y. Deng, and H. Wang: Changes to the North Atlantic subtropical high and its role in the intensification of summer rainfall variability in the southeastern United States, J. Climate, 24, 1499-1506, 2011.

Li, L., W. Li, Y. Kushnir: Variation of North Atlantic Subtropical High western ridge and its implication to the Southeastern US summer precipitation, Clim. Dyn., 39:1401-1412, 2012.

Li, Q., D. J. Jacob, T. D. Fairlie, H. Liu, R. M. Yantosca, and R. V. Martin: Stratospheric versus pollution influences on ozone at Bermuda: Reconciling past analyses, J. Geophys. Res., 107, 2002.

Lin, M., A. M. Fiore, L. W. Horowitz, O. R. Cooper, V. Naik, J. Holloway, B. J. Johnson, A. M. Middlebrook, S. J. Oltmans, I. B. Pollack,T. B. Ryerson, J. X. Warner, C. Wiedinmyer, J. Wilson, B. Wyman: Transport of Asian ozone pollution into surface air over the western United States in spring, J. Geophys. Res., 117, 2012.

Lin, M., A. M. Fiore, L. W. Horowitz, A. O. Langford, S. J. Oltmans, D. Tarasick, H. E. Rieder: Climate variability modulates western US ozone air quality in spring via deep stratospheric intrusions, Nature Communications, 2015.

McDonald-Buller, E. C., D. T.Allen, N. Brown, D.J. Jacob, D. Jaffe, C.E. Kolb, A.S. Lefohn, S. Oltmans, D.D. Parrish, G. Yarwood, L. Zhang : Establishing Policy Relevant Background (PRB) Ozone Concentrations in the United States, Environ. Sci. Technol., 45, 9484- 9497, 2011.

Gammon, N., J. Tobin, A. McNeel, and G. Li: A conceptual model for eight-hour ozone exceedances in Houston, Texas, Part I: Background ozone levels in eastern Texas, 52 pp.,

Houston Adv. Res. Cent., Houston, Tex, 2005.

Ngan. F., D. Byun: Classification of Weather Patterns and Associated Trajectories of High-Ozone Episodes in the Houston- Galveston-Brazoria Area during the 2005/06 TexAQS-II, J. Appl. Meteor. Climatol., 50,485-499, 2011.

Ortegren, J. T., P. A. Knapp, J. T. Maxwell, W. P. Tyminski, and P. T. Soule: Ocean–atmosphere influences on low frequency warm-season drought variability in the Gulf Coast and southeastern United States, J. Appl. Meteor. Climatol., 50, 1177-1186, 2011.

Rappenglück, B., R. Perna, S. Zhong, and G. A. Morris: An analysis of the vertical structure of the atmosphere and the upper-level meteorology and their impact on surface ozone levels in Houston, Texas, J. Geophys. Res., 113, D17315, 2008.

Reidmiller, D. R., A. M. Fiore, D.A. Jaffe, D. Bergmann, C. Cuvelier, F.J. Dentener, et al.: The influence of foreign vs. North American emissions on surface ozone in the US, Atmos. Chem. Phys., 9, 5027-5042, 2009.

Shen, L., L. J. Mickley, A. P. K. Tai: Influence of Synoptic Patterns on Surface Ozone Variability over the Eastern United States from 1980 to 2012, Atmos. Chem. Phys. Discuss., 15, 13073-13108, 2015.

Pakalapati, S., S. Beaver , J. A. Romagnoli , A. Palazoglu: Sequencing diurnal air flow patterns for ozone exposure assessment around Houston, Texas, Atmos. Environ., 43, 715-723, 2009.

Stahle, D. W., M. K. Cleaveland: Reconstruction and analysis of spring rainfall over the southeastern U.S. for the past 1000 years, Bull. Amer. Meteor. Soc., 73, 1947-1961, 1992.

Tai, A.P.K., L.J. Mickley, D.J. Jacob: Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change, Atmos. Environ., 44, 3976-3984, 2010.

Venables, W.N., B.D. Ripley: Modern Applied Statistics with S. Springer, New York, NY, USA, 2003.

Wang, X., D. L. Mauzerall: Characterizing distributions of surface ozone and its impact on grain production in China, Japan and South Korea: 1990 and 2020, Atmos. Environ., 38, 4383-4402, 2004.

Zhang, L., D.J. Jacob, N.V. Smith-Downey, D.A. Wood, D. Blewitt, C.C. Carouge, A. van Donkelaar, D.B.A. Jones, L.T. Murray, Y. Wang: Improved estimate of the policy-relevant background ozone in the United States using the GEOS-Chem global model with $1/2^{o}$x$2/3^{o}$ horizontal resolution over North America, Atmos. Environ., 45(37), 6769-6776, 2011.

Zhu, J., X. Liang: Impacts of the Bermuda High on Regional Climate and Ozone over the United States, J. Climate, 26, 1018-1032, 2013.