Statistical Comparison of Observed and Multi-Resolution CMAQ Modeled Hourly Ozone Concentration

Center for Integrating Statistical and Environmental Science

Li Chen
The University of Chicago
Joint work at CISES with Michael L. Stein, Alexis Zubrow and V. R. Kotamarthi

This research is funded by U.S. EPA - Science To Achieve Results (STAR) Program Grant # R-82940201-0

Abstract

CMAQ usually runs at multiple resolutions for research and policy purposes. This presentation compares observed hourly ozone concentrations to the CMAQ modeled hourly ozone concentrations at different spatial resolutions in Chicago area and in Atlanta area. Fractional bias (FB) and root normalized mean squared error (RNMSE) are calculated. The results show that high resolution CMAQ model output does not necessarily provide smaller FB and RNMSE than lower resolution runs. However, when high resolution output is aggregated to reduce small scale spatial fluctuations, one generally obtains better agreement than either the unaggregated high resolution model output or the low resolution model output in terms of RNMSE.We also decompose the total variation into components depending on hour, day and location and their interactions to better understand the statistical behavior of CMAQ model output at different resolutions. The temporal variation is captured reasonably well by CMAQ model output, but spatial variation and space-time interactions are not.

1 Case Study One: Chicago Area

1.1 Data

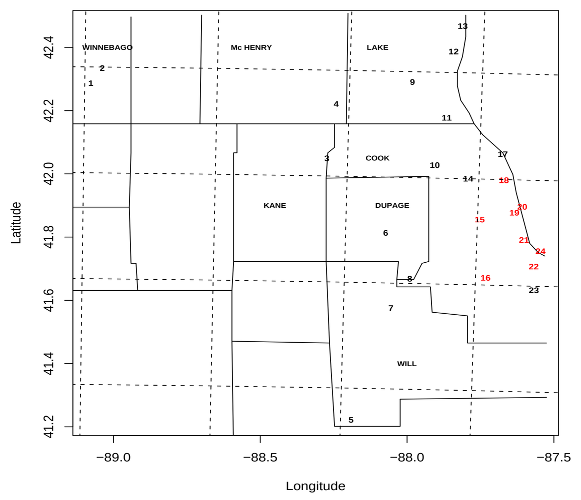

Figure 1: The Chicago area.

- Two sets of nested 36km, 12km and 4km CMAQ model output: one with low PBL input and the other with high PBL input.
- Observations at 24 monitoring sites.
- Time period: 06/24/1996 - 08/01/1996.
- Two regions: the Urban Region (the 36 km grid cell which contains monitors in red) and the Rural Region (everywhere else).

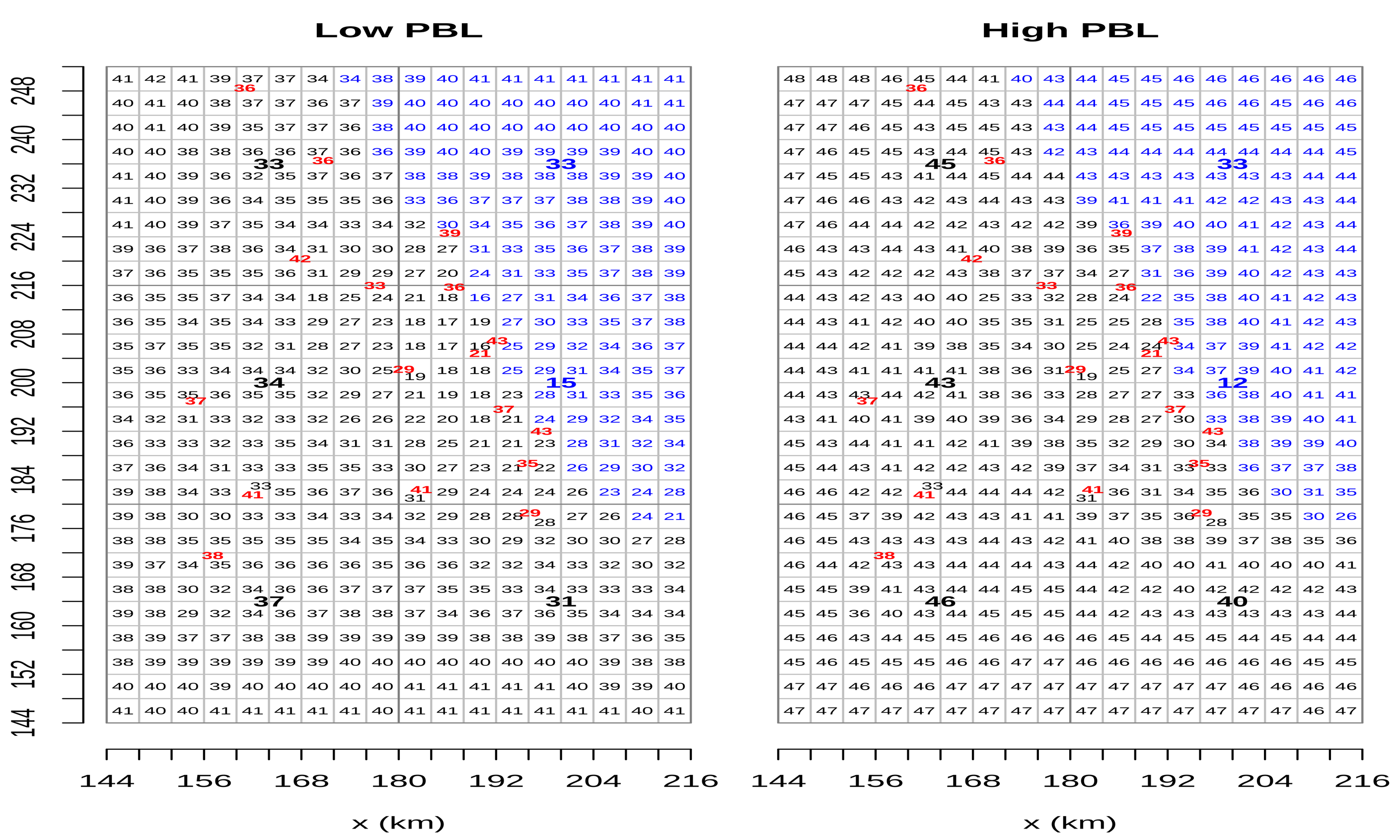

Figure 2: Day time average hourly ozone concentration (ppb).

1.2 FB and RNMSE

Write X_{ijk} the observed ozone at location i hour k on day j. Similarly, M_{ijk} is the model output from some version of the CMAQ model at location i hour k on day j (Shao et al. 2005). FB and RNMSE (Canepa and Irwin 2005) are defined as

$$FB_i = \frac{\overline{M}_{i..} - \overline{X}_{i..}}{(\overline{M}_{i..} + \overline{X}_{i..})/2},$$

$$RNMSE_i = \sqrt{\frac{\frac{1}{N_i}\sum_{j,k}(M_{ijk} - X_{ijk})^2}{\overline{M}_{i..}\overline{X}_{i..}}},$$

where $\overline{M}_{i..} = \frac{1}{N_i}\sum_{j,k} M_{ijk}$, $\overline{X}_{i..} = \frac{1}{N_i}\sum_{j,k} X_{ijk}$, and N_i is the total number of non-missing observations at location i. The missing observations and the corresponding model output are not included in this step.

Table 1. FB and RNMSE for the Chicago area study ($\times 10^{-2}$).

	FB					RNMSE				
	M^{36}	M^{12}	A^{12}	M^4	A^4	M^{36}	M^{12}	A^{12}	M^4	A^4
RL	3	1	9	5	7	53	56	51	58	50
RH	20	20	25	14	23	52	53	51	53	50
UL	-27	-40	-9	-54	-8	89	93	69	109	68
UH	-27	-14	11	-24	10	91	73	62	81	61

1.3 Analysis of variance (ANOVA)

Let Z be a general notation for the quantity of interest. We decompose Z_{ijk} as,

$$Z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + r_{ijk}. \quad (1)$$

- Every term sums to 0 when summed over any index, for example, $\sum_i (\alpha\beta)_{ij} = 0$ for all j and $\sum_j (\alpha\beta)_{ij} = 0$ for all i.
- α_i represents the site effect at location i; β_j is the j-th day effect and γ_k is the hourly effect (diurnal pattern).
- $(\alpha\beta)_{ij}$ is for the effect at site i on day j, $(\alpha\gamma)_{ik}$ is for the effect at site i hour k, and $(\beta\gamma)_{jk}$ is for the effect on day j at hour k.
- In this study, we group overall mean and γ_k as the diurnal effect, since this is the dominant source of variation in the data.

We do this analysis of variance for the differences between CMAQ model output and observations, and compare to the corresponding decomposition for the observations. If CMAQ is able to capture some variation, we would expect to see smaller number in the decomposition of differences than in the observations.

Table 2. ANOVA for the Chicago area study ($\times 10^3$ per site).

	M^{36}	M^{12}	A^{12}	M^4	A^4	X	M^{36}	M^{12}	A^{12}	M^4	A^4	X
	Hour						Day					
RL	45	41	45	36	38	864	35	40	34	45	35	91
RH	52	49	71	31	60		38	36	34	34	32	
UL	111	103	38	148	33	737	60	55	49	63	51	109
UH	88	23	15	45	14		70	57	51	56	45	
	Site						Day×Hour					
RL	8	11	10	7	8	15	48	51	47	50	47	57
RH	13	12	9	9	9		58	60	56	58	53	
UL	30	17	25	16	25	24	88	91	84	90	82	85
UH	39	18	25	16	25		104	100	98	97	93	
	Site×Day						Site×Hour					
RL	21	24	20	25	19	13	5	3	3	4	3	3
RH	21	20	16	22	16		6	5	4	5	4	
UL	17	24	16	26	16	13	13	10	11	10	11	11
UH	18	23	15	26	14		14	11	11	10	11	
	Total						Residual					
RL	207	224	201	227	193	1079	45	54	43	61	43	35
RH	238	245	236	228	219		49	63	46	70	46	
UL	353	339	256	401	252	1010	33	40	33	47	33	31
UH	368	277	250	305	238		36	45	34	55	35	

2 Case Study Two: Atlanta Area

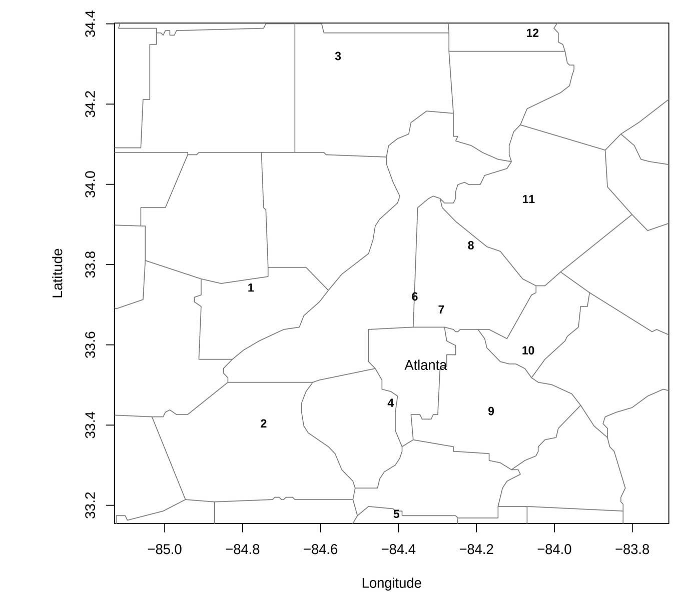

Figure 3: The Atlanta area.

- Hourly ozone concentrations from nested 32km, 8km and 2km CMAQ model output.
- Observed hourly ozone concentrations at 12 monitoring sites.
- Time period: 08/01/1999 - 08/24/1999.

Table 3. FB and RNMSE ($\times 10^{-2}$).

	M^{32}	M^8	A^8	M^2	A^2
FB	38	30	30	29	30
RNMSE	51	51	48	49	46

Table 4. ANOVA for Atlanta area study ($\times 10^4$)

Effect	M^{32}	M^8	A^8	M^2	A^2	X
Hour	342	216	235	201	231	1972
Day	36	53	49	27	25	41
Site	27	46	28	39	25	20
Hour× Day	49	49	46	45	42	59
Site×Hour	35	34	20	27	18	31
Site×Day	23	34	24	31	22	19
Residuals	68	114	73	114	70	61
Total	580	545	475	485	431	2202

3 Summary

- The results of Atlanta area study have the similar pattern as shown in the Rural Region in Chicago area study.
- High resolution CMAQ model output does not necessarily predict hourly ozone concentration better than low resolution CMAQ model output in terms of FB and RNMSE.
- But the aggregated model output from high resolution to low resolution does help to improve the on average performance, especially for the Urban Region.
- The analysis of variation shows that CMAQ has great capability to model both diurnal pattern and day to day variation.
- Aggregation helps to capture more diurnal variation in the Urban Region. Moreover aggregation helps in capturing the daily effect for both regions.
- For site related variations, CMAQ does not capture them well.

References

Canepa, E. and J. Irwin (2005). Evaluation Of Air Pollution Models. In P. Zannetti (Ed.), *Air Quality Modeling - Theories, Methodologies, Computational Techniques, and Available Databases and Software. Vol II - Advanced Topic*. The EnviroComp Institute and the Air & Waste Management Association

Shao, X., Stein, M. L. and Ching, J. (2005).Statistical comparisons of methods for interpolating the output of a numerical air quality model, *Technical Report at the Center for Integrating Statistical and Environmental Science 23*.

Although the research described herein has been funded wholly or in part by the United States Environmental Protection Agency through STAR Cooperative Agreement #R-82940201-0 to the University of Chicago, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.