7.6

MONTE CARLO ESTIMATES OF UNCERTAINTIES IN OUTPUTS OF REGIONAL OZONE MODELS DUE TO UNCERTAINTIES IN INPUTS TO THE BEIS3 BIOGENICS EMISSIONS MODEL

Steven R. Hanna*
Hanna Consultants, Kennebunkport, Maine

A.G. Russell and J. Wilkinson
Georgia Tech Research Corp., Atlanta, GA

J. Vukovich
Carolina Environmental Program, UNC, Chapel Hill, NC

D. A. Hansen
EPRI, Palo Alto, CA

1. INTRODUCTION

A Monte Carlo (MC) probabilistic approach is used to estimate uncertainties of the emissions outputs of the Biogenics Emissions Inventory System Version 3 (BEIS3) model (Pierce, 2001) and subsequent predictions by three chemical transport models (CTMs) due to uncertainties in BEIS3 model parameters and input variables. BEIS3 has been developed by the U.S. EPA to estimate emissions of isoprene, monoterpenes, and oxygenated volatile organic compounds (OVOCs) due to biological activity in or on plant tissues, and to estimate emissions of biogenic nitric oxide (BNO) due to biological activity in soils. Outputs of CTMs, based on the BEIS3 emissions estimates, are then used to set policies concerning emission reductions needed from industrial plants and other man-made sources. Biogenic VOC emissions are estimated to be of the same order of magnitude as man-made VOC emissions in many parts of the U.S. The results briefly described in this paper are discussed in detail in two reports (Hanna *et al*., 2002 and 2003).

As is the case with most environmental model parameterizations, those in BEIS3 are based on limited observations over a narrow range of conditions. Uncertainties are likely to grow for geographic regions and for combinations of weather conditions and vegetation conditions outside of the central range of conditions used in model derivation. For this reason, this BEIS3 uncertainty study covers a range of seasons and geographic locations. The study uses three episodes (24-29 May, 11-15 July and 4-9 September 1995) that have been extensively investigated using several CTMs. The geographic domain for all three time periods covers most of the Eastern U.S. and parts of the Midwest. Input files for these three episodes are well-established as a result of previous studies by the EPA and others, and are used to define the median inputs for this project.

There has been a rapid growth in the use of MC probabilistic uncertainty analysis. Cullen and Frey (1999) describe a framework for uncertainty analysis and give examples of the MC approach applied to various power plant and air quality risk consequence issues. In previous MC uncertainty projects applied to regional ozone problems (e.g., Hanna *et al*., 1998 and 2001), the effects of uncertainties in biogenic emissions on the predictions of CTMs were included in the comprehensive uncertainty analysis. For example, Hanna *et al*. (1998) showed that uncertainties in biogenic VOC emissions had one of the largest effects on uncertainties in ozone predictions by the CTM, UAM-IV, although the biogenic VOC emissions were not broken down into components and there was no dependency on inputs such as LAI, PAR, temperature, or specific model formulation parameters. As another example, Hanna *et al*., (2001) focused on the CTM, UAM-V, and found that uncertainties in the ozone predictions were most strongly correlated with uncertainties in the NO_2 photolysis rate. Also important were wind speed and direction, relative humidity, cloud cover, and biogenic VOC emissions. Bergin et al. (1999) applied MC methods with Latin Hypercube Sampling (LHS) to a Lagrangian photochemical air pollution model in Southern California. They accounted for meteorological variability by using several solutions of a mass-consistent wind model, run with random data-withholding assumptions, and found that the variability in predicted ozone was about ± 30 to 50%.

2. APPROACH

Numerous reports and journal articles (e.g., Lamb *et al.*,1999, and Guenther *et al*., 1993, 1999 and 2000) have been written on the biogenic VOC and biogenic NO model formulations that are in BEIS3 (Pierce, 2001).

Short-term variations in isoprene emissions from vegetative species are influenced by leaf temperature

---

Corresponding author address: Steven R. Hanna, 7 Crescent Ave., Kennebunkport, ME 04046; e-mail: hannaconsult@adelphia.net

and by the amount of PAR (photosynthetically active radiation) reaching the leaves. In BEIS3, Pierce (2001) uses a slightly modified form of the Guenther *et al.* (1993) formulations to estimate isoprene emissions based on leaf temperature and PAR:

$$E = x \cdot E_s \cdot C_L^A \cdot C_T \cdot A \qquad (1)$$

where E (ì g· hr^{-1}) is the total emission rate of isoprene from area A; *î* (dimensionless) is the seasonal adjustment coefficient, which is 1.0 for the spring, summer, and fall periods studied in this paper; E_s (ì g· ha^{-1}· hr^{-1}) is the plant species-specific emissions flux at 30ºC and 1000 ì mol· m^{-2}· s^{-1} PAR (i.e., the biogenic emissions factor); C_L^A (dimensionless) is the canopy-adjusted PAR, which accounts for changes in PAR as attenuated by the leaves; C_T (dimensionless) is the temperature correction factor that accounts for changes in leaf temperature; and A (ha) is the areal extent of the plant species in the location modeled.

Canopy Adjusted PAR. C_L^A (dimensionless) is computed via a canopy model that accounts for the effects of variations of PAR with height in the leaf canopy. The canopy model in BEIS3 is based on a leaf energy balance and knowledge of the Leaf Area Index, LAI. C_L^A, or the canopy adjusted PAR, for the sunlit and shaded leaves, is calculated based on parameterizations of L_S^f (dimensionless), which is the fraction of sunlit leaves; L_D^f (dimensionless), which is the fraction of shaded leaves; C_L, which is the light correction factor; PAR_D (ì mol· m^{-2}· s^{-1}), which is the amount of PAR on the shaded leaves; and PAR_S (ì mol· m^{-2}· s^{-1}), which is the amount of PAR on the sunlit leaves.

The light correction factor, C_L (dimensionless) is calculated as:

$$C_L = \frac{a \cdot c_{L1} \cdot L}{\sqrt{1 + a^2 \cdot L^2}} \qquad (2)$$

where *á* = 0.0027 ì mol^{-1}· m^{2}· s and c_{L1} = 1.06 (dimensionless) are empirical coefficients; and L is the PAR flux rate (ì mol· m^{-2}· s^{-1}). Variations in *á* and c_{L1} are included in our MC uncertainty study. L (i.e., PAR) is not directly sampled in our MC study; instead the total incoming solar radiation, *I,* is sampled. The variation of C_L with L (i.e., PAR) is based on values of *á* = 0.0027 ì mol^{-1}· m^{2}· s and c_{L1} = 1.06 (Guenther *et al.* 1993).

The current MC uncertainty analysis uses I (W/m^2) as a fundamental parameter for resampling. Additional equations in BEIS3 concerning the partitioning of solar radiation can be found in Pierce (2001) and Hanna *et al*. (2002).

Temperature Correction Factor C_T. To estimate the temperature correction factor, C_T, BEIS3 follows Guenther *et al.* (1993):

$$C_T = \frac{\exp\left(\dfrac{c_{T1} \cdot (T - T_S)}{R \cdot T_S \cdot T}\right)}{1 + \exp\left(\dfrac{c_{T2} \cdot (T - T_M)}{R \cdot T_S \cdot T}\right)} \qquad (3)$$

where c_{T1} = 95000 J· mol^{-1}, c_{T2} = 230000 J· mol^{-1}, and T_M = 314 K are empirical coefficients which are varied in our MC uncertainty study; R is the ideal gas constant (8.314 J· K^{-1}· mol^{-1}); T_S is an empirical normalizing temperature (303 K); and T is the leaf temperature (K), which is assumed to be the ambient temperature (K).

Monoterpene emissions occur from coniferous species and some deciduous species. BEIS3 follows Guenther *et al*. (1993) to estimate monoterpene emissions, E (µg/hr):

$$E = x \cdot E_s \cdot A \cdot C_T \qquad (4)$$

x (dimensionless) is the seasonal adjustment factor, which equals 1.0 for the three episodes studied. E_S (µg/ha/hr) is the species-specific emissions flux at 303 K (30 C), and is varied in our MC uncertainty analysis. The temperature correction factor, C_T, is dimensionless and is estimated based on leaf temperature, which is assumed to equal ambient temperature, T:

$$C_T = \exp(b\,[T - T_S]) \qquad (5)$$

where *â* = 0.09 K^{-1} is the empirical coefficient that can also be thought of as an inverse temperature scale. The coefficient *â* is one of the parameters varied in our MC study. T_S, is assumed to equal 303 K and is not varied in the MC study. Unlike the isoprene temperature correction factor, the monoterpene emissions temperature correction factor, which is also the OVOC temperature correction factor, monotonically increases with increasing temperature. The exponential form of Equation (5) is based purely on a statistical fit of measured emissions rates. Similar exponential forms are used for the C_T term for isoprene and BNO emissions.

It is well known that a wide variety of oxygenated and other BVOCs, which are collectively named OVOC, are emitted from vegetation. In BEIS3, Pierce (2001) uses the monoterpene formulas (Equations 4 and 5) to estimate OVOC emissions, although E_S is different for OVOCs.

Biogenic nitric oxide (BNO) is emitted as a result of microbial nitrification-denitrification activities in soil and is enhanced through nitrogen-based fertilizer application (Williams *et al*., 1992), stubble burning, and soil tilling. Soil NO emissions factors range over two orders magnitudes or more. In general, wetlands and tundra have very low soil NO emissions, forests have moderate soil NO emissions, and agricultural and grasslands have the highest soil NO emission rates. BEIS3 uses the empirical model of Williams et al. (1992) to model BNO emissions, E (µg/hr), from soils:

$$E = E_s \cdot A \cdot C_T \tag{6}$$

where E_S (µg/ha/hr) is the species-specific emissions flux at 303 K (30 C), and is varied in our MC uncertainty analysis; and C_T (dimensionless) is the temperature correction factor that is estimated based on soil temperature, T_{soil}:

$$C_T = \exp\left(T_3\left[T_{soil} - 30\right]\right) \tag{7}$$

where T_3 = 0.071°C^{-1} is an empirical scaling parameter that describes the rate of increase of BNO emissions with soil temperature; 30 C is a constant identical to the temperature scale T_S; and T_{soil} (C) is the parameterized soil temperature which is determined as follows:

$$T_{soil} = T_1\left(T - 273.15K\right) + T_2 \tag{8}$$

where T_1 (dimensionless) = 0.72 and T_2 = 5.8 C are empirical parameters that relate soil temperature to ambient temperature; T (C) is the leaf temperature which is taken to be ambient temperature. The reader should note that the units of temperature switch back and forth from C to K in these equations depending on the particular convention followed in the sets of papers that derive the empirical relations.

Seventeen BEIS3 model parameters and model data input variables have been assumed to vary and have been sampled in the current Monte Carlo (MC) uncertainty study. The model parameters refer to the various coefficients and integral scales that are not input by users but whose values are part of the BEIS3 code. Each MC sample of the variation of one of these model parameters is assumed to be uniform in time (over the entire episode) and space (over the entire geographic domain). The model data input variables refer to the required user inputs to BEIS3. Each MC sample of the variation of ambient temperature, *T*, and of total incoming solar energy, *I*, is assumed to vary in time and space. However, each MC sample of the variation of plant species-specific model data inputs (e.g., LAI and biogenics emissions factors, E_S) is assumed to be uniform in time and space.

Table 1 summarizes the seventeen BEIS3 model parameters and data input variables whose random variations or uncertainties were sampled. Each parameter or input variable is listed plus the following attributes: its mean value; its uncertainty range; the type of distribution from which MC samples were drawn; an indicator as to the parameter's variability with respect to time, space, and plant species; and the chemical species that the parameter affects.

The Simple Random Sampling (SRS) technique (Morgan and Henrion, 1990) is used to draw one thousand random and independent samples from the assigned distributions of each model parameter listed in Table 1. Although the sampling software can accommodate correlations among variables, insufficient data were available to allow estimation of the correlation matrix. Therefore, in this analysis, the model parameters are assumed to be independent.

Because the uncertainty in CTM predictions due to uncertainties in biogenic emissions may be dependent upon the period(s) chosen for simulation, a suite of three episodes is used, representing different meteorological conditions. The correlations among MC-selected BEIS3 input parameters CTM outputs, are preserved, since each MC run with a CTM uses a set of BEIS3 outputs from a MC run carried out in the earlier phase of the study. The 20 MC runs of each CTM were made by selecting 20 emissions sets from the available sets of BEIS3 MC outputs.

Three CTMs are used in this study: 1) The Multiscale Air Quality Simulation Platform (MAQSIP) (Odman and Ingram 1996, MCNC 1997); 2) The Variable Grid Urban Airshed Model (UAM-V) version 1.30 (SAI, 1999); and 3) The Urban-to-Regional Multiscale (URM) model (Kumar and Russell, 1996; Boylan *et al.,* 2002). A variety of simulation periods were available, representing episodes from April to September, including the July 1995 period. MAQSIP runs are available for the entire period, and URM and UAM-V runs are available for July. Model inputs and results for each of those have been thoroughly evaluated. Table 2 contains the dates and other details for the three episodes that were selected.

Since MAQSIP is applied to all three episodes, the differences in the uncertainty of that CTM over several time periods could be assessed. Since three CTMs are applied to the same July episode, the differences in the uncertainties across the CTMs could be assessed. A map of the geographic domain is given in Figure 1, which also includes the locations of ozone monitoring stations, known as Aerometric Information Retrieval System (AIRS) sites, used for part of the analysis of the MC outputs. Because of the very large size of the domain seen in Figure 1, the analysis has considered four subdomains: Midwest, Northeast, Southeast, and Texas-Gulf Coast. If this division is not made, it would be possible for the predicted domain-wide maximum concentration to occur in Maine and the observed maximum concentration to occur in Texas.

Table 1. Summary of uncertain BEIS3 parameters used in Monte Carlo study.

Parameter	Mean	Uncertainty (1 s)	Distribution	Variable[(1)]	:hemical Species[(2]
α	0.0027 $\mu mol \cdot m^{-2} \cdot s^{-1}$	0.0015 $\mu mol \cdot m^{-2} \cdot s^{-1}$	Lognormal	SI, TI, PI	I
c_{L1}	1.06	0.2	Normal	SI, TI, PI	I
c_{T1}	95,000 $J \cdot mol^{-1}$	20,000 $J \cdot mol^{-1}$	Lognormal	SI, TI, PI	I
c_{T2}	230,000 $J \cdot mol^{-1}$	150,000 $J \cdot mol^{-1}$	Lognormal	SI, TI, PI	I
T_M	314 K	3 K	Normal	SI, TI, PI	I
β	0.09 K^{-1}	0.02 K^{-1}	Lognormal	SI, TI, PI	T, O
T_3	0.071 C^{-1}	0.007 C^{-1}	Normal	SI, TI, PI	N
T_2	5.8 C	2.9 C	Lognormal	SI, TI, P	N
T_1	0.72	0.36	Lognormal	SI, TI, P	N
E_S	Mean value is plant-species and pollutant dependent	$\pm$25%	Normal	SI, TI, P	I, T, O, N
LAI (summer)	Mean value depends on plant species	$\pm$12.5%	Normal	SI, TI, P	I
LAI (Sp/F/W)	Mean value depends on plant species	$\pm$1 unit	Normal	SI, TI, P	I
T	Mean value is based on local grid obs	1.9 K	Normal	S, T, PI	I, T, O, N
I	Mean value is based on local grid obs	$\pm$12.5%	Normal	S, T, PI	I

[1]SI is space invariant; TI is time invariant; S is space variant; T is time variant; PI is plant-species invariant; and P is plant-species variant. [2]I is isoprene; T is monoterpenes; O is OVOCs; and N is biogenic nitric oxide.

Table 2. Episodes used in MC analysis. Dates, CTM names, project references, and grid resolutions are listed.

Episode	CTM	Project	Grid Resolution
May 24-29, 1995	MAQSIP[1]	SMRAQ-SESARM[4]	36 km
July 11-15, 1995	UAM-V[2]	EPRI MC[5]	12 km
	MAQSIP[1]	SMRAQ-SESARM[4]	36 km
	URM[3]	SAMI[3]	12-24-48-96-192 km
Sept. 4-9, 1995	MAQSIP[1]	SMRAQ-SESARM[4]	36 km

[1]Odman and Ingram, 1996; [2]SAI, 1999; [3]Boylan *et al*., 2002, [4]SMRAQ 1997, Kasibhatla and Chameides, 2000; [5]Hanna *et al*., 2001

3. RESULTS

The results of the MC uncertainty analysis are presented below for the BEIS3 emissions model in Section 3.1 and for the CTMs in Section 3.2.

3.1 Emissions results

The prescribed uncertainties for BEIS3 model parameters and input variables listed in Table 1 were used to generate 1000 Monte Carlo (MC) simulations for three separate time periods (11-15 July 1995, 4-9 September 1995, and 24-29 May 1995). The results of the comprehensive analysis are described by Hanna *et al.* (2002), and include Cumulative Distribution Functions (CDFs) of daily domain-wide emissions. Plots are presented containing gridded (36 km) mean daily emissions, coefficients of variation (CVs), and correlation coefficients between the model parameters and daily emissions estimates, by pollutant. For the purposes of this paper, detailed results are given only for one day of one episode (11 July 1995). Results for the other days are briefly summarized and similarities and differences with 11 July 1995 are described.

11-15 July 1995 Episode

Figure 2a shows the mean daily, gridded (36 km) temperature (*T*) field, and Figure 2b shows the mean daily, gridded PAR field that result from taking the average of the 1000 MC samples at each grid location for 11 July 1995. Although total solar radiation (*I*) was the parameter that was sampled, the following discussion refers to the analyses in terms of PAR. By directly varying total solar radiation, PAR is indirectly varied as was discussed in Section 2. Figure 2a demonstrates that much of the eastern United States experienced warm weather with temperatures in excess of 25 C. On 11 July, the upper Midwest United States, Canada, and the Northeast United States experienced cooler weather with temperatures below 21 C and in some cases as low as 11 C. This trend persisted, in particular over Lake Superior, for the duration of the 11-15 July 1995 episode. Figure 2b shows that, on 11 July, mean daily PAR values were basically saturated (i.e., the mean daily PAR is greater than 90 $W \cdot m^{-2}$) over much of the land-based domain. Note that, when PAR is saturated, its variations have little effect on biogenic emissions. On the last two days of the episode, PAR values were much smaller due to the development of clouds.

<u>Analysis of Total Uncertainty</u> - CDFs were constructed by summing the daily BEIS3 emissions estimates by pollutant across the domain for each MC sample. The lognormal distribution provided a best fit to most of the CDFs. The fundamental conclusions concerning the uncertainties and the CDFs for the total domain-wide daily emissions over the domain are:

- Total domain-wide daily isoprene emissions estimates have a 95% uncertainty range of almost one order of magnitude;
- Total domain-wide daily BNO emissions estimates have a 95% uncertainty range of over one order of magnitude; and
- Total domain-wide daily OVOC and monoterpene emissions estimates have a 95% uncertainty range that is more tightly distributed within about 15% of the mean.

Figure 3 shows, for July 11, the gridded (36 km) plots of the mean daily isoprene emissions (Figure 3a), and the Coefficient of Variation (CV), or standard deviation divided by the mean of the daily isoprene emissions (Figure 3b) for the 1000 MC samples. Consistent with previous findings (e.g., Guenther *et al*., 2000), maximum isoprene emissions are seen to occur in the southern United States. In the latter part of the 11-15 July episode, high isoprene emissions are also predicted over parts of the northeastern states, due to increasing temperature and PAR. The CV in Figure 3b is seen to be about 0.5 in the southern U.S., where emissions are highest. High CVs, in excess of about 0.6, are found over much of the northern part of the domain for the duration of the episode, and are due to relatively low predicted isoprene emissions coupled with relatively high standard deviations. High CVs are also observed over Texas and Louisiana in the latter part of the episode when both temperature and PAR decrease, and predicted isoprene emissions decrease.

Maximum daily monoterpene emissions on July 11 are predicted to occur in the coniferous forests of the southern United States and northern New England, consistent with previous findings (e.g., Guenther *et al*., 2000),. High monoterpene emissions are also predicted in the coniferous forests of upper Minnesota and upper Wisconsin throughout much of the episode. The monoterpene CVs are generally lower than those for isoprene, with typical values of about 0.2 and with no CV exceeding 0.42. The predicted OVOC emissions on July 11 are more uniformly distributed throughout the domain than the predicted monoterpene emissions. This is because OVOC emissions are suspected to be ubiquitous to plant species (Lamb *et al*., 1999). As with the monoterpene CVs, the OVOC CVs are generally relatively low with values of about 0.1 in areas with maximum OVOC emissions. These results for monoterpenes and OVOCs are consistent with the finding that the total uncertainty in the OVOC emissions is relatively low compared to the total uncertainty that exist in the isoprene emissions estimates.

The highest BNO emissions on July 11 are predicted to be in the farm belt of the United States. Because large uncertainties exist in BNO emissions, relatively large CVs in excess of 1.0 are found, with values of about 2 in the upper Midwest. A large tail (i.e., at extreme values) in the distribution of BNO emissions is the major cause of the large uncertainties. The BNO CVs are less than 1.0 only where temperatures are predicted to be low (i.e., the Northeast United States and Canada on the first two days of the episode).

Analysis of Correlations between Variations in Inputs and Variations in Output Emissions - To investigate the relations between variations in BEIS3 inputs (seeTable 1) and variations in outputs of emissions for the 1000 MC runs, pair-wise Pearson linear correlation coefficients were computed. In order for it to be concluded, with 95% confidence, that a calculated correlation coefficient is significantly different from 0.0, its magnitude should exceed about 0.06.

The following discussions describe a few of the input variables' correlation coefficients for 11 July 1995. Hanna *et al*. (2002) present dozens of correlation plots for all inputs and episodes. For example, on 11 July, the correlations between variations in c_{T1} and variations in the estimated isoprene emissions are weakly correlated with magnitudes less than 0.2. It is interesting to note that c_{T1} has negative correlations (about –0.3) in New England, and positive correlations (about 0.5) in Kansas and Oklahoma. An analysis by Hanna (2004) confirms that the relative contributions of the c_{T1} term to the total uncertainty vary strongly with ambient temperature and therefore will vary strongly with geographic region. In contrast, *á* is correlated to the isoprene emissions estimates throughout the domain with correlation coefficients in excess of 0.7. *á* strongly impacts the PAR attenuation. Correlations between variations in PAR and variations in daily isoprene emissions are small, probably because isoprene emissions are very insensitive to increasing PAR after PAR reaches its saturation point (about 1000 ì mol· m^{-2}· s^{-1}).

The correlation between variations in C_T and variations in the monoterpene emissions varies with geographic position. There are strong correlations (about 0.7) throughout much of the Midwest and Northeast United States and Canada and weaker correlations (0.0 to 0.3) in the South and the Midwest. The correlation patterns for *â* are very similar to those for C_T, except the signs are switched. In general, variations in mean daily temperature, T, show relatively weak correlation with the variations in monoterpene emissions. Since C_T depends on both *â* and T, this result implies that variations in *â* are controlling variations in C_T. The magnitudes and spatial distributions of the correlations of variations of C_T, T, *â*, and E_S with variations of OVOC emissions estimates follow patterns that are similar to those for monoterpene emissions.

Correlation coefficients were also calculated between variations in model parameters and inputs and variations in BEIS3 estimates of BNO emissions. Variations in C_T show a strong correlation with variations in BNO emissions over the domain. A breakdown of the correlation shows that variations in T_1 are the most important contributor to the correlation of variations in C_T with BNO emissions. Variations of T_2, T_3, and the mean daily temperature show much weaker correlations with variations in BNO emissions. Further, the variations in the area weighted BNO emissions factor, E_S, are weakly correlated with variations in BNO emissions.

4-9 September 1995 Episode

The analysis of the 4-9 September 1995 time period follows the procedures for 11-15 July 1995. However, because the results are so similar, they are only briefly summarized here. Much of the eastern United States experienced typical late summer conditions with temperatures in the low twenties and high teens (degrees C) at the start of the episode with the exception of Oklahoma and Texas where temperatures were much warmer. As the episode progressed, there was a cooling trend throughout much of the Midwest and Northeast United States as a cold front passed through from the north-northwest. The South remained relatively warm for the duration of the episode with temperatures in the mid- to upper-twenties (degrees C).

CDFs were generated for the domain-wide daily isoprene, OVOC, monoterpene, and BNO emissions in a manner similar to those generated for the 11-15 July 1995 episode. The fundamental conclusion for the 4-9 September 1995 episode is that the uncertainty ranges are similar to those for the 11-15 July 1995 period. The BNO range is slightly less and the OVOCs and monoterpenes ranges are slightly larger in September than in July. The main difference is that the effects of the cold front can be seen on emissions behind the front for the 4-9 September period.

24-29 May 1995 Episode

During the 24-29 May 1995 episode, daily average temperatures of about 28 C occur in the Deep South, similar to conditions in the July and September episodes. However, the daily averaged temperatures in the May episode are much cooler (i.e., less than 16 C) north of a stationary front that extends from the panhandles of Oklahoma and Texas through New York. The stationary cold front persisted for the duration of the episode, dipping somewhat to the south at the end of the period. CDFs were generated for domain-wide emissions, verifying that the uncertainty ranges for the May episode are comparable to the uncertainty ranges for the July and September episodes. The exception is that the uncertainty range for the BNO emissions distribution in May is about 30% less than that for July or September, possibly attributable to the low temperatures over the Midwest, where BNO emissions are typically the largest.

Because the mean BNO emissions in May were much lower in Iowa and other farm-belt areas north of the stationary front, the correlation patterns shifted such that the variations in BNO emissions became nearly completely determined by variations in E_S. Variations of C_T have less of an effect at low temperatures.

3.2 Results of uncertainties in CTM outputs

Total uncertainty

As described earlier, one component of the analysis of the MC results is focused on the total uncertainty in the CTM predictions due to uncertainties in the BEIS3 emissions estimates. Another component is focused on the identification of the emissions uncertainties that are chiefly responsible for the total uncertainty through correlation analysis. To shorten the discussions, only one model (MAQSIP) and one episode (July 1995) are presented in the tables. Furthermore, a figure for only one day is presented. Figure 4 contains the base run predictions for 11 July 1995 time period for MAQSIP. Hanna *et al*. (2003) discuss in detail the results for all episode days for all CTMs.

Table 3 summarizes the findings concerning the uncertainties in predicted unpaired peak ozone (1 and 8 hour average) for each subdomain (Midwest, Northeast, Southeast, Texas-Gulf Coast). The table contains the observed peak ozone concentration and its location, the predicted peak ozone concentration in the base case and its location, the minimum and maximum predictions as derived from the twenty MC runs of MAQSIP, the normalized range, and the peak estimation accuracy statistic for each subdomain and averaging time for the July 1995 episode. The predicted peak ozone concentration for the so-called "base case run" uses the model run with the median values of MAQSIP inputs and parameters. Note that the peak predicted concentration is usually not paired in time or space with the observed value. The only restriction is that the predicted peak must occur on the same day and within the same domain (e.g., Northeast or Midwest) as the observed peak. Also note that the observed peak must naturally occur at an AIRS monitoring location (about 100 to 200 in each domain), while the predicted concentration can occur anywhere on the domain, usually at grid squares where there are no AIRS monitors. In about half of the cases, the predicted and observed peaks are located very near to each other. It is seen that, for the July episode, MAQSIP predicts the peak ozone concentration to be over Lake Michigan (for the Midwest domain) and in the Atlantic Ocean south of Long Island and east of New Jersey (for the Northeast domain). In July, Lake Michigan and the Atlantic Ocean are colder than the air being advected over the water, causing a stable boundary layer to be simulated by the model, with minimal dispersion.

A measure of model performance is whether the range of the 20 MC-predicted peak ozone concentrations encompasses the value of the observed peak ozone concentration. To be strictly correct, the observed and predicted values should be from the same grid square and the same time. We make the assumption that the distribution of predicted peak ozone concentrations, unpaired in space and time at various grids across the domain, is similar to the distribution of predicted peak ozone concentrations at the location of the AIRS observation. If the observed ozone concentration is in the minimum to maximum range of the ozone predictions, then it can be concluded that the observed value could be a member of the population of predicted concentrations. For the data in Table 3, for MAQSIP for the July episode, 75 % of the observations are within the max-min range. Hanna et al. (2003) show that this percentage is about the same for MAQSIP for the other two episodes. However, the percentage drops to 15% for the median over all three models for the July episode, since there is a moderate overprediction bias from UAM-V and an underprediction bias from URM.

A measure of the relative scatter or variability of the MC predictions is also given for MAQSIP for the July episode in Table 3, where the Normalized Range (i.e., [max-min]/base*100) is listed for the predicted peak ozone concentrations. The overall median value of the Normalized Range is 20%. Hanna et al. (2003) show that there is no major variation of the Normalized Range with subdomain, model, episode, or averaging time. The only clear dependency in Table 3 is that the relative variability tends to be larger for the subdomains (Midwest and Northeast) where the predicted maximum occurs over the water.

It should be noted that, since the AIRS monitors tend to be sited in NOx-limited regions, it is anticipated that the total uncertainty range in the CTM ozone predictions at those sites due to uncertainties in BEIS3 inputs would be less than elsewhere on the domain.

The peak estimation accuracy, A_{ts}, is a standard EPA performance measure, and has been calculated for each model and each MC run:

$$A_{ts} = \langle (C_p - C_o)/C_o \rangle \qquad (9)$$

where < > indicates an average over all monitors and all time. C_p and C_o are the peak predicted and observed at a given location and time. The standard guidance is that the magnitude of A_{ts} should be less than 20 % for episodes where the one hour ozone NAAQS is being studied. The final three columns in Table 3 list the values of A_{ts} for the set of maxima and the set of minima of the 20 MC runs, and for the base run, respectively. For the base run results in Table 3, the criterion that A_{ts} should be less than $\pm$20% is met 50% of the time for the one hour and eight hour standards for MAQSIP for the July episode, with the discrepancy primarily due to the previously mentioned comparisons in the Midwest and Northeast domains. For the "minimum" values of the 20 MC runs, the $\pm$20% criterion is met about 62% of the time for the one hour and eight hour standards, which is expected since the slight overprediction bias of the base runs would be mollified. Also as expected, for the "maximum" values of the 20 MC runs, the $\pm$20% criterion is met only about 38% of the time, due to an exacerbation of the overprediction tendency.

Table 3. Analysis of predicted and observed peak ozone concentrations for MAQSIP for the July episode. The location and the magnitude of the peak concentration within the domain are given. Domains are Midwest (MW), Northeast (NE), Southeast (SE), and Texas-Gulf Coast (TG).

T_a	Domain	Observed Peak Location	Predicted Peak Location	Obs Peak (ppb)	Predicted Peak (ppb)			Normalized Range (%)	Peak Est. Acc % (A_{ts})		
					Base	Max	Min		Max	Min	Base
8 hr	MW	Benton Harbor, MI	Lake MI just W of Benton Harbor	163	223	289	155	60	77.3	-4.9	36.8
1 hr	MW	Benton Harbor, MI	Lake MI just W of Benton Harbor	178	287	393	206	65	120.8	15.7	61.2
8 hr	NE	Baltimore	In Ocean E of NJ and S of Long Island	152	261	287	209	30	88.8	37.5	71.7
1 hr	NE	Philadelphia	In Ocean E of NJ and S of Long Island	184	291	312	245	23	69.6	33.2	58.2
8 hr	SE	Atlanta	Knoxville	133	155	167	141	17	25.6	6.0	16.5
1 hr	SE	Atlanta	Knoxville	166	165	174	152	13	4.8	-8.4	-0.6
8 hr	TG	Houston	LA coast by TX	121	129	138	122	12	14.0	0.8	6.6
1 hr	TG	Houston	Biloxi, Memphis, Houston	171	150	158	133	17	-7.6	-22.2	-12.3

Evaluations of Whether the Range of the 20 Monte Carlo Predictions Include the AIRS Observation

The 20 MC predictions of ozone concentration at every AIRS monitor have been analyzed for every CTM, subdomain, time period, and averaging time. If the 20 MC runs represent the range of uncertainty in model predictions, and the observed concentration is within the bounds of the 20 MC runs, then it can be concluded that the CTM is performing adequately. Therefore, the maximum and minimum ozone predictions of the twenty MC runs can be considered to represent the 95% confidence bounds. It is found that, over all three episodes, MAQSIP overpredicts ozone by a median value of 21% with a range from +6.6% to +65%. UAM-V (July only) underpredicts ozone by a small amount (median value of 4%) with a range from -13% to +17%. URM (July only) underpredicts ozone by a median value of 37% with a range from -50% to -19%. In aggregate, the models slightly overpredict ozone by a median value of 11% with a range from –50% to +65%.The overall "percent in range" of AIRS observations within the minimum to maximum C_p range of the 20 MC predictions has a median of 11% with a range from 1% to 47%. This percentage depends on (1) the mean bias (if the model has a large mean bias, the observation has little chance of being within the MC range of predictions), and (2) the range of the 20 MC predictions (the smaller the range, the less the chance the range has of capturing the observation).

The difference of the maximum and minimum ozone predictions divided by the minimum of the 20 MC runs, has a median of 13% with a range from 5% to 27%, again with not much dependence on model, domain, episode, or averaging time. This is a major finding of the study – the uncertainties in BEIS3 inputs are causing about a 13% uncertainty in CTM predictions, with little variation with external variables.

Correlation analysis for CTMs

As described in Section 2, a component of the analysis of the MC results is focused on the correlations that exist between the variations in the BEIS3 input parameters and the variations in the CTM-predicted one and eight-hour averaged ozone concentrations. Because there were 20 MC runs for each CTM, the correlations were made for 20 pairs of numbers. For 20 pairs of numbers, statistical confidence limits tests suggest that the correlations with magnitudes less than 0.44 are not significantly different from zero, at the 95% confidence level. The results of the CTM correlation analysis are not as conclusive as the results on the total uncertainties. This lack of correlation is perhaps caused by the fact that the total relative scatter or uncertainties in hourly averaged predicted ozone concentrations due to uncertainties in BEIS3 input parameters is only about 13%. The correlations between variations in individual BEIS3 input parameters and ozone predictions are generally small and non-significant. It is concluded that there are a few significant correlations, but the number of MC runs (N = 20) was not large enough to allow confidence to be attached to these results.

5. CONCLUSIONS

The plots of Monte Carlo (MC)-estimated uncertainties in the total daily emissions of isopreme, monoterpenes, OVOCs and BNO over the geographic domain allowed some general conclusions to be reached. For example, it was found that the estimated total uncertainties were nearly the same over the three time periods. The 95 % confidence ranges on the calculated isoprene, monoterpene, OVOC, and BNO emissions covered ± factor of 10, ± 15%, ± 15%, and ± factor of 10, respectively. The only major difference between the episodes was seen for BNO for the May 1995 time period, when there was cold air north of a stationary front from Texas to New York and the BNO emissions were relatively low in Iowa and other parts of the farm belt where BNO emissions usually reach their maximum.

Correlations were calculated between the 1000 MC samples of pairs of variations in model parameters or inputs and variations in emissions estimates, with the result that there were some correlations greater than about 0.5 for some of the assumed internal model parameters, such as a, but that the correlations were small for meteorological inputs such as the ambient temperature and the photosynthetically active radiation (PAR). Variations in the model parameters that affect PAR attenuation (i.e., C_L, a, and c_{L1}) were strongly correlated (0.7 to 1.0) with variations in the BEIS3 estimates of daily isoprene emissions.

The correlation patterns and conclusions for isoprene, monoterpenes, and OVOCs were similar for the three time periods studied. Variations in the monoterpenes and OVOCs emissions factors, E_S, had the strongest correlations with variations in the predicted daily monoterpenes and OVOCs emissions over the warmer part of the geographic domain. In the cooler parts of the geographic domain, the variations in the temperature correction factor, C_T, had a relatively high correlation with variations in the predicted daily OVOCs and monoterpenes emissions. These effects could be anticipated from the exponential form of the single emissions equation. For all the time periods with one exception, the fluctuations in one of the BNO input parameters, T_1, had the strongest correlation (0.7) with the BEIS3 predicted fluctuations in daily BNO emissions estimates.

The analyses of MC-estimated uncertainties in the CTM-predicted 1 and 8-hour averaged ozone concentrations over the geographic subdomains and at AIRS monitoring sites showed that the total uncertainties are nearly the same for the three CTMs over the three time periods and for the 1-hour and 8-hour averages. The 95% confidence ranges on the calculated ozone concentrations cover approximately 15 to 20% for all combinations of models, geographic domains, time periods, and averaging times. Focusing on the peak ozone concentration in each subdomain, about 25% of the observed peaks are within the uncertainty range of the 20 MC predictions. On the other hand, when the observed and predicted ozone concentrations at each AIRS monitor are compared for the 20 MC runs, about 12% of the observations are within the uncertainty range of the predictions.

As mentioned above, because the AIRS monitors were deliberately placed in NOx-limited regions, it is expected that the MC-estimated uncertainty ranges due to uncertainties in BEIS3 inputs would be less than the ranges elsewhere on the domain.

Some correlations calculated between the 20 MC samples of pairs of variations in BEIS3 model parameters or inputs and variations in ozone estimates were greater than about 0.5 for some of the inputs. However the results were just barely significant, suggesting that it would have helped the significance issue if we had carried out more than 20 MC runs.

Acknowledgements: This research was sponsored by EPRI under contract EP-P5243/C2592. The authors acknowledge guidance received from the BEIS3 model developer, Mr. Thomas Pierce of the U.S. Environmental Protection Agency. The authors also appreciate the detailed suggestions and review comments received from Prof. H. Christopher Frey of North Carolina State University.

REFERENCES

Bergin, M.S., G.S. Noblet, K. Petrini, J.R. Dhieux, J.B. Milford and R.A. Harley, 1999: Formal uncertainty analysis of a Lagrangian photochemical air pollution model, *Env. Sci. Tech., 33*, 1116-1126.

Boylan, J.W., M.T. Odman, J.G. Wilkinson, A.G. Russell, K.G. Doty, W.B. Norris and R.T. McNider, 2002: Development of a comprehensive, multiscale "one-atmosphere" modeling system: application to the Southern Appalachian Mountains, *Atmos. Environ., 36*, 3721-3734.

Cullen, A.C. and H.C. Frey, 1999: *Probabilistic Techniques in Exposure Assessment. A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, Plenum Press, New York, 335 pp.

Guenther, A., B. Baugh, G.Brasseur, J. Greenberg, P. Harley, L. Klinger,D. Serca and L.Vierling, 1999: Isoprene emission estimates and uncertainties for the Central African EXPRESSO study domain, *J. Geophys Res*., *104*, 30625-30639.

Guenther, A., C. Geron, T. Pierce, B. Lamb, P. Harley and R. Fall, 2000: Natural emissions of non-methane volatile organic compounds, carbon monoxide, and oxides of nitrogen from North America, *Atmos. Environ., 34*, 2205-2230.

Guenther, A., P. Zimmerman, P. Harley, R. Monson and R. Fall, 1993: Isoprene and monoterpene emission rate variability: Model evaluation and sensitivity analysis, *J. Geophys. Res., 98*, 12609-12617.

Hanna, S.R., J.C. Chang and M.E. Fernau, 1998: Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables, *Atmos. Environ., 32*, 3619-3628.

Hanna, S.R., Z. Lu, H.C. Frey, N. Wheeler, J. Vukovich, S. Arunachalam, M. Fernau and D.A. Hansen, 2001: Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, *Atmos. Environ*., *35*, 891-903.

Hanna, S.R., A.G. Russell, J. Wilkinson and J. Vukovich, 2002: Review of BEIS3 Formulation and Consequences Relative to Air Quality Standards: Estimation of Uncertainties in BEIS3 Emissions Outputs, EPRI Tech. Report 1005159, EPRI, 3412 Hillview Ave., Palo Alto, CA 94304.

Hanna, S.R., A.G. Russell, J. Wilkinson and J. Vukovich, 2003: Review of BEIS3 Formulation and Consequences Relative to Air Quality Standards: Estimation of Effects in Uncertainties in BEIS3 Emissions on Uncertainties in Ozone Predictions by Chemical Transport Models, EPRI Tech. Report 1005244, EPRI, 3412 Hillview Ave., Palo Alto, CA 94304.

Kasibhatla, P. and W.L. Chameides, 2000: Seasonal modeling of regional ozone pollution in the eastern United States, *Geo. Res. Let.*, *27*, 1415–1418.

Lamb, B., D. Grosjean, B. Pun and C. Seigneur, 1999: Review of the Emissions, Atmospheric Chemistry, and Gas/Particle Partition of Biogenic Volatile Organic Compounds and reaction products. Doc No CP051-1b-99, CRC Proj No A-23, CRC, 219 Perimeter Center Parkway, Suite 400, Atlanta, GA 30346, 206 pages, available at www.crcao.com/reports/recentstudies00-01/A-23%20AER%20final%20report.pdf.

Morgan, M.G. and M. Henrion, 1990: *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, New York.

Odman, M.T. and C.L. Ingram, 1996: Multiscale Air Quality Simulation Platform (MAQSIP): Source code documentation and validation, MCNC Tech. Report ENV-96TR002-v1.0, RTP, NC.

Pierce, T.E., 2001: BEIS3 version 0.9. ftp://ftp.epa.gov/amd/asmd/beis3v09/.

SAI, 1999: http://www.uamv.com/documents/uam-v_1.31_user%27s_guide.pdf, UAM-V User's Guide, Systems Application International Inc.

SMRAQ, 1997: The SMRAQ project: development and application of a seasonal model for regional air quality, Available online at http://envpro.ncsc.org/SMRAQ.

Williams, E.J., A. Guenther and F.C. Fehsenfeld, 1992: An inventory of nitric oxide emissions from soils in the United States, *J. Geophys. Res., 97*, 7511-7519.

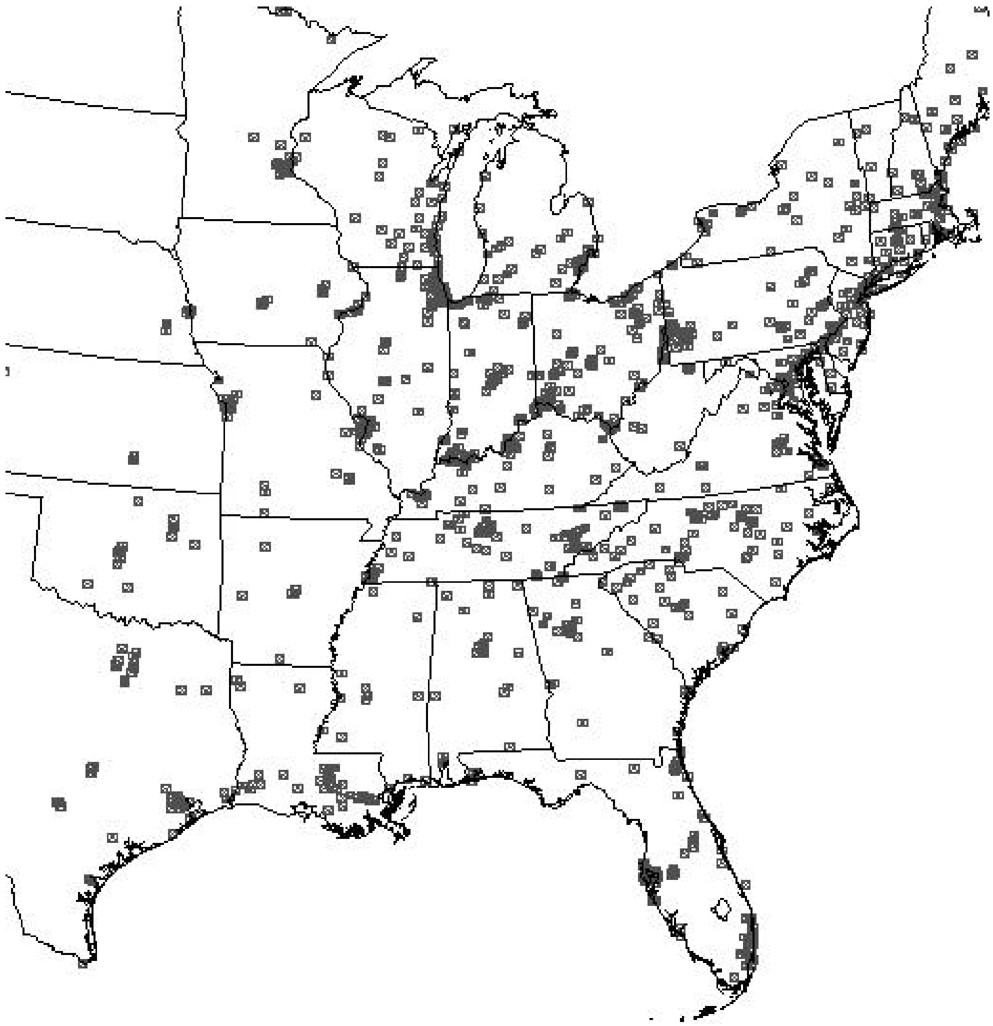

Figure 1. Map of the BEIS3 and CTM air quality modeling domain showing the locations of AIRS stations.

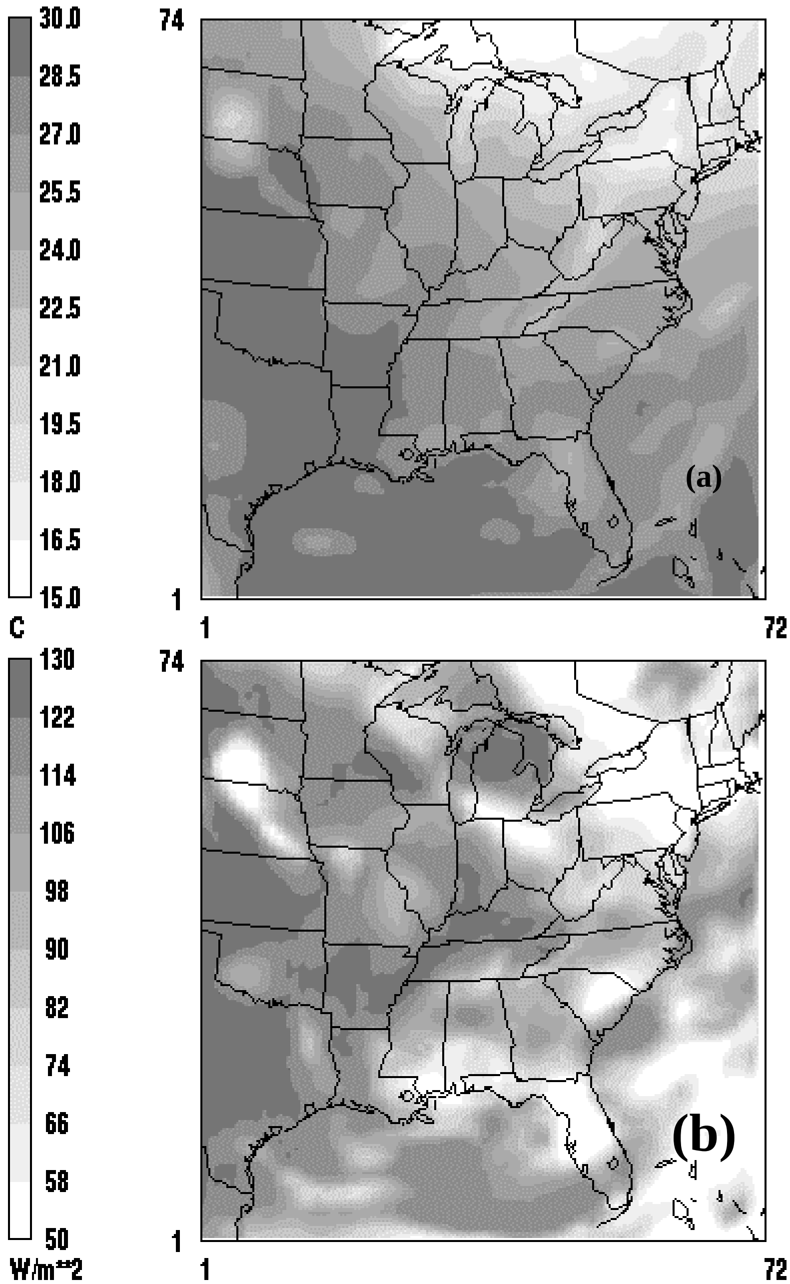

Figure 2 – Observed mean daily temperatures (Figure 2a, in C) and mean daily PAR values (Figure 2b, in W/m2) for 11 July 1995. Values represent averages over a 36 km by 36 km grid.

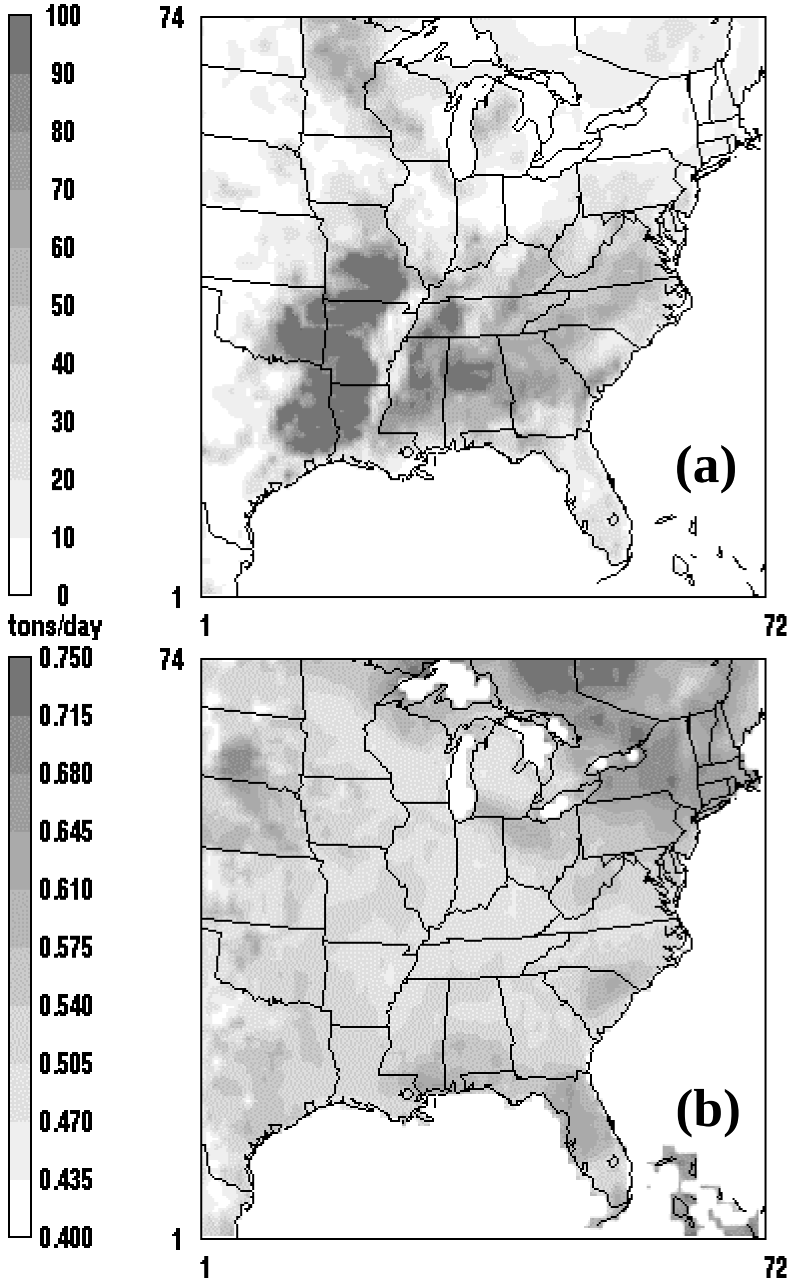

Figure 3 – Results of analysis of 1000 MC estimates of daily BEIS3 isoprene emissions for 36 km by 36 km grid squares for 11 July 1995. The mean is given in part a (tons/day), and the CV, or ratio of standard deviation to mean, is given in part b.

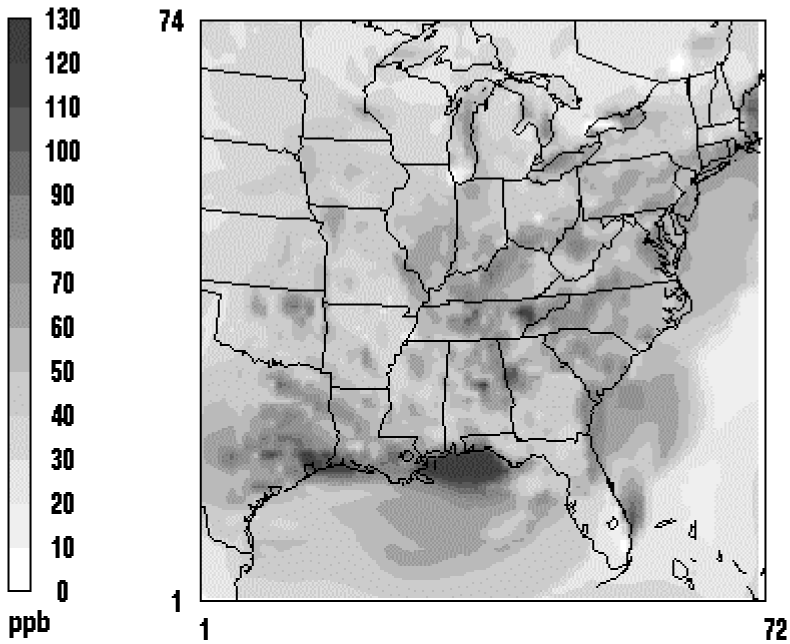

Figure 4 – Base run MAQSIP predictions of daily maximum one-hour averaged ozone concentration (ppb) for 36 km by 36 km grid squares for 11 July 1995.