Simulation of Within–Field Yield Variability in a Four–Crop Rotation Field Using SSURGO Soil–Unit Definitions and the EPIC Model

J. F. Perez–Quezada, J. Cavero, J. Williams, A. Roel, R. E. Plant

Abstract. *Soil data were collected from a 30 ha commercial field using a 60 m sampling grid. Monitored yield data were also collected in this field between 1996 and 1999, when it had a wheat–processing tomato–bean–sunflower crop rotation. A comparison between SSURGO–NRCS soil–unit definition and field–measured soil data showed that in this field the former are a good approximation and starting point for precision agriculture studies and management. In a second test, the EPIC model, using the SSURGO database soil type definitions, was found to reproduce the yield variability within this field with reasonable accuracy. The model's performance was not as good when tested with data from soil samples, apparently due to the way EPIC simulates water holding capacity from texture information and the lack of some key variables (not sampled), such as water content at field–capacity (FC), wilting–point (WP), and soil saturated conductivity. A set of runs was performed to simulate the yield at 13 point–locations in the field using FC, WP, and bulk density. Although the accuracy of the simulation did not improve greatly, the model reproduced the yield trend of two of the crops (wheat and sunflower).*

***Keywords**. EPIC model, Management zones, Precision agriculture, Spatial variability, SSURGO data, Temporal variability.*

Computer modeling can be used to address many questions that would need a great expenditure of resources to test experimentally. The Erosion–Productivity Impact Calculator (EPIC) cropping systems model (Williams et al., 1984) simulates conditions of weather, irrigation, fertilization, tillage, and management at the field level. EPIC has been tested for the study of complex crop rotations in southern France (Cabelguenne et al., 1990), simulating growth and yield of corn, grain sorghum, sunflower, soybean, and wheat. After calibrating and validating the model with two years of data, Cabelguenne et al. (1990) concluded that EPIC was able to simulate complex rotations with acceptable accuracy. Although this was not a replicated experiment, it attempted to measure the accuracy of the model's predictions by using 28 pairs of plots that were coincident in terms of year of harvest, crop, preceding crop, and input level. Among these comparisons between field research plots, 85% had yields within 20% of each other. When computer simulations were compared with measurements of yield, 81% of the simulated yields were within 20% of the observed yields (76 plot–years were analyzed). They concluded that EPIC was almost as good a predictor of plot yield as the yield of a similar (paired) plot in the same experiment.

Bryant et al. (1992) used EPIC to measure yield response of corn to changes in irrigation timing, finding that it was not only able to simulate the effect of total amount of water but also the effects of the distribution of irrigation events. Their results after simulating three years of data showed coefficients of determination between the observed and simulated yields that ranged from 0.72 to 0.86. The authors were able to improve the highest value to 0.91 by changing a parameter that simulated the effect of a hailstorm. This indicates that the EPIC model is very versatile but also that many years of data may be needed to get a calibration that adequately accounts for the variability caused by climatic conditions from year to year. EPIC has not been considered as necessarily providing an accurate simulation of a particular crop in a given field and a given year (Steduto et al., 1995).

Working under research station conditions, Cavero et al. (2001) sampled intensively a 27 × 27 m field, measuring bulk density, infiltration rate, and soil texture at 91 points. Soil water retention at wilting point, field capacity, and soil depth were measured at 100 points. The advance and recession time of an irrigation front (which when combined define the irrigation opportunity time) and soil surface elevation were measured at 361 points. Corn yield measurements were made in 73 1.5 × 1.5 m areas. Cavero et al. (2001) found that when they used the estimated values of irrigation depth at each location as input to the EPICphase crop model, the best correlation with the measured yield variability occurred when they used only the spatial variability of infiltration rate (leaving opportunity time fixed), obtaining a coefficient of determination of 0.51. EPICphase is a modification of EPIC, especially improved for water and N stress

Article was submitted for review in December 2002; approved for publication by the Soil & Water Division of ASAE in June 2003.

The authors are **Jorge F. Perez–Quezada,** Graduate Student, Graduate Group in Ecology, University of California, Davis, California; **José Cavero,** Tenured Scientist, Dep. Genética y Producción Vegetal, Estacion Experimental de Aula Dei (CSIC), Zaragoza, Spain; **Jimmy Williams,** Research Scientist, Texas Agricultural Experiment Station, Grassland Soil and Water Research Laboratory, Temple, Texas; **Alvaro Roel,** Graduate Student, Graduate Group in Ecology, University of California, Davis, California; and **Richard E. Plant, ASAE Member Engineer,** Professor, Departments of Agronomy and Range Science and Biological and Agricultural Engineering, University of California, Davis, California. **Corresponding author:** Richard E. Plant, Dept. of Agronomy and Range Science, University of California, 1 Shields Avenue, Davis, CA 95616; phone 530–752–1705; e–mail: replant@ucdavis.edu.

modeling (Cabelguenne et al., 1999). As a second approach, Cavero et al. (2001) used the simulated irrigation depth at each location obtained with the irrigation model B2D (Playan et al., 1996) considering the spatial variability of the infiltration rate and of the surface elevation within the field as inputs for the EPICphase crop model. Unexpectedly, the authors obtained a slightly better match between observed and EPICphase calculated yield with the simulated infiltration depth than with the measured depth. As the authors explain, this may be due to the fact that the simulated irrigation depth value with the B2D irrigation model was the result of averaging nine nodes within each yield sampling area, whereas the estimated irrigation depth was derived from only one point measured within the yield sampling area. These results, although not spectacular, show the possibility of predicting within–field yield variability by using simulation models.

To address the spatial variability at national and regional scale, Priya and Shibasaki (2001) developed a "Spatial EPIC" model to simulate crop yield in India. They also proposed a methodology for generating fine–resolution data from the coarse–resolution data available at the national and regional scale. With an addition of dynamic adaptations, this model generates simulations on a pixel–by–pixel basis following a row and column sequence with multiple soil, climate, and management information provided in the form of geographic information system layers.

There have been relatively few studies of the ability of EPIC to accurately simulate within–field variability in soil properties. Since it requires greater precision, this is a more difficult task for a simulation model than simulating typical yield over a whole field. The capacity to accurately reproduce yield values at the within–field scale would, however, be useful in precision agriculture research. Of particular interest is the capability of the model to carry out these simulations without intensive soil sampling of the field in question. The default for characterization of soil properties in EPIC are the SSURGO (Soil Survey Geographic) descriptions of soil units provided by the Natural Resource Conservation Service (NRCS). There have been concerns about the accuracy and precision of these data at the field scale, and whether they can be useful for site–specific farming purposes (Stermitz et al., 1998). The ability of EPIC to simulate within–field yield variations depends on the accuracy of SSURGO soil descriptions in representing actual soil properties. The objectives of this study are to determine the accuracy of the SSURGO soil descriptions in a highly variable field and to test the ability of EPIC to simulate within–field yield variability in a complex crop rotation.

Methods

General Characteristics

We carried out detailed studies of EPIC model simulations of crop yield variability in a commercial field in the Sacramento Valley, California, over a 4–year period, using both the default SSURGO data and using data collected from intensive soil sampling. According to the information published on its website (www.ftw.nrcs.usda.gov/ssur_data.html), the SSURGO database meets national map accuracy standards for soil maps throughout the U.S. As provided by the NRCS, the digital maps resemble the original soil survey maps at scales that range from 1:12,000 to 1:63,360. In order to determine the accuracy and precision of the SSURGO data for our study field, we compared the SSURGO soil–unit definitions (the field contained three different soil types) and the intensively sampled soil data collected in the field prior to carrying out the EPIC simulations. We then tested EPIC for its ability to reproduce the yield patterns observed in the study field at different spatial scales. The map unit of the SSURGO soil–type definition was used to represent the scale at which growers make management decisions. In addition, we explored the ability of EPIC to generate yield estimations from soil data sampled at the point level in the field.

The study area is a commercially managed field in the Sacramento Valley, California (latitude 38° 32′ N, longitude 121° 58′ W). The climate is Mediterranean, with an average annual rainfall of 54.7 cm, almost all of which occurs during the winter. Summer crops are fully irrigated, and winter crops receive supplemental irrigation, generally during the late season. The study field is 30 ha in area, laser leveled, and is comprised of three different soil types: Capay silty clay (Ca), Brentwood silty clay loam (Br), and Yolo silty loam (Ya) (Andrews, 1972) (fig. 1). Soil Survey Geographic (SSURGO) database information was downloaded from the National Resource Conservation Service website.

Soil samples were taken during the first season (a winter wheat crop planted in 1995 and harvested in 1996) using a 60 m square grid, resulting in 86 samples at a 30 cm depth (fig. 1). Soil characteristics measured included texture components, pH, and soil organic matter. A set of 13 points was chosen for a more detailed texture analysis in a north–south transect, as shown in figure 1. Samples were collected from this transect in 1997, from soil layers at increasing depths from 30 cm to 1.5 m. Soil particle size was measured using the Standard Pipette Method (USDA–NRCS, 1996). In August 2001, a set of samples was taken from these

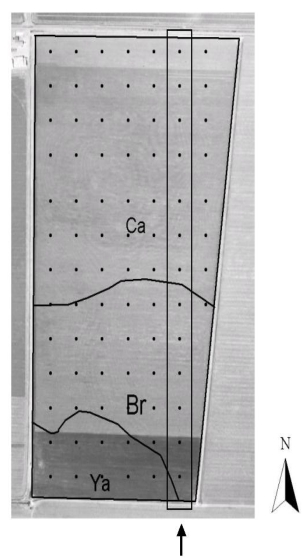

Figure 1. Aerial photograph of the study area (30 ha) in August–1998 during the bean crop. The polygons of black borders show the three soil types (from the SSURGO database): Ca = Capay, Br = Brentwood, and Ya = Yolo. The black points represent the location of soil samples, and the arrow and block show the transect of 13 points where more detailed measurements were taken. The field measures approximately 750 m in length and 375 m in width at the north end.

same 13 points in order to estimate bulk density, moisture content at field capacity and wilting point, and organic matter content. Three samples were taken at each point from depths of 30, 60, and 90 cm. At the time that the samples were taken, the field was in its second year of alfalfa. This may have altered the values of the variables measured, especially bulk density and organic matter content; however, the primary interest was in the relative relation between the sample points and the representation of the yield gradient observed in this field. The field had a wheat–processing tomato–bean–sunflower rotation from 1995–96 to 1999, from which yield monitor data were available. Plant et al. (1999) provide a detailed analysis of the wheat crop grown the first year. Perez–Quezada et al. (2003) conducted a statistical analysis of spatial and temporal yield patterns observed in this field during the 4–year period. Yield data were not collected from every point in the field during every season due to technical problems with the yield monitor or global positioning system. In 1998 and 1999, the grower planted a different crop at the south end of the field, as shown in figure 1. Yield maps were intersected with the soil sample locations to obtain the yield estimates for each crop at each of the 86 sample points for which data were available in that year.

Analysis of Suitability of SSURGO Data for Management Zone Definition

In order to test the accuracy and precision of SSURGO soil type definition, the data of the three textural components were used to define the soil textural classes for each of the 86 samples collected. The soil textural classes were defined based on the percentage of textural components following the texture triangle (Brady, 1984). These were compared with the NRCS definition (Andrews, 1972). To be able to map the texture defined from soil samples, K–means cluster analysis, with the value of K set to 3, was used to obtain the three most distinct soil types based on texture. K–means clustering is an algorithm based on the square error clustering method (Jain and Dubes, 1984), in which each sample is assigned to one of the clusters, so the variance within groups is minimized and the variance between groups is maximized. The K–means clustering process was carried out using the values of soil sand, silt, and clay content without considering spatial information. Cluster analysis represents a mathematical algorithm that reproduces the methodology that was originally used to define these soil types in the sense that it identifies groups so as to maximize between–group differences and minimize within–group variation (Andrews, 1972).

To test the coincidence of SSURGO data with yield, the wheat data were selected as the most complete data set. The 86 data points of wheat yield were re–classified into three classes using the natural breaks scale (Robinson et al., 1984), which creates groups by maximizing the difference between groups and minimizing the difference within groups. This is the same basic idea as K–means cluster analysis but works with only one set of data, in this case wheat yield. The result of this process was three classes of yield: low, medium, and high. The percentage match was calculated by counting the points whose classification coincided with the soil type definition (Ca, Br, and Ya). We mapped the redefined soil types in the field using a Thiessen polygon extrapolation method (Robinson et al., 1984). Geographic data analysis was carried out using ArcView and ArcInfo software (ESRI, Redlands, Cal.).

EPIC Simulations

Version 8120 of the EPIC model (Williams, 1995; Williams et al., 1984) was used for all simulations. Average weather data were obtained from the model itself and consisted of data from Sacramento, which is the weather station nearest (35 km) to the study field available in the EPIC database. These values were used only when detailed data were not available or when the random weather generator was used. Detailed weather data were obtained from the University of California Statewide Integrated Pest Management Project database. Two different files were created from the Winters station, which is located 5 km from the field. One contained 5–year data for the study period (1995–1999), and a second contained data for eight full–rotation periods from 1955 through 1994. The 5–year file was used to calibrate the model, and the 40–year file was used to simulate long–term behavior. The 5–year file contained data on solar radiation, maximum and minimum temperature, rainfall, relative humidity, and wind velocity, whereas the 40–year file contained only temperature and rainfall information. In the latter case, EPIC estimated the missing variables from the average values or from other existing variables.

Characteristics of the three soil types were obtained from the MUUF program (Baumer et al., 1994). This program is a database of NRCS soil types that gives output in EPIC format, which may then be used as input to define model soil properties. Within each soil type there is a list of possible subtypes that are defined in terms of the soil texture in the surface layer. A file was created by choosing the subtype according to the texture defined by SSURGO for each soil type present in the field. In subsequent runs, we also used other files obtained from MUUF according to the texture defined from direct soil sampling, and then we modified the files using the values of the soil variables sampled. Data from the 13 deep soil samples were used to create separate soil files for each sampled location. These files were used in simulations to test the ability of EPIC to function at the small (point) scale. Crop parameter values in the model primarily were based on default parameters contained in the EPIC database. Values different from default EPIC data were taken from Steduto et al. (1995) for wheat, Cavero et al. (1997) for processing tomato, and Cabelguenne et al. (1990) for sunflower. Management data corresponded to the grower's actual practices over the 4–year period.

Results and Discussion

Analysis of SSURGO Data Suitability for Management Zone Definition

Table 1 gives the basic statistics of the soil traits measured in the field in 1995. Among the textural components, silt showed a small variation (CV = 6.3%), whereas sand showed the highest within–field variability (CV = 26%). The pH values were consistent over the field (CV = 2.1%) and were not considered for the cluster analysis to determine the three most distinct soil types. Although organic matter content varied more (CV = 17.3%), its inclusion in the cluster analysis did not change the final result. When the textural class was calculated for each of the 86 points and compared with the SSURGO soil type definition, the percentage of matching points was very low (2.3%) (table 2). However, if we considered the textural classes defined by SSURGO to be

Table 1. Basic statistics of sampled soil variables for 0–30 cm depth.

Variable	N[a]	Average	Max.	Min.	SD	CV (%)
pH	86	5.8	6.0	5.6	0.1	2.1
Sand (%)	86	25.3	44.6	17.3	6.6	26.0
Clay (%)	86	37.3	41.5	23.2	6.3	16.8
Silt (%)	86	37.5	42.8	30.8	2.4	6.3
OM[a]	86	2.0	2.5	0.8	0.3	17.3

[a] N = Number of sample points.
[b] OM = organic matter.

clay, clay loam, and loam for Capay, Brentwood, and Yolo, respectively, the percentage of matching points increased to (80.2%) (table 2).

Figure 2 shows the relationship between soil classes as defined by SSURGO and by the cluster analysis. The correspondence between the spatial distribution of soil types as defined by SSURGO and by the cluster analysis is reasonably good. However, the textures as computed based on the soil triangle are different from the ones defined by the NRCS classification. The area originally mapped as Yolo was found to have a predominantly loamy texture, instead of silty loam. The Brentwood area had a predominantly clay loam texture, instead of silty clay loam. The Capay area was subdivided into two different textures, clay and clay loam.

In summary, soil definition from SSURGO at the field level was found to be fairly accurate in delineating zones spatially, but not in defining soil types according to their measured textural composition. The failure of the SSURGO soil type definitions to match the textural classes obtained through soil sampling may be due to the fact that the soil sample textural classes were based on points obtained from a single field. The cluster analysis may have yielded a different result if it were based on samples obtained on a landscape level basis, matching the extent of the original soil map.

Table 2 also shows the correspondence of SSURGO soil type with a subdivision of wheat yield into three classes by the natural breaks method. In this instance the correspondence was fairly good (82.5%). Our results are not consistent with those of Stermitz et al. (1999), who used SSURGO maps to predict yield response in eight highly variable fields in north–central Montana. They found that the maps did not explain the within–field yield variability and that the description of yields did not improve when reported by soil unit. It is not unexpected that the accuracy of SSURGO information when used in this context would be highly location specific, nor that the optimal conditions for application of SSURGO data would be a highly controlled Mediterranean environment cropping system, such as that found in California's Central Valley.

EPIC Simulations

Table 3 contains the basic statistics of the yields observed in the 4–year period along with crop–related factors that were measured during the first year of the experiment (the wheat year). Each of the factors (stand, weed level, and disease level) was estimated visually on a scale of 1 to 5 at each sample point by an expert agronomist. Wheat stand is a measure of crop density. Disease level is a visual measure of leaf symptoms. Table 4 gives the summary yield statistics for each crop when the data are stratified by SSURGO soil type (fig. 1). The trend of increasing yield toward the southern end of the field can be observed in the table.

Table 5 summarizes the results of the simulations (S1 to S5) in EPIC for the four crops using the 5–year file. The first three rows contain the measured median yield values over each soil type for each crop (the median was used because of its lower sensitivity to extreme values than the mean). The distribution of yield values at points within each soil classification at which data were collected, extracted from yield monitor data as described in the Methods section, were used to test the accuracy of the simulation. The probability values shown in the table correspond to the results of the sign test (Steel and Torrie, 1980), a non–parametric test of the null hypothesis that the median of the distribution of yield values from which the samples were obtained is equal to the simulated yield value.

In simulation 1 (S1), the original crop–parameter file from EPIC was used, along with the soil files obtained from MUUF, according to the texture classification from Andrews (1972). The simulated yield values are reasonably similar to the observed ones for the Capay soil. However, for this soil type, the only crops that had simulated yield values not significantly different from the observed were wheat and bean. Bean also had a good result in Brentwood soil. There was no difference between the simulated values of Brentwood and Yolo, whereas the observed median wheat yields were 46% higher than the simulated values in Yolo. This suggested that the model was not able to reproduce the yield variability observed from two soil types of loamy texture (silty clay loam for Brentwood and silty loam for Yolo). When analyzing the results by crop, the best match between observed and simulated yields was that of bean. There are no data for beans in Yolo soil in 1998 and 1999 because, as mentioned in the Methods section, the cooperating farmer planted a different crop in that portion of the field (fig. 1). The simulated yields for the 1998 and 1999 crops in Yolo soil are reported to help determine the model's ability to generate variability between soil types. The results of S1 for all four crops are graphically represented in a box and whisker plot in figure 3. This figure shows that, for example, the value simulated for tomato in Capay soil, although significantly different from the median, was a fairly good approximation, given the data distribution. Similarly, other results fall within the range of observed data.

Table 2. Number and percentage of matching points of textural class and wheat yield in the study field.

Trait	Sample	SSURGO	SSURGO Reclassified	Match (points)	Total	Match (%)
Texture[a]	C, CL, SIC, L	SIC, SICL, SIL	—	2	86	2.3
	C, CL, SIC, L	—	C, CL, L	69	86	80.2
Yield	Low, Med, High	Ca, Br, Ya[b]	—	71	86	82.5

[a] C = clay, L = loam, CL = clay–loam, SIC = silty clay, SIL = silt loam, SICL = silty clay–loam.
[b] Ca = Capay, Br = Brentwood, Ya = Yolo.

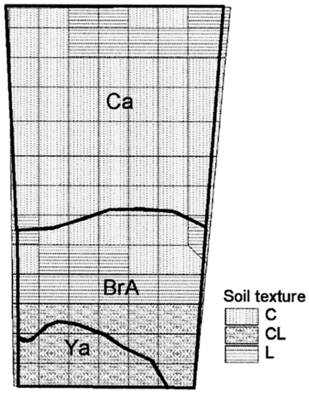

Figure 2. Soil type reclassification (from cluster analysis) using soil sample data of sand, silt, and clay content. Square polygons were constructed from sample points using Thiessen polygons: C = clay, CL = clay loam, and L = loam. SSURGO soil types (regions defined by thick black lines) are Capay (Ca), Brentwood (BrA), and Yolo (Ya).

In S2, the only change from S1 was that the soil files taken from MUUF were selected according to the texture defined from the soil samples obtained in the field (table 5). The area originally mapped as Capay silty clay actually contains clay soil according to the soil texture triangle (fig. 2). Changing soil types did not improve the performance of the model in any of the crops. Indeed, the simulation of wheat in Capay became significantly different from the median value (2.7 tons/ha in S2 vs. 2.2 tons/ha in S1). In the Brentwood soil, the silty loam texture was not available from MUUF so the soil file for silty clay loam was used, and for this reason the values do not change from the previous simulation (S1). In Yolo soil, the texture was changed to loam, according to the reclassification discussed above (fig. 2), but again this did not improve the performance of the model, except for wheat, which became differentiated from Brentwood, but only slightly. These simulations suggest that the yield response of the EPIC model is not highly sensitive to changes in the parameters defined by soil texture.

In S3, the soil files from S1 were modified using information from the soil samples. This information was texture, pH, and soil organic matter. This change was initially made using the average values of each sampled variable. The results of using the updated soil files (denoted Capay–up, Brent–up, and Yolo–up) showed no difference from the S1 results, which confirmed that these variables do not have an important effect in EPIC on the yield simulation, at least in this case. To further test sensitivity to texture, pH, and organic matter, the extreme measured values of each parameter were tested in a simulation. Field observations indicated a relatively constant silt level in the field (Plant et al., 1999), so in the tests silt was maintained constant while either sand or clay was adjusted to its maximum possible value. No change in simulated yield was observed, nor was any observed when pH or organic matter values were varied. Tests using the same procedure but adjusting the values of other parameters such as water content at wilting point (WP), field capacity (FC), and saturated conductivity (SC) indicated that these parameters have a large effect on the simulated yield.

Based on the observations reported in the previous paragraph, in S4, the values of FC and WP were removed from the soil files used in S3 (the updated values) to allow the model to estimate FC and WP from the texture information. The only change in simulated yields between S3 and S4 was

Table 3. Basic statistics of yield (kg/ha) and yield–related parameters (only for wheat) from the same locations where soil samples were taken.

	Wheat Yield (kg/ha)	Stand[a] (wheat year)	Weeds[a] (wheat year)	Disease[b] (wheat year)	Tomato Yield (kg/ha)	Bean Yield (kg/ha)	Sunflower Yield (kg/ha)
Average	3065	3.1	2.9	2.1	56723	1468	2064
Minimum	1417	2.0	1.0	1.0	27374	850	1613
Maximum	6728	5.0	5.0	3.0	105443	2073	2893
SD	1356	0.7	1.2	0.7	14258	328	190
CV (%)	44.3	22.5	40.7	33.7	25.1	22.4	9.2
Moisture (%)	11				94	9	9

[a] Observational rating of plant (stand) and weed density: 1 = low, 2 = med–low, 3 = med, 4 = med–high, 5 = high)
[b] Disease rating: 0 = 0%, 1 = 0%–3%, 2 = 4%–14%, 3 = 15%–29%, 4 = 30%–49%, 5 = necrosis flag/penult leaves.

Table 4. Basic statistics of crop yield (kg/ha) within each soil type, using data from the same locations where soil samples were taken.

Soil Type	Crop	N[a]	Median (kg/ha)	Mean (kg/ha)	Min. (kg/ha)	Max. (kg/ha)	SD (kg/ha)	CV (%)
Capay	Wheat	50	2187	2232	1417	3830	476	21.3
	Tomato	36	45750	45737	27374	63350	8063	17.6
	Bean	50	1490	1421	850	2073	308	21.7
	Sunflower	50	2000	2033	1613	2893	198	9.8
Brentwood	Wheat	27	3879	3688	1436	6005	1045	28.3
	Tomato	27	69441	69732	54353	105443	9512	13.6
	Bean	6	1925	1858	1404	2059	231	12.5
	Sunflower	24	2050	2130	1774	2461	157	7.4
Yolo	Wheat	9	5728	5823	4837	6728	636	10.9
	Tomato	9	63235	61638	42851	71429	9181	14.9
	Bean	0	—	—	—	—	—	—
	Sunflower	0	—	—	—	—	—	—

[a] N = Number of sample points.

Table 5. Results of EPIC yield (t/ha) simulations of a four–crop rotation in three different soil types within a field.[a]

Simulation	Weather	Crop File[b]	Soil File	Wheat	p[c]	Tomato	p	Bean	p	Sunflower	p
			Capay	2.2	—	45.7	—	1.5	—	2.0	—
Median Observed Yield Values:			Brentwood	3.9	—	69.4	—	1.9	—	2.1	—
			Yolo	5.7	—	63.2	—	–	—	–	—
S1	5–year	O	Capay SIC	2.2	0.672	41.7	0.004	1.5	1.00	2.5	0.000
	5–year	O	Brent. SICL	4.6	0.006	45.0	0.000	1.7	0.219	2.6	0.000
	5–year	O	Yolo SIL	4.6	0.004	45.0	0.039	1.7	—	2.6	—
S2	5–year	O	Capay C	2.7	0.000	40.0	0.000	1.5	1.00	2.6	0.000
	5–year	O	Brent. SICL	4.6	0.006	45.0	0.000	1.7	0.219	2.6	0.000
	5–year	O	Yolo L	4.7	0.004	45.0	0.039	1.7	—	2.7	—
S3	5–year	O	Capay–up	2.2	0.672	41.7	0.004	1.5	1.00	2.5	0.000
	5–year	O	Brent–up	4.6	0.006	45.0	0.000	1.7	0.219	2.6	0.000
	5–year	O	Yolo–up	4.6	0.004	45.0	0.039	1.7	—	2.6	—
S4	5–year	O	Capay–up	4.6	0.000	45.0	0.618	1.7	0.000	2.6	0.000
	5–year	O	Brent–up	4.7	0.002	45.0	0.000	1.7	0.219	2.6	0.000
	5–year	O	Yolo–up	4.7	0.004	45.0	0.039	1.7	—	2.6	—
S5	5–year	C	Capay–up	2.7	0.000	70.0	0.000	1.5	1.00	2.4	0.000
	5–year	C	Brent–up	5.2	0.000	76.7	0.000	1.7	0.219	2.7	0.000
	5–year	C	Yolo–up	5.2	0.180	76.7	0.004	1.7	—	2.7	—
S6	40–year	C	Capay–up	5.4	0.000	52.1	0.000	3.2	0.000	4.6	0.000
	40–year	C	Brent–up	6.4	0.000	61.7	0.000	3.2	0.000	5.3	0.000
	40–year	C	Yolo–up	6.4	0.180	61.9	1.00	3.2	—	5.4	—
S7	40–year	O	Capay–up	4.5	0.000	30.6	0.000	3.2	0.000	4.6	0.000
	40–year	O	Brent–up	5.6	0.000	36.7	0.000	3.2	0.000	4.8	0.000
	40–year	O	Yolo–up	5.6	0.508	36.7	0.004	3.2	—	4.8	—

[a] First three rows contain the median values of the observed yields in each soil type.
[b] O = original; C = calibrated.
[c] p values represent the probability of observing a simulated yield difference at least this great from the observed median (sign test) if the data have the same median as the simulated value.

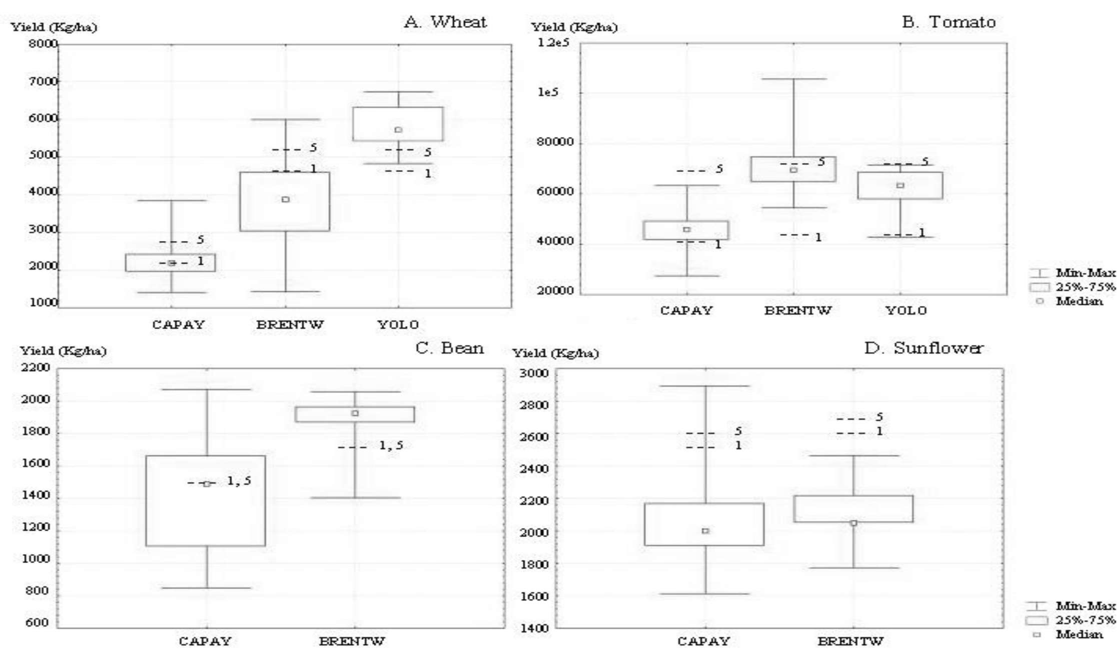

Figure 3. Result of simulations 1 and 5 (dashed lines) and its location relative to the median of observed wheat yield data in different soil types.

observed for Capay soil (table 5). This change produced an improvement in the simulation of tomato in Capay, but also a corresponding decrease in accuracy of simulated wheat yield. In EPIC, the available water for the crop depends on soil water content, rooting depth, and soil properties (FC and WP). In the case that these soil properties are not measured, the soil water holding capacity is estimated from texture data according to the Ritchie method (Ratliff et al., 1983). This estimation could lead to some deviation from the real values, but it is based on field–measured water holding capacities. In any case, Cabelguenne et al. (1999) demonstrated a better simulation of crop growth and yield by improving the modeling of water extraction capacities of the crops in EPIC. Newer versions of EPIC have replaced the Ritchie method with the option of using Rawls's (Rawls and Brakensiek, 1985) or Baumer's (Baumer et al., 1994) models, both of which are related to organic carbon so they allow the model to update wilting point and field capacity annually.

In S5, the crop parameter file was modified following previously published calibrations of the EPIC model for wheat, tomato, and sunflower. The comparison of S5 with S4 gives an indication of the sensitivity of EPIC to variety parameters. The calibrations were done by Steduto et al. (1995), who worked with wheat in a Mediterranean climate; Cavero et al. (1997), who calibrated EPIC for processing tomato at a University of California field station less than 20 km from the study site; and Cabelguenne et al. (1990), who calibrated the model for sunflower in southern France. No calibration was available for bean, although some changes were required in the crop parameter file in order to make the model run under the conditions of this study. These changes were to adjust the potential heat units to 1400 and to change the maximum leaf area index to 5, similar to other crops grown in the area (J. Williams, personal observation). The results of simulation S5 (table 5 and fig. 3) indicate that the calibration for wheat improved the simulation only for the Yolo soil, whose simulated yield value (5.2 tons/ha) is closer to the median of the observed yield than those of previous simulations. In terms of relative difference between soil types, EPIC correctly distinguishes Capay from Brentwood and Yolo, although the latter two are not differentiated from each other. Simulated tomato yield increased dramatically (to 76.7 tons/ha), attaining a level closer to but exceeding the observed median values. The effect of the modification of the tomato parameters may be due to the fact that the variety of processing tomato used by Cavero et al. (1997) was the same as the one used by the grower in this study. In addition, the texture of the soil used by Cavero et al. (1997) was sandy loam, closer to the texture of the Brentwood and Yolo soils in this study.

Simulation S6 retained all the conditions from S5 but used the long–term 40–year weather file for temperature and rainfall data. Simulation S7 also used the 40–year weather file and used the crop and soil parameters of simulation S4. The values given in table 5 represent the average of the eight 5–year rotations that were simulated. In most cases, the simulated yield over the eight rotations was fairly consistent, with a CV of less than 10%. The primary exceptions were wheat in Capay soil, which had a CV of 30% in S6 and 22% in S7. The lack of difference between the simulated yields for Brentwood and Yolo is also present in this simulation. Crop yields in general increased except for tomato, which was highest in the simulation for Yolo soil. When the crop file was set back to its default values in S7, yield estimation improved slightly for wheat, whereas tomato yields were clearly underestimated, and bean and sunflower yields were overestimated. The eight–rotation simulations of bean and sunflower were particularly inaccurate. We did not pursue the reason for this in detail, but we speculate that it may be at least partially due to an incorrect calculation of evapotranspiration when data on relative humidity and wind velocity are not available.

A more detailed simulation was carried out for wheat because more information was available for this crop, specifically the weed, stand, and disease levels reported in table 3. Since the EPIC model version 8120 does not account for the effect of weed infestation, only those points where weed infestation level was low were considered. Table 6 gives the basic statistics of wheat yield in the three different soils present in the field, according to the measured weed infestation level. In Capay soil, this differentiation caused little effect on the average yield, and the CV of the low infested area (15.2%) is much lower than that of the highly infested area (23.5%). In Brentwood soil, the difference between observed average and simulated yields (869 kg/ha) was more dramatic, and again the CV was lower for the low infested area. The same trend was observed in Yolo soil, where the average yield of the low infested area (6002 kg/ha) was higher than that of the highly infested area (5465 kg/ha), and the coefficient of variation (CV) was lower.

The field had a generally low disease level (table 3), and the effect of disease on yield was much less than that of weed competition (Plant et al., 1999). Therefore, disease was not considered as a factor in the simulation. To take into account the effect of plant density, a set of simulation runs was carried out for each soil type, adjusting the plant density in the parameter files and using the original soil files (as in S1). The original (maximum) plant density of the simulation was 250 plants/m^2. To approximate the effects of plant density, this value was reduced in each soil type in proportion to the average stand score in each soil type (at those points where the weed score was less than or equal to 2). This resulted in

Table 6. Basic statistics of observed wheat yields (1995–1996) in different soil types, depending on weed competition level.[a]

Soil	Weed Rating[b]	N[c]	Median (kg/ha)	Average (kg/ha)	Min. (kg/ha)	Max. (kg/ha)	SD (kg/ha)	CV (%)
Capay	All (1–5)	50	2187	2232	1417	3830	476	21.3
	≤2	14	2194	2257	1809	3155	343	15.2
	>2	36	2170	2222	1417	3830	523	23.5
Brentwood	All (1–5)	27	3878	3688	1436	6005	1045	28.3
	≤2	13	3882	4139	3025	6005	874	21.1
	>2	14	3134	3270	1436	6005	1043	31.9
Yolo	All (1–5)	9	5728	5823	4837	6728	636	10.9
	≤2	6	6026	6002	5125	6728	599	10.0
	>2	3	5407	5465	4837	6151	658	12.1

[a] Data from 86 locations where soil samples were collected.
[b] Weed rating based on visual scale: 1 = lowest, 5 = highest.
[c] N = Number of sample points.

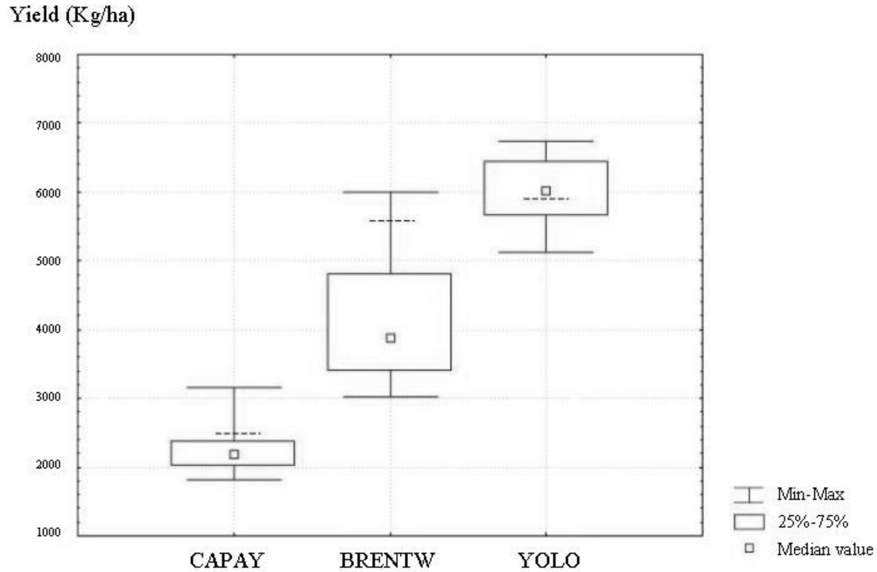

Figure 4. Simulated wheat yield (dashed lines) for three different soil types relative to the median of observed yield data. Simulations used plant density adjusted for measured values, with measurements restricted to points of low weed infestation.

plant density values of 132 plants/m^2 for Capay, 157 plants/m^2 for Brentwood, and 207 plants/m^2 for Yolo. A few other adjustments were also made in the crop parameter files. The plant population parameters PPC1 and PPC2 were changed to 125.60 and 250.95, respectively, following the recommendation of Kiniry et al. (1992). These parameters indicate the relationship between plant population and the maximum LAI that can be attained. The number to the left of the decimal point is the number of plants per square meter, and the number to right is the fraction of the maximum LAI at that population. Thus, with 125 plants/m^2, the LAI that can be attained is 60% of maximum, and with 250 plants/m^2, the LAI that can be attained is 95% of maximum. The maximum root depth was set to 1.5 m, and the lower limit of soil nitrate concentration was set to 3 ppm, according to Cavero et al. (1998). Winter dormancy was set to 0.1 to simulate the mild winters that occur in this area. The results of this set of runs are given in table 7, and figure 4 shows their position relative to the median yield, measured in those sampled locations that had low weed infestation (score $\leq$ 2). Comparison of these results with those of simulations 1 and 5 (table 5 and fig. 3A) indicates that the use of weed infestation information improved the ability of EPIC to simulate yield differences between Capay soil (2.5 t/ha) and the other two soil types (table 7). It also differentiated Brentwood (5.6 t/ha) from Yolo (5.9 t/ha), although this difference was not as large as the measured value (fig. 4).

The EPIC–simulated yields are based on the effect of five types of stress: temperature, aeration, water, nitrogen, and phosphorous. EPIC calculates the number of days during the course of the season at which the level of each of these factors is sufficiently different from optimal to reduce yield. The calculated numbers of days of each type of stress for the wheat simulation are shown in table 7. Temperature and aeration stresses were responsible for the increased differentiation of simulated yields for each of the three soil types. The other types of stress showed little difference between soil types. EPIC failed to generate the higher aeration stress level found by Plant et al. (1999) in Brentwood soil during the wheat crop. Plant et al. (1999) identified aeration stress as the primary cause of reduced yield in the wheat crop in this part of the field.

Table 7. Observed and simulated wheat yields (t/ha) and total number of days of stress, using crop density values adjusted to approximate the median measured values in each soil type.

Soil Type	Yield (t/ha)		Stress (days)[a]				
	Obs.	Sim.	WS	NS	PS	TS	AS
Capay SIC	2.2	2.5	5	0	0	9	101
Brent. SICL	3.9	5.6	1	6	0	19	16
Yolo SIL	5.7	5.9	1	0	0	21	20

[a] Stress types: WS = water, NS = nitrogen, PS = phosphorus, TS = temperature, AS = aeration.

Overall, our results indicate that the EPIC model adequately reproduced observed yield variability at the spatial scale of the soil survey map unit in this field. In the absence of detailed information about soil properties, EPIC was able to simulate yields adequately by estimating soil factors based on SSURGO soil type definitions. However, inclusion of measured soil properties in place of those obtained from predefined soil types did not improve EPIC's value as a predictor of yield. This could be due to the fact that measured water holding capacity on disturbed soil cores does not represent accurately the field–measured limits of water availability (Ritchie, 1981). The results indicate that biotic determinants such as weed infestation and crop stand must be taken into account if one is to obtain a more accurate simulation of yield distribution.

To test the ability of EPIC to simulate yield at a finer spatial scale than the soil survey map unit, we carried out a set of simulations along a transect of 13 sample points in the north–south direction, as shown in figure 1. The direction of this transect is that of the predominant direction of variability in this field, which is that of its long axis, north to south (Perez–Quezada et al., 2003). Field capacity (FC), wilting point (WP), bulk density (BD), and textural components were measured up to 1 m depth. Saturated conductivity was not measured; therefore, the model estimated its value. Figure 5 shows the trend of the observed and simulated crop yields for each of the 13 points in the transect (fig. 1). The low yield

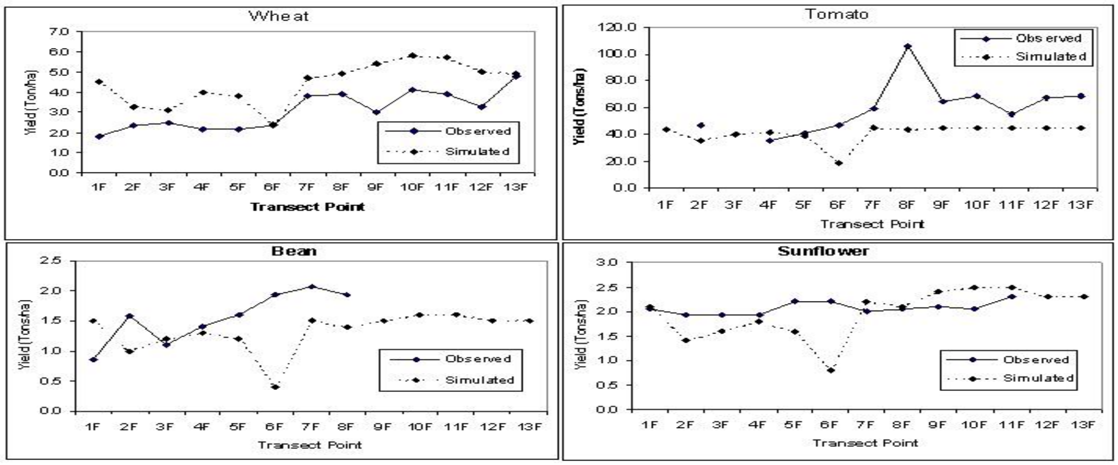

Figure 5. Observed and simulated yields at points along the north–south transect shown in figure 1.

values obtained for all the crops at point 6 (the sixth point from the northern end of the transect) was apparently due to an error in lab analysis of the soil material. It was found that the lab measurements of FC were significantly different from those of a blind duplication (paired t–test, $P < 0.038$). EPIC is highly sensitive to changes in available water in the profile, which is determined by FC relative to WP.

Figure 5 shows that EPIC was able to reproduce the general trend of increasing wheat yields from north to south. The simulated tomato yield values did not vary much, and from points 9 to 13 they were stagnant. For the bean and sunflower crops, the simulated values followed the same trend as for wheat, although there were fewer observed points in the south end of the field. Without considering point 6, bean yield values ranged from 1000 to 1600 kg/ha, which is a good approximation compared with the observed range of 900 to 2100 kg/ha. In the case of sunflower, simulated yields ranged from 1400 to 2500 kg/ha, which is in good agreement with observed yields of 1900 to 2300 kg/ha, if point 6 is not considered.

CONCLUSIONS

In the field in which these tests were carried out, soil definitions from SSURGO were found to be relatively accurate in defining the spatial extent of yield zones. These definitions were not precise in delineating soil types according to their textural composition, although they were very similar and reproduced the spatial trend. Working at the field level, our results indicate that SSURGO data may be a good source of information for the researcher for planning either a management trial or directed sampling as part of a site–specific research project. They may also be useful to the commercial grower as a basis for organizing a site–specific crop management strategy.

The EPIC model was found to be a generally good tool to reproduce the yield variability within this field, using the SSURGO soil type definition. When weed infestation information was considered, the ability of EPIC to distinguish between wheat yield level of the Capay soil and that of the other two soils improved, although it was still difficult to differentiate between Brentwood and Yolo soils. The performance of the model was not as good at reproducing yields when tested with soil data from sampled locations in the field. Proper calibration of the soil variables water content at field capacity and wilting point was identified as a key component of the modeling process. Saturated conductivity was also found to have a strong effect on the simulations, although its measurement is more complex. Attempts to simulate long–term average yield behavior were considerably less successful, possibly due to the inability of the model to accurately simulate meteorological data when these data are not available as model inputs. Our results suggest the possibility of using SSURGO data and the EPIC model to assess the intra–field yield variability in areas where multi–crop rotations are used.

ACKNOWLEDGEMENTS

This research was supported in part by the Fertilizer Research and Education Program of the California Department of Food and Agriculture. The authors are grateful to Jayou Deng and Julie Young for expert technical support.

REFERENCES

Andrews, W. F. 1972. *Soil Survey of Yolo County, California*. Washington, D.C.: USDA Soil Conservation Service.

Baumer, O., P. Kenyon, and J. Bettis. 1994. *MUUF v214 User's Manual*. Lincoln, Neb.: National Soil Survey Center.

Brady, N. C. 1984. *The Nature and Properties of Soils*. 9th ed. New York, N.Y.: Macmillan.

Bryant, K. J., V. W. Benson, J. R. Kiniry, J. R. Williams, and R. D. Lacewell. 1992. Simulating corn yield response to irrigation timings – Validation of the EPIC model. *J. Production Agric.* 5(2): 237–242.

Cabelguenne, M., C. A. Jones, J. R. Marty, P. T. Dyke, and J. R. Williams. 1990. Calibration and validation of EPIC for crop rotations in southern France. *Agric. Systems* 33(2): 153–171.

Cabelguenne, M., P. Debaeke, and A. Bouniols. 1999. EPICphase, a version of the EPIC model simulating the effects of water and nitrogen stress on biomass and yield, taking account of developmental stages: Validation on maize, sunflower, sorghum, soybean, and winter wheat. *Agric. Systems* 60(3): 175–196.

Cavero, J., R. E. Plant, C. Shennan, and D. B. Friedman. 1997. The effect of nitrogen source and crop rotation on the growth and yield of processing tomatoes. *Nutrient Cycling in Agroecosystems* 47(3): 271–282.

Cavero, J., R. E. Plant, C. Shennan, J. R. Williams, J. R. Kiniry, and V. W. Benson. 1998. Application of EPIC model to nitrogen cycling in irrigated processing tomatoes under different management systems. *Agric. Systems* 56(4): 391–414.

Cavero, J., E. Playan, N. Zapata, and J. M. Faci. 2001. Simulation of maize grain yield variability within a surface–irrigated field. *Agronomy J.* 93(4): 773–782.

Jain, A. K., and R. C. Dubes. 1984. *Algorithms for Clustering Data*. Upper Saddle River, N.J.: Prentice–Hall.

Kiniry, J. R., J. R. Williams, P. W. Gassman, and P. Debaeke. 1992. A general, process–oriented model for two competing plant species. *Trans. ASAE* 35(3): 801–810.

Perez–Quezada, J. F., G. S. Pettygrove, and R. E. Plant. 2003. Spatial–temporal analysis of yield and soil factors in two four–crop–rotation fields in the Sacramento Valley, California. *Agronomy J.* 95(3): 676–687.

Plant, R. E., A. Mermer, G. S. Pettygrove, M. P. Vayssieres, J. A. Young, R. O. Miller, L. F. Jackson, R. F. Denison, and K. Phelps. 1999. Factors underlying grain yield spatial variability in three irrigated wheat fields. *Trans. ASAE* 42(5): 1187–1202.

Playan, E., J. M. Faci, and A. Serreta. 1996. Modeling microtopography in basin irrigation. *J. Irrigation and Drainage Eng., ASCE* 122(6): 339–347.

Priya, S., and R. Shibasaki. 2001. National spatial crop yield simulation using GIS–based crop production model. *Ecological Modeling* 136(2–3): 113–129.

Ratliff, L. F., J. T. Ritchie, and D. K. Cassel. 1983. Field–measured limits of soil–water availability as related to laboratory–measured properties. *SSSA J.* 47(4): 770–775.

Rawls, W. J., and D. L. Brakensiek. 1985. Prediction of soil water properties for hydrologic modeling. In *Proc. Symposium Sponsored by Committee on Watershed Management/Irrigation and Drainage Division*, 293–299. St. Joseph, Mich.: ASAE.

Ritchie, J. T. 1981. Soil–water availability. *Plant and Soil* 58(1–3): 327–338.

Robinson, A., R. Sale, J. Morrison, and P. C. Muehrcke. 1984. *Elements of Cartography.* New York, N.Y.: Wiley.

Steduto, P., V. Pocuca, A. Caliandro, and P. Debaeke. 1995. An evaluation of the crop–growth simulation submodel of EPIC for wheat grown in a Mediterranean climate with variable soil–water regimes. *European J. Agronomy* 4(3): 335–345.

Steel, R. G. D., and J. H. Torrie. 1980. *Principles and Procedures of Statistics: A Biometrical Approach.* New York, N.Y.: McGraw–Hill.

Stermitz, R. J., G. A. Nielsen, and D. S. Long. 1998. Testing quality of soil survey geographic (SSURGO) database for precision farming. In *Proc. 4th International Conference on Precision Agriculture*, 319–326. Madison, Wisc.: American Society of Agronomy.

USDA–NRCS. 1996. *Soil Survey Laboratory Methods Manual*. Soil Survey Investigations Report No. 42. Lincoln, Neb.: National Soil Survey Center.

Williams, J. R. 1995. The EPIC model. In *Computer Models of Watershed Hydrology.* V. P. Singh, ed. Highlands Ranch, Colo.: Water Resources Publications.

Williams, J. R., C. A. Jones, and P. T. Dyke. 1984. A modeling approach to determining the relationship between erosion and soil productivity. *Trans. ASAE* 27(1):129–144.