JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, D08301, doi:10.1029/2008JD011224, 2009

An analysis of simulated wet deposition of mercury from the North American Mercury Model Intercomparison Study

O. Russell Bullock Jr.,[1] Dwight Atkinson,[2] Thomas Braverman,[3] Kevin Civerolo,[4] Ashu Dastoor,[5] Didier Davignon,[5] Jia-Yeong Ku,[4] Kristen Lohman,[6] Thomas C. Myers,[7] Rokjin J. Park,[8] Christian Seigneur,[6] Noelle E. Selin,[9] Gopal Sistla,[4] and Krish Vijayaraghavan[6]

Received 1 October 2008; revised 17 December 2008; accepted 6 February 2009; published 18 April 2009.

[1] A previous intercomparison of atmospheric mercury models in North America has been extended to compare simulated and observed wet deposition of mercury. Three regional-scale atmospheric mercury models were tested: the Community Multiscale Air Quality (CMAQ) model, the Regional Modeling System for Aerosols and Deposition (REMSAD), and the Trace Element Analysis Model (TEAM). These models were each employed using three sets of lateral boundary conditions to test their sensitivity to intercontinental transport of mercury. The same meteorological and pollutant emission data were used in each simulation. Observations of wet deposition were obtained from the National Atmospheric Deposition Program's Mercury Deposition Network. The regional models can explain 50–70% of the site-to-site variance in annual mercury wet deposition. CMAQ was found to have slightly superior agreement with observations of annual mercury deposition flux in terms of the mean value for all monitoring sites, but REMSAD showed the best correlation when measured by the coefficient of determination (r^2). With the exception of one CMAQ simulation, all of the models tended to simulate more wet deposition of mercury than was observed. TEAM exceeded the observed average annual wet deposition by 50% or more in all three of its simulations. CMAQ and REMSAD were better able to reproduce the observed seasonal distribution of mercury wet deposition than was TEAM, but TEAM showed the highest correlation for weekly wet deposition samples. An analysis of model accuracy at each observation site showed no obvious geographic patterns for correlation, bias, or error. Adjusting simulated mercury deposition on the basis of the difference between observed and simulated precipitation data improved the correlation and error scores for all of the models.

Citation: Bullock, O. R., Jr., et al. (2009), An analysis of simulated wet deposition of mercury from the North American Mercury Model Intercomparison Study, *J. Geophys. Res.*, *114*, D08301, doi:10.1029/2008JD011224.

1. Introduction

[2] This work is a continuation of the North American Mercury Model Intercomparison Study (NAMMIS) described in the work of *Bullock et al.* [2008], where three regional-scale atmospheric mercury (Hg) models were applied in a tightly constrained testing environment with a focus on North America. Each model simulated the entire year of 2001 using common input data sets for initial conditions, meteorology, emissions, and boundary values. Each model was also applied to the same horizontal modeling grid. *Bullock et al.* [2008] provided a detailed description of the study design and model-to-model comparisons of simulated air concentration and wet and dry deposition patterns of Hg. This paper continues the analysis of results with a comparison of simulated Hg wet deposition to observations from the Mercury Deposition Network (MDN) described by *Vermette et al.* [1995] and from event-based monitoring at the Proctor Maple Research Center (PMRC) near Underhill, VT [*Keeler et al.*, 2005].

[3] Three regional-scale atmospheric Hg models are the prime subjects of the study: the Community Multiscale Air

[1]National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA.
[2]Office of Water, U.S. Environmental Protection Agency, Washington, D.C., USA.
[3]Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA.
[4]New York State Department of Environmental Conservation, Albany, New York, USA.
[5]Air Quality Research Division, Environment Canada, Dorval, Quebec, Canada.
[6]Atmospheric and Environmental Research, Inc., San Ramon, California, USA.
[7]ICF International, San Rafael, California, USA.
[8]School of Earth and Environmental Sciences, Seoul National University, Seoul, Korea.
[9]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

0148-0227/09/2008JD011224$09.00

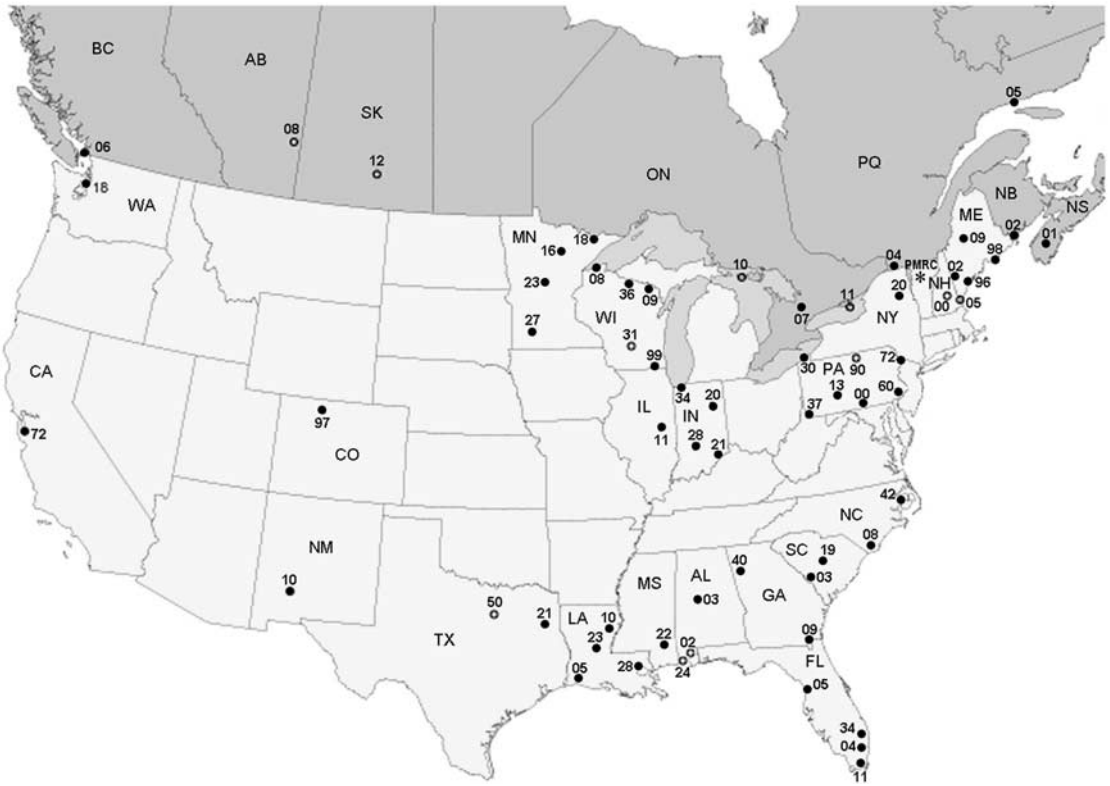

Figure 1. Mercury Deposition Network (MDN) monitors providing observational data for this study. Solid circles indicate where valid measurements cover 50% or more of all four seasons during the 2001 study period. Numbers and state/province labels indicate MDN station identifiers. The location of the Proctor Maple Research Center (PMRC) monitor is also shown.

Quality (CMAQ) model, the Regional Modeling System for Aerosols and Deposition (REMSAD), and the Trace Element Analysis Model (TEAM). These regional-scale models were each applied using three separate initial condition/boundary condition (IC/BC) data sets derived from different global-scale models: the Chemical Transport Model for Hg (CTM-Hg) described by *Shia et al.* [1999] and *Seigneur et al.* [2001], the Goddard Earth Observing System–Chem (GEOS-Chem) model described by *Selin et al.* [2007], and the Global-Regional Atmospheric Heavy Metal (GRAHM) model described by *Dastoor and Larocque* [2004] and *Ariya et al.* [2004]. All regional-scale model simulations used the same set of meteorological input data derived from simulations of the Penn State/NCAR Mesoscale Meteorological Model–Generation 5 (MM5), which is described by *Grell et al.* [1994].

[4] The CMAQ, REMSAD, and TEAM models are all described by *Bullock et al.* [2008, and references therein]. These models all simulate air concentrations and deposition of three species of Hg: elemental Hg (Hg^0), reactive gaseous Hg (RGM), and particulate Hg (PHg). Hg^0, RGM, and PHg are known to have very different atmospheric behaviors and scientific uncertainties about Hg transformations in air and cloud water lead to modeling uncertainties that need to be better understood [*Lindberg et al.*, 2007]. CMAQ, REMSAD, and TEAM have each been used to estimate source attribution for mercury deposition in the United States in assessments conducted by private and/or governmental organizations. However, this is the first time they have been employed using the same input data and compared in a relative sense against observations of mercury wet deposition on the time scale of those observations.

2. Description of the Observational Data

[5] The MDN was well established by the 2001 test period and 50 or more weekly integrated total-Hg wet deposition measurements are typically available for any time during this period. Observed Hg wet deposition data for this study were obtained from the MDN website (http://nadp.sws.uiuc.edu/mdn). Figure 1 provides the location of the MDN monitors that were operational during the NAMMIS study period. Monitors indicated with a solid circle reported valid data for at least 50% of the weekly sampling periods within all four seasons of 2001. Similar Hg wet deposition monitoring was also conducted during 2001 at the Proctor Maple Research Center (PMRC) near Underhill, Vermont, as reported by *Keeler et al.* [2005]. These samples were taken at the end of each precipitation event instead of on a fixed weekly basis.

[6] A data survey performed by the New York State Department of Environmental Conservation (NYSDEC) found that no standardized network observation of speciated-Hg air concentration or total-Hg dry deposition was performed within the study domain during 2001. Unfortunately, wet deposition is only part of the removal pathway for atmospheric mercury and it remains impossible to show that any model has a complete and accurate treatment for the entire mercury cycle. Some total gaseous Hg (TGM) air

Table 1. Observed and Simulated Mean and Standard Deviation for Annual Hg Wet Deposition at All Monitors and Model Correlations to All Observations of Annual Hg Wet Deposition

		CMAQ			REMSAD			TEAM		
	Observed	CTM	GEOS-Chem	GRAHM	CTM	GEOS-Chem	GRAHM	CTM	GEOS-Chem	GRAHM
Mean	9.09	9.08	11.33	11.08	12.15	13.40	11.14	15.80	17.88	15.35
σ	4.34	3.85	5.02	5.03	6.46	7.17	5.95	5.51	6.56	5.30
r^2		0.522	0.639	0.686	0.643	0.650	0.624	0.504	0.613	0.592

concentration measurements were available, but the applicability of TGM data to model evaluation is limited owing to the prevalence of Hg^0 in those samples and the strength to which Hg^0 air concentrations at the boundary drive regional model TGM concentrations across the domain.

3. Model Comparisons to Observed Annual Hg Wet Deposition

[7] Hg wet deposition during 2001 as simulated by the three regional-scale models was compared with observations from the MDN and at the PMRC using two methods. In the first method of comparison, data from MDN monitors were qualified for use only when the monitor reported valid weekly samples for at least one half of all weeks during each of the four seasons of the year. Seasons were defined on the basis of the following sets of months: winter (January, February, and December), spring (March, April, and May), summer (June, July, and August), and autumn (September, October, and November). Fifty-one of the 62 MDN monitors met this qualification. For the second method of comparison, all valid Hg wet deposition reports from the 62 MDN monitors were used regardless of the fraction of time for which valid samples were missing. For both methods of comparison, model-simulated wet deposition totals for each observation site were calculated by taking into account the actual collection period for each valid sample. The PMRC monitor recorded Hg wet deposition for individual precipitation events during 2001 and the annual deposition flux reported by *Keeler et al.* [2005] was used as one additional observation in each method of comparison. Simulation results were extracted on the basis of the model grid cell containing the location of each monitor. These two methods yielded essentially the same results in our statistical comparisons of annual Hg wet deposition flux. Thus, the second, more inclusive method of comparison using all valid wet deposition samples was adopted for all of the annual results shown in this work.

[8] Table 1 shows a comparison of the mean and standard deviation of the observed Hg wet deposition for 2001 at all 63 monitors (62 MDN and 1 PMRC) to the mean of the Hg wet deposition simulated by the three regional-scale models using each of the three IC/BC data sets. With one exception, all regional model simulations produced higher mean annual wet deposition than was observed. However, it should be noted that the precipitation simulated by the MM5 meteorological model was 12.3% greater than observed when averaged for all valid MDN samples. Averaged among the three IC/BC cases, the CMAQ simulation of Hg wet deposition was only about 15% above the observed value. REMSAD produced about 35% more Hg wet deposition than was observed. TEAM predicted considerably more Hg wet deposition than the other two models, exceeding the observed average by about 80%. This is due, at least in part, to less chemical reduction of Hg^{II} to Hg^0 in the TEAM cloud water chemistry mechanism compared to the mechanisms of CMAQ and REMSAD. Concentrations of HO_2 in clouds are reduced by a factor of 5 in TEAM to account for uncertainties regarding the aqueous reaction of Hg^{II} with HO_2 discussed by *Gårdfeldt and Jonsson* [2003]. All three models simulate Hg^0 as being sparingly soluble in water, so it is largely transferred from cloud droplets to air when it is produced by cloud water chemistry. Less reduction of Hg^{II} to Hg^0 in TEAM leads to more total Hg in cloud water and more simulated Hg wet deposition. For standard deviation, the CMAQ results were generally comparable to the observations while REMSAD and TEAM exceeded observations.

[9] The response to differing boundary concentrations is not the same for all of the regional models. CMAQ simulates the lowest total Hg wet deposition using the CTM boundary concentrations, whereas the other two regional models simulate the lowest total Hg wet deposition using the GRAHM boundary concentrations. However, all three regional models simulate the highest total annual Hg wet deposition using the GEOS-CHEM boundary concentrations. *Bullock et al.* [2008] showed the three global models produced quite different air concentrations of Hg^0, RGM, and PHg at the regional model boundaries. Table 1 suggests the effect of these variations in lateral boundary concentrations is about as important as the differences seen between the CMAQ and REMSAD simulations when using the same input data.

[10] Table 1 also shows the coefficient of determination (r^2) for simulated versus observed annual wet deposition flux of Hg at all 63 observation sites for each of the nine simulation test cases. The r^2 values range roughly between 0.5 and 0.7, implying that the regional models can explain between 50% and 70% of the site-to-site variation in annual Hg wet deposition. The CMAQ correlation statistics show the greatest sensitivity to changes in boundary conditions whereas the REMSAD correlation statistics show the least sensitivity. Again, the precipitation data used by these models has a bearing on the accuracy of their simulation of wet deposition. The r^2 value for annual precipitation from MM5 versus MDN was 0.52. Since the input data for precipitation are accounting for only about half of the observed variance, the accuracy that can be achieved by the CMAQ, REMSAD, and TEAM models in their wet deposition simulations is certainly limited.

4. Model Comparisons to Observed Seasonal Hg Wet Deposition

[11] To examine the effect of seasonal changes in temperature, precipitation and other meteorological factors, the

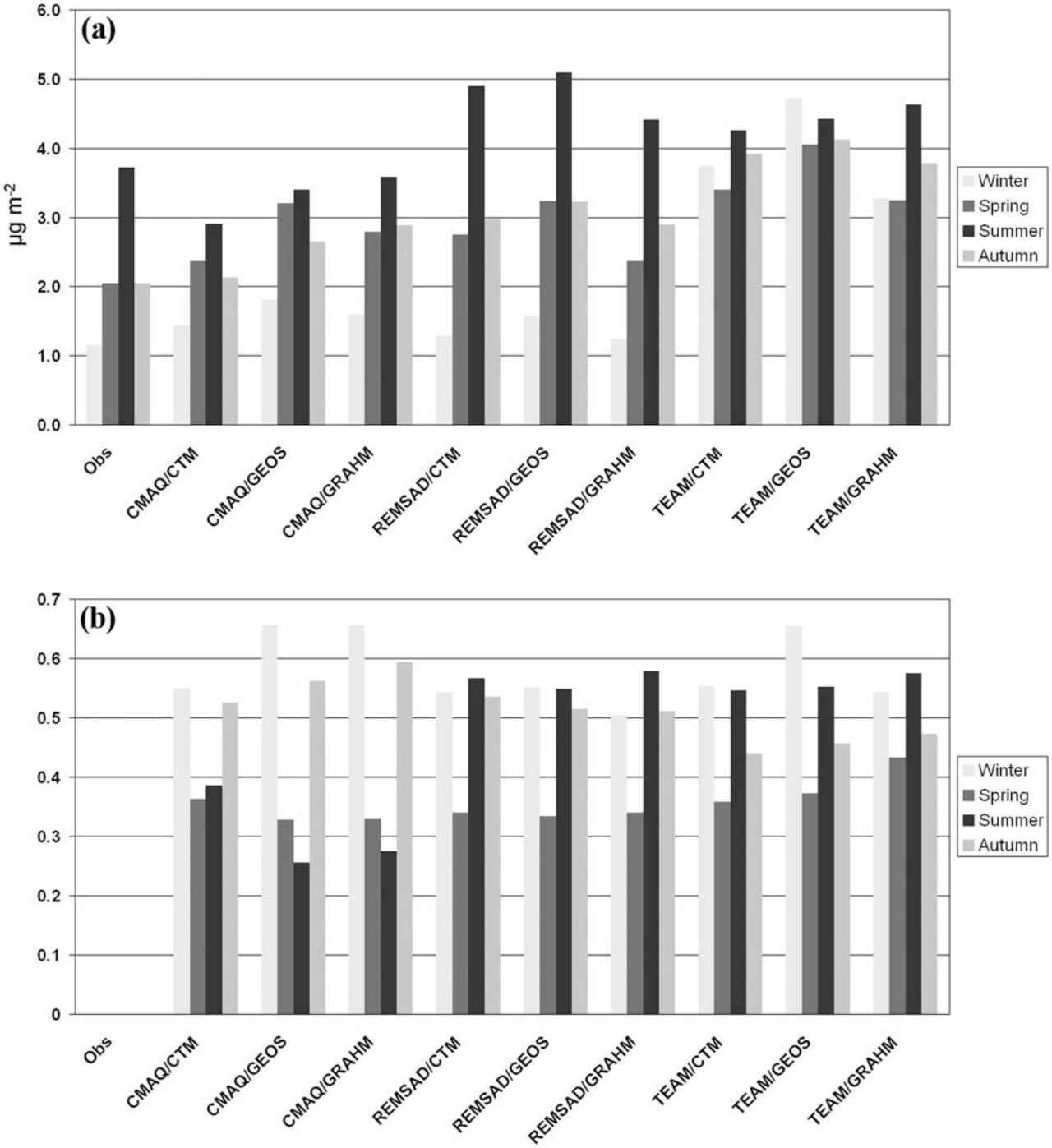

Figure 2. Seasonal values for (a) mean observed and simulated Hg wet deposition and (b) coefficient of determination of simulated versus observed Hg wet deposition. CMAQ, Community Multiscale Air Quality; CTM, Chemical Transport Model; GEOS, Goddard Earth Observing System; GRAHM, Global-Regional Atmospheric Heavy Metal; REMSAD, Regional Modeling System for Aerosols and Deposition; TEAM, Trace Element Analysis Model.

comparative analysis above using all valid MDN and PMRC data (second method described above) was repeated for each of the four seasons in 2001. Figure 2a shows the mean of the observed Hg wet deposition total for each season along with the comparable simulated value from each of the nine test cases (three regional models; three IC/BC sets). The observations show that the greatest Hg deposition flux occurs during summer whereas the lowest occurs during winter. CMAQ and REMSAD generally reproduce this seasonal pattern regardless of the boundary conditions being used, whereas TEAM generally fails to do so.

[12] Figure 2b presents r^2 values for seasonal Hg wet deposition totals at all MDN monitors reporting valid

Table 2. Definition of Statistical Performance Metrics for Hg Deposition[a]

Performance Metric	Equation
Mean bias (ng m^{-2})	$MB = \frac{1}{N}\sum_{i=1}^{N}(D_m - D_o)$
Mean error (ng m^{-2})	$ME = \frac{1}{N}\sum_{i=1}^{N}\lvert D_m - D_o\rvert$
Normalized mean bias (−1 to +∞)	$NMB = \frac{\sum_{i=1}^{N}(D_m - D_o)}{\sum_{i=1}^{N}D_o}$
Normalized mean error (0 to +∞)	$NME = \frac{\sum_{i=1}^{N}\lvert D_m - D_o\rvert}{\sum_{i=1}^{N}D_o}$
Mean fractional error (0 to +2)	$MFE = \frac{1}{N}\sum_{i=1}^{N}\frac{\lvert D_m - D_o\rvert}{\left(\frac{D_o + D_m}{2}\right)}$

[a]D_m is model value and D_o is observed value.

samples during a season. The spring and summer seasons present an exceptionally difficult situation for realistic simulation of wet deposition for Hg, and for any other wet-deposited substance for that matter. Precipitation in the warm seasons over North America is largely convective in nature and this presents special difficulties for the meteorological simulation on which the air-quality models rely to estimate wet deposition. For all models in general, the lowest value is obtained for spring with r^2 ranging from 0.32 (CMAQ/GEOS-Chem) to 0.42 (TEAM/GRAHM), but for CMAQ its lowest correlation of 0.27 is in the summer season (CMAQ/GEOS-Chem). Performance is moderate for winter with r^2 ranging from 0.50 (REMSAD/GRAHM) to 0.66 (CMAQ/GRAHM). Performance for summer is lowest for CMAQ (r^2 = 0.27 to 0.40) and considerably higher for REMSAD (r^2 = 0.57 to 0.60) and TEAM (r^2 = 0.56 to 0.59). Performance for fall is lowest for TEAM (r^2 = 0.47 to 0.51), somewhat higher for REMSAD (r^2 = 0.54 to 0.56), and highest for CMAQ (r^2 = 0.56 to 0.63).

5. Model Comparisons to Individual MDN Samples

[13] To put the regional-scale models to the most stringent test possible with the available data, individual weekly MDN samples have been compared to the corresponding regional-scale modeling results. The term "weekly" is used here to mean one sampling period from the MDN where a valid Hg wet deposition measurement was taken. The normal MDN sampling period is one week in length, but actual sample integration times varied from less than one day to as many as 21 days. Simulated wet deposition from the CMAQ, REMSAD, and TEAM models was analyzed taking into account these varying sampling periods. Pairs of observed and simulated data were only used in the generation of weekly deposition statistics if both values were nonzero. When no precipitation is indicated by the meteorological input data, none of the three regional models can (or should) simulate wet deposition. However, they do simulate some Hg wet deposition whenever any nonzero precipitation amount is indicated. Thus, the statistics generated are indicative of the air-quality models' ability to reproduce an observed wet deposition flux whenever any precipitation is indicated.

[14] Table 2 describes the statistical performance metrics used to investigate the ability of the regional models to reproduce individual (generally weekly) MDN wet deposition measurements. Table 3 shows these statistics for the CMAQ, REMSAD, and TEAM Hg wet deposition results for all three IC/BC test cases and for the MM5-simulated precipitation amounts as compared to the MDN observations. Table 3 also shows r^2 statistics for least squares linear regressions of simulated Hg wet deposition and precipitation amount versus those observed at all of the MDN monitors. The most desirable outcomes for Hg wet deposition statistics are shown in bold. Given that the MM5 model was only able to resolve about 35% of the observed variance in weekly precipitation amount (r^2 = 0.35), the ability of the CMAQ, REMSAD, and TEAM models to reproduce the coincident wet deposition of Hg is obviously limited. Indeed, their r^2 values for weekly Hg wet deposition are quite low (0.12 to 0.18) as compared to those for annual and seasonal Hg wet deposition. Highest r^2 was obtained by the TEAM model using the IC/BC data set derived from the GRAHM model. An investigation of CMAQ-simulated ammonium wet deposition by *Davis and Swall* [2006] showed a similar dependence on the accuracy of input precipitation data with r^2 for wet deposition being significantly lower than the r^2 for precipitation. The TEAM model shows a strong positive bias and the largest errors in general. The most desirable results for all bias and error statistics were obtained by the CMAQ model using the CTM-derived IC/BC data set. Taking all IC/BC cases into account, the CMAQ model generally shows the best results for bias and error. In all cases, the best normalized and fractional errors obtained for Hg wet deposition are similar to those inherent in the precipitation data used.

[15] Figure 3 shows an analysis of the coefficient of variation (CV) for observed weekly Hg wet deposition and the CV in the corresponding results obtained from the

Table 3. Various Model Performance Metrics Based on Weekly MDN Sample Collections in 2001[a]

	CMAQ			REMSAD			TEAM			
	CTM	GEOS-Chem	GRAHM	CTM	GEOS-Chem	GRAHM	CTM	GEOS-Chem	GRAHM	MM5 Precip.
r^2	0.15	0.12	0.14	0.16	0.15	0.16	0.14	0.12	**0.18**	0.35
Mean bias (ng m^{-2})	**−12.2**	46.9	40.2	67.8	100.2	41.3	164.2	220.3	155.2	1.9 (mm)
Mean error (ng m^{-2})	**178.1**	213.0	207.0	226.0	248.7	213.8	278.8	326.3	264.8	15.3 (mm)
Normalized mean bias	**−0.049**	0.187	0.160	0.270	0.399	0.164	0.653	0.876	0.617	0.078
Normalized mean error	**0.708**	0.847	0.823	0.899	0.989	0.850	1.109	1.298	1.053	0.620
Mean fractional error	**0.725**	0.771	0.759	0.839	0.861	0.835	0.885	0.928	0.867	0.641

[a]The most desirable outcomes for Hg wet deposition statistics are shown in bold.

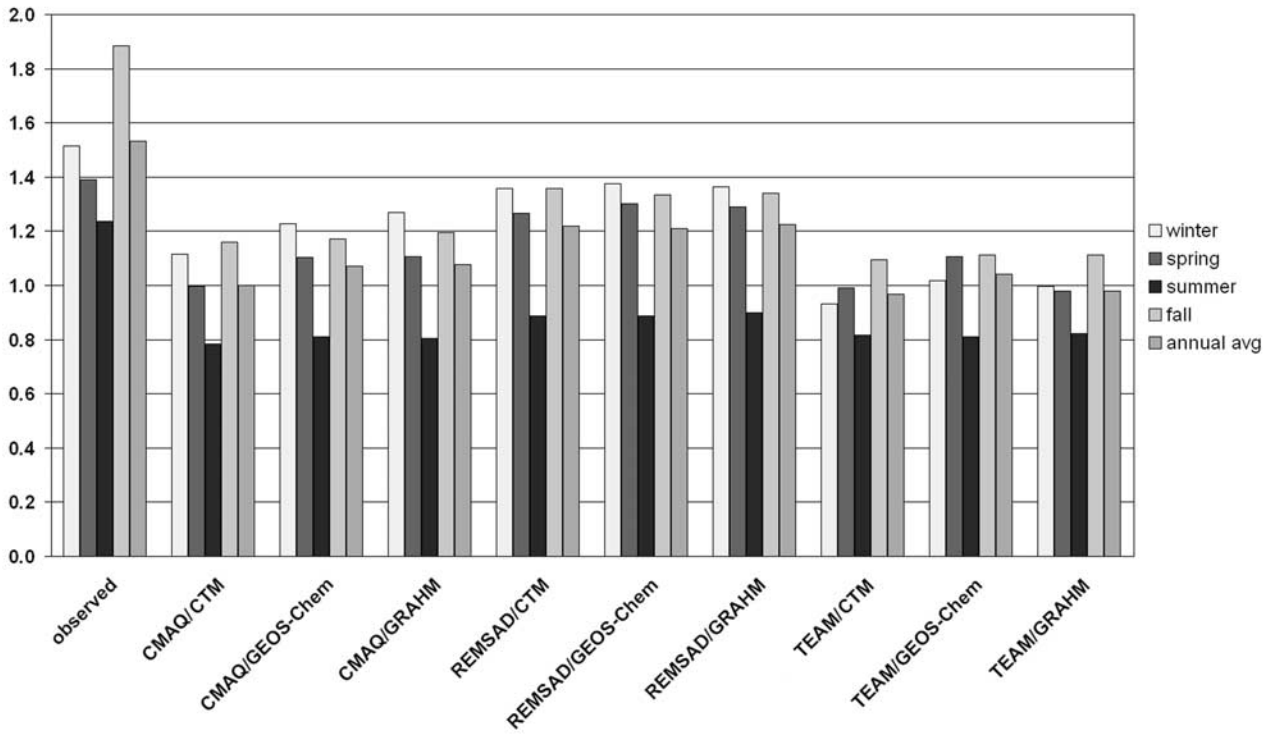

Figure 3. Coefficient of variation for the observed and simulated weekly Hg wet deposition data for each of the four seasons and for the entire annual period.

nine test simulations. Statistics for each season and for the entire annual period are shown. The observed CV is lowest in summer. Because of the more convective nature of precipitation in the warm seasons, one might expect weekly wet deposition of mercury to be more variable than during the cold seasons. However, the mean deposition amount is highest in summer, thus reducing the CV even though the standard deviation is higher than for the other seasons. The depth of the atmosphere through which convective precipitation can incorporate mercury may also be a factor since mercury concentrations in the upper troposphere are likely to be more homogeneous than near the surface where emission sources exist. *Selin and Jacob* [2008] argue that convective precipitation is a major contributor to Hg deposition in the United States because it scavenges globally distributed RGM from high altitude. All nine regional model simulations produced lowest CV in summer and produced relative rankings for the other seasons that were similar to observation. The models consistently show a lower CV than the observations, which is a reasonable outcome when considering the spatial averaging inherent in all finite-grid models. REMSAD generally produced the highest CV (most realistic) and TEAM produced the lowest CV (least realistic).

[16] Figure 4 shows $X-Y$ scatterplots for Hg wet deposition versus precipitation amount including all of the individual MDN sampling periods in 2001 where deposition was observed and simulated by all of the regional models (p = 2023). Figure 4a shows the observed Hg wet deposition versus the observed precipitation. Figures 4b–4d show the average simulated Hg wet deposition from each regional model (averaged among all three IC/BC sets) versus the MM5-simulated precipitation. The dashed lines in each plot indicate the concentration of Hg associated with particular $X-Y$ ratios. The "Observed" plot (Figure 4a) shows a significant number of samples with very high or very low Hg concentrations. Nearly all of the data points in the plot of TEAM-simulated samples (Figure 4d) lie between the 5 ng L^{-1} and 40 ng L^{-1} concentration lines with the lower limit most clearly defined. The CMAQ-simulated samples (Figure 4b) also show an upper concentration limit near 40 ng L^{-1}, but none of its samples show more than 2000 ng m^{-2} of deposited Hg mass. CMAQ shows a weakly defined lower concentration limit that is similar to the observations. The REMSAD samples (Figure 4c) are the most widely scattered among all models, even more scattered than the observations and with many samples indicating nearly zero Hg concentration.

6. Geographic Patterns of Model Performance

[17] In addition to domain-wide assessments of model performance, a quantitative analysis of the model's ability to reproduce observed week-to-week variation of Hg deposition at the individual measurement sites was performed. Figure 5 shows maps of the r^2 statistic for simulated-versus-observed Hg wet deposition based on individual MDN and PMRC samples. In all of the nine regional model simulations, the r^2 value at a majority of the monitors is less than 0.2. CMAQ has better success matching observations around the Great Lakes, Ohio Valley, and Northeast regions

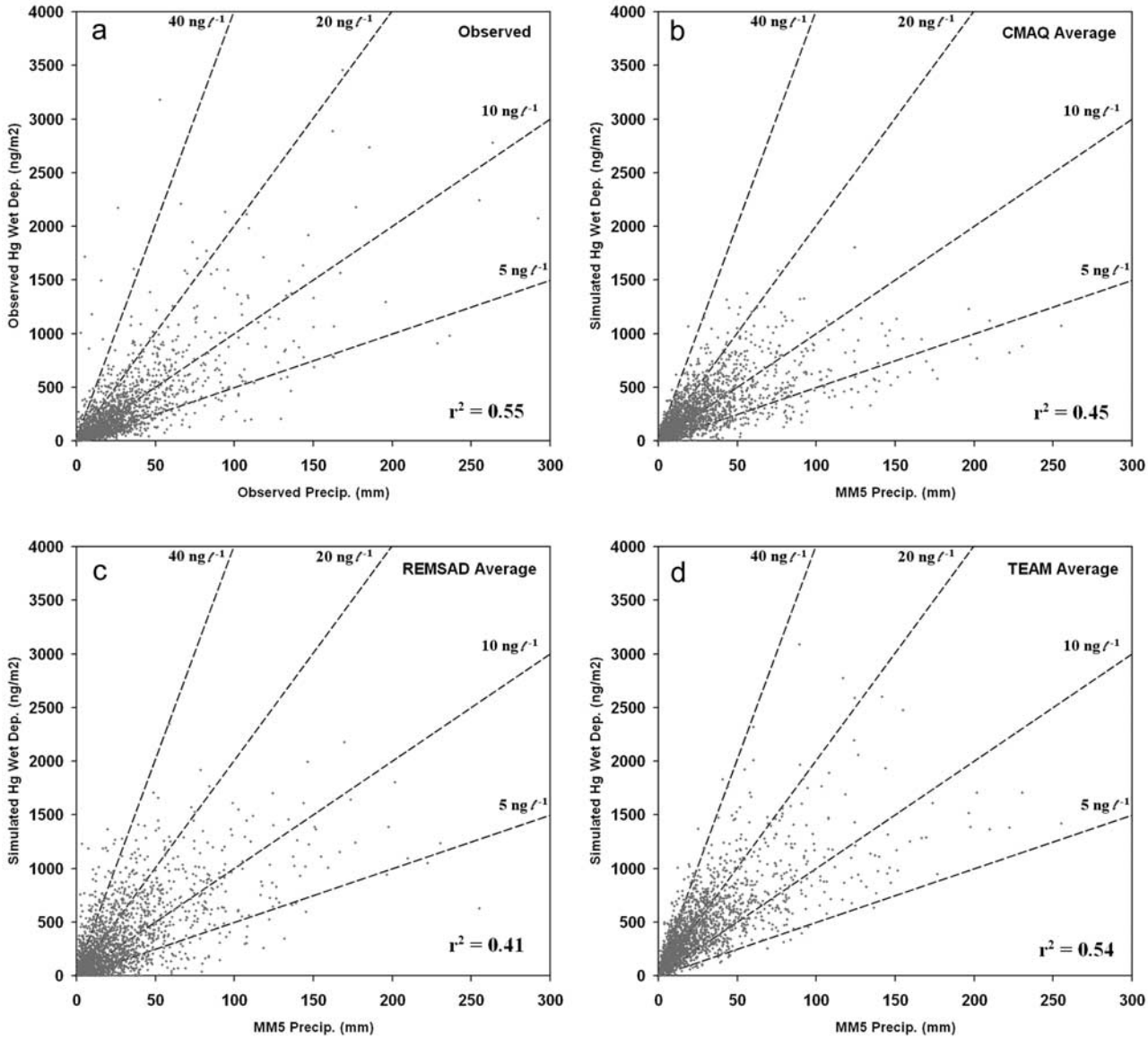

Figure 4. *X–Y* scatterplots for Hg wet deposition versus precipitation for all MDN sampling periods where deposition was observed and simulated by all of the regional models. MM5, Mesoscale Meteorological Model–Generation 5.

than in other locations. The area of best performance for REMSAD is from western New York and Pennsylvania through Illinois. TEAM performs best at widely scattered locations. All of the models do quite well for the MDN monitor in Saskatchewan (SK12), which only reported data from 31 May to 15 August. As was the case for CV, the choice of IC/BC data set is not a significant factor in any of the models' geographic pattern for the r^2 statistic. Figure 6 shows a map of the r^2 statistic for MM5-derived precipitation versus observation. In general, the input precipitation data is in better agreement with observation than the Hg wet deposition simulated by any of the regional models tested. There are some site-by-site correlations of high r^2 for input precipitation and high r^2 for simulated Hg wet deposition, but these correlations are limited in number.

[18] Figure 7 shows maps of the normalized mean error (NME) for simulated Hg wet deposition at each of the MDN and PMRC monitors. Figure 8 shows the NME for MM5-simulated precipitation at each monitor. Note that the scale in Figure 8 is shifted downward by 0.2 (20%) compared to the scale in Figure 7, making the NME in Hg wet deposition from CMAQ appear roughly equal to the NME in the input precipitation data. In accordance with the findings in Table 3, CMAQ shows generally lower NME than REMSAD or TEAM, especially for the CTM IC/BC test case. As was the case for r^2, there is no obvious geographic pattern to NME for any of the regional model simulations that would suggest a systematic modeling error related to climatic regime and meteorological interactions with mercury deposition processes. However, there does appear to be a somewhat stronger site-by-site correlation

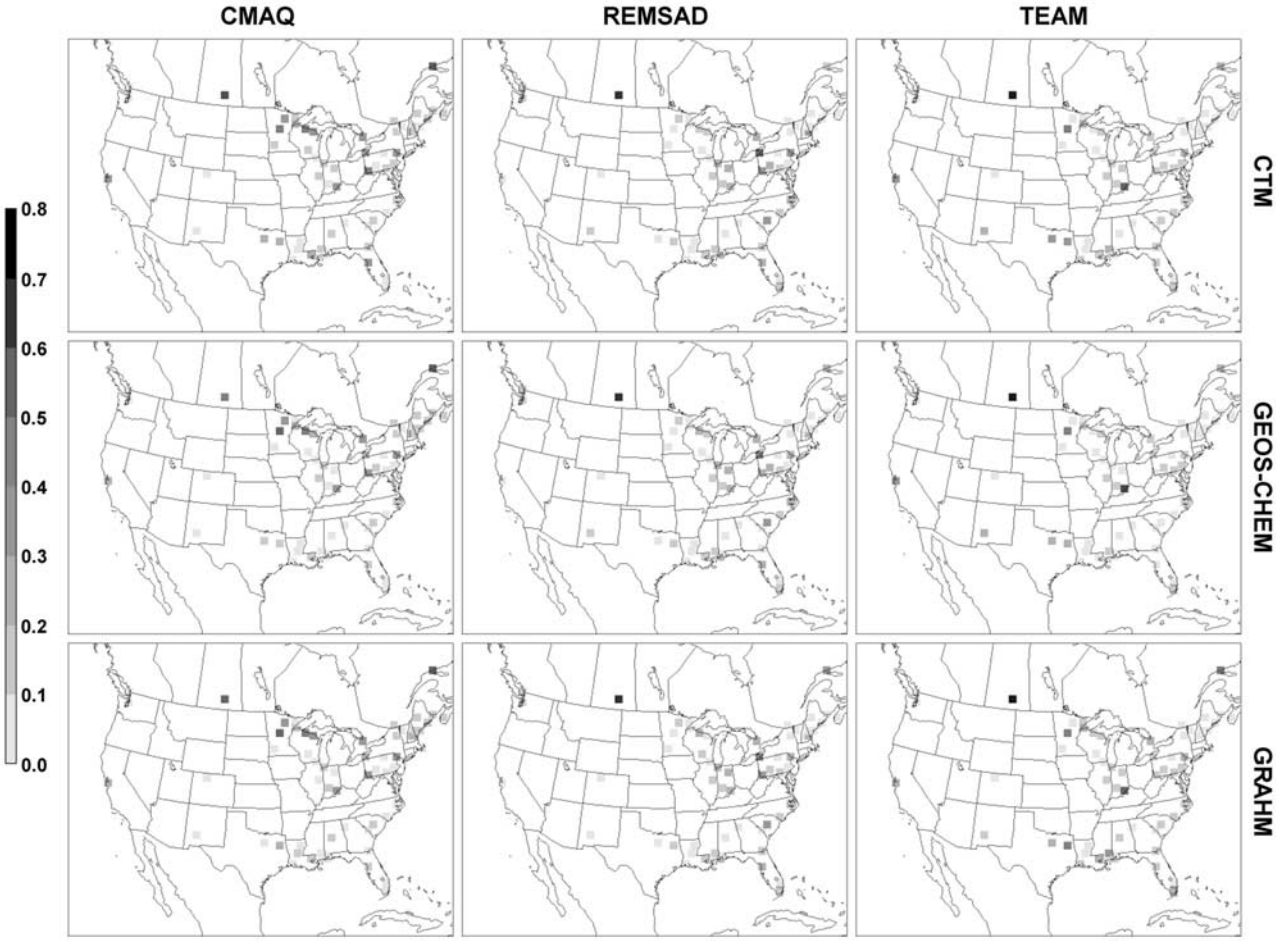

Figure 5. Maps of the coefficient of determination (r^2) for simulated Hg wet deposition versus samples collected at the individual MDN and PMRC monitors.

between the NME for simulated Hg wet deposition and the NME for the input precipitation data than was found for r^2. Also, the choice of IC/BC data set does have a bearing on the general magnitude of the NME scores at the individual monitors, certainly more so than for r^2. This would suggest that the influence of Hg transported into the modeling domain is rather evenly distributed across the entire area. For those monitors where the NME for Hg wet deposition is often much greater than the NME for precipitation regardless of the model or IC/BC data employed (e.g., BC06), an unresolved emission source or geophysical feature may be a contributory factor.

[19] Figures 9 and 10 show maps of the normalized mean bias (NMB) for simulated Hg wet deposition and MM5-simulated precipitation, respectively. The NMB scores for Hg wet deposition are generally greater than zero, especially for the REMSAD and TEAM models. This agrees with the domain-wide model results shown on Table 3. As was the case for r^2 and NME, there is no obvious geographic pattern to NMB for any of the models. The choice of IC/BC data set affects the general magnitude of NMB for each model but has little effect on its geographic pattern. The pair of MDN sites near Vancouver and Seattle (BC06 and WA18), and the pair of sites in northeast Texas (TX21 and TX50) provide examples of how bias in the input precipitation data affect bias in simulated Hg wet deposition. In each pair, the sites are in close proximity to each other and have a similar climate, with one of the two having considerably higher bias in its precipitation data. In both cases, NMB for Hg wet deposition correlates well with NMB for precipitation.

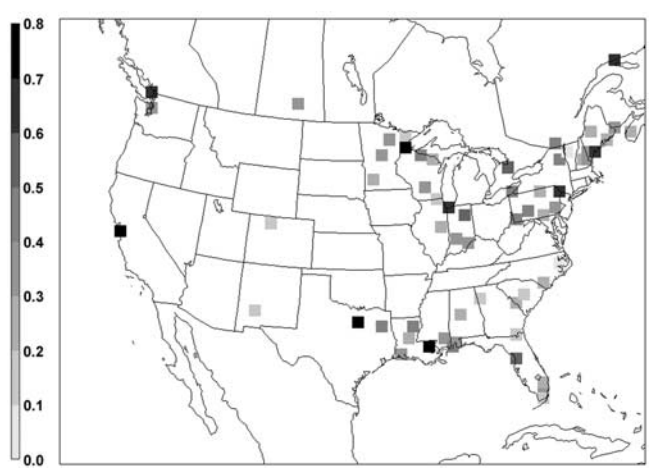

Figure 6. Map of the coefficient of determination (r^2) for MM5-simulated precipitation versus samples collected at the individual MDN and PMRC monitors.

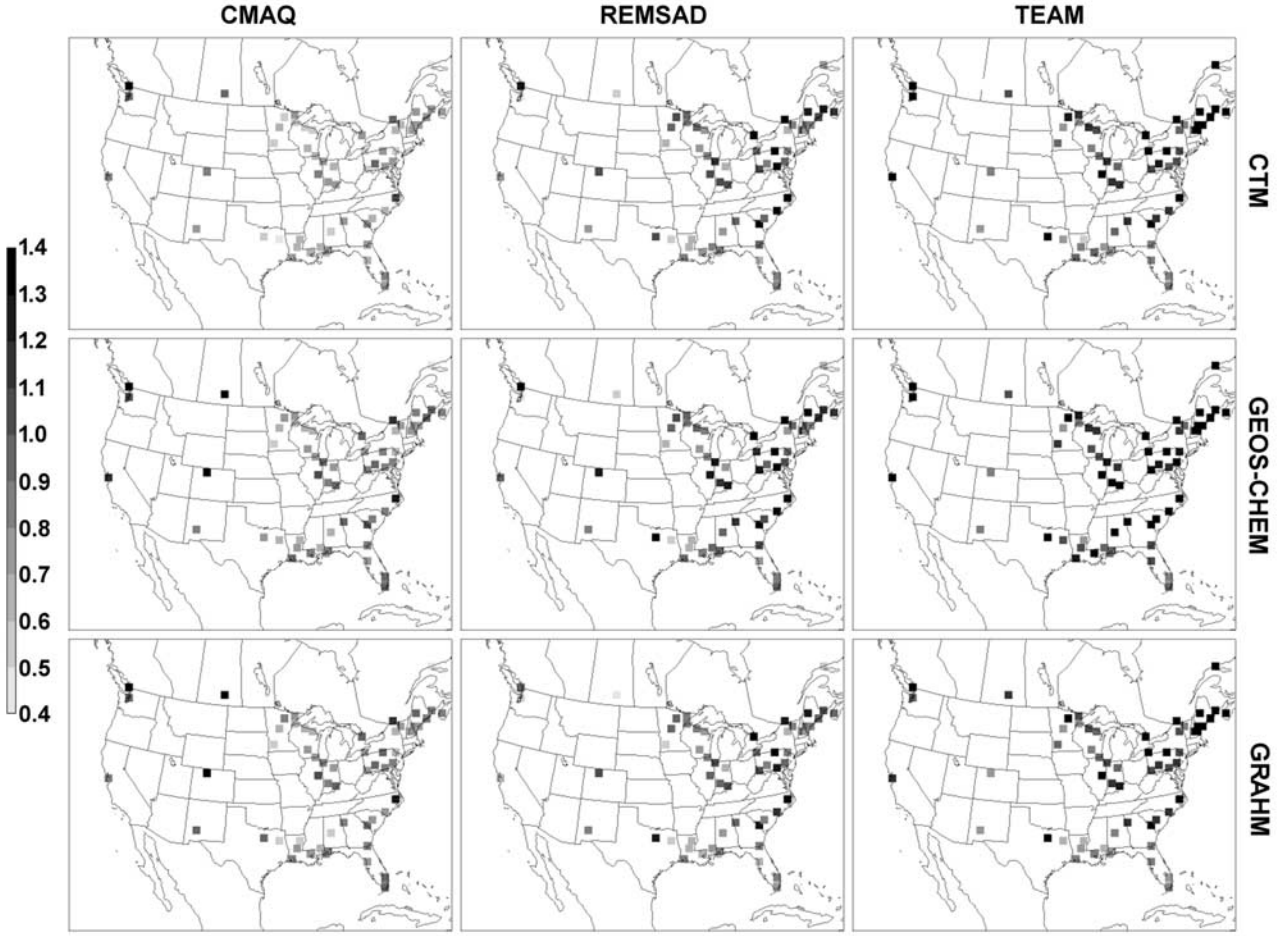

Figure 7. Maps of the normalized mean error (NME) for simulated Hg wet deposition at each of the MDN and PMRC monitors.

Figures 9 and 10 show a number of other examples where this correlation between precipitation bias and Hg wet deposition bias is evident. On the other hand, the MDN site in New Mexico (NM10) shows a high bias for precipitation but a relatively low bias for Hg wet deposition. One might suspect that emissions of readily deposited Hg from sources in or near that area may have been underestimated in the emission inventory used by all of the regional models.

7. Scaling of Simulated Results to Account for Precipitation Errors

[20] The finding of site-by-site correlations between simulated Hg wet deposition and input precipitation data led us to further investigate the importance of precipitation error. We experimented by scaling the Hg wet deposition simulated for each observed sample on the basis of the ratio of the observed precipitation to the MM5-derived precipitation. It is worth noting here that the scaling was made on the modeling results based on precipitation ratios, but that response of the models to changes in precipitation is not necessarily linear. In fact, the results described below indicate a more complex response from all models.

[21] In addition to simply multiplying the simulated deposition samples by the ratio of the observed and MM5-derived precipitation samples (scaling factor = 1.0), the scaling factor was varied in increments of 0.1 from zero to 1.2 times the precipitation ratio. Figure 11 shows the effect of this scaling on the r^2 statistic for each model based on the average of all three IC/BC test cases. Scaling

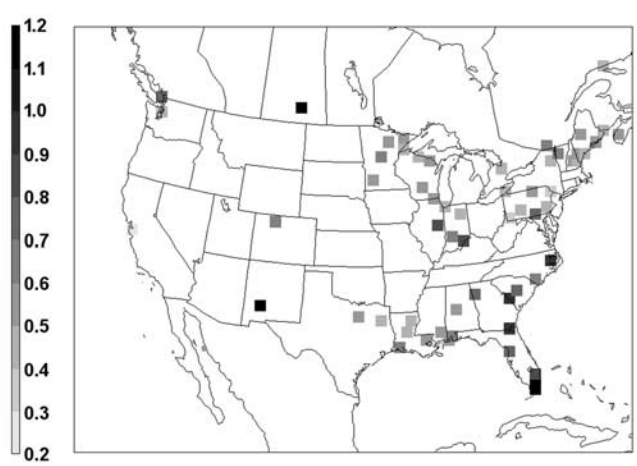

Figure 8. Map of the NME for MM5-simulated precipitation at each of the MDN and PMRC monitors.

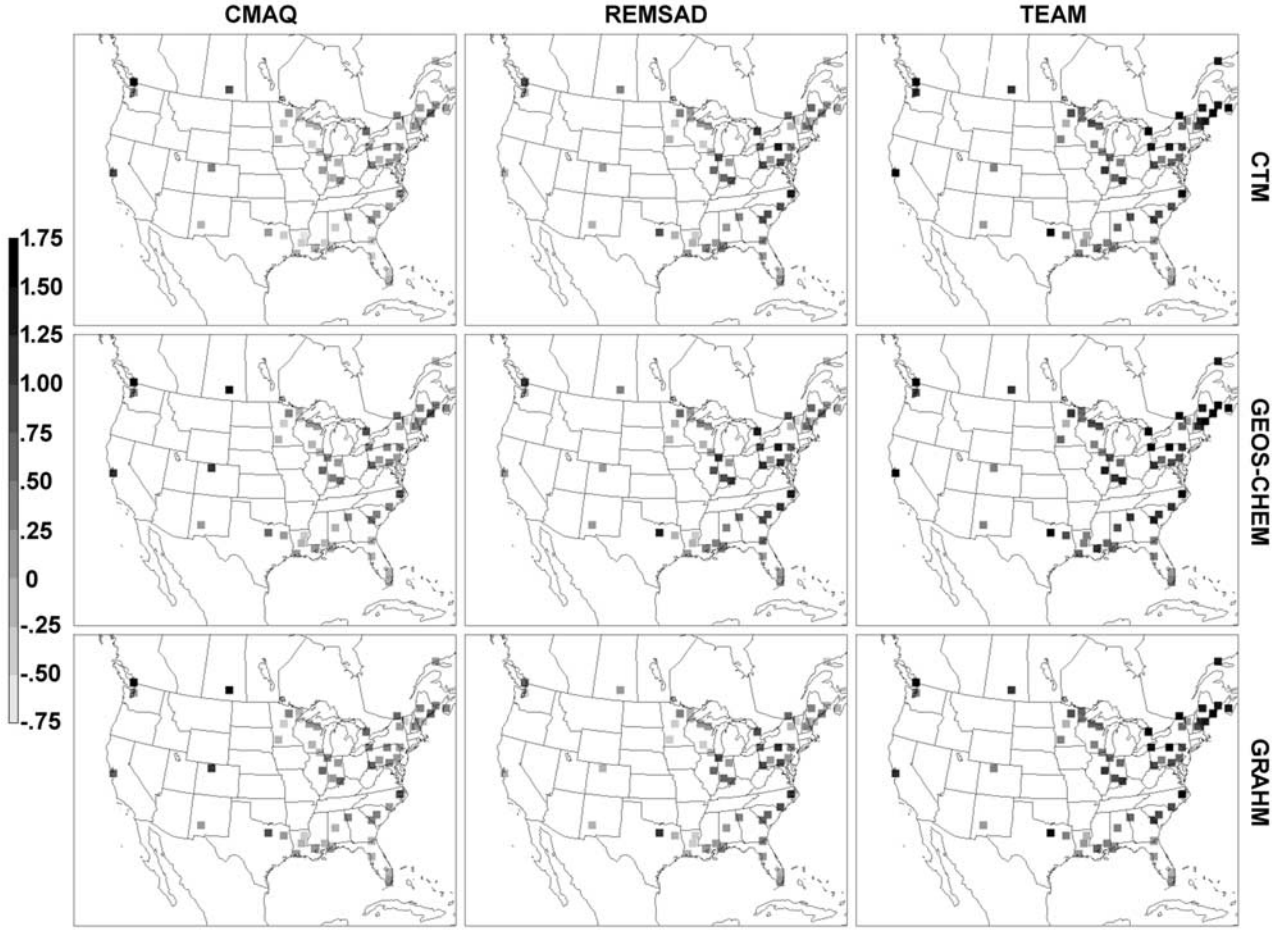

Figure 9. Maps of the normalized mean bias (NMB) for simulated Hg wet deposition at each of the MDN and PMRC monitors.

simulated Hg wet deposition on the basis of errors in the input precipitation data resulted in significantly higher r^2 for all of the models, with the maximum benefit being obtained at scaling factors of less than one. Application of this Hg wet deposition scaling had the greatest benefit to the CMAQ model, which scored the lowest r^2 without any scaling, but

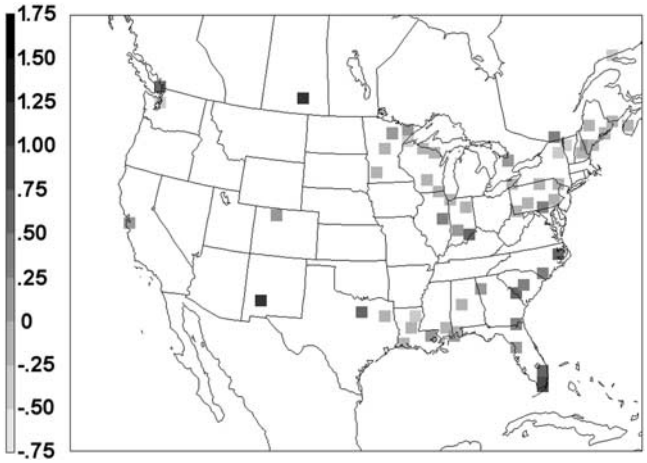

Figure 10. Map of the NMB for MM5-simulated precipitation at each of the MDN and PMRC monitors.

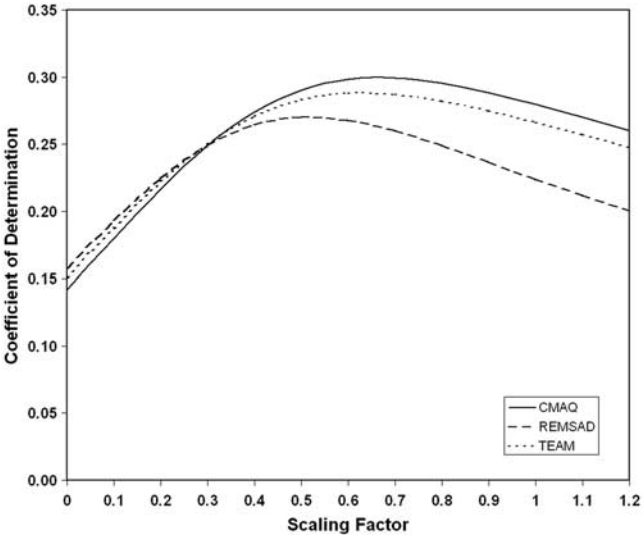

Figure 11. Graph showing the effect on r^2 scores for each model (based on the average of all three initial condition/boundary condition (IC/BC) test cases) when simulated Hg wet deposition samples are scaled on the basis of precipitation error.

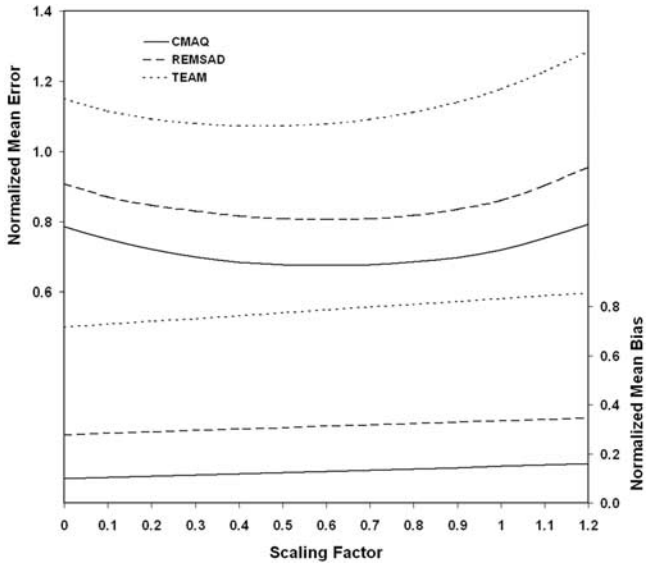

Figure 12. Graph showing the effect on NME and NMB scores for each model (based on the average of all three IC/BC test cases) when simulated Hg wet deposition samples are scaled on the basis of precipitation error.

highest with a scaling factor of 0.4 or greater. Figure 12 shows the effect of this scaling on the overall NME and NMB averaged among the three IC/BC test cases. It results in lower NME for all of the models when the scaling ratio is greater than zero but less than about one. The benefit of scaling on NME diminishes to zero before the scaling factor reaches one for TEAM and before it reaches 1.2 for all models. For NMB, the effect of scaling was somewhat surprising. NMB actually increased for all three models as the strength of the adjustment to simulated Hg wet deposition was increased. Bias for any model-to-observed comparison is limited on the negative side by the observed value, but virtually unlimited on the positive side. Even when precipitation was relatively light, the models simulated considerable wet deposition of Hg where RGM or particulate Hg existed in significant concentrations. If the observed precipitation was actually heavy, the scaling resulted in extremely high adjusted values for wet deposition that skewed the overall NMB toward higher values. In essence, this exemplifies the nonlinear nature of the Hg wet deposition process.

[22] Table 4 shows the effect of applying a scaling factor of 0.6 to the domain-wide model performance statistics previously discussed from Table 3. The model-IC/BC combinations resulting in the most desirable scores did not change. All of the scores for correlation and error were improved. However, the scores for bias were not improved, except for the one case (CMAQ/CTM) where the original results showed a slightly low bias.

[23] The results above indicate that the regional models could certainly benefit from more accurate precipitation data. It is interesting that r^2 was maximized in all of the models using a scaling factor of less than one and each model had a different optimum value. If wet deposition were mostly a product of in-cloud oxidation of Hg and precipitation of the oxidation products, one might expect the most benefit from a scaling factor of exactly one. If it were mostly from subcloud scavenging of highly soluble RGM, whether from industrial sources or from gas-phase oxidation, the amount of precipitation should not be a major factor as depletion should occur rapidly. This suggests that even with perfect precipitation data, modeling uncertainties regarding the processes leading to the capture and deposition of Hg in precipitation would continue to produce significant error.

8. Summary

[24] Comprehensive evaluation of atmospheric mercury models is not currently possible because of a lack of observational data for Hg dry deposition and Hg-speciated air concentrations. Previously, model intercomparison has been employed in Europe and in North America to gauge modeling uncertainty. Here, we have compared modeling results for Hg wet deposition to standardized observations over a range of sample integration times from weekly to annual. The regional models tested were able to resolve 50 to 70% of the observed site-to-site variation in annual Hg wet deposition. However, correlations to observation trended downward as the time scale of comparison was reduced to one week. The r^2 correlation for simulated weekly Hg wet deposition varied from 0.12 to 0.18, while the r^2 correlation for the simulated precipitation data used as input was 0.35, suggesting that considerable improvement in process modeling is possible given that all relevant processes are eventually identified and adequately described.

[25] An analysis of model agreement to all weekly observations of Hg wet deposition found REMSAD scored highest for r^2 correlation when averaged among the three test cases, but CMAQ scores were superior for all bias and error statistics. For normalized mean bias, CMAQ's average score was 0.10, with REMSAD and TEAM scoring 0.28 and 0.72, respectively. This compares to a normalized mean bias of 0.08 in the precipitation data used by all of the

Table 4. Model Performance Metrics From Table 3 When a Scaling Factor of 0.6 is Used to Adjust Simulated Hg Wet Deposition on the Basis of Precipitation Error[a]

	CMAQ			REMSAD			TEAM			
	CTM	GEOS-Chem	GRAHM	CTM	GEOS-Chem	GRAHM	CTM	GEOS-Chem	GRAHM	MM5 Precip.
r^2	0.32	0.26	0.30	0.28	0.25	0.27	0.28	0.24	**0.34**	0.87
Mean bias (ng m^{-2})	**−5.7**	55.7	47.5	76.0	110.5	48.6	181.4	240.6	170.0	0.8 (mm)
Mean error (ng m^{-2})	**149.7**	187.8	178.6	199.1	224.2	187.7	260.4	312.8	243.3	6.1 (mm)
Normalized mean bias	**−0.023**	0.221	0.189	0.302	0.439	0.193	0.722	0.957	0.676	0.031
Normalized mean error	**0.595**	0.747	0.710	0.792	0.892	0.746	1.036	1.244	0.967	0.248
Mean fractional error	**0.623**	0.695	0.678	0.762	0.794	0.754	0.833	0.885	0.813	0.315

[a]The most desirable outcomes for Hg wet deposition statistics are shown in bold.

models. For normalized mean error averaged among all test cases, CMAQ scored 0.79, REMSAD scored 0.91, and TEAM scored 1.15. This compares to a normalized mean error in the precipitation data of 0.62.

[26] When r^2 correlation, NMB and NME were analyzed for weekly Hg wet deposition at the individual observation sites, no obvious geographic patterns were found for any of the regional model simulations. However, there was evidence of superior NME and NMB scores at locations where the precipitation data was more accurate. Scaling the weekly Hg wet deposition simulated by each model at every observation site on the basis of the corresponding error in weekly precipitation resulted in significantly improved scores for r^2 correlation and normalized mean error. However, scores for normalized mean bias did not improve. The models generally had a positive bias in Hg wet deposition before this scaling was applied. Scaling caused normalized mean bias scores to increase because of the unbounded nature of upward adjustments to simulated wet deposition and the nonlinear relationship between wet deposition and precipitation. The best improvements in r^2 correlation and normalized mean error were generally obtained when a scaling factor of 50–70% was applied.

[27] When considering this analysis of simulated and observed wet deposition of Hg, it is important to remember that Hg is also deposited to aquatic ecosystems in the absence of precipitation. Any model could be matching observed wet deposition fluxes because it is simulating too much or too little dry deposition. As new technology for measuring Hg dry deposition is developed and deployed, our ability to better test these models will be greatly enhanced. Comparisons of atmospheric mercury models to observations on shorter time scales suggest that current modeling could be improved. More accurate information regarding the location and intensity of precipitation would certainly help.

[28] **Acknowledgments.** The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. It has been subjected to the Agency's administrative review and approved for publication. The work presented here does not necessarily reflect the policies or views of the other participating agencies.

References

Ariya, P., et al. (2004), Arctic: A sink for mercury, *Tellus, Ser. B*, *56*, 397–403.

Bullock, O. R., et al. (2008), The North American Mercury Model Intercomparison Study (NAMMIS): Study description and model-to-model comparisons, *J. Geophys. Res.*, *113*, D17310, doi:10.1029/2008JD009803.

Dastoor, A. P., and Y. Larocque (2004), Global circulation of atmospheric mercury: A modeling study, *Atmos. Environ.*, *38*, 147–161, doi:10.1016/j.atmosenv.2003.08.037.

Davis, J. M., and J. L. Swall (2006), An examination of the CMAQ simulations of the wet deposition of ammonium from a Bayesian perspective, *Atmos. Environ.*, *40*, 4562–4573, doi:10.1016/j.atmosenv.2006.04.007.

Gårdfeldt, K., and M. Jonsson (2003), Is bimolecular reduction of Hg(II) complexes possible in aqueous systems of environmental importance, *J. Phys. Chem. A*, *107*, 4478–4482, doi:10.1021/jp0275342.

Grell, G., J. Dudhia, and D. Stauffer (1994), A description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5), NCAR/TN-398+STR, 138 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.

Keeler, G. J., L. Gratz, and K. Al-Wali (2005), Influences on the long-term atmospheric mercury wet deposition at Underhill, Vermont, *Ecotoxicology*, *14*, 71–83, doi:10.1007/s10646-004-6260-3.

Lindberg, S., R. Bullock, R. Ebinghaus, D. Engstrom, X. Feng, W. Fitzgerald, N. Pirrone, E. Prestbo, and C. Seigneur (2007), A synthesis of progress and uncertainties in attributing the sources of mercury in deposition, *Ambio*, *36*, 19–32, doi:10.1579/0044-7447(2007)36[19:ASOPAU]2.0.CO;2.

Seigneur, C., P. Karamchandani, K. Lohman, K. Vijayaraghavan, and R.-L. Shia (2001), Multiscale modeling of the atmospheric fate and transport of mercury, *J. Geophys. Res.*, *106*(D21), 27,795–27,809, doi:10.1029/2000JD000273.

Selin, N. E., and D. J. Jacob (2008), Seasonal and spatial patterns of mercury wet deposition in the United States: Constraints on the contribution from North American anthropogenic sources, *Atmos. Environ.*, *42*, 5193–5204, doi:10.1016/j.atmosenv.2008.02.069.

Selin, N. E., D. J. Jacob, R. J. Park, R. M. Yantosca, S. Strode, L. Jaegle, and D. Jaffe (2007), Chemical cycling and deposition of atmospheric mercury: Global constraints from observations, *J. Geophys. Res.*, *112*, D02308, doi:10.1029/2006JD007450.

Shia, R. L., C. Seigneur, P. Pai, M. Ko, and N. D. Sze (1999), Global simulation of atmospheric mercury concentrations and deposition fluxes, *J. Geophys. Res.*, *104*(D19), 23,747–23,760, doi:10.1029/1999JD900354.

Vermette, S., S. Lindberg, and N. Bloom (1995), Field tests for a regional mercury deposition network: Sampling design and preliminary test results, *Atmos. Environ.*, *29*, 1247–1251, doi:10.1016/1352-2310(94)00321-B.

D. Atkinson, Office of Water, U.S. Environmental Protection Agency, Washington, DC 20460, USA.

T. Braverman, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA.

O. R. Bullock Jr., National Exposure Research Laboratory, U.S. Environmental Protection Agency, Mail Drop E243-03, Research Triangle Park, NC 27711, USA. (bullock.russell@epa.gov)

K. Civerolo, J.-Y. Ku, and G. Sistla, New York State Department of Environmental Conservation, Albany, NY 12233-0001, USA.

A. Dastoor and D. Davignon, Air Quality Research Division, Environment Canada, 2121 Autoroute Transcanadienne, Dorval, QC H9P 1J3, Canada.

K. Lohman, C. Seigneur, and K. Vijayaraghavan, Atmospheric and Environmental Research, Inc., San Ramon, CA 94583, USA.

T. C. Myers, ICF International, San Rafael, CA 94903, USA.

R. J. Park, School of Earth and Environmental Sciences, Seoul National University, Seoul 151-747, Korea.

N. E. Selin, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA.