Ecological Applications, 7(1), 1997, pp. 46–58

UNITED STATES LAND USE INVENTORY FOR ESTIMATING BIOGENIC OZONE PRECURSOR EMISSIONS

ELLEN KINNEE,[1] CHRIS GERON,[2] AND THOMAS PIERCE[3,4]

[1]*Dyntel, Research Triangle Park, North Carolina 27709 USA*
[2]*National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711 USA*
[3]*Air Resources Laboratory, National Oceanic and Atmospheric Administration, Research Triangle Park, North Carolina 27711 USA*

Abstract. The U.S. Geological Survey's (USGS) EROS (Earth Resources Observation System) Data Center's (EDC) 1-km classified land cover data are combined with other land use data using a Geographic Information System (GIS) to create the Biogenic Emissions Landcover Database (BELD). The land cover data are being used to estimate biogenic emissions in the contiguous United States. These emissions include volatile organic compound (VOC) emissions from vegetation and nitric oxide (NO) from soils. The EDC data are used predominately in the western United States, while other sources, such as the U.S. Department of Agriculture's Census of Agriculture and the U.S. Forest Service Eastwide Forest Inventory and Analysis Database (EWDB), are used in the eastern United States. The EROS Data Center land cover classifications must be used with caution in heterogenous areas. Emission factors vary drastically by specific crop and tree genera, and mixed classes in the EDC scheme may not always accurately reflect the actual crop/genus mix. However, future use of satellite-derived vegetation indices and other land cover characteristics may prove useful in understanding geographic distributions of foliar mass and seasonal variation in Leaf Area Index (LAI), which are important drivers in biogenic emission models.

Key words: biogenic emissions; foliar mass; GIS (Geographic Information System); isoprene; land cover; monoterpene; nitric oxide.

INTRODUCTION

Emission inventories of biogenic volatile organic compounds (BVOC) from vegetation and nitric oxide (NO) from soils are important inputs to models that simulate photochemical smog. Results from these models are often used in developing strategies for controlling anthropogenic sources of VOC and NO (Roselle 1994). Although they are not subject to controls, emissions from biogenic sources must be accurately estimated to develop effective control strategies for anthropogenic sources, such as motor vehicles and power-generation facilities. Averaged over the entire U.S., estimated BVOC emissions are greater than anthropogenic VOC emissions. Biogenic NO emissions from agricultural soils can also be regionally significant (Novak and Pierce 1993). Soil NO and BVOC emission models for the United States are described in Williams et al. (1992), Lamb et al. (1993*a*), Guenther et al. (1994), and Geron et al. (1994). A generalized approach typically used in these systems is illustrated in Fig. 1.

The land use, soils, and vegetation types used by biogenic emission models are aggregated into classes consistent with available emission rate data. The emission rate factors are usually expressed as mass of compound per unit time per unit area or plant dry mass. Therefore, spatial and temporal estimates of leaf area and foliage biomass are necessary. Emissions of specific compounds are estimated independently as functions of air and soil temperature and solar radiation. Emissions from specific soil or vegetation types are estimated on an hourly basis according to inputs of environmental variables, and then aggregated to appropriate grid or county scales for input into air quality simulation models. Although most of the uncertainty associated with BVOC and soil NO emission models is due to variation in measured emission rates and response to environmental factors, accurate characterization of land cover is very important. Individual vegetation types can emit drastically different types and quantities of compounds. In this paper, we will focus on the development of updated and refined land use/land cover data for BVOC emission models. Effects of

Manuscript received 3 January 1995; revised and accepted 21 November 1995; final version received 5 March 1996. For reprints of this Invited Feature, see footnote 1, p. 1.

[4] On assignment to the National Exposure Research Laboratory, U.S. Environmental Protection Agency.

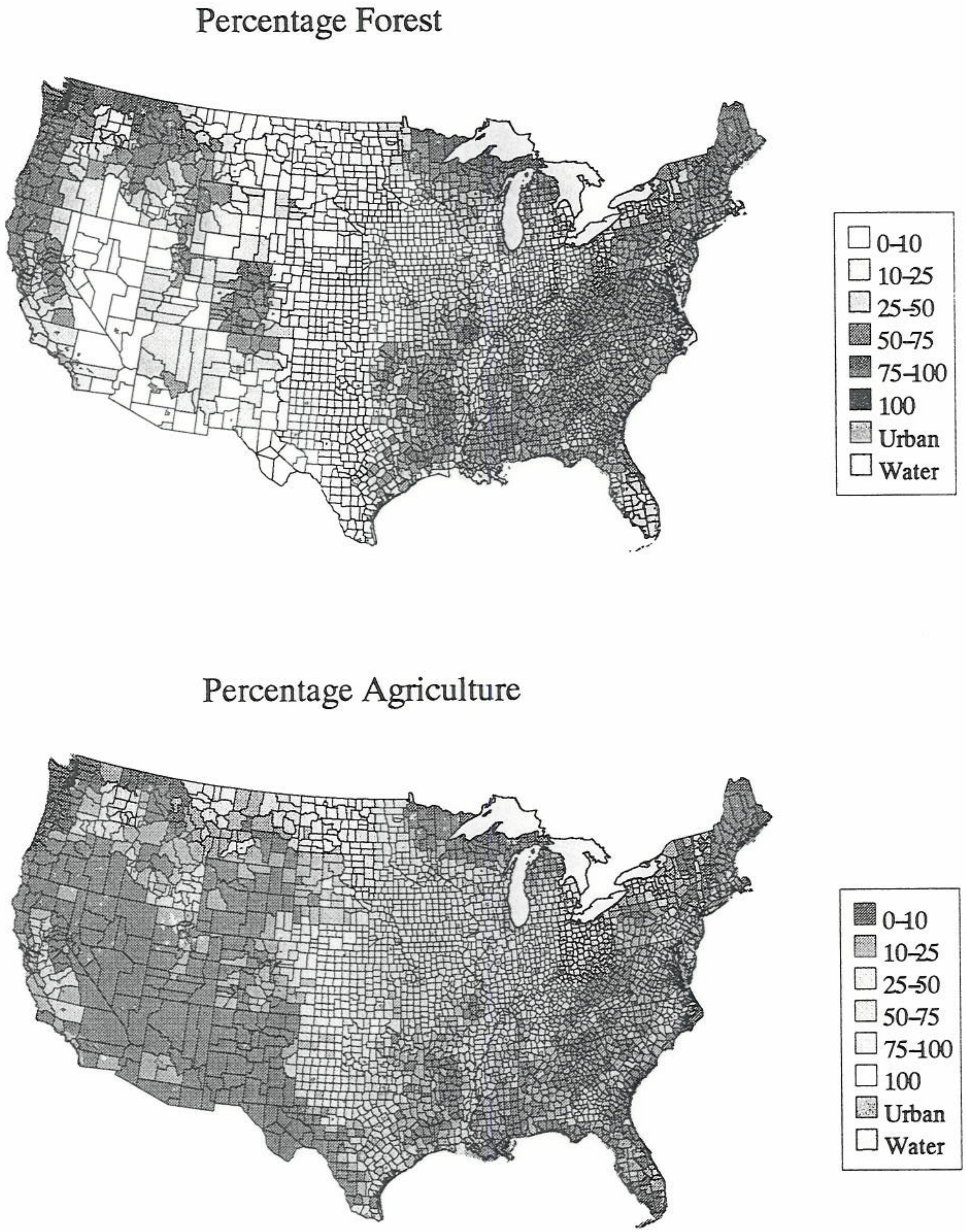


PLATE 1. Biogenic Emissions Landcover Database (BELD); percentage county area occupied by forest and agriculture and areal distribution of urban areas and water.

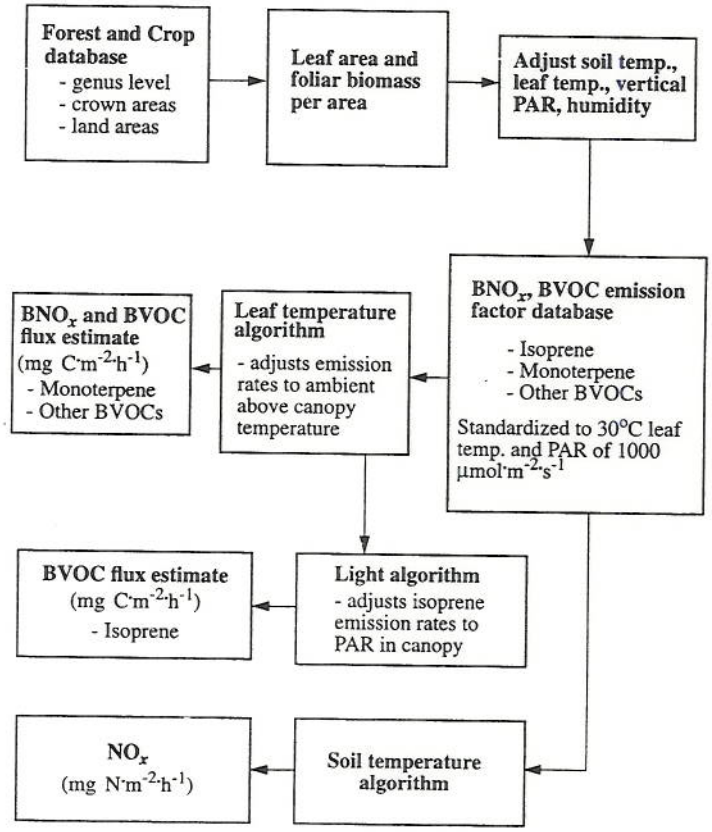


FIG. 1. General scheme for estimating emissions of biogenic area source emissions in the United States. BNO_x = biogenic NO_x; BVOC = biogenic volatile organic compound; NO_x = the oxides of nitrogen, NO and NO_2; PAR = photosynthetically active radiation.

land use classification on calculated emissions will also be explored.

METHODS

Biogenic Emissions Landcover Database (BELD)

Hourly county level and gridded estimates of BVOC and NO are needed for photochemical air quality models. The initial development of a land characteristics database used by the Environmental Protection Agency (EPA) for estimating biogenic area source emissions in the contiguous United States will aggregate vegetation types and land use to the county level. In the second stage of this work, a gridded database will be constructed from similar data sources. The process used to develop the county-level database is described below.

The Biogenic Emission Landcover Database (BELD) consists of county level data for nine broadly defined land use classes for every county in the contiguous United States. These classes include urban areas, coastal and inland water, forest, urban forest, agriculture, barren, scrub, grass, and other. Each of these land use classes can include many different land use types; up to 158 specific land use types are included in the database. These data were obtained from sources listed in Table 1 and separately processed to obtain the county-level area for each land use type. They were then merged into a Geographic Information System (GIS) layer of county boundaries using a hierarchy of rules, listed in Table 1, that emphasizes the relative importance of each land use type in the database. For example, Level I is the U.S. Forest Service Eastwide Database (EWDB) of forest crown covers by genus. This land use is merged first to ensure that the forest classes needed for biogenic emissions modeling are preserved. The next level added is water, then urban areas, and so on until either the county is 100% full or the areas from all land use data sources are exhausted. Each county in the completed database consists of one polygon with total county area and percentage of each land use type associated with that county. Quality assurance checks in the form of GIS plots were created after each land cover class was added to the database to ensure that the amount and distribution of land use classes were reasonable across the country. A comparison of the land use distribution in BELD with similar land use inventories is discussed in the *Results*.

Processing of data

The completed BELD database contains land use areas for the 3109 county polygons comprising the conterminous United States. A series of steps was required to define the county boundaries, to process separate sources of data, and to merge the data into each county.

TABLE 1. Sources of data and hierarchy of rules used in building the Biogenic Emissions Landcover Database (BELD).

Level	Description of raw data	BELD vegetation class
I.	U.S. Forest Service East-wide database (≈1990). Contains 0.4-ha plot survey of ≈97000 forest locations in eastern U.S. Allometric equations used diameter-at-breast-height measurements to estimate crown cover by genus.	Total forest land area Forest crown cover by tree genus
II.	USGS Land Cover Characteristics Data Set (AVHRR ≈1990); 1-km classified pixels for conterminous U.S.	Inland water area
III.	U.S. Census Bureau. 1990 urbanized area boundaries.	Total urban area
IIIa.	U.S. Forest Service (D. Nowak, *personal communication*). Fraction of urban area assumed to be forested. Based on potential natural vegetation and relative percentages of forest/tree genera in surrounding area.	Total urban forest land area
IV.	U.S. Department of Commerce (1987), Bureau of the Census. 1987 Census of Agriculture.	Total agricultural area
IVa.	Specific crop areas.	Area by crop type
IVb.	Nondesignative cropland.	Area assumed as miscellaneous crop
V.	USGS Land Cover Characteristics Data Set (see II). Land use classes (except water) assigned to Guenther et al. (1994) emission categories. Used almost exclusively in the western U.S. and southern Florida, except for barren, scrub, and grass classes used for entire U.S.	Area of generalized western forest classes, i.e., barren, scrub and grass
VI.	Undesignated/other. Area remaining in a county that was not allocated above.	Area of other

During creation, separate county boundary files were developed for eastern states and western states because they use different data sources and processing schemes. Two county boundary files were created for each side. The first contains boundaries from USGS Digital Line Graph (DLG) county boundary files, which are clipped to the coastal shorelines (i.e., do not extend beyond the shorelines). The second file contains county boundaries from the U.S. Census Bureau's 1990 Topologically Integrated Geographic Encoding and Referencing (TIGER) line files. These boundaries extend out into coastal and inland waters and represent the true county areas. The coastal water area was calculated by overlaying the TIGER county polygons with the clipped USGS county polygons. The difference between the two areas was considered to be coastal water.

The inland water polygons were obtained from the USGS EROS Data Center Land Cover Characteristics database (Loveland et al. 1991, referred to hereafter as the EDC data). The water image extracted from the EDC data was first converted into a GIS grid and clipped to coastal boundaries. This raster grid file was converted into vector polygons and overlaid on the county boundaries. The inland water polygons areas were then summed by county. Total water area and its percentage of total county area were calculated. The total water area in each county is the sum of the inland and coastal water areas.

The percentage of urbanized area per county was calculated using the U.S. Census urbanized-area polygons obtained from the U.S. EPA GRIDS database (1993). Urbanized areas are defined as "a central city or cities and surrounding closely settled territory areas that together have a minimum population of 50,000" (Bureau of the Census 1989). It was decided if an urban polygon and a water polygon overlapped, the area would be considered water. With this in mind, the urban polygons were first made to conform to the U.S. coastal boundaries. Next, the inland water polygons were used to erase the urban polygons so that any water areas on the edge of or within urban polygons were preserved. The final urban polygons were overlaid with the county boundaries to get the urban area within each county.

A set of crude assignments was used for tree cover in urban areas. A map of potential vegetation was used to calculate the relative amounts of forest, grass, and desert that would be present in each county if it were undisturbed. If a given urban area fell in a forested region, then 32% of that urban area was said to be covered by tree canopy composed of genera found in the surrounding region. This percentage fell to 22 and 10% in grass and rangeland/desert classes, respectively. These figures are in general agreement with those from Dave Nowak (*personal communication,* USDA Forest Service). Percentages of forest, grass, and desert were weighted by relative area and an urban forest factor calculated for each county. This factor was then used to multiply the total urban area to get the area of urban forest.

Once the coastal and inland water and urbanized areas were completed, the processing for the 37 eastern states and 11 western states diverged. The total forest area and individual areas of 12 forest classes for counties in the western U.S. were obtained from the 1-km

TABLE 2. The relationship between EDC, Anderson Level II, and BELD data classes.†

EDC class	Anderson Level II class	BELD class
1–21	dryland cropland	miscellaneous crops/grass
22–32	irrigated cropland	miscellaneous crops/grass
33–34	mixed cropland	miscellaneous crops/grass
35–39	grassland cropland	miscellaneous crops/grass
40–54	woodland cropland	woodland cropland
55–65	grassland	grass
66–71	desert cropland	scrub/barren
72–83	mixed shrub/grass	scrub/grass
84–85	chaparral	scrub/grass
86–89	savannah	scrub/grass
90–93	northern deciduous forest	hardwood forest
94–95	southeastern deciduous forest	southern hardwood forest
96–97	western deciduous forest	hardwood forest
98–99	southeastern coniferous forest	coniferous forest
100–125	western coniferous forest	western coniferous forest
126–132	western woodlands	western woodlands
133–137	northern mixed forest	northern mixed forest
138–141	eastern mixed forest	southeastern mixed forest
142–148	western mixed forest	western mixed forest
149–149	water	water
150–152	wetlands	wetland forest
153–154	forested wetlands	wetland forest
155–155	barren	barren
156–156	subalpine forest	boreal forest
157–159	alpine tundra	boreal forest

† EDC = EROS (Earth Resources Observation System) Data Center; BELD = Biogenic Emissions Landcover Database.

USGS EDC land cover image. This image aggregates the original 159 EDC classes into 25 modified Anderson Level II classes (reference the 1990 Conterminous U.S. Land Cover Characteristics Data Set for classification). The Anderson classes were further aggregated into classes suitable for biogenic emission modeling. The relationships between the original 159 EDC classes, Anderson Level II classes, and the final BELD classes are illustrated in Table 2. For each of the 23 western forest classes, it is assumed that a constant proportion of plant composition is present within each cell wherever the cell is identified. For example, if a cell is classified as western coniferous forest it is assumed that the forest is distributed evenly across that cell. For many western cover types, spatial heterogeneity is minimal with respect to the 1-km pixel size, and therefore use of the EDC in these regions is probably justified. In regions with complex vegetation mixes, however (such as montane forests), it should be used with caution.

Western forest cover areas were calculated by first converting the Land Cover Characteristics image into a GIS grid format. A mask of the western counties was imposed on the land cover grid to clip out western counties. Next, a grid was created of the western county boundary polygons whose cell size and grid origin matched the land cover grid exactly. The cells of the county boundary grid contained the state/county Federal Information Processing Standards (FIPS) codes. The county grid was added to the land cover grid resulting in grid cells with both state/county FIPS and land cover class. The number of cells in each class were then multiplied by the cell area (1 000 000 m^2) and summed by state/county FIPS to obtain the area of each forest class per county. An identical methodology was applied to the EDC image of the entire U.S. to obtain area values for the barren, scrub, and grass classes.

The forest area for the eastern counties was obtained from the Forest Inventory and Analysis (FIA) Eastwide Data Base (EWDB, Hansen et al. 1992). The forest coverage data used for the eastern 37 states are described in Geron et al. (1994). Forest extent, species composition, and tree diameter distribution data were obtained from the most recently available EWDB data. The data are collected on ≈10-yr cycles by the United States Department of Agriculture Forest Service FIA units. Using the EWDB as opposed to the EDC data in the eastern region has two advantages: (1) areal coverages, which are estimated from aerial photos, are likely to provide more accurate estimates of forested areas, and (2) estimated crown coverages in forested areas can be separated into specific tree genera, allowing use of the genus-level emission rate factors from Guenther et al. (1994). These factors provide more accurate emission rates of BVOC from forested environments. However, the EDC data provide indices of vegetation biomass that may be useful in the future for estimating vegetation density (Guenther et al. 1996).

Data from the EWDB used in the BELD database were processed into two separate data files. The first data file, containing the total forest land area per county, was merged directly into the county polygons. The second data file contains crown cover for up to 67

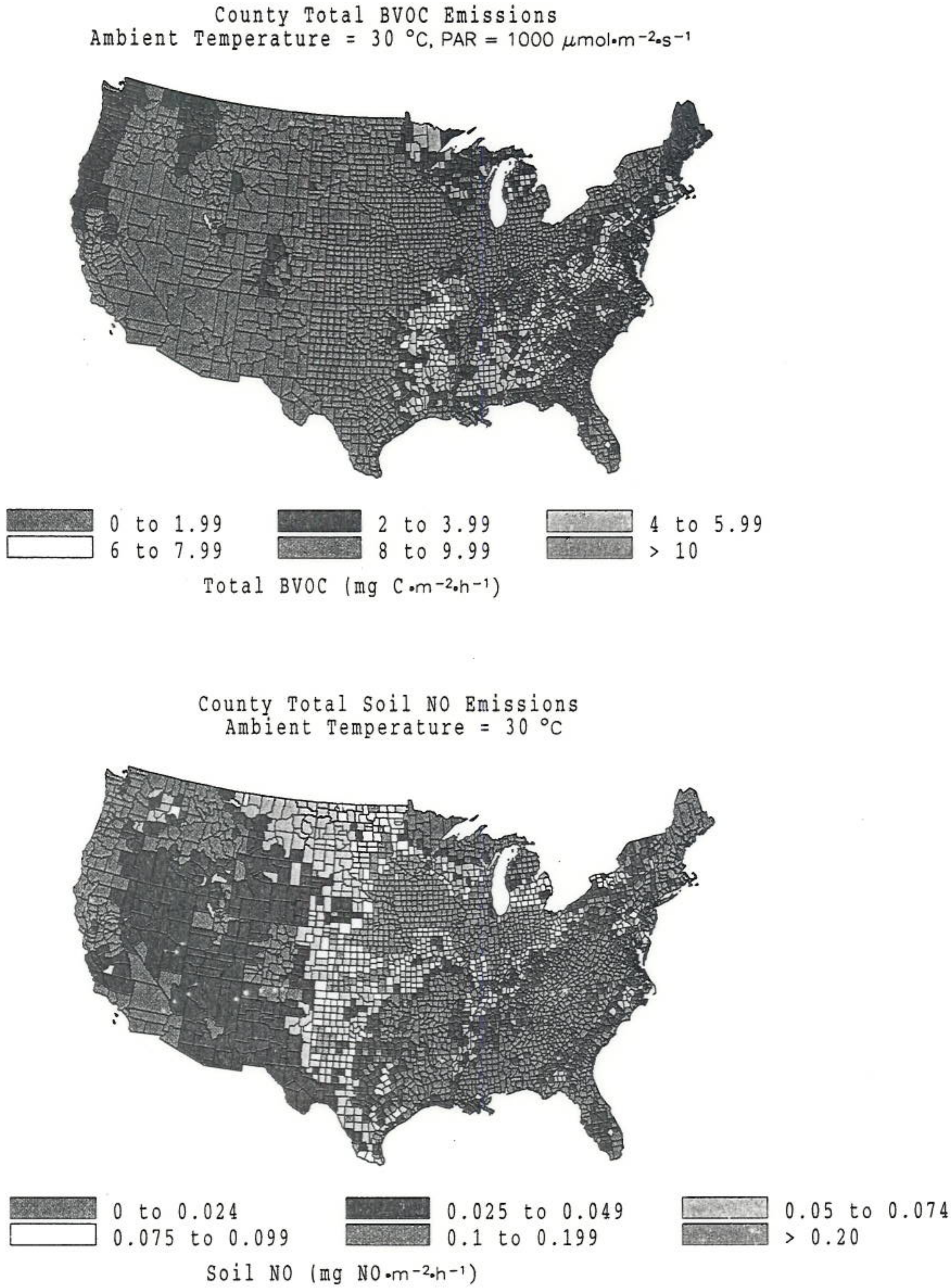


PLATE 2. (Top) Total county biogenic volatile organic compound (BVOC) emissions (mg $C \cdot m^{-2} \cdot h^{-1}$) and (bottom) soil NO emissions (mg $NO \cdot m^{-2} \cdot h^{-1}$) modeled using BELD (Biogenic Emissions Landcover Database) from Geron et al. (1994).

different genus types within each county. The large number of genus types dictated that crown cover data be stored in a separate data file and accessed by state and county FIPS code as needed.

Areas for specific crops and nondesignative cropland were extracted from the Census of Agriculture data (U.S. Dept. of Commerce 1987). County acreage for a total of 15 individual crops as well as total cropland and total farmland were read into a data file. The total cropland area was subtracted from total farmland and the difference added into the miscellaneous crop category to obtain the total agricultural area.

Once all the data sources were processed and county level areas calculated across the United States, the land use types were merged into the county polygons using the hierarchy of rules shown in Table 1. This hierarchy for filling in the land use for a county ensured that the total land use area did not exceed 100% of the county area. For counties that did not have 100% of the area filled by individual land cover types, the remaining area was designated as Other.

Biogenic emission rates

Each land cover type developed in the BELD database can be associated with a corresponding emission factor. For biogenic emission modeling, the area of each land use type is multiplied by a standardized emission flux to obtain a standardized emission rate by county. This emission rate is then adjusted to account for environmental conditions.

Standardized emission rates

Guenther et al. (1994) and Geron et al. (1994) provide a basis for the emission rates of isoprene and monoterpenes for dominant tree genera and EDC mixed-forest classes in North America. These emission rates have been standardized for bright sunlight (1000 μmol photons$\cdot$m$^{-2}\cdot$s^{-1}) and a leaf temperature of 30°C. Standardized emission rates of isoprene and monoterpenes for species within a genus generally have been found to fall within $\pm$ 50% of these emission rates. A third emission rate class aggregates all other volatile organic compounds (OVOC) that have a typical atmospheric lifetime of $<$1 d. Standardized emission rates for soil NO were adapted from Williams et al. (1992).

Emission rates for specific agricultural crop types, with the exception of corn, are taken from Lamb et al. (1993*b*). The negligible VOC emission factors for corn are based on more recent field data by Sharkey et al. (1992). These agricultural emission factors assume a constant biomass that reflects peak foliation during the growing season. Table 3 lists examples of the land use cover types and their standard BVOC and soil NO emission rates. These emission rates are broken down by the individual forest genera as available from the EWDB and for the mixed classes available from the EDC. Emission rates for the EDC land use types are derived from a vegetation species mix described by Guenther et al. (1994).

Environmental correction

BVOC emission rates increase exponentially with increasing leaf temperature (up to $\approx$38°C for isoprene), while the isoprene emission rate also increases with increasing light intensity up to $\approx$800 μmol photons$\cdot$m$^{-2}\cdot$s^{-1} (Tingey et al. 1979, 1980, Guenther et al. 1991). The equations of Guenther et al. (1993) were used to simulate the effects of leaf temperature on isoprene and monoterpene emission rates.

Since photosynthetically active radiation (PAR) strongly controls isoprene emission rate, a simple algorithm to reduce PAR at lower levels within forest canopies was applied. Specific leaf mass (SLW, in g/m^2 dry mass) is also adjusted vertically through broadleaf canopies. These effects are discussed in detail in Geron et al. (1994) and Lamb et al. (1993*a*, *b*).

Emission rates from other types of vegetation are expressed as emission rate per unit area, with a constant peak growing season biomass assumed. The light and temperature corrections are applied, but no canopy model is used. Soil NO emissions are estimated as functions of soil temperature, which are assumed to vary as a function of air temperature as suggested by Williams et al. (1992).

Results

BELD

Plate 1 illustrates the county level land cover distribution for inland waters, urban areas, forests, and agriculture from BELD. The figures appear to yield reasonable spatial patterns for each broad land use class, for example, high percentages of forest follow the mountainous regions, and agriculture is concentrated in the midwest. However, the areas of all land use classes when summed are often much less than the total county land area. This is likely due to problems with the source data. For instance, in the Agricultural Census data, summation of the areas in individual crops per county often yields total land areas that are much lower than total cropland area. This problem can be resolved by using the EDC for crop land areas in the same way that forest areas were calculated in the western U.S. This would entail estimation of emission factors for the mixed-crop classes in the EDC data. The EWDB data used for forest in the east is collected only for tracts over 0.4 ha in size. The smaller stands of trees left uncounted can often add up to a significant portion of county area. Also the EWDB does not always account for forested wetlands, which are important in localized areas.

Comparison of BELD to other land use inventories

Land use areas in the BELD county level database were compared with three other land use data sets de-

TABLE 3. Examples of isoprene, monoterpene, other VOC (volatile organic compounds), and soil NO emission rates ($\mu g \cdot m^{-2} \cdot h^{-1}$). Forest genera and mixed-EDC (EROS Data Center) forest class rates are adapted from Guenther et al. (1994) and Geron et al. (1994). Crop category rates are derived from Lamb et al. (1993) except for corn, which is from more recent field data. NO rates for all classes are derived from Williams et al. (1992). All emission rates are standardized for bright sunlight (1000 μmol photons$\cdot m^{-2} \cdot s^{-1}$) and 30°C.

Code	Isoprene	Mono-terpene	Other VOC	NO	Description
Common individual forest genera from the EWDB†					
Abie	170.0	5100.0	2775.0	4.5	*Abies* (fir)
Acer	42.5	680.0	693.7	4.5	*Acer* (maple)
Betu	42.5	85.0	693.7	4.5	*Betula* (birch)
Cary	42.5	680.0	693.7	4.5	*Carya* (hickory)
Fagu	42.5	255.0	693.7	4.5	*Fagus* (beech)
Frax	42.5	42.5	693.7	4.5	*Fraxinus* (ash)
Liqu	29750.0	1275.0	693.7	4.5	*Liquidambar* (sweetgum)
Liri	42.5	85.0	693.7	4.5	*Liriodendron* (poplar)
Pice	23800.0	5100.0	2775.0	4.5	*Picea* (spruce)
Pinu	79.3	2380.0	1295.0	4.5	*Pinus* (pine)
Plat	14875.0	42.5	693.7	4.5	*Platanus* (sycamore)
Popu	29750.0	42.5	693.7	4.5	*Populus* (aspen)
Pseu	170.0	2720.0	2775.0	4.5	*Pseudotsuga* (douglas fir)
Quer	29750.0	85.0	693.7	4.5	*Quercus* (oak)
Robi	5950.0	85.0	693.7	4.5	*Robinia* (black locust)
Sali	14875.0	42.5	693.7	4.5	*Salix* (willow)
Tsug	79.3	158.7	1295.0	4.5	*Tsuga* (eastern hemlock)
Crop categories					
Alfa	19.0	7.6	11.4	12.8	alfalfa
Barl	7.6	19.0	11.4	256.7	barley
Corn	0.5	0.0	0.0	577.6	corn
Cott	7.6	19.0	11.4	256.7	cotton
Gras	56.2	140.5	84.3	57.8	grass
Hay	37.8	94.5	56.7	12.8	hay
Mscp	7.6	19.0	11.4	12.8	miscellaneous crops
Oats	7.6	19.0	11.4	256.7	oats
Pean	102.0	255.0	153.0	12.8	peanuts
Pota	9.6	24.0	14.4	192.5	potato
Rice	102.0	255.0	153.0	0.2	rice
Rye	7.6	19.0	11.4	12.8	rye
Scru	37.8	94.5	56.7	57.8	scrub
Sorg	7.8	19.5	11.7	577.6	sorghum
Soyb	22.0	0.0	0.0	12.8	soybean
Toba	0.0	58.8	235.2	256.7	tobacco
Whea	15.0	6.0	9.0	192.5	wheat
Mixed-cover types derived from EDC data					
Borf	910.0	713.0	755.0	4.5	boreal foest
Conf	1550.0	1564.0	1036.0	4.5	conifer forest
Harf	8730.0	436.0	882.0	4.5	hardwood forest
Nmxf	10150.0	1100.0	850.0	4.5	northern mixed forest
Shrf	10750.0	530.0	910.0	4.5	southeastern/western deciduous forest
Smxf	17000.0	1500.0	1250.0	4.5	southeastern mixed forest

EWDB = U.S. Forest Service Eastwide Forest Inventory and Analysis Database.

veloped for the purpose of estimating biogenic emission of VOC. These include Advanced Very High Resolution Radiometer (AVHRR)-based data for a three-state area in the southeastern U.S. done in 1990, a foliar biomass inventory compiled for the Oak Ridge, Tennessee area using Landsat Thematic Mapper (TM) imagery in 1991/1992, and an evaluation of vegetation biomass and emission factors conducted in the South Coast Air Basin surrounding Los Angeles, California in 1990.

A study conducted at the Computer Graphics Center (CGC) at North Carolina State University for the U.S. EPA focused on North Carolina, South Carolina, and Georgia in development of a methodology for mapping the current (1990) status of land use types (CGC 1994). Land use classification was based on the use of AVHRR data calibrated with classified Landsat TM imagery, aerial photography, and USDA FIA field data. The study area was divided into three distinct physiographic regions: (1) Coastal Plain; (2) Piedmont, and (3) Appalachian Ridge and Valley prior to classification. Initially, the spatial distribution of six general cover types was studied using Landsat TM data. The land cover types included urban, water, nonforest, hardwood forest, conifer forest, and mixed forest. The resulting land cover files were then used to characterize the spectral response of these six types within the AVHRR data for the three state study area. A least-squares linear mixture model that predicted the occurrence of the six cover types within a given AVHRR cell based upon that pixel's spectral characteristics was developed using the observed relationships between the classified Landsat TM data and the unclassified AVHRR data. Cover type occurrence data at the 1-km resolution of the AVHRR data were aggregated so as to summarize the data at the 20 × 20 km resolution of the EPA modeling grid.

For comparison purposes, the six land use types were aggregated to the physiographic province level. This resulted in three separate data sets. Because the CGC data contained a mixed-forest class, which could not be accurately broken down into hardwood and conifer, the forest classes were summed to a total forest value. The county-level BELD data were intersected with physiographic boundaries and land areas for the five classes aggregated for the three provinces. Table 4A lists the amount of each of the five land use types by province for both BELD and the CGC study. The BELD database shows significantly less urban area for each of the three provinces. Because the AVHRR urban class is dependent on spectral signatures, the AVHRR data would tend to overestimate urban area due to the inclusion of land use such as barren and desert whose spectral signatures are similar to urban areas. The water area in BELD is also much smaller than in the CGC data. This is probably due to the difference in spatial resolution between the EDC-derived water in BELD (1 km) and the Landsat TM data in the CGC data (30 m). The nonforested area in BELD includes agriculture, barren, scrub, grass, and other land use types. It is higher than the CGC data in physiographic regions 1 and 3 and about equal in region 2. Because of the complex nature of this class it is difficult to assess the cause of any differences. Finally, the total forest area in BELD is larger for all the regions with a more pronounced difference in region 2. From the unaggregated data this seems to be due primarily to a lesser amount of conifer in the CGC data.

The Oak Ridge, Tennessee land use data comes from a Landsat TM classification image created using multidate imagery and supervised classification techniques (Baugh et al. 1996). Eleven land use classes were identified, with an emphasis on oak tree concentrations since they are the dominant BVOC (isoprene) emitters in the region. The classes are as follows: conifer forest, mixed-pine forest, young/dense conifer, high oak deciduous (50–80% oak), medium oak deciduous (25–50% oak), low oak deciduous (<25% oak), shrubs and grasses, agriculture, water, bare soil, and urban. The satellite image data was collected by the Landsat Thematic Mapper (TM) on 17 December 1991, 23 April 1992, and 30 September 1992 and covers a square region ≈ 105 km on a side. The supervised classification method used in this analysis is called the Spectral Angle Mapper (SAM). This method treats image spectra as n-dimensional vectors (where n is the number of bands). A test spectrum (from a pixel) is compared with a reference spectrum (the desired class) and a smaller angle between the spectral vector indicates a better match. A threshold angle is used to determine whether a spectrum is classified as a specific class. The power of the SAM method is that it is very sensitive to small spectral differences, yet it is highly resistant to illumination differences (Kruse et al. 1993).

In order to compare the land cover types of the two

TABLE 4. Comparison of BELD (Biogenic Emissions Landcover Database) land use areas to areas from other land use inventories.

A) BELD vs. North Carolina state—southeastern United States

Land use type	BELD (ha)	North Carolina state (ha)
Physiographic province 1: Coastal Plain		
Urban	381 813	510 375
Water	326 460	1 430 333
Nonforest	7 109 131	6 107 476
Forest	12 312 357	12 089 857
Physiographic province 2: Piedmont		
Urban	680 669	772 844
Water	81 171	824 568
Nonforest	3 906 147	3 908 803
Forest	7 387 750	6 549 954
Physiographic province 3: Appalachian Ridge and Valley		
Urban	65 400	183 143
Water	9 895	171 326
Nonforest	942 139	677 700
Forest	2 774 891	2 759 979
Total area	35 977 823	35 977 358

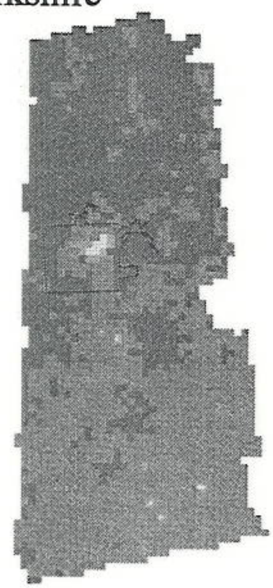
B) BELD vs. Oak Ridge TM (Thematic Mapper)—eastern Tennessee

Land use type	BELD (ha)	Oak Ridge (ha)
Urban	53 911	61 145
Agriculture	140 745	75 847
Forest	604 146	700 447
Water	13 151	31 408
Scrub/grass	4 088	72 395
Barren	0	7 574
Other	181 225	48 449
Total area	997 266	997 265

C) BELD vs. SoCAB (South Coast Air Basin)—southern California

Land use type	BELD (ha)	SoCAB (ha)
Urban other	604 328	561 665
Urban forest	100 024	82 099
Grass	335 611	114 396
Scrub	269 061	1044 531
Barren	887 849	664 466
Inland water	22 885	0
Agriculture	143 777	204 000
Wood/cropland	8 406	0
Forest	601 461	293 390
Total area	2 973 402	2 965 047

a) Berkshire



BELD	
Isoprene Emission Flux 3.51 mg·m^{-2}h^{-1}	
LandCover Percentages	
Forest 76.59	Grass 0.0
Agri 5.30	Scrub 0.0
Urban 5.10	Water 0.04
Other 12.97	

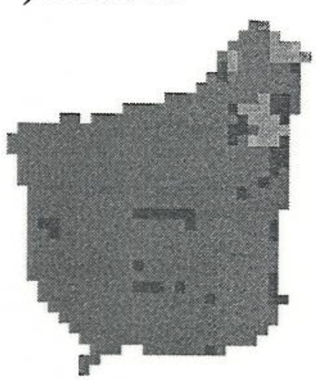
EDC	
Isoprene Emission Flux 5.06 mg·m^{-2}h^{-1}	
LandCover Percentages	
Forest 94.31	Grass 0.1
Agri 5.55	Scrub 0.0
Urban 0.0	Water 0.04
Other 0.0	

b) Fulton

BELD	
Isoprene Emission Flux 7.47 mg·m^{-2}h^{-1}	
LandCover Percentages	
Forest 40.89	Grass 0.44
Agri 2.06	Scrub 1.56
Urban 55.05	Water 0.0
Other 0.0	

EDC	
Isoprene Emission Flux 6.59 mg·m^{-2}h^{-1}	
LandCover Percentages	
Forest 74.14	Grass 2.63
Agri 22.97	Scrub 0.26
Urban 0.0	Water 0.0
Other 0.0	

c) Franklin



BELD	
Isoprene Emission Flux 4.52 mg·m^{-2}h^{-1}	
LandCover Percentages	
Forest 50.58	Grass 0.0
Agri 21.41	Scrub 0.0
Urban 0.0	Water 0.0
Other 28.01	

EDC	
Isoprene Emission Flux 8.67 mg·m^{-2}h^{-1}	
LandCover Percentages	
Forest 97.72	Grass 0.05
Agri 2.23	Scrub 0.0
Urban 0.0	Water 0.0
Other 0.0	

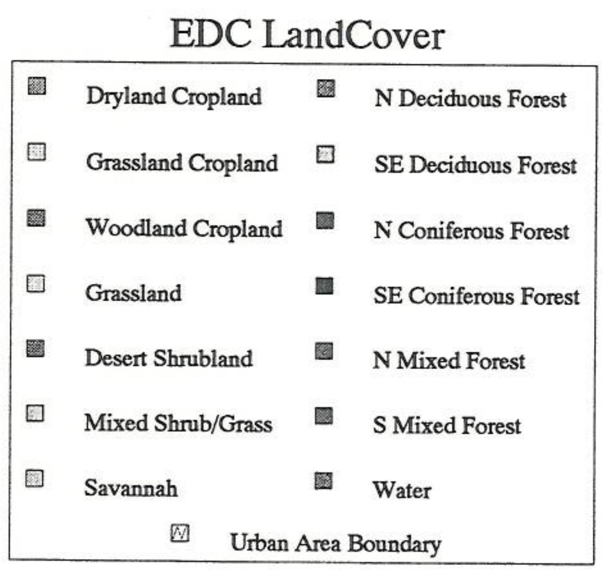


PLATE 3. Vegetation cover derived from EDC (EROS Data Center) data for (a) Berkshire County, Massachusetts, (b) Fulton County, Georgia, and (c) Franklin County, Georgia. Tables compare isoprene emission fluxes calculated using BELD (Biogenic Emissions Landcover Database) data vs. those calculated using the EDC data.

databases, the Oak Ridge conifer and oak forest classes were combined into one forest class. This resulted in the seven land use types listed in Table 4B whose areas were summed over the entire image. The boundary of the study area was extracted and then used to clip the BELD county level land use. BELD classes were then summed for the study area with the results also listed in Table 4B. Both the urban and forest classes match up reasonably well across the region, while the agriculture in BELD is approximately twice that of the Oak Ridge data. The TM classification may have put some crop/pasture categories, which are considered agriculture in BELD, into the Oak Ridge scrub/grass class. The total amount of water is much lower in BELD and is again probably due to differences in the spatial resolution of the input data. The higher amount of the OTHER class in BELD indicates a larger degree of uncertainty in the data but not necessarily a more inaccurate classification.

The South Coast Air Basin (SoCAB) in southern California was the site of another land use inventory (Horie et al. 1990). This project was designed to produce basic data on leaf biomass in the SoCAB and incorporate emission factors for plant species composing this biomass. The main focus of the study was placed on characterization of species composing the urban vegetation and development of a spatially resolved biomass inventory for both urban and natural areas. Using readily available aerial photographic resources and street maps, a detailed land use map was constructed for urbanized portions of SoCAB. A low aerial photographic survey yielded vegetation type and characteristics including size, number, and tree species for 13 land use categories. A total of 70 sites were visited by a ground survey team to determine a representative mix of plant species and stem volume to area ratio for each tree species. Vegetation in natural areas was inventoried using data, and maps delineating plant communities on USGS quadrangles were used to create a gridded inventory or land cover by plant community.

For purposes of comparison the urban land use types in the SoCAB study were merged into two classes, urban other and urban forest. The natural area land use types were mapped into existing BELD land use classes and all were summed across the study area. Study boundaries were derived from an existing coverage of the Urban Airshed Model modeling domain used to define the SoCAB area. The BELD data were clipped and all land use types summed with the results shown in Table 4C. The BELD data show slightly more area in the urban classes, which could be due to differences in the Census definition of urban area and the delineation of developed areas on aerial photographs used for the SoCAB study. The barren, scrub, and grass categories all show differences, but when summed together into one class are similar. One interesting feature is the lack of an inland water class in the SoCAB study. It is unclear where water was included as a class within the natural area definition. The total agricultural area in BELD is much less than in SoCAB, while the overall forest area is over twice as large. This could be due to the SoCAB study classifying natural areas as scrub, while BELD put these areas into scrub forest categories.

Emission fluxes

The forest coverage provided by the EWDB data in BELD provides genus-level data for the eastern U.S. and is well suited to experimental BVOC emission rate information available for modeling. Plate 2 (top) illustrates the pattern of total BVOC emissions derived from the BELD landuse. Maximum emission rates are found in the Ozark and Appalachian Mountains and adjoining piedmont regions. This is attributed largely to the abundance of high-isoprene-emitting oak trees in these areas. The soil NO patterns (Plate 2) (bottom) show highest emissions in the midwest where fertilized croplands dominate the rural landscapes. Forested lands exhibit the lowest soil NO rates, with grasslands being intermediate between these two classes. Hourly emission rates are dependent upon timing of fertilization, moisture, and other factors. These temporal factors, as well as spatial uncertainty in current soils databases, induce high levels of uncertainty in NO emissions models.

Forested areas produce the highest levels of biogenic VOC emissions. Therefore, large differences in emission fluxes can result when using the broadly defined forest classes and forest areas from the EDC as opposed to the genus-level EWDB data in BELD. Three counties are used to illustrate this point in Plate 3. In Franklin County, Georgia, the isoprene emission flux calculated using EDC data is 8.67 $mg \cdot m^{-2} \cdot h^{-1}$, nearly a factor of two higher than the flux of 4.52 $mg \cdot m^{-2} \cdot h^{-1}$ calculated when using BELD. This roughly corresponds to the higher percentage of forest area reported in the EDC data as opposed to BELD. Likewise, Berkshire County, Massachusetts shows isoprene emission estimates over one-third higher when using the EDC land use percentages as opposed to the EWDB. This is again attributed to the much greater forest area determined from the EDC. In contrast, isoprene emissions for Fulton County (Atlanta), Georgia estimated using the EWDB are higher, even though the estimated forest area is lower using this data. This is due to a substantial amount of oak (*Quercus*), a high isoprene emitter, in the area as determined from the EWDB as opposed to the EDC. In the two rural counties, the estimated percentage of forest from EDC data may actually be more reasonable than the EWDB, since a substantial percentage (28 and 13%, respectively) of the land area in these counties is classified as "other" in BELD. It is likely that a significant proportion of this unclassified area is covered by smaller tracts of tree cover (i.e., <0.4 ha), which are not accounted for in the EWDB source data but are correctly classified in the EDC data. However, many of the EDC classes contain joint definitions such as cropland/woodland, which prevent accurate estimates of forest cover and lack tree species identifications. Thus, while the EWDB in BELD provides genus-level coverage of forest land, perhaps the EDC data provide a more accurate assessment of the total land area of tree cover.

Discussion

In urban areas and western ecosystems, we must rely on the EDC classifications and other databases that are not as compatible with current biogenic emissions modeling schemes. Urban areas are especially important since they are often the focus of ozone (O_3) air quality modeling exercises. In addition, soils databases that include variables such as texture, pH, albedo, bulk density, and moisture capacity are also needed for more accurately modeling soil NO fluxes. Soils data from the USDA Natural Resources Conservation Service (NRCS) STATSGO (State Soil Geographic) database are being explored for driving soil NO emissions, water balance, and transpiration rates.

Work is underway to improve the biogenic emission modeling procedure by handling seasonal changes in leaf biomass and LAI. Currently the model uses just one emission factor table for summer emissions and one for winter emissions. Genus types are assigned a fixed emission rate and LAI value independent of season and phenology. The improved model will take advantage of phenology metrics that have been calculated at the EROS data center using 5-yr average NDVI (Normalized Difference Vegetation Index) values and their relation to foliar biomass and LAI. These metrics include onset of greenness, peak of greenness, and end of greenness for each 1-km grid cell. Rates between the onset, peak, and end have also been calculated.

For the eastern United States, the EWDB genus-level data have initially been aggregated to the county level. Individual plot data are available, but statistically the county-level data is more reliable (e.g., Thompson 1989). County labels will be moved to the centroid of the county land area using GIS. Individual genus crown areas associated with each county centroid will then be contoured in order to eliminate false boundaries at county lines that are due to sampling and aggregation procedures. Once a continuous surface of each genus type has been created, it will be gridded to match the 1-km NDVI grid. NDVI metrics will be joined to each cell, thereby establishing a relationship between FIA genus crown areas and phenology. Because the LAI for each genus type at onset, peak, and end of greenness is well established, these values can be associated with the corresponding metric data for each cell. A leaf biomass value at monthly intervals will be calculated based on the dates and rates of emergence and senescence. These biomass values will be used by the model to more accurately reflect the increase and decrease of leaf biomass as the seasons progress.

Results presented in this paper and in Geron et al. (1994) indicate that relatively few genera are estimated to be responsible for a large proportion of BVOC flux (especially in the case of isoprene) on a regional basis. Data describing the geographic extent and seasonal foliation patterns of these major genera (e.g., *Quercus, Liquidambar, Populus, Picea,* and *Pinus*) could further improve BVOC emission models.

Recent work by Sharkey et al. (1992), Kuzma and Fall (1993), and Monson et al. (1993) indicates that there are strong seasonal effects on emission rates, even after leaves have fully emerged in the spring. Isoprene emission rates are very low for several weeks following leaf out, and likewise decrease during leaf senescence. Seasonal vegetation characteristics data are needed to quantify the seasonal variation in leaf area and foliage biomass.

Since biogenic emission estimates for urban areas are important in local air quality assessments, more resolved land use and vegetation cover data are needed for these areas. The FIA data set is of limited utility in urban areas because it contains only samples of forested areas that are at least 0.4 ha in size as determined by aerial photography. While this may include some suburban areas, many sparsely forested areas are omitted. Likewise, the Census of Agriculture data are limited to county level resolution and tracts at least 0.4 ha in size. Urban vegetation surveys for individual urban centers can help address this problem, although such surveys can be prohibitively expensive. Remotely sensed data such as aerial photography and satellite imagery are more useful in describing forest extent. LANDSAT Thematic Mapper (TM) data have been used in spatially complex areas (A. B. Guenther et al., *unpublished manuscript*). Spectral characteristics from TM data can be coupled with survey data available from the FIA or other urban forest inventories derived from aerial photography. Vegetation indices that are correlated with Leaf Area Index (LAI) can also be derived from remotely sensed data. Such inputs into biogenic emissions models will allow determination of sunlight and leaf temperature profiles, both of which control biogenic emission rates. As national survey databases of soils, vegetation, and remotely sensed information become more readily available, the accuracy of national biogenic emission models should improve.

The improvements suggested above may assist regulatory decisions being made to control smog production in many areas of the country. For instance, in southern regions such as the Atlanta area, the relative abundance of high-isoprene-emitting vegetation from a few specific tree types influences the decision to control NO_x sources, if NO_x emissions are indeed found to be the limiting chemical factor in photochemical smog production. On the other hand, northeastern and southwestern regions of the U.S. may be VOC limited. If this is determined to be the case, accurate natural emission models will be needed to verify this and to approximate the impacts of controlling anthropogenic VOC sources.

Midwestern states may be particularly sensitive to biogenic sources of both VOC and NO_x, since intensive agriculture and forest cover area are usually intertwined in these landscapes. Natural and agricultural sources of these compounds must be estimated in an accurate and scientifically valid manner, since the estimates may ultimately affect important regional policy and economic decisions.

Acknowledgments

The information in this document has been funded wholly or in part by the U.S. Environmental Protection Agency. At this time, it has not been subjected to Agency review and approval for publication.

Literature Cited

Baugh, W. M., L. Klinger, and A. Guenther. 1996. Oak tree identification with Landsat TM and SPOT Data for the purpose of estimating biogenic isoprene emissions in Tennessee. Remote Sensing of the Environment, *in press.*

Bureau of the Census. 1989. Census of Population and Housing, Tabulation and Publication Program, p.51. Data Users Services Division, Washington, D.C., USA.

CGC. 1994. Vegetation canopy and landuse characterization. Final report submitted to U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, 18 October 1994. Computer Graphics Center, North Carolina State University, Raleigh, North Carolina, USA.

Geron, C., A. Guenther, and T. E. Pierce. 1994. An improved model for estimating biogenic volatile organic compound emissions from forests in the eastern United States. Journal of Geophysical Research **99**:12773–12792.

Guenther, A. B., R. K. Monson, and R. Fall. 1991. Isoprene and monoterpene emission rate variability: observations with eucalyptus emission rate algorithm development. Journal of Geophysical Research **96**:799–810.

Guenther, A. B., P. R. Zimmerman, P. C. Harley, R. Monson, and R. Fall. 1993. Isoprene and monoterpene emission rate variability: model evaluations and sensitivity analyses. Journal of Geophysical Research **98**:12609–12617.

Guenther, A. B., P. R. Zimmerman, L. Klinger, J. Greenburg, C. Ennis, K. Davis, H. Westburg, G. Allwine, and C. Geron. 1996. Estimates of regional natural volatile organic compound fluxes from enclosure and ambient concentration measurements. Journal of Geophysical Research **1** (D1): 1345–1359.

Guenther, A. B., P. R. Zimmerman, and M. Wildermuth. 1994. Natural volatile organic compound emission rate estimates for U.S. woodland landscapes. Atmospheric Environment **28**:1197–1210.

Hansen, M. H., T. Frieswyck, J. F. Glover, and J. F. Kelly. 1992. The eastwide forest inventory database: user's manual. U.S. Forest Service Department of Agriculture General Technical Report **NC-151**.

Horie, Y., S. Sidwai, and R. Ellefson. 1990. Inventory of leaf biomass and emission factors for vegetation in California's South Coast Air Basin. South Coast Air Quality Management District (SCAQMD) Contract number 90163, Valley Research Corporation, Van Nuys, California, USA.

Kruse, F. A., A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz. 1993. The Spectral Image Processing System (SIPS)—interactive visualization and analysis of imaging spectrometer data. Remote Sensing of the Environment **44**:145–163.

Kuzma, J., and R. Fall. 1993. Leaf isoprene emission rate is dependent on leaf development and level of isoprene synthase. Plant Physiology **101**:435–440.

Lamb, B., E. Allwine, S. Dilts, H. Westburg, T. Pierce, C. Geron, D. Baldocchi, A. Guenther, L. Klinger, P. Harley, and P. Zimmerman. 1993*a*. Evaluation of forest canopy models for estimating isoprene emissions. *In* Proceedings, Regional Photochemical Measurement and Modeling Studies. Volume 2. Results and status of modeling. Air and Waste Management Association, Pittsburgh, Pennsylvania, USA.

Lamb, B., D. Gay, H. Westburg, and T. Pierce. 1993*b*. A biogenic hydrocarbon emission inventory for the U.S. using a simple forest canopy model. Atmospheric Environment **27A**:1673–1690.

Loveland, T. R., J. W. Merchant, D. O. Ohlen, and J. F. Brown. 1991. Development of a land-cover characteristics database for the conterminous U.S. Photogrammetric Engineering and Remote Sensing **57**:1453–1463.

Monson, R., P. Harley, M. Litvak, M. Wildermuth, A. Guenther, P. Zimmerman, and R. Fall. 1994. Environmental and developmental controls over the seasonal pattern of isoprene emission from aspen leaves. Oecologia **99**:260–270.

Novak, J., and T. Pierce. 1993. Natural emissions of oxidant precursors. Water, Air, and Soil Pollution **67**:57–77.

Roselle, S. J. 1994. Effects of biogenic emission uncertainties on regional photochemical modeling of control strategies. Atmospheric Environment **28**:1757–1772.

Sharkey, T. D., P. J. Vanderveer, and F. Loreto. 1992. Biogenic hydrocarbons: measurements on corn and kudzu in 1992. *In* Proceedings of the Conference on Emission Inventory Issues in the 1990s, Durham, North Carolina. Air and Waste Management Association, Pittsburgh, Pennsylvania, USA.

Thompson, M. T. 1989. Forest statistics for Georgia. U.S. Forest Service, Department of Agriculture Research Bulletin **SE-109**.

Tingey, D. T., M. Manning, L. C. Grothaus, and W. F. Burns. 1979. The influence of light and temperature on isoprene emission rates from live oak. Plant Physiology **47**:112–118.

Tingey, D. T., M. Manning, L. C. Grothaus, and W. F. Burns. 1980. Influence of light and temperature on monoterpene emission rates from slash pine. Plant Physiology **65**:797–801.

U.S. Dept. of Commerce. 1987. Census of agriculture. County data file. Volume 1. Geographic data series. Bureau of the Census, Data User Services Division, Washington, D.C., USA.

Williams, E. J., A. B. Guenther, and F. C. Fehsenfeld. 1992. An inventory of nitric oxide from soils in the United States. Journal of Geophysical Research **97**:511–519.