Memorandum

To: Hoke Kimball

CC: Joelle Burleson, George Bridgers

From: Wayne L. Cornelius

Date: 2007-05-09 (revised 2008-12-16 and 2009-04-14)

Re: **Mendenhall PM2.5 Data Imputation for 4Q2006**

Contents

List of Tables

List of Figures

Introduction

The North Carolina Division of Air Quality (DAQ) is in nonattainment for Davidson County for PM2.5 from the monitoring station in the city of Lexington (AQS Site ID: 370570002). DAQ has performed design value calculations with the PM2.5 data for 2005-2008 for Lexington. These calculations indicate that the design value for this monitor will be in attainment for this time period and thus DAQ will be applying for PM2.5 nonattainment redesignation. Redesignation requires assessing the PM2.5 data from PM2.5 monitors at Mendenhall (Guilford Co., ID 370810013), Lexington (Davidson Co., ID 370570002) and Hattie (Forsyth Co., ID 370670022).

The design value calculations for 2005-2008 for the Mendenhall site are incomplete because no valid PM2.5 data were collected during the fourth quarter of 2006. This happened because of major complications in having to move the site. DAQ moved the site about 100 yards because a 2 story field house that was constructed

immediately adjacent to the monitoring site (unpublished letter to Artra Cooper, 12 December 2006). The construction was started without DAQ's knowledge. When DAQ realized what was happening it was too late to stop the project, the new field house was built, and the site no longer met ambient siting criteria.
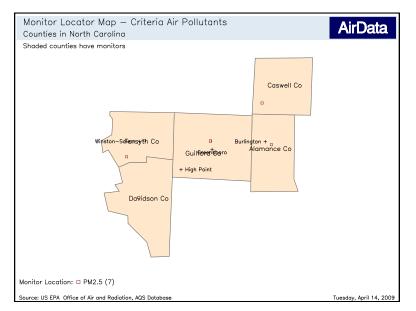

Figure 1 Map of Monitor Locations

Since there were no 4Q2006 data with which to calculate a proper Mendenhall Design value, the DAQ Planning and Ambient staff decided to present to EPA Region IV an estimate of what the missing Mendenhall PM2.5 sample values would have been if they had been properly observed, along with the resulting Design value summary statistics. The estimate is based on linear regression using data acquired during the four years, 2002 through 2005 at surrounding sites including those in the MSA and also Hopedale (Alamance County, ID 370010002) and Cherry Grove (Caswell Co., ID 370330001). These monitor locations are shown in Figure 1 (an extraneous PM2.5 monitoring site at Clemmons, southwest Forsyth County, is also shown for reference but was not used in the analysis).
.

Methods

The estimation procedure is as follows:

1. Fit a linear regression to the 2002-2005 PM2.5 data of the regressors to determine equation coefficients
2. Estimate missing sample values for Mendenhall by substituting the corresponding observed PM2.5 data in 4Q2006 into the regression equation
3. Compute quarterly averages for Mendenhall including the imputed 4Q2006 data using actual data where available and imputed data where provided by the regression procedure
4. Compute weighted averages for each year
5. Compute the completed Design value for Mendenhall derived by averaging the weighted annual means

Results

I applied two regression fits to the data, starting with the most inclusive possible model, using Lexington, Cherry Grove, Hopedale and Hattie Avenue all as predictors. Estimates from this model are shown in Table 1. In this combination, Lexington, Cherry Grove and Hattie Avenue are not significant predictors for Mendenhall.

Table 1 Regression Analysis using the Lexington, Cherry Grove, Burlington and Hattie Avenue PM2.5 data

Call: lm(formula = MH ~ LX + UC + HD + HA, data = MH4q.md3, na.action = na.exclude)				
Residuals:				
Min	1Q	Median	3Q	Max
−3.763	−1.161	−0.3787	0.5814	11.34
Coefficients:				
Regressor	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	0.6105	0.5347	1.1417	0.2565
LX	0.1615	0.1165	1.3861	0.1690
UC	0.0791	0.0947	0.8346	0.4061
HD	0.4782	0.1495	3.1982	0.0019
HA	0.2466	0.1414	1.7445	0.0844
Residual standard error: 2.111 on 93 degrees of freedom				
Multiple R-Squared: 0.8798				

The second regression removed Cherry Grove and Hattie Avenue from the model. The resulting model had a residual standard error of 2.094 and R^2= 0.877. Both Lexington and Hopedale were significant in this regression, but the intercept term was not significant, so I fit the model with its intercept forced to zero. This model's estimates are shown in Table 2. The regression equation is shown as equation (1)

$$MH = 0.3464 * LX + 0.6322 * HD \text{ (Equation 1)}$$

I fit (1) to the Lexington and Mendenhall data values acquired during 4Q2006. Table 3 shows the regressors for the 22 days with valid data for both regressors, and the resulting Mendenhall estimates. The average of the 22 imputed samples is 12.92.

Table 2 Regression Analysis using the Lexington and Hopedale PM2.5 data

Call: lm(formula = MH ~ -1 + LX + HD, data = MH4q.md3, na.action = na.exclude)				
Residuals:				
Min	1Q	Median	3Q	Max
−4.258	−0.9024	−0.1271	0.915	11.67
Coefficients:				
Regressor	Value	Std. Error	t value	Pr(>|t|)
LX	0.3464	0.0891	3.8873	0.0002
HD	0.6322	0.0957	6.6042	0.0000
Residual standard error: 2.103 on 107 degrees of freedom				

Table 4 shows the quarterly averages for 2004, 2005 and 2006, including the imputed value for 4Q2006 and the 11 actual values for the remaining quarters. Finally Table 5 shows the 3 annual means and the overall Design value result that obtains from them, 14.01.

Table 3 Imputed Raw Data

Sampling_Date	Mendenhall	Hopedale	Lexington
10/02/2006	11.485813	11.1	12.9
10/05/2006	23.900103	24.6	24.1

Sampling_Date	Mendenhall	Hopedale	Lexington
10/08/2006	6.785218	7.5	5.9
10/11/2006	13.710670	14.4	13.3
10/20/2006	11.761229	11.7	12.6
10/26/2006	11.848665	11.4	13.4
11/01/2006	16.130242	16.2	17.0
11/04/2006	12.507779	13.1	12.2
11/07/2006	10.703944	12.0	9.0
11/16/2006	4.536147	4.6	4.7
11/22/2006	3.983631	4.0	4.2
11/25/2006	16.255683	14.7	20.1
11/28/2006	18.157377	15.9	23.4
12/01/2006	4.790717	4.4	5.8
12/04/2006	9.893171	8.8	12.5
12/10/2006	16.526730	14.8	20.7
12/13/2006	25.615500	24.3	29.6
12/19/2006	18.599245	18.9	19.2
12/22/2006	8.990912	9.4	8.8
12/25/2006	6.246493	6.1	6.9
12/28/2006	14.394680	13.4	17.1
12/31/2006	17.455206	17.2	19.0

Table 4 Quarterly Summaries

Period	CY2004	CY2005	CY2006
1Q	11.76	11.45	10.55
2Q	14.40	13.12	13.71
3Q	16.54	19.25	19.07
4Q	13.19	12.21	12.92

Table 5 Weighted Annual Means and Design Value

Period	CY2004	CY2005	CY2006	2004-2006 D.V.
Mean	13.97	14.01	14.06	14.01

Discussion

I maintain that the estimated Design value presented in Table 5 is an accurate prediction of the result that would have been obtained from Mendenhall for 2004-2006, had siting conditions not changed during 4Q2006. The imputed average is also the most accurate and appropriate value to use for the 2006-2008 Design value calculations at Mendenhall to assist with the redesignation package for the Lexington site.

Recommendations

Design value calculations for the Greensboro-Winston-Salem-High Point MSA (or any subsequently redefined area that includes Greensboro) for any group of years that includes 2006 should use the imputed 4Q2006 value as a surrogate for the missing "actual" 4Q2006 at the Mendenhall site.

For future consideration, we can *apply* (1) to data acquired after 2006 from Lexington, Hopedale and Mendenhall. We can also repeat the regression fitting exercise using data acquired from the regressor sites in 2007 and later instead of 2002-2005 Either of these actions can be used to demonstrate how well the moved site location "represents" the original location.