GIS-based solute transport modeling applications: Scale effects of soil and climate data input

JOHN P. WILSON,* WILLIAM P. INSKEEP, JON M. WRAITH, and ROBERT D. SNYDER

*J.P. Wilson, Dep. of Earth Sciences, Montana State Univ., Bozeman, MT 59717; W.P. Inskeep and J.M. Wraith, Dep. of Plant, Soil, and Environmental Sciences, Montana State Univ., Bozeman, MT 59717 R.D. Snyder, Geographic Information and Analysis Center, Montana State Univ., Bozeman, MT 59717. Received 28 Apr. 1995. *Corresponding author (uesjw@msu.oscs.montana.edu).*

Abstract:

The Weather Generator (WGEN) and Chemical Movement through Layered Soils (CMLS) computer models were modified and combined with two sets of soil and climate inputs to evaluate the impact of input data map resolution on model predictions. The basic soil and climate inputs required by WGEN and CMLS were acquired from either: (i) the USDA-NRCS State Soil Geographic Database (STATSGO) database; (ii) the USDA-NRCS (County) Soil Survey Geographic (SSURGO) database; (iii) the Montana Agricultural Potential System (MAPS) database (which divides Montana into approximately 18,000 twenty square kilometer cells and stores more than 200 different land and climate characteristics for each of these cells); and (iv) a series of fine-scale monthly climate surfaces developed by the authors ($0.55 km^2$ cell size) using thin-plate splines, published climate station records, and USGS Digital Elevation Models (DEMs). Fifteen years of daily precipitation and evapotranspiration (ET) values were generated and combined with soil and pesticide inputs in CMLS to estimate the depth of picloram (4-amino-3,5,6-trichloro-2-pyridinecarboxylic acid) movement at the end of the growing season for every unique combination (polygon) of soil and climate in a $320 km^2$ area in Teton County, Montana. Results indicate that: (i) the mean depths of picloram movement predicted for the study area with the SSURGO (county) soils and MAPS (coarse-scale) climate information and the two model runs using the fine-scale climate data were significantly different from the values predicted with the STATSGO (state) soils and MAPS climate data (based on a new variable containing the differences between the depths of leaching predicted with the different input data by soil/climate map unit and testing

whether the mean difference was significantly different from zero at the 0.01 significance level); and (ii) CMLS identified numerous (small) areas where the mean center of the picloram solute front was likely to leach beyond the root zone when the county soils information was used. This last measure may help to identify areas where potential chemical applications are likely to contaminate groundwaters.

Abbreviations:

WGEN, weather generator; CMLS, Chemical Movement through Layered Soils; STATSGO, State Soil Geographic Database; SSURGO, Soil Survey Geographic Database; MAPS, Montana Agricultural Potential System; DEMs, Digital Elevation Models; PFBA, pentafluorobenzoic acid; DFBA, difluorobenzoic acid; BD, bulk density; OC, organic carbon; OM, organic matter; AWC, available water-holding capacity; ET, evapotranspiration; PET, potential evapotranspiration; SAS, Statistical Analysis System.

Approximately 50% of Montana's population obtains its drinking water from groundwater wells. Consequently, the transport of agricultural chemicals through soils into groundwater represents a potential threat to water quality in Montana. The Environmental Management Division of the Montana Department of Agriculture reported detectable quantities of picloram (4-amino-3,5,6-trichloro-2-pyridinecarboxylic acid), 2,4-D [(2,4-dichlorophenoxy) acetic acid], MCPA [(4-chloro-2-methylphenoxy acetic acid], dicamba (3,6-dichloro-2-methoxybenzoic acid), and aldicarb [2-methyl-2 (methylthio) propionaldehyde 0-methylcarbamoyl oxime] in groundwater around the state (DeLuca *et al.*, 1989). These pesticides may be transported to groundwater because of their low solubility, the large amount of pesticide applied, and/or the presence of seasonally high water tables, shallow soils, and high permeability rates (sandy soils).

The GIS-based solute transport models may offer simple, user-friendly tools for identifying sites that are likely to experience pesticide contamination problems, and Wilson *et al.* (1993) have described one such methodology that combines modified versions of the CMLS model (Nofziger and Hornsby, 1986, 1987) and WGEN weather generator (Richardson and Wright, 1984) with the STATSGO (Bliss and Reybold, 1989; Reybold and TeSelle, 1989) and MAPS (Nielsen *et al.*, 1990) databases. Wilson *et al.* (1993) also digitized several weed infestation maps and overlaid them in ARC/INFO with the CMLS model results for picloram to demonstrate how their methodology could be used to illustrate the threat to groundwater posed by current herbicide applications in Teton County, Montana.

The CMLS is a one-dimensional solute transport model that uses a piston flow approach to simulate the vertical movement of selected chemicals through the agricultural root zone on a layer-by-layer basis. Although this model was written primarily as a management and educational tool, it has been tested favorably and used with several different types of input data in many parts of the country (Hornsby *et al.*, 1988; Jones and Hanks, 1988; Mulla *et al.*, 1989; Zhang *et al.*, 1990; Inskeep *et al.*, 1996). The choice of input

data resolution for GIS-based solute transport modeling studies is usually justified in terms of the geographic extent of the problem (study area), and coarse input data (i.e., single values representing large mapping units or raster cells) are often used in place of actual field data in regional studies covering entire watersheds, counties, or states because of computer resource and/or data availability limitations (e.g., Zhang *et al*., 1990; Petach *et al*., 1991; Hutson, 1993; Wilson *et al*., 1993). The consequences of taking this type of approach are not always obvious and more work is needed to document the effects of data quality and scaling issues on model predictions (Merchant, 1994).

Our recent work has tackled these problems at two different scales. In one study, Inskeep *et al*. (1996) compared predicted and observed pentafluorobenzoic acid (PFBA), 2,6-difluorobenzoic acid (DFBA) and dicamba travel times at a single field site near Manhattan, MT. CMLS and LEACHM (Wagenet and Hutson, 1989) predictions were generated using: (i) detailed site-specific measurements (both models); (ii) conductivity and retentivity functions estimated from the USDA-NRCS SSURGO database (LEACHM); and (iii) volumetric water contents estimated from textural data in the SSURGO database and daily precipitation and ET estimated with the WGEN weather generator and MAPS database (CMLS model). Comparison of observed and simulated mean travel times showed that: (i) both LEACHM and CMLS performed adequately with site-specific inputs (the CMLS predicted mean travel times, for example, were within 3.5 to 38% of observed data over two growing seasons under a variety of crop and fallow conditions); and (ii) the CMLS predictions were less sensitive to data input resolution than the LEACHM predictions due in part to the fact that CMLS provides an oversimplified description of transport processes. Inskeep *et al*. (1996) concluded that the use of the SSURGO and MAPS databases with CMLS may provide a reasonable approach for classifying the susceptibility of map units in terms of solute movement, although the potential applicability of this approach is limited to areas with digital copies of the SSURGO (selected counties scattered throughout the country) and MAPS databases (Montana only). The sensitivity of CMLS model predictions to other soil and climate databases was not examined at the Manhattan site.

The current paper reports the results from a second study in which four readily available sets of soil and climate inputs were used in CMLS to estimate the depth of picloram (4-amino-3,5,6-trichloro-2-pyridinecarboxylic acid) movement at the end of the growing season for every unique combination (polygon) of soil and climate in a 326 km^2 area in Teton County, Montana. A large study area was needed to evaluate the effects of using four different combinations of: (i) the USDA-NRCS State Soil Geographic Database (STATSGO) database; (ii) the USDA-NRCS (County) Soil Survey Geographic (SSURGO) database; (iii) the Montana Agricultural Potential System (MAPS) database (which divides Montana into approximately 18,000 twenty km^2 cells and stores more than 200 different land and climate characteristics for each of these cells); and (iv) a series of fine-scale (high spatial resolution; 0.55 km^2 cell size) monthly climate surfaces developed by the authors using thin-plate splines, published climate station records and USGS DEMs on CMLS model predictions. The ARC/INFO GIS was used in this study to: (i) generate and organize the input data for the CMLS model runs; and (ii) prepare maps showing different predicted picloram leaching depths.

Models and data sources

Chemical movement through layered soils (CMLS) model

The CMLS model was developed by Nofziger and Hornsby (1986, 1987) to interactively simulate chemical movement through soil with easily obtained soil, chemical, and weather inputs. CMLS divides the soil into as many as 20 layers and estimates the position of the chemical in the soil at different times using an algorithm first proposed by Rao *et al.* (1976). This algorithm incorporates six key assumptions: (i) all soil water residing in pore spaces participates in the transport process; (ii) water entering the soil redistributes instantaneously to "field capacity"; (iii) water is removed by ET from each layer in the root zone in proportion to the relative amount of water available in that layer; (iv) upward movement of water does not occur anywhere in the soil profile; (v) the adsorption process can be described by a linear, reversible, equilibrium model; and (vi) the half-life for biological degradation of the chemical is constant with time. Nofziger and Hornsby (1986) explain why these assumptions are valid for many soils and when they are likely to be violated.

The soil properties affecting water and chemical movement (soil texture, bulk density, "field capacity" and "permanent wilting point" volumetric water contents, soil organic C content) may vary among the layers, but are assumed to be uniform within each layer. Two chemical properties (the partition efficient (K_{oc}) normalized to soil organic C, and degradation half-life) and the climatic and cultural factors known to affect water and chemical movement (plant rooting depth, daily rainfall, irrigation and ET amounts) are also required by the model. The interactive PC version of CMLS acquired from the model developers was ported to a VMS mainframe and modified by the authors to read from and write to a series of digital data files because of the large variety of sod and weather conditions experienced in the study area. The model itself was not altered and the VMS version required the same inputs for each location (i.e., each soil layer in each unique climate and soil map unit) as the original interactive version.

Four sets of CMLS predictions were prepared with different combinations of soils and weather data. One run used STATSGO soil data and weather data compiled from U.S. Weather Service climate records for the Town of Choteau and the MAPS database. The second run used the STATSGO soil database and a series of higher resolution (0.55 km^2 cell size compared to 20 km^2 for the MAPS database) climate surfaces developed by the authors using thin-plate splines, published climate station data, and USGS 1:250,000-scale DEMs. The final two runs used these two sets of weather data with the higher resolution SSURGO soils data. The weather files were combined with the soil records (one per layer) for each of these runs in ARC/INFO and a series of FORTRAN programs written by the authors. The chemical data file was compiled from published literature (Jury *et al.*, 1987; Nofziger and Hornsby, 1987). The final overlays and maps were prepared with ARC/INF0. The overall approach is summarized in Fig 1 and the individual steps are described in more detail in the four subsections that follow.

Soil data

The STATSGO database was developed at a scale of 1:250,000 by generalizing more detailed county soil survey maps and is used primarily for regional resource planning management, and monitoring (Reybold and TeSelle, 1989). This database divides the landscape into map units (i.e., polygons) and percentage composition of the soil series ($\leq$21 soil series) that occur in these general map units is recorded. A series of geocodes is provided for linkage with the nation Soil Interpretations Record database that provides detailed information on various properties for each soil, usually ranges of high and low values by soil layer (Bliss and Reybold, 1989). The soils that compose each map unit will have generally formed in similar kinds of parent material and have a similar repeating pattern of landforms, but will vary in one or more characteristics (texture, available water-holding capacity, etc.). This arrangement means that simple maps cannot be used to present information on a specific soil series attribute because there is no map delineation for the locations of individual soil series that make up each STATSGO map unit.

Bliss and Reybold (1989) have described the general process that should be used for linking soil attributes to soil maps and also provide a series of examples to illustrate the wealth of attribute data that can be accessed via the STATSGO map units. Our approach, summarized in Fig. 1, generally follows their schematic diagram linking GIS and the two STATSGO database components.

Our work with the STATSGO database started with the Teton County ARC/INFO coverage used by Wilson *et al.* (1993). Four contiguous polygons (representing two different STATSGO map units) that covered parts of the USGS 1:24,000-scale Pishkun Reservoir and nine adjacent map sheets were extracted and used to define the study area (Fig. 2a). The SSURGO database was constructed by the authors by one or other of two methods: (i) converting USDA-NRCS soil map unit boundaries stored as Digital Line Graph (DLG) files to a series of ARC/INFO line coverages or (ii) scanning USDA-NRCS mylar separates (map sheets) containing the soil map unit boundaries and converting these data to a series of ARC/INFO line coverages. The ARC/INFO coverages were then edited, labeled, edgematched, and map joined to form one coverage and this product was clipped to match the extent of the four STATSGO polygons that defined the study area (Fig. 2b).

SSURGO soil maps are prepared in the field using aerial photographs as a map base by USDA-NRCS staff. These maps, which are usually organized by county, offer increased spatial resolution compared to the STATSGO maps (compare Fig. 2a and 2b) and are used primarily for farm and ranch conservation planning and other local resource planning and management activities (Reybold and TeSelle, 1989). However, this database is similar to the STATSGO database in that it divides the landscape into map units and the percentage composition of the soil series ($\leq$5 soil series) that occur in these map units is recorded. A series of geocodes is provided for linkage with the national Soil Interpretations Record database discussed above. Hence, we were able to use the same programs for the STATSGO and SSURGO model runs because of the structural similarities between these two databases.
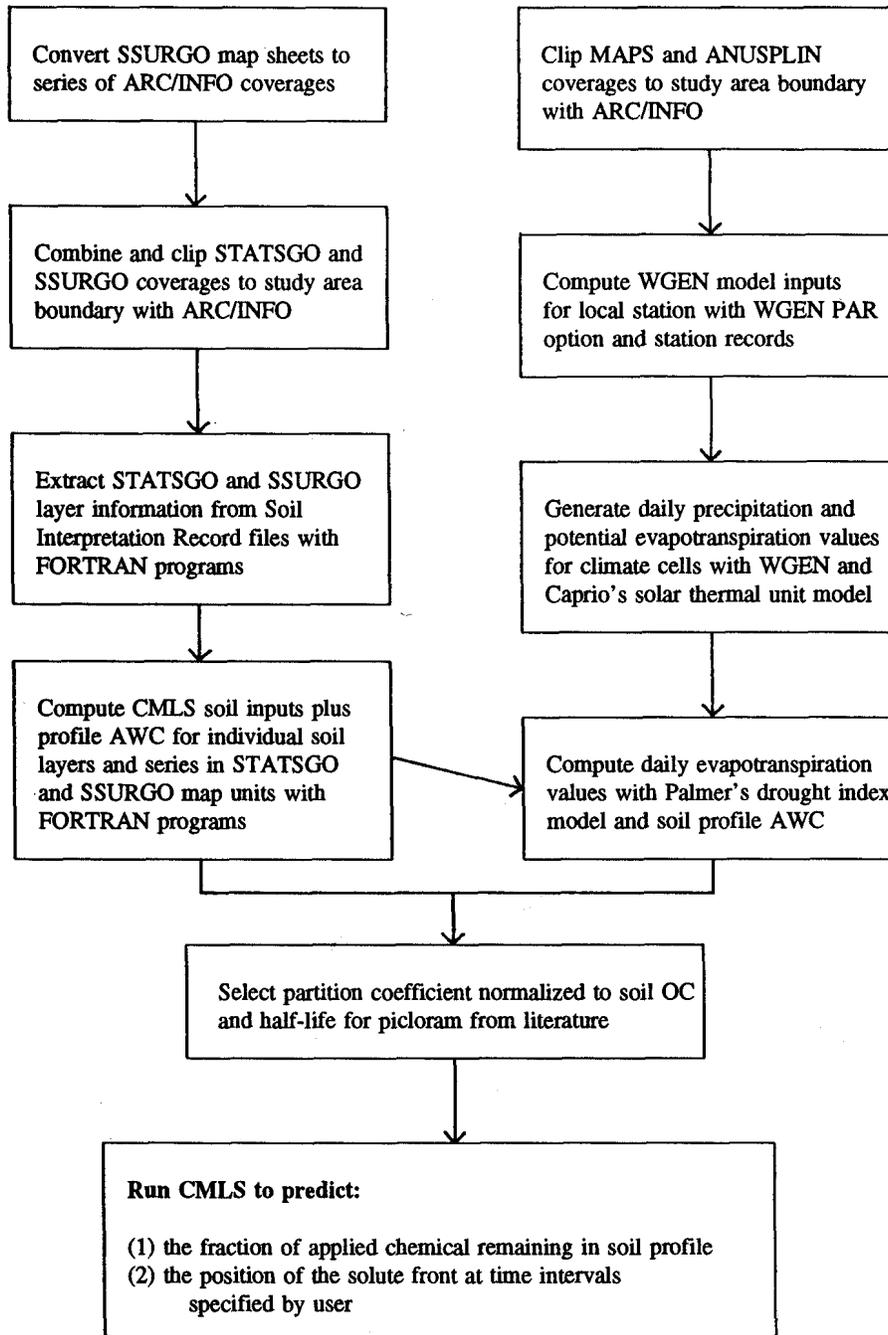
```
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Convert SSURGO map sheets to│        │ Clip MAPS and ANUSPLIN      │
│ series of ARC/INFO coverages│        │ coverages to study area     │
│                             │        │ boundary with ARC/INFO      │
└──────────────┬──────────────┘        └──────────────┬──────────────┘
               ↓                                       ↓
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Combine and clip STATSGO and│        │ Compute WGEN model inputs   │
│ SSURGO coverages to study   │        │ for local station with WGEN │
│ area boundary with ARC/INFO │        │ PAR option and station      │
│                             │        │ records                     │
└──────────────┬──────────────┘        └──────────────┬──────────────┘
               ↓                                       ↓
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Extract STATSGO and SSURGO  │        │ Generate daily precipitation│
│ layer information from Soil │        │ and potential               │
│ Interpretation Record files │        │ evapotranspiration values   │
│ with FORTRAN programs       │        │ for climate cells with WGEN │
│                             │        │ and Caprio's solar thermal  │
│                             │        │ unit model                  │
└──────────────┬──────────────┘        └──────────────┬──────────────┘
               ↓                                       ↓
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Compute CMLS soil inputs    │        │ Compute daily               │
│ plus profile AWC for        │───────▶│ evapotranspiration values   │
│ individual soil layers and  │        │ with Palmer's drought index │
│ series in STATSGO and SSURGO│        │ model and soil profile AWC  │
│ map units with FORTRAN      │        │                             │
│ programs                    │        │                             │
└──────────────┬──────────────┘        └──────────────┬──────────────┘
               └──────────────────┬───────────────────┘
                                  ↓
              ┌─────────────────────────────────────┐
              │ Select partition coefficient        │
              │ normalized to soil OC and half-life │
              │ for picloram from literature        │
              └──────────────────┬──────────────────┘
                                 ↓
              ┌─────────────────────────────────────┐
              │ Run CMLS to predict:                │
              │                                     │
              │ (1) the fraction of applied chemical│
              │     remaining in soil profile       │
              │ (2) the position of the solute front│
              │     at time intervals               │
              │         specified by user           │
              └─────────────────────────────────────┘
```

Fig. 1. Schematic diagram showing linkages between individual models and databases.

A FORTRAN program developed by Wilson *et al.* (1993) was modified and used to: (i) read the STATSGO/SSURGO Component data files and write the STATSGO/SSURGO soil map unit codes and soil series sequence (component) numbers, codes and names to a new file; (ii) combine this file with the STATSGO/SSURGO Layer files based on soil mapping unit and soil series codes; and (iii) extract and write the soil

layer attributes required to calculate CMLS soil inputs to a new file. A second FOR-TRAN program was used to calculate the actual CMLS model inputs. This program checked the layer records for missing data and computed CMLS inputs for each layer and soil series with complete information. This approach meant that a small number of STATSGO and SSURGO map units were rejected because of incomplete information and excluded from the model comparisons discussed below. These polygons covered <10 km^2 (3%) of the study area in all instances.
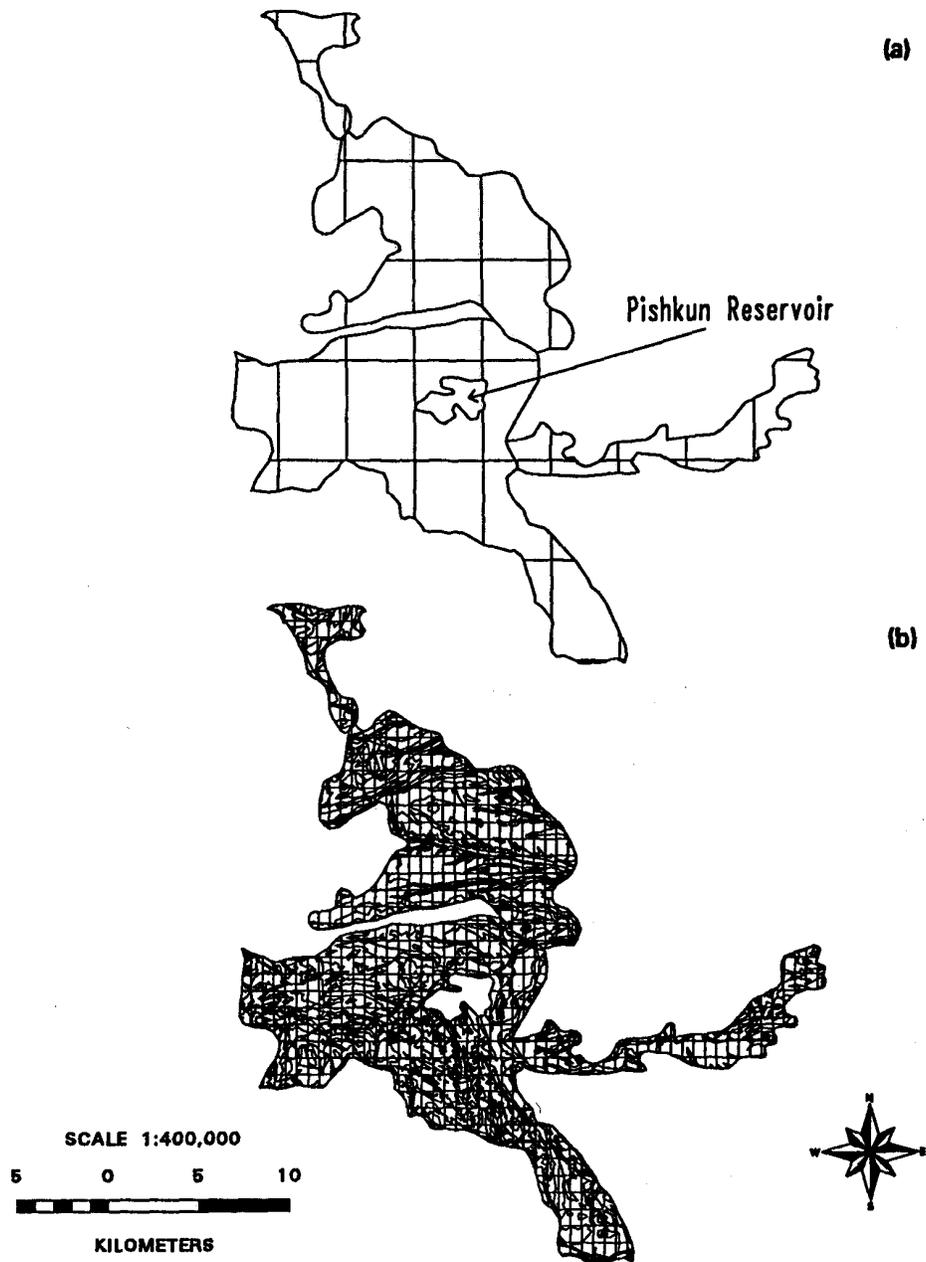
Fig. 2. Maps of study area showing (a) MAPS cells overlaid on STATSGO mapping units, and (b) ANUSPLIN cells overlaid on SSURGO mapping units.

Eight soil inputs (percents clay, silt, sand; bulk density; organic matter content; wilting point, field capacity, and saturation: volumetric water contents) were computed for the VMS version of the CMLS model as follows.

Bulk density (BD) and percent clay were obtained by computing the midpoints of the ranges specified for each soil layer in the SSURGO and STATSGO databases. Percent sand was computed as:

$$\% \text{ sand} = 100[(n_{10} - n_{200})/n_{10}] \tag{1}$$

where n_{10} and n_{200} are the average percentages of soil passing through the no. 10 and 200 soil sieves, respectively. The percent clay and percent san10d were then subtracted from 100 to estimate percent silt. Bulk densities were estimated from soil texture information and the Grossman/Baumer ternary diagram (reproduced in Wilson *et al.*, 1992) for those soil layers missing recorded BD ranges.

Surface and subsurface soil layer organic C contents (OC) were computed by a combination of methods. The average organic matter content (OM) was multiplied by 0.5 to obtain OC values for surface layers. Because SSURGO and STATSGO only report OM contents for surface layers, subsurface OC values were estimated by the following methods.

The Montana Soil Pedon Database (Jersey and Nielsen, 1992) was used to access laboratory determined soil OC values by profile depth for more than 60 agricultural soils in Montana. Linear regression analysis of soil OC values with profile depth produced Eq. [2] ($R^2 = 0.71$), which was used to compute subsurface OC values for soils with surface OC values exceeding 0.8%:

$$OC_i = 1.71 + 0.117(OC_s) - 0.367[\ln(D_i)] \tag{2}$$

where OC_i is the percent organic C content of the *i*th layer, OC_s. is the average percent organic C content of the surface layer, and D_i is the average depth of the *i*th layer in meters times 100 divided by profile depth in meters (Wilson *et al.*, 1993).

For soils with surface OC values below 0.8 %, two equations developed by Jury *et al.* (1987) were found to best estimate subsurface OC values (based on comparisons with the Montana Soil Pedon data) (Wilson *et al.*, 1993). These equations, originally developed to estimate subsurface microbial populations, are:

$$OC_2 = OC_{s,}e^{-\gamma(Z-L)} \tag{3}$$

$$OC_i = OC_s\, e^{-\gamma(H-L)} \tag{4}$$

where OC_2 is the organic C content of the second layer, γ was treated as a depth constant of 3 m^{-1} for qualifying soils in the study area, Z is the average depth of the second layer in meters, L is the depth of the surface layer in meters, OC_i is the organic C content in hori-

zons below the second layer, and H is the depth from the surface to the bottom of the ith layer in meters.

Standard equations reported by Rawls and Brakensiek (1989) were used to compute the three volumetric water contents needed by CMLS as follows:

$$\theta_{v.1} = (n_{10}/100)*[0.4118 - 0.0030(\%sand) + 0.0023(\%clay) + 0.0317(\%OM)] \tag{5}$$

$$\theta_{v.15} = (n_{10}/100)*[0.026 + 0.005(\%clay) + 0.0158(\%OM)] \tag{6}$$

$$\theta_{vSAT} = (n_{10}/100)*[0.7899 - 0.0037(\%sand) + 0.01(\%OM) - 0.1315(BD)] \tag{7}$$

where $\theta_{v.1}$, $\theta_{v.15}$, and θ_{vSAT} are the field capacity (-0.01 MPa), wilting point (- 1.5 MPa), and saturation volumetric water contents, respectively, n_{10} is the average percentage of soil particles passing through a no. 10 soil sieve, %OM is average percent organic matter content, and BD is the average bulk density in Mg m^{-3}.

These eight soil parameters were computed for each soil layer and printed to a file for input to the VMS version of CMLS. The same FORTRAN program was also used to compute the available water-holding capacity (AWC) of the entire soil profile for each soil series as this information was needed to compute the daily ET values required by CMLS. The profile AWC values were estimated by multiplying the depths of the individual soil layers (cm) by their AWCs (cm of water per cm of soil) and summing these sub-totals.

Weather data

The first two CMLS model runs required daily precipitation and ET totals for the 53 twenty km^2 MAPS cells that overlapped the study area (Fig. 2a). The MAPS cell boundaries were converted to a polygon coverage, and attribute tables containing mean monthly precipitation totals and minimum/maximum temperatures were transferred from the MAPS database (Nielsen *et al.*, 1990) to ARC/INFO attribute tables. The STATSGO and SSURGO soil coverages were then overlaid with the MAPS coverage in ARC/INFO to produce new coverages containing 58 and 1595 climate and soil map units (polygons), respectively. Several climate and soil map unit combinations appeared more than once, so that the MAPS/STATSGO and MAPS/ SSURGO coverages divided the study area into 51 and 548 unique climate and soil polygons, respectively.

The fine-scale climate data used in this study were extracted from a series of mean monthly and annual climate surfaces that have been developed by one of the authors for Montana and the Dakotas with ANUSPLIN (Hutchinson, 1995). This program fits multi-dimensional Laplacian thin-plate splines to spatial data with the degree of smoothing determined by minimizing the predictive error of the surface through generalized cross-validation. Latitude, longitude, and elevation usually serve as independent variables for

this routine, which has been successfully applied to several areas, including two USGS 1: 250,000-scale map sheets in Montana (Custer *et al.*, 1996). The ANUSPLIN program provides an objective and repeatable method for generating spatially variable estimates of mean monthly and annual precipitation and temperature from DEMs and published climate station data.

The final two CMLS runs required the same daily precipitation and ET values as the earlier MAPS runs for the 719 0.55 km^2 ANUSPLIN cells that overlapped the study area (Fig. 2b). The STATSGO and SSURGO soil coverages were overlaid with the ANUSPLIN coverage in ARC/INFO to produce new coverages containing 758 and 4448 climate and soil map units, respectively. Some of these climate and soil map units also appeared more than once, so that the ANUSPLIN/STATSGO and ANUSPLIN/SSURGO coverages divided the study area into 748 and 3061 unique climate and soil polygons, respectively.

The WGEN weather generator, which provides simulated daily precipitation, minimum temperature, maximum temperature, and solar radiation values with the same statistical characteristics as the actual weather for user-specified periods at given locations, was modified and expanded to generate the weather parameters required by CMLS. The WGEN PAR option was used to generate parameters from the long-term daily records available for the Choteau (in Teton County) and Great Falls (the nearest station with daily total solar radiation data) climate stations maintained by the U.S. Weather Service. WGEN PAR reads daily precipitation, minimum temperature, maximum temperature and total solar radiation values, and writes the generation parameters (probabilities, long-term averages, etc.) required by WGEN to another data file.

These data were used with monthly means to generate the daily precipitation but not the daily ET values needed by CMLS. Two additional steps were required to generate spatially variable daily ET values: (i) daily potential evapotranspiration (PET) was estimated with the solar thermal unit model developed by Caprio (1971) using the daily mean temperature and total solar radiation values calculated with WGEN as inputs; and (ii) daily ET values were estimated from these PET values and the profile AWC for every unique combination of climate and soil mapping unit. A separate FORTRAN program was written for this last step as the ARC/INFO "joinitem" command could not achieve the desired geographic linkage because the common (relate) attribute was not unique in at least one of the two files. The PET values were then read one cell at a time and the appropriate soil parameters for that cell were accessed sequentially and used with the *Et al*gorithm from the Palmer (1965) drought index model to generate 15 years of daily ET data for each accessed record. The ET data were written to a file that also contained the daily precipitation data generated by WGEN. This file served as the climate input for the CMLS model runs.

Chemical data

The chemical attributes required for CMLS input include the partition coefficient normalized to soil organic C (K_{oc}, L kg^{-1}) and the half-life ($t_{1/2}$, days). For picloram, we used a K_{oc} of 48 L kg^{-1} and a surface half-life value of 100 d (Jury *et al.*, 1987; Nofziger and

and Hornsby, 1987). For soil horizons below the surface layer we used a biodegradation depth function developed by Jury *et al.* (1987):

$$t_{1/2}(z) = t_{1/2}(0)e^{\gamma(Z-L)} \qquad\qquad [8]$$

where $t_{1/2}(z)$ is the half-life with respect to depth z in meters, $t_{1/2}(0)$ is the half-life at the soil surface, γ was treated as a depth constant equal to 1.5 m^{-1} for this study, Z is the average depth of the ith layer in meters, and L is the depth of the surface layer in meters. Using a γ value of 1.5 m^{-1} resulted in maximum $t_{1/2}$ values 10 times the surface horizon values. This relationship is generally consistent with the effects of soil depth on herbicide degradation rates ($t_{1/2}$) observed in other studies (Moorman and Harper, 1989; Pothuluri *et al.*, 1990; Boesten and van der Linden, 1991).

Results and discussion

Four sets of CMLS model predictions were prepared with different input data to evaluate the impact of input data map resolution on model outcomes. Wilson *et al.* (1993) noted how information on both the spatial variability and depth of movement of applied chemicals is required to evaluate the threat to groundwater posed by current herbicide applications and the results presented below address both of these types of measurements. The CMLS model predictions prepared with the MAPS and STATSGO databases are examined first because these model runs used the most generalized climate and soil information and the results from these model runs provide reference values against which the predictions from the other model runs can be compared.

The two pairs of STATSGO mapping units contained 12 and 15 soil series, respectively and the GIS overlays of MAPS cells and STATSGO map units divided the study area into 58 polygons (Fig. 2a). Three polygons labeled as water (Pishkun Reservoir) and one polygon with missing soil information in the STATSGO database were excluded from the model comparisons. CMLS was used to predict the position of the picloram solute at the end of the growing season for each of the remaining 54 polygons averaged over all soil series (weighted by area) and all 15 weather years (Table 1). The study area mean of 18.4 cm predicted with CMLS indicates a moderate leaching hazard, and the range (6.7 cm), weighted standard deviation (3.8) and coefficient of variation (0.2) provide measures of the spatial variability for this set of CMLS model predictions.

The MAPS/SSURGO model runs combined 53 MAPS cells and 102 SSURGO soil map units (198 soil series) into 1595 climate and soil polygons, so that each of the MAPS/STATSGO polygons reproduced in Fig. 2a was divided into approximately 30 MAPS/SSURGO polygons. The CMLS model predicted more solute movement and greater spatial variability (measured in terms of minimum and maximum predicted depths, the weighted standard deviation and the coefficient of variation) with the MAPS and SSURGO information compared to the MAPS and STATSGO data inputs (Table 1).

Table 1. Predicted depth of picloram solute front at end of growing season for different Chemical Movement through Layered Soils (CMLS) model runs.

Model run	No. of polygons†	Predicted solute front depth, cm			
		Min.	Max.	Mean‡	SD§
MAPS/STATSGO	54	15.26	21.94	18.37	3.76
MAPS/SSURGO	548	11.77	631.74	20.52	5.03
ANUSPLIN/STATSGO	749	14.59	19.84	16.51	2.57
ANUSPLIN/SSURGO	3061	11.79	562.56	19.14	4AS

†Number of unique climate cells and soil mapping units (see Fig. 2a and 2b for Montana Agricultural Potential System/State Soil Geographic Database (MAPS/STATSGO) and splinted climate data/Soil Survey Geographic Database (ANUSPLIN/SSURGO) examples).
‡CMLS predictions averaged over all soil series (weighted by areal extent) and 15 weather years.
§Weighted standard deviations.

The MAPS/SSURGO and MAPS/STATSGO predictions were then aggregated using the overlay functions in ARC/INFO and the difference between each of the MAPS/SSURGO and MAPS/STATSGO weighted mean depths (in each unique MAPS/STATSGO/SSURGO polygon) was calculated. The Statistical Analysis System (SAS) was used to calculate a t statistic and probability value for the null hypothesis that the mean difference is equal to zero (SAS Institute, 1988). The results indicate that the average difference is significantly different from zero at the 0.01 level of significance (Table 2). The first of the three frequency histograms reproduced in Fig. 3a shows the differences between the individual MAPS/ STATSGO and MAPS/SSURGO polygons in 5-cm depth increments (-7.5 to -2.5 cm, -2.5 to 2.5 cm, 2.5 to 7.5 cm, 7.5 to 12.5 cm, etc.) weighted by area. Negative values in Fig. 3 indicate that the MAPS/SSURGO, ANUS-PLIN/STATSGO, and ANUSPLIN/SSURGO input data predicted lower leaching depths than the MAPS/ STATSGO input data. The results reproduced in Fig. 3a show that: (i) the two sets of CMLS predictions varied by less than ± 2.5 cm and ± 7.5 cm. in 57 and 94.9% of the study area, respectively; and (ii) the predicted depth of the solute front exceeded 100 cm in 0.08% (0.24 km^2) of the study area when the SSURGO soils information was used in place of the STATSGO soils information.

Table 2. T test results comparing the Montana Agricultural Potential System/Soil Survey Geographic (MAPS/SSURGO), ANUSPLIN/STATSGO, and ANUSPLIN/SSURGO CMLS model predictions with the MAPS/STATSGO CMLS model predictions.

Model run	Mean depth, cm	Std. error	T	Prob > |T|
MAPS/SSURGO	2.52	0.43	5.80	0.0001
ANUSPLIN/STATSGO	-1.37	0.19	-7.12	0.0001
ANUSPLIN/SSURGO	1.65	0.47	3.49	0.0010

The ANUSPLIN/STATSGO model runs combined 719 ANUSPLIN cells and the four STATSGO polygons into 758 climate and soil polygons. These CMLS model runs predicted more spatially uniform and generally less solute movement than the MAPS/STATSGO model runs (Table 1). This result occurred because the ANUSPLIN climate data differed from the MAPS data in two important respects: (i) the values used

for monthly means (ANUSPLIN inputs were interpolated by the authors with a model that incorporated 1961 to 1990 climate station data, latitude, longitude, and elevation whereas the MAPS values had been estimated by hand (?) from published data and local knowledge some years earlier); and (ii) the spatial resolution (the ANUSPLIN and MAPS cells measured 30 arc s and 3 min on a side, respectively). The average difference was again significantly different from zero (Table 2), although the histogram reproduced in Fig. 3b shows that the differences between the MAPS/STATSGO and ANUS-PLIN/STATSGO predictions were less than ± 2.5 cm in 66.2 % of the study area and that these differences were bounded by -7.5 cm (i.e., ANUSPLIN/STATSGO predicted depths < MAPS/STATSGO predicted depths) and + 2.5 cm. (i.e., ANUSPLIN/STATSGO predicted depths > MAPS/STATSGO predicted depths).

The final CMLS models used the ANUSPLIN climate and SSURGO soils information as inputs. These data sources divided the study area into 4448 polygons because many of the 3061 unique polygons occurred more than once in the study area (Fig. 2b). The CMLS model predictions generated with these data sources were more variable and larger in magnitude than those produced with the MAPS climate and STATSGO soils information (Table 1). The average difference in area-weighted solute movement was again significantly different from zero (Table 2) even though the effects of substituting SSURGO in place of STATSGO soils data and ANUSPLIN in place of MAPS climate data tended to cancel each other out. The histogram reproduced in Fig. 3c has the same general form as the histogram generated with the MAPS/SSURGO data (Fig. 3a), and reaffirms the importance of using SSURGO soils information to identify those locations where the depth of solute movement is likely to exceed the rooting depth.

The results described above show: (i) how a GIS can be used to compile several different sets of soil and climate input data; and (ii) the impact of soil and climate input data map resolution on CMLS model predictions. The ARC/INFO GIS and a series of FORTRAN programs written by the authors were used to integrate the soils information vertically over the layers in each soil series in each soil/climate map unit (Fig. 1) and horizontally based on the percentage of that soil/climate map unit occupied by that soil series. This approach generated large numbers of map units (polygons) and retained the spatial variability inherent in the input data (Fig. 2) in the model predictions (Tables 1 and 2). Colored maps of the CMLS model predictions could have been produced relatively easily with ARC/INFO and used to illustrate this variability as well. The sensitivity of the CMLS model predictions to the methods and map resolution used to represent the soil and climate input data (Fig. 3) demonstrates the need to exercise care in choosing model inputs for solute transport modeling applications.

Further work of this type is needed to: (i) extend the results obtained in this study to other models and/or environments; and (ii) determine the accuracy of model predictions based on different input data map resolutions compared to observed leaching behavior. Use of the GIS approach discussed in this study with other models may yield significantly different model predictions of mean travel times because of differential effects of parameter estimation routines on model calculations. For example, Inskeep *et al.* (1996) showed that LEACHM predicted mean travel times were significantly higher (slower transport) than CMLS values when estimation routines (e.g., pedotransfer functions) were used to generate necessary model input parameters. They noted that estimated saturated
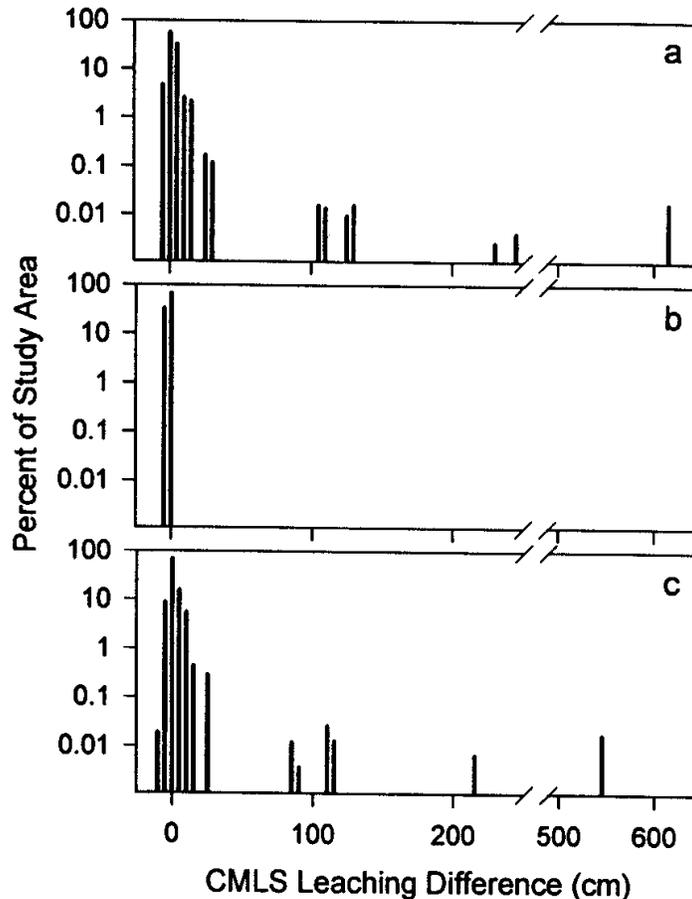
Fig. 3. Histograms showing differences in depth of solute movement between CMLS model predictions based on MAPS/STATSGO databases and (a) MAPS/SSURGO, (b) ANUS-PLIN/STATSGO, and (c) ANUSPLIN/SSURGO databases.

hydraulic conductivities (necessary for LEACHM) obtained from regression equations based on texture and porosity were lower than measured values, and resulted in higher mean travel times of several nonreactive tracers than observed. Although LEACHM is a more detailed process model than CMLS, it requires a correspondingly greater number of input parameters. Consequently, LEACHM requires the estimation of a greater set of input parameters for climate/soil map unit applications, and what might be gained in process description may be sacrificed in the estimation of required input parameters. Results from the current study and those obtained by Inskeep *et al.* (1996) show that CMLS is a useful transport model for GIS applications, and that predicted mean travel times over several different weather years could be used to classify map units based on their susceptibility to solute transport.

Summary

The CMLS model was chosen for this study because of favorable reviews and past work by Zhang *et al*. (1990) and Wilson *et al*. (1993) in linking this simplified one-dimensional solute transport model to GIS data layers. Four sets of CMLS model picloram-leaching predictions were prepared with different input data to evaluate the sensitivity of model outcomes to different input data map scales. The CMLS model predictions prepared with the most generalized climate (MAPS) and soil (STATSGO) information provided reference values against which the predictions from the other model runs were compared. The MAPS/SSURGO (coarse climate and fine-scale soil data), ANUSPLIN/STATSGO (fine-scale climate and coarse-scale soil data), and ANUSPLIN/SSURGO (fine-scale climate and soil data) results were combined with the MAPS/STATSGO predictions using the overlay functions in ARC/INFO and the difference between each of the MAPS/SSURGO, ANUSPLIN/ STATSGO; and ANUSPLIN/SSURGO weighted mean depths and the MAPS/STATSGO weighted mean depths (for each unique polygon) was calculated. The Statistical Analysis System (SAS) was used to calculate a t statistic and probability value for the null hypothesis that the mean difference is equal to zero. The average difference was significantly different from zero in all three instances at the 0.01 level of significance. The two CMLS model runs that used the SSURGO soil inputs in place of the STATSGO soil inputs also identified numerous small areas where the predicted depth of the picloram solute front exceeded 100 cm. These results demonstrate that the CMLS model predictions vary with the choice of climate and soil inputs and that high resolution SSURGO soil information is needed if the goal is to identify those areas where potential chemical applications are likely to contaminate groundwater. Additional research is required to extend these results to other models and/or environments.

Acknowledgments

Literature cited

Bliss, N.B., and W.U. Reybold. 1989. Small-scale digital soil maps for interpreting natural resources. J. Soil Water Conserv. 44:30-34.

Boesten, J.J., and A.M. van der Linden. 1991. Modeling the influence of sorption and transformation on pesticide leaching and persistence. J. Environ. Qual. 20:425-435.

Caprio, J.M. 1971. The solar-thermal unit theory in relation to plant development and potential evapotranspiration. Montana Agric. Exp. Stn. Circ.251, Bozeman, MT.

Custer, S.G., P. Farnes, J.P. Wilson, and R.D. Snyder. 1996. A comparison of hand and spline-drawn precipitation maps for mountainous Montana. Water Resour. Bull. 32(2). (in press.)

DeLuca, T., J. Larson, L. Thorma, and G. Algard. 1989. A survey of pesticide residues in groundwater in Montana. Montana Dep. Agric., Environ. Manag. Div. Tech. Rep. 89-1, Helena.

Hornsby, A.G., K.D. Pennell, R.E. Jessup, and P.S.C. Rao. 1988. Modeling environmental fate of toxic organic chemicals in soils. p. 7-39. *In* Florida Dep. Environ. Regulation Final Rep. WH-1491 Tallahassee, FL.

Hutchinson, M.F. 1995. Interpolating mean rainfall using thin plate smoothing splines. Int. J. Geogr. Info. Syst. 9(4):385-403.

Hutson, J.L. 1993. Applying one-dimensional deterministic chemical fate models on a regional scale. Geoderma 60:201-212.

Inskeep, W.P., J.M. Wraith, J.P. Wilson, R.D. Snyder, R.E. Macur, and H.M. Gaber. 1996. Input parameter and model resolution effects on predictions of solute transport. J. Environ. Qual. 25:453-462 (this issue).

Jersey, J.K., and G.A. Nielsen. 1992. Montana soil pedon database reference manual. Montana Agric. Exp. Stn. Rep. SR-043, Bozeman.

Jones, R.L., and R.J. Hanks. 1988. Review of unsaturated zone leaching models from a user's perspective. p. 129-139. *In* Int. Symp. on Water Quality Modeling of Agricultural Non-Point Sources, Logan, UT. 19-23 June 1988. USDA-Agric. Res. Serv., Washington, DC.

Jury, W.A., D.D. Focht, and W.J. Farmer. 1987. Evaluation of pesticide groundwater pollution potential from standard indices of soil chemical adsorption and biodegradation. J. Environ. Qual. 16:422-428.

Moorman, T.B., and S.S. Harper. 1989. Transformation and mineralization of metribuzin in surface and subsurface horizons of a Mississippi Delta soil. J. Environ. Qual. 18:302-306.

Merchant, J.W. 1994. GIS-based groundwater pollution hazard assessment: A critical review of the DRASTIC model. Photogram. Eng. Remote Sens. 60:1117-1127.

Mulla, D.J., H.H. Cheng, G. Tuxhorn, and R. Sounhein. 1989. Management of ground water contamination in Washington's Columbia Basin. Washington Water Res. Center Rep. 72, Pullman.

Nielsen, G.A., J.M. Caprio, P.A. McDaniel, R.D. Snyder, and C. Montague. 1990. MAPS: A GIS for land resource management in Montana. J. Soil Water Conserv. 45:450-453.

Nofziger, D.L., and A.G. Hornsby. 1986. A microcomputer-based management tool for chemical movement in soil. Appl. Agric. Res. 1:50-56.

Nofziger, D.L., and A.G. Hornsby. 1987. CMLS: Chemical Movement through Layered Soils Model Users Manual. Univ. of Florida, Gainesville.

Palmer, W.C. 1965. Meteorological drought. Weather Bureau Res. Paper 45. U.S. Dep. of Commerce, Washington, DC.

Petach, M.C., R.J. Wagenet, and S.D. DeGloria. 1991. Regional water flow and pesticide leaching using simulations with spatially distributed data. Geoderma 48:245-269.

Pothuluri, J.V., T.B. Moorman, D.C. Obenhuber, and R.D. Wauchope. 1990. Aerobic and anaerobic degradation of alachlor in samples from a surface to groundwater profile. J. Environ. Qual. 19:525-530.

Rao, P.S.C., J.M. Davidson, and L.C. Hammond. 1976. Estimation of nonreactive and reactive solute front locations in soils. p. 235-241. *In* Proc. of Hazardous Wastes Res. Symp. USEPA-Rep. 600/19-76-015. USEPA, Tucson, AZ.

Rawls, W.J., and D.L. Brakensiek. 1989. Estimation of soil water retention and hydraulic properties. p. 275-300. *In* Unsaturated flow in hydrologic modeling: Theory and practice. Kluwer Academic, Amsterdam.

Reybold, W.U., and G.W. TeSelle. 1989. Soil geographic data bases. J. Soil Water Conserv. 44:28-29.

Richardson, C.W., and D.A. Wright. 1984. WGEN: A model for generating daily weather variables. USDA-ARS Rep.ARS-8, Washington, DC.

SAS Institute. 1988. SAS user's guide. Statistics. Version 6.03 ed. SAS Inst., Cary, NC.

Wagenet, R.J., and J. Hutson. 1989. LEACHM: Leaching estimation and chemistry model. Water Resource Inst., Cornell Univ., Ithaca, NY.

Wilson, J.P., K.E.S. Gerhart, G.A. Nielsen, and C.M. Ryan. 1992. Climate, soils, and crop yield relationships in Cascade County, Montana. Appl. Geogr. 12:261-279.

Wilson, J.P., W.P. Inskeep, P.R. Rubright, D. Cooksey, J.S. Jacobsen, and R. D. Snyder. 1993. Coupling Geographic Information Systems and Models for Weed Control and Groundwater Protection. Weed Technol. 7:255-264.

Zhang, H., D.L. Nofziger, and C.T. Haan. 1990. Interfacing a root-zone transport model with GIS. ASAE Paper 903034. ASAE, St. Joseph, MI.