REGIONAL ECOLOGICAL ANALYSIS AND DATA BASE APPLICATIONS

Martha K. Nungesser and Richard J. Olson, Oak Ridge National Laboratory

Abstract

The regional ecological analysis program at Oak Ridge National Laboratory requires the ability to access, manipulate, and analyze large quantities of ecological data. The Geoecology Data Base provides diverse ecological information on a county unit level for the conterminous United States. It includes data on vegetation, soils, bedrock, wildlife, climate, and environmental pollutants. Management of this extensive Data Base involves several stages. SAS is used for compilation of data from existing sources, editing data for errors or inconsistencies, storage and retrieval, and data analysis and display.

Introduction

The Geoecology Data Base, developed at Oak Ridge National Laboratory (ORNL), was designed to facilitate evaluating the impact of environmental energy development at a regional scale (Olson and Strand 1978, Olson, Kleopatek, and Emerson 1979, Olson, Emerson, and Nungesser 1980). The Data Base contains county-level data for the continental United States. The environmental information contained in the Data Base is quite diverse and includes information on soils, bedrock and geological structures, vegetation (both existing and potential), agriculture, forestry, land use, energy, population, climate (rainfall, temperatures, etc.), wildlife (mammals and birds), park and wilderness areas, and environmental pollutants. Data are stored for common spatial units (counties) so data sets are easily combined for display and analysis.

Since 1970, when Olson and Strand (1978) discussed the Geoecology Data Base, it has been updated and extended. Coverage for many data sets has been expanded from a regional to a national level. Data recently added include wilderness areas, national parks, and water quality. Other thematic sectors have been expanded: vegetation now includes both ecoregions (Bailey 1976) and potential natural vegetation (Küchler 1964, Küchler 1966). Climatic data now include temperature and precipitation parameters for state climatic divisions. Recently, the Data Base has been converted to metric units of measure.

Data Base Management

The Geoecology Data Base is managed using SAS (Barr et al. 1979) (Fig. 1). Currently 80 SAS data sets are stored on a single OS disk file occupying 12 million bytes. In addition, several larger data sets are stored offline as SAS tape data sets. Each thematic data set is usually created from a single source such as a map or survey. Most data sets contain observations for the 3071 county units in the conterminous U.S. The data sets contain consistent county codes to facilitate merging the files for analysis. Collectively, the data sets contain over 1000 variables.

A significant facet of the project is maintaining up-to-date and adequate documentation. Throughout the Data Base variable names are uniform and labels are unique and meaningful. Thus, minimal online documentation is available for the casual user via the SAS CONTENTS procedure. SAS does not provide the capability to manipulate the directory information other than with the CONTENTS procedure. To overcome this limitation, the SAS variable names, variable labels and data set labels were extracted from the data set directories to form a dictionary data set. It is stored as a directory file on-line. This dictionary has been supplemented with information defining spatial and temporal aspects of each data set. The dictionary is used both to insure consistency and completeness within the data base and to generate reports describing the contents of the Data Base.

A second major facet of managing the Data Base is data set editing to insure consistency and quality control. SAS's flexible input statements, numeric-character conversions, sorting, merging, updating, and searching capabilities allow new data sets to be reformatted and cross-checked. The Geoecology Data Base contains a number of independent data sets which share variables related to similar county attributes. Merging and comparing such data sets facilitates checking for data set errors and inconsistencies.

The conversion of data from English to
The large size of the Data Base has required reducing the storage length of each variable to minimize storage requirements. Codes for data characteristics are used within the data sets; dictionaries to translate codes (i.e., species names) are stored as separate data sets. An in-house procedure, FORMAT, is used for direct replacement of county codes with the actual county names.

Cartographic display is used both for editing and for displaying analytical results. Errors in a Geoecology data set may result in a spatial incongruity in a map, so detection of certain types of errors can be greatly simplified by visual inspection of the mapped data sets. After data are analyzed, they can be effectively displayed using computer mapping systems. The input for the mapping program is often a system tape or disk file that is created as a formatted output file by SAS. An in-house system for mapping called "EZ-MAP" has been developed at Oak Ridge National Laboratory to produce outlines of the country, 48 states, and county boundaries, as well as other regions. These polygons may be shaded as desired or may be printed with values inside the polygon (see Figs. 2-4). In addition to maps and standard SAS output, procedure DISPLA (Olson and Kumar 1980) provides an interface between SAS and the DISPLA software to produce high quality line plots.

SAS has proven very useful in managing and analyzing the Geoecology Data Base. It is used in all stages of data management, from compiling new data, through editing and storing it, to analysis and display of the output.
SO₂ Impacts on Natural Vegetation

An example of the applications of the Geocology Database is determining regions of the U.S. where vegetation is at risk from sulfur dioxide (SO₂) and other air pollutants. A major source of SO₂ is coal combustion.

As the United States attempts to decrease its dependence on foreign oil imports, coal is becoming more important as a fuel. Much of the available coal is high in sulfur content and when it is burned, several byproducts are released. These residuals are primarily sulfur oxides, nitrogen oxides, and particulates. A byproduct of combustion is ozone, a photochemical oxidant. Sulfur dioxide (SO₂) has been studied extensively because it is known to damage vegetation. Some plants tolerate SO₂ well while others are injured at low levels. Although knowledge of how SO₂ affects particular species is important, it may be more important to know how it affects ecosystems over large areas and over long periods of time.

The impact of air pollution on vegetation is complex. Factors to be considered include different sensitivities of plants to pollutants, timing and duration of exposure, levels of pollutants and interactions with other environmental factors, and overall ecosystem responses. At low levels, sulfur and nitrogen oxides can act as growth stimulants if nitrogen or sulfur are limiting in the environment. At lower levels, some air pollutants may have beneficial effects upon plants. At higher levels, however, these same gases may reduce growth and cause visible injury. Even if effects do not include visible injury to leaves, plant growth or productivity may be reduced, reproduction may be impaired, and resistance to disease may be lowered (West et al. 1980, National Environmental Research Center 1973, Davis and Wihour 1976).

Not all vegetation responds similarly to SO₂ and other air pollutants. Some species are very sensitive while others are resistant (see Fig. 2). Most research has attempted to discern the lower limits of tolerance to short-term (a few hours), higher level doses of SO₂ or other contaminants. Visible injury is often used as the criteria. Tolerance to long-term, lower level exposures to SO₂ has not yet been established for most species.

Research conducted at Oak Ridge National Laboratory using a mathematical model (McLaughlin et al. 1975, West et al. 1980) has predicted long-term (500-year) impacts of chronic air pollution in eastern deciduous forests. The model indicated that responses of particular species in a forest situation may greatly differ from predictions where competition is absent. Factors most important in determining forest responses were the levels of sensitivity of component species and the species interactions. Even though one species may be harmed by air pollution, it may gain competitive advantage in the forest because other species may suffer more. In this way, then, the long-term structure and species composition of a forest may be altered by chronic air pollution stress.

Regional Impact Assessment

To investigate regions of the southern U.S. with conflicts of SO₂ and vegetation, county data on SO₂ levels and SO₂-sensitive vegetation were overlaid. This technique identified natural communities and tree species that run the greatest risk of SO₂ injury. To determine natural communities and trees that risk being harmed by SO₂ levels, we used several data sets in the Geocology Database. Levels of sulfur dioxide and the distribution of sensitive natural vegetation communities were compared in order to assess conflicts between ecological and energy concerns. This useful analysis was conducted for both natural vegetation (Fig. 3) and for eastern white pine, Pinus strobus, which is sensitive to sulfur dioxide and commercially important in the eastern U.S. (Fig. 4).

Vegetation sensitivity was classified based on the dominant species in the community, such as major trees, shrubs, or grass species. Researchers at Argonne National Laboratory (Ballou et al. 1979, Irving and Ballou 1979) surveyed extensive literature to determine the relative SO₂ sensitivity of individual species. They classified general vegetation types based on the sensitivities of the dominant species. The vegetation types they used were those described by A. W. Kuchler (1964) as potential natural vegetation.

The sulfur dioxide data are historic data collected from 1973 to 1976 by the U.S. Environmental Protection Agency for Air Quality Control Regions. Another data set associates these regions with county designations through an index file. The white pine and vegetation data already existed at the county level, so the data were compatible for analysis. Most vegetation types have been reduced in area through conversion of the land to other uses. Adjusted land area was calculated by subtracting current land uses, such as agriculture, pasture, range, and urban land, from the original land area (Klopatek et al. 1979).

Based on the results of our analysis of the South, the areas with the most significant conflict between natural vegetation and elevated SO₂ levels are West Virginia and eastern Kentucky, the area of the Texas-Louisiana border, and scattered portions of Louisiana, Oklahoma, and North and South Carolina (Fig. 3). Changes in natural vegetation communities may have economic consequences, but are primarily ecological in significance. However, injury to white pine, a commercially valuable timber species, may result in economic
VEGETATION WITH LOW TOLERANCE TO SO2 (RELATIVE SENSITIVITY)

Figure 2.

VEGETATION AT RISK FROM SO2 LEVELS

COUNTIES WITH SENSITIVE VEGETATION AND HIGHER SO2 LEVELS
HECTARES OF REMAINING NATURAL VEGETATION

Figure 3.
loss. Areas where white pine and SO2 levels conflict appear in Fig. 4.

If stricter standards of emissions and air pollution control are maintained, SO2 alone will not be as much of a problem for vegetation as it was in the early and middle 1970's. However, there is increasing evidence that SO2 interacts with ozone, a common photochemical product of hydrocarbon and fossil fuel combustion, and this interaction damages vegetation significantly even at low levels (Dochinger and Heck 1969). As the synergistic impacts become better defined, the Geoecology Data Base capabilities can contribute to determining regions where the ozone-SO2 interactions cause significant impacts.

Conclusions

Regional ecological analysis at ORNL has utilized SAS capabilities significantly in both data management and analysis. It reduces the difficulty of handling and analyzing large quantities of ecological data in the Geoecology Data Base. One application of SAS capabilities is environmental conflict analysis such as the sulfur dioxide-vegetation conflicts, where problem areas can be detected. Similar analyses are planned for other air pollutants and other environmental issues.

REFERENCES


