Dynamic Model Performance Evaluation Using Weekday-Weekend and Retrospective Analyses

Jim Smith* and Mark Estes
Texas Commission on Environmental Quality, Austin, Texas, USA

1. INTRODUCTION

Model performance evaluation (MPE) is commonly accomplished by comparing modeled concentrations for a base period with concentrations measured during the same period. While this is a necessary part of MPE, it does not address a fundamental aspect of the modeling: its ability to predict concentration changes as a function of changes in model inputs. Although critical to assessing the future state of an airshed, model responsiveness is evaluated much less frequently than base-case performance because it is often difficult and resource-intensive to accomplish.

In this paper the authors describe two approaches to dynamic MPE used by the Texas Commission on Environmental Quality (TCEQ) in a recent State Implementation Plan (SIP) revision for the Houston/Galveston/Brazoria (HGB) ozone nonattainment area [TCEQ, 2010a]: Weekday/Weekend Analysis and Retrospective Analysis. In the former, the model's response to changes in emissions between weekdays and weekends (primarily traffic) are compared with observed responses, and in the latter the model is used to predict ozone design values (DVs) in a prior year, then the modeled year-to-year design value changes are compared with the observed design value changes. The results suggest that the photochemical model is not as responsive to emission changes as is the real airshed.

2. MODELING FOR THE HGB SIP

The HGB area historically exhibited some of the highest ozone levels in the nation, but in recent years has seen a remarkable improvement in air quality - in 2009 the ozone DV for every regulatory monitor in the area met the 1997 ozone National Ambient Air Quality Standard (NAAQS) of 0.08 parts per million. Modeling has played a key role in developing a number of SIP revisions over the past twenty years which have led to this favorable outcome. In 2010 the TCEQ submitted an attainment demonstration (AD) SIP revision to the United States Environmental Protection Agency (EPA) demonstrating that the area will achieve attainment of the 1997 NAAQS by the area's attainment date of 2018.

The modeling was conducted using the Comprehensive Air quality Model with Extensions (CAMx) with a 36-12-4-2 km nested grid focusing on the HGB area, shown in Figure 1.

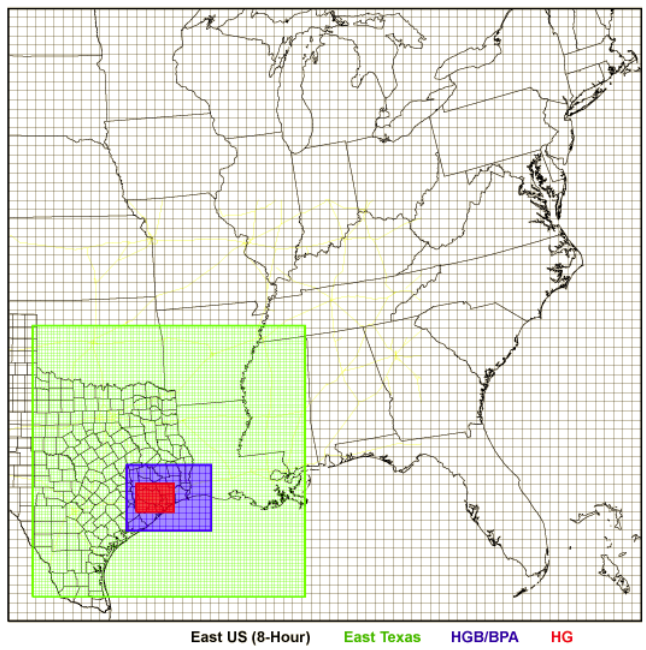

Figure 1: HGB nested grid showing 36 km (black), 12 km (green), 4 km (blue), and 2 km (red) nested grids

2.1 The 2006 Base Case MPE

The base case for the 2010 HGB AD SIP consisted of 96 days from several ozone episodes in 2005 and 2006. This period was chosen largely to coincide with the Second Texas Air Quality Study (TexAQS II) [Parrish 2009], which provided a rich database of aerometric observations from a variety of platforms including four aircraft, a variety of sondes, a research vessel, and several special-purpose monitoring sites. The extensive base

**Corresponding author:* Jim Smith, Texas Commission on Environmental Quality, 12100 Park 35 Circle, MC-164, Austin, TX 78711-3087; e-mail: jismith@tceq.state.tx.us

case MPE not only showed that the model replicated observed ozone concentrations reasonably well, but also provided an opportunity to evaluate the model's inner workings against observations of rarely-measured intermediate species in three dimensions.

While the base-case MPE provided an incredible wealth of information about the model and its relation with the airshed, it still was a *static* evaluation in that it did not directly examine the model's response to modifications to its inputs. But the real reason for performing regulatory modeling is to help understand how the environment will respond to changes in its inputs, specifically emissions that grow over time or are modified through established or proposed controls, hence the need for *dynamic* MPE.

2.2 Attainment Modeling

Upon completion of the standard (static) MPE, a baseline inventory was created from the base case 2006 emissions by replacing hour-specific electricity-generating unit emissions from the Acid Rain Program Data Base, along with hour- specific emissions of certain Highly Reactive Volatile Organic Compounds (HRVOCs) - collected between August 15 and September 15, 2006 in a special inventory [TCEQ, 2010b] – with average profiles, as recommended in the EPA guidance [EPA, 1997]. Replacing these emissions with average values accomplishes two purposes: first, it assures that all units are represented fairly in the analysis, even those that may have temporarily suspended operations during the base case period, and second it greatly expedites conducting modeling analyses involving modifying point source emissions; otherwise a separate point source emissions file would need to be prepared for each day of the simulation. Other emissions variability, such as day-of-week variation in mobile source emissions and meteorologically-driven variability in biogenic emissions were left unchanged. The base case modeling was repeated, except that the base case emissions were replaced with the baseline emissions.

Again following EPA guidance, a future base inventory was constructed by projecting anthropogenic emissions to 2018 using appropriate growth factors and all applicable controls. The future base emissions were then modeled, and relative response factors (RRFs) were calculated at each regulatory monitor in the HGB area to estimate the monitor's change in daily peak 8-hour ozone concentration resulting from future growth and controls. A baseline design values (DV_b) was next calculated for each site by averaging the site's 2006, 2007, and 2008 design values, and its future predicted DV (DV_p) was calculated by multiplying the RRFs by the corresponding DV_b.(we use the notation 'predicted DV' or 'DV_p' instead of the usual 'future DV' or 'DV_f' for reasons which will become obvious). The actual attainment test consisted of comparing the DV_p for each site with the 1997 ozone NAAQS.

A map of the monitoring sites referred to in this paper is provided in Figure 2.

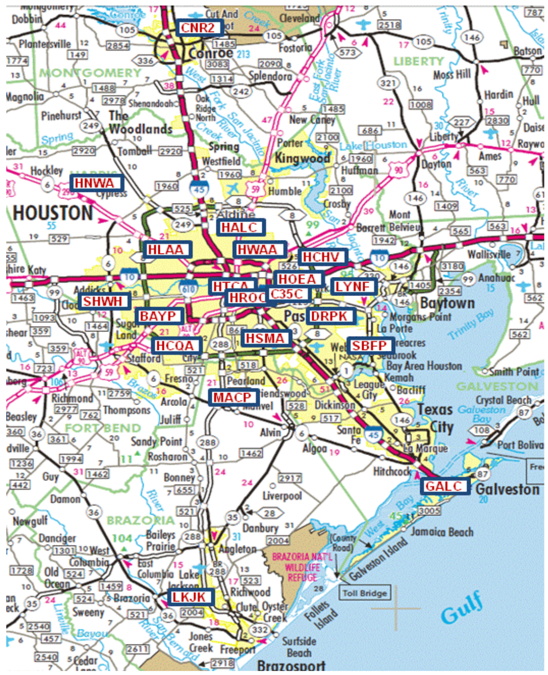

Figure 2: Subset of monitoring locations in the HGB area referred to in this paper

3. RETROSPECTIVE MODELING

The principle behind retrospective analysis is simple: use the model to predict prior year ozone concentrations, then compare the model predictions with what was actually observed in that year. In practice, retrospective modeling is not often performed because a significant effort is required to develop a modeling inventory for a particular year, and most modelers focus their efforts on the base and future years instead of past years. In the case of HGB, however, a baseline modeling inventory for the year 2000 had already been developed on the same grid as that used in the 2010 HGB ozone AD for an earlier AD submitted in 2004[1] [TCEQ, 2004]. The

[1] The 36 km grid used in the 2010 AD was larger than that used for the earlier AD, so the 2006 emissions

retrospective analysis was conducted by modeling the 2000 baseline inventory just as if it were a future inventory, calculating (retrospective) RRFs, and calculating predicted 2000 DV_Ps for each regulatory monitor.

3.1 Results

In the standard attainment test, the DV_b is not really a design value, but is the average of three design values each containing the baseline year. Therefore, the DV_p is not specifically a predictor of a single future year's DV but rather of a three-year average. Consequently, we calculated a 2000 DV_b for each regulatory monitor by averaging its 2000, 2001, and 2002 DVs and compared these to the 2000 DV_ps.

Table 1 shows the regulatory monitors for which both 2000 and 2006 DV_bs could be calculated. The second column of the table shows each monitor's 2006 DV_b, followed by the observation-based 2000 DV_b. The fourth column is just the ratio of the 2000 DV_b to the 2006 DV_b, which represents the actual RRF. The fifth column shows the modeled RRFs, and the final column of Table 1 shows the 2000 DV_p, which is the product of columns two and five. In this case, since we are predicting a value from a prior year, both the actual and modeled the RRFs are all greater than one.

3.2 Analysis

Comparing the 2000 DV_ps with the 2000 DV_bs shows that overall, the model did a good job of predicting ozone DVs in 2000. The two monitors with the highest 2000 DV_bs, Bayland Park (BAYP) and Deer Park (DRPK), were both predicted within 1 ppb. One monitor, Monroe (HSMA), was over-predicted by 8.5 ppb, and Westhollow (SHWH) was over-predicted by 2.6 ppb. The remaining monitors were all under-predicted by between 3.4 and 10.8 ppb.

To better quantify the model's responsiveness to the 2000-to-2006 emission changes, we inverted the average modeled and actual RRFs to examine them in a prospective context. We see that the predicted 2000-to-2006 average RRF is 0.877, while the actual RRF was 0.843, meaning that the actual improvements in ozone air quality were overall better than those predicted by the model.

Table 1: Comparison of observed and modeled RRFs and observed and predicted 2000 DV_bs

Monitor Code	2006 DV_b	2000 DV_b	2006-to-2000 RRF		2000 DV_p
			Actual	Model	
BAYP	96.7	107.0	1.11	1.11	107.0
C35C	79.0	97.0	1.23	1.18	93.5
DRPK	92.0	107.7	1.17	1.18	108.1
GALC	81.7	98.3	1.20	1.11	90.7
HALC	85.0	108.7	1.28	1.15	97.9
HCQA	87.0	105.3	1.21	1.13	98.6
HLAA	77.7	90.0	1.16	1.11	86.4
HNWA	89.0	104.7	1.18	1.13	100.4
HOEA	80.3	102.0	1.27	1.17	94.0
HROC	79.7	95.0	1.19	1.15	91.6
HSMA	90.3	96.3	1.07	1.16	104.8
HNWA	76.3	97.3	1.28	1.14	86.9
SHWH	92.3	100.3	1.09	1.11	102.9
Avg.	85.2	100.7	1.19	1.14	97.1

4. WEEKDAY/WEEKEND ANALYSIS

Weekday/weekend analysis relies on the intrinsic differences between weekday and weekend traffic patterns, which provide a kind of natural laboratory where the airshed can be examined in two (or more) different states. It has often been used as an observation-based modeling technique to indicate whether an airshed is NO_X- or VOC- limited (or in-between) [Blanchard, 2005]. Some areas have seen increases in weekend peak ozone concentrations, indicating that the higher weekday concentrations of NO_X are inhibiting ozone production (at least locally), and that peak ozone in these areas may be VOC-sensitive. Conversely, ozone concentrations that decrease on the weekends would suggest NO_X-sensitive ozone formation.

Weekday-weekend analysis is less often performed for dynamic MPE, largely because the technique requires a fairly large number of modeled days to produce robust conclusions. In this section, we use the relatively large number of modeled days (87) from the 2010 AD to perform one analysis, then use a series of sensitivity runs to expand the number of days used in an alternative analysis [also see Smith, 2007].

4.1 Weekday/Weekend Emission Patterns

For weekday/weekend analysis to be meaningful there must be significantly different emission patterns between weekdays and weekends. Figure 3 shows 2005 modeled 6 AM Wednesday, Saturday, and Sunday NO_X anthropogenic emissions by category summed

from some distant areas (more than 1000 kilometers from Houston) were not replaced.

over the 8-county HGB nonattainment area. On-road mobile source emissions drop by almost two-thirds from Wednesday to Saturday, and by another 50% from Saturday to Sunday – this is why we chose to analyze Saturday and Sunday separately. Non-road mobile sources, which include construction equipment, also show declines on the weekend. Modeled on-road VOC emissions decline similarly, but overall VOC emissions show much more modest day-of-week variation, due to increases in non-road VOC emissions from gasoline-powered recreational vehicles and boats. The large area source VOC component includes about 9 tons/hour from oil and gas production activities every day, but other sources such as solvent usage cause this category to decrease on the weekends as well.

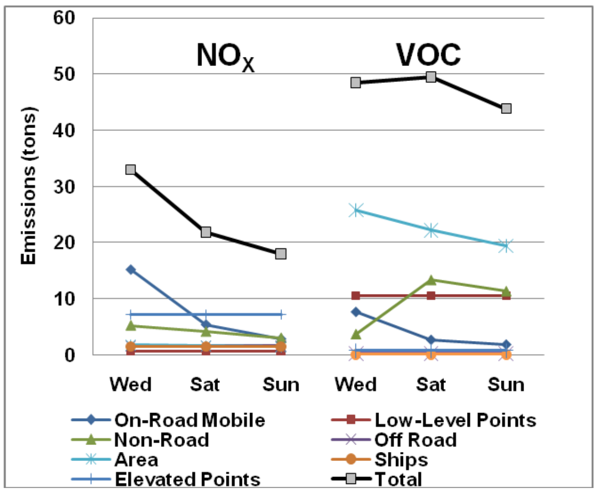

Figure 3: Modeled 6 AM anthropogenic NO_X and VOC emissions by source category, HGB 8-county totals

Figure 4 shows median observed and modeled 6 AM NO_X concentrations for episode days at all monitors which measured NO_X during the period. For comparison purposes, all concentrations are shown as a percent of the Wednesday values. At every site but one, observed NO_X concentrations declined from Wednesday to Saturday and at all sites further declined from Saturday to Sunday. Modeled NO_X concentrations for most sites show behavior similar to the observations, but in general do not drop off as rapidly as do the observed concentrations. Both observed and modeled NO_X concentrations show a large variation from site-to-site, which is at least partially due to relatively small samples: 11 each Saturdays and Sundays and 16 Wednesdays.

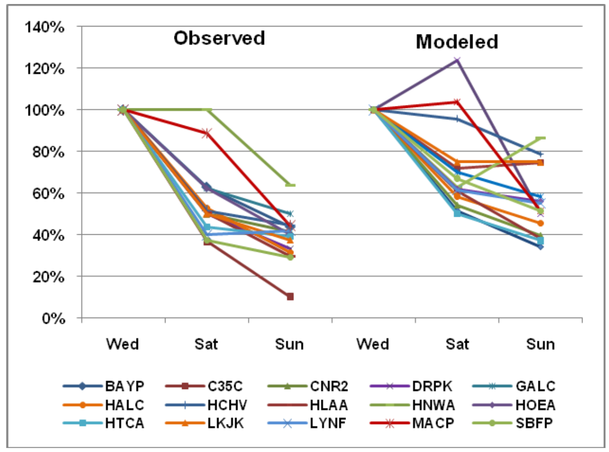

Figure 4: Median observed and modeled 6 AM NO_X concentrations as a percent of Wednesday, 2005 & 2006 episode days

4.1 Weekday/Weekend Ozone Patterns

Given the differences between the observed and modeled morning NO_X concentration response to day-of-week influences, it is logical to expect differences in the behavior of ozone concentrations as well. Figure 5 shows median observed and modeled Wednesday, Saturday, and Sunday daily peak 8-hour ozone concentrations as a percent of Wednesday for the ozone monitors that were active in 2005-6. There is no consistent day-of-week effect evident in the observed concentrations, suggesting that 8-hour peak ozone is neither predominantly NO_X- nor VOC–sensitive throughout the area, although a similar analysis conducted for one-hour peak ozone (not shown here) did suggest that the one-hour peaks are generally somewhat NO_X-sensitive. It is worth noting that for the sites with the highest design values (BAYP, DRPK), there is little weekend variation in median peak 8-hour ozone.

The median modeled concentrations in Figure 5, on the other hand, shows that the model exhibits a distinctive tendency towards increasing ozone concentrations on the weekends, with 17 of 20 modeled Saturday concentrations and all 20 Sunday concentrations higher than the Wednesday values. This suggests that the modeled 8-hour peaks may be more VOC-sensitive than the atmosphere itself.

4.2 Weekday/Weekend Sensitivity Analyses

These small sample sizes using only the modeled days do not allow for robust conclusions to be drawn, particularly considering the role of highly-variable coastal meteorology in the area.

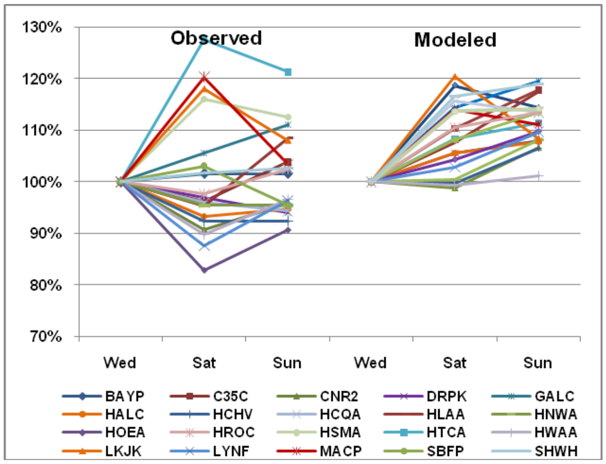

Figure 5: Median observed and modeled 8-hour peak ozone concentrations as a percent of Wednesday, 2005 and 2006 episode days.

To increase the number of samples, we performed three sensitivity modeling runs as follows: First, daily anthropogenic emissions from the baseline were replaced with Wednesday emissions, so that each day's emissions were the same as the previous day's (excluding the day-specific biogenic emissions). Then daily emissions were replaced with Saturday, then Sunday emissions, providing 96 instances of each type of day. For comparison, we calculated median observed NO_X and ozone concentrations from five recent ozone seasons, May 15 through October 15, 2005 through 2009, which provided 110 observations for each day type at each monitor except Galveston (66).

Figure 6 compares median observed NO_X concentrations from five ozone seasons with median modeled NO_X concentrations from the all-Wednesday, all Saturday, all-Sunday (all-WSS) runs. The figure shows patterns generally similar to those seen when only episode days were considered (Figure 4). In particular, the modeled concentrations are more consistent from monitor to monitor since the all-WSS test removes the confounding effect of weather patterns. Again, the observed weekend NO_X concentrations relative to Wednesday tend to be lower than the modeled counterparts with most monitors' Sunday concentrations clustered between 30 and 50 percent of Wednesday, while the modeled Sunday concentrations are spread fairly evenly between 30 and 80 percent of Wednesday.

Figure 7 compares median observed and modeled ozone concentrations for the 2005-9 ozone seasons with the all-WSS modeled concentrations.

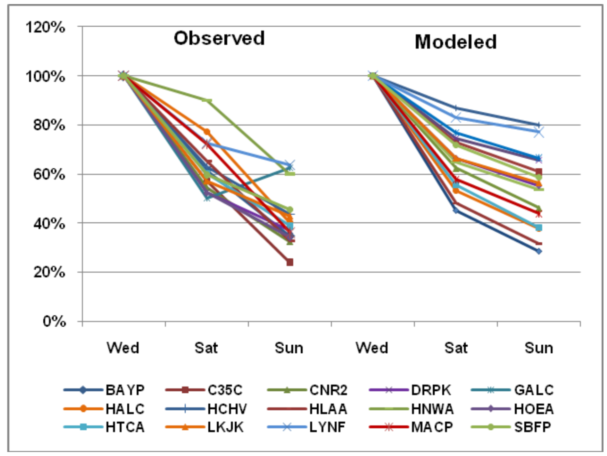

Figure 6: Median observed and modeled NO_X concentrations as a percent of Wednesday. Observed data from May 15-Oct. 15, 2005-9; Modeled data from all-WSS runs

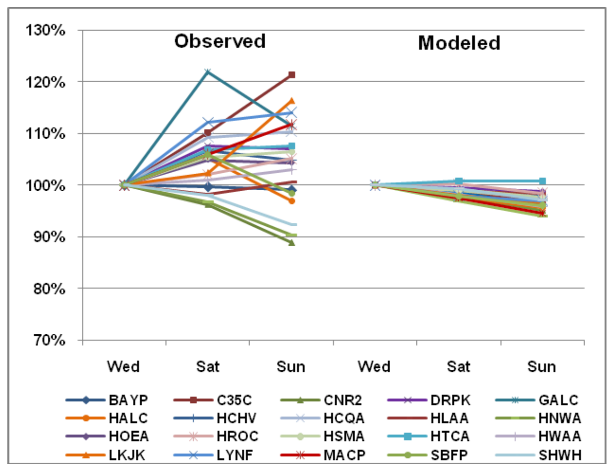

Figure 7: Median observed and modeled ozone concentrations as a percent of Wednesday. Observed data from May 15-Oct. 15, 2005-9; Modeled data from all-WSS runs

Like Figure 5, Figure 7 shows little discernible trend among the observed data. A Kruskal-Wallis test (Conover, 1971) for differences among day-type medians yielded no significant differences for any monitor despite the relatively large sample sizes afforded through five ozone season's data. The most striking feature of Figure 7 is the very close grouping of the modeled values for each day type, and contrary to the apparent increasing trend seen in Figure 5, the modeled ozone concentrations, with one exception, all decrease from Wednesday to Saturday to Sunday. Further investigation revealed the reason for these seemingly contradictory signals is simply sampling error; the particular set of Wednesdays modeled had lower modeled ozone concentrations than the modeled weekend days not because of emission differences but because of random meteorological effects. The all-WSS analysis, along with increasing the sample sizes, also factors out the random meteorological effect by using identical meteorology for each day type.

The observational data depicted in Figure 7, besides showing no statistically significant day-of-week effect also failed to show patterns suggesting whether or not the model is responding to day-of-week emission changes appropriately. But the modeling episodes are specifically selected to represent periods of high ozone concentration, so may not be well-represented by using the entire population of ozone-season days, since many ozone-season days are cloudy, windy, or otherwise not conducive to ozone formation. A better comparison is to look at the upper quantiles of the observed data. Figure 8 shows the 75th and 90th percentiles of the 2005-9 ozone season data, and a pattern begins to emerge.

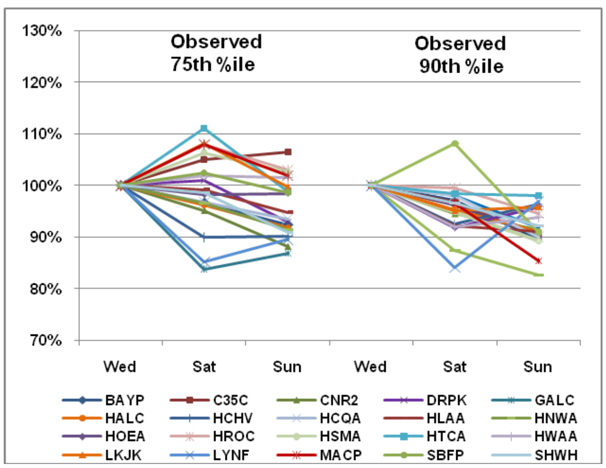

Figure 8: 75th and 90th percentiles of observed 8-hour peak ozone concentrations as a percent of Wednesday, May 15 – Oct. 15, 2005-9.

Figure 8 shows that the higher ozone concentrations appear to be more NO_X-sensitive than the lower or mid-range concentrations, and at the 90th percentile all but one monitor's weekend ozone concentrations were lower than the Wednesday value. Overall, both the 75th and 90th percentile observed values show consistency with the model's behavior shown in Figure 7, and the 90th percentile values appear to be more responsive to day-of-week effects than the model as shown in Figure 7. This latter result suggests that the model is not as responsive to emissions reductions as the atmosphere, at least for higher ozone concentrations.

5. CONCLUSIONS

Both the retrospective modeling and the weekday-weekend analysis show that the modeling conducted for the HGB AD behaves in a manner consistent with the atmosphere, and both suggest that the model may be somewhat less responsive than reality. The weekday-weekend analysis highlighted the potential problems associated with small sample sizes, but this paper provides a technique that can be used to greatly increase the number of modeled days for analysis.

6. REFERENCES

EPA (1997), Guidance on the Use of Models and Other Analyses for Demonstrating Attainment of Air Quality Goals for Ozone, PM2.5, and Regional Haze, www.epa.gov/ttn/scram/guidance/guide/ final-03-pm-rh-guidance.pdf

Blanchard, C.L. and Tanenbaum, S. (2005), Final Report - Weekday/Weekend Differences in Ambient Concentrations of Primary and Secondary Air Pollutants in Atlanta, Baltimore, Chicago, Dallas-Fort Worth, Denver, Houston, New York, Phoenix, Washington, and Surrounding Areas, www.ladco.org/reports/ technical_support_document/references/ nrel_final_report_es04-1.pdf

Conover, W. J. (1971) Practical Nonparametric Statistics, John Wiley & Sons.

Parrish, D. D., et al. (2009), Overview of the Second Texas Air Quality Study (TexAQS II) and the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS), *J. Geophys. Res.,* 114, D00F13, doi: 10.1029/2009JD011842.

TCEQ (2004), Revisions to the State Implementation Plan (SIP) for the Control of Ozone Air Pollution Houston/Galveston/Brazoria Ozone Nonattainment Area

TCEQ (2010a), Revision To The State Implementation Plan For The Control Of Ozone Air Pollution, Houston-Galveston-Brazoria 1997 Eight-Hour Ozone Standard Nonattainment Area
TCEQ (2010b), ...Appendix B: Emissions Modeling

(All TCEQ SIP Revisions referenced herein are available at: www.tceq.state.tx.us/implementation/ air/sip/sipplans.html)

Smith, J., Estes, M., Mercado, F., and Mellberg, J (2007) Assessing Model Performance Using Weekday/Weekend Analysis, Presented at the 6th Annual CMAS Conference, Chapel Hill, NC, October 1-3, 2007