WATER RESOURCES RESEARCH, VOL. 29, NO. 6, PAGES 1579–1587, JUNE 1993

Metamodels and Nonpoint Pollution Policy in Agriculture

AZIZ BOUZAHER AND P. G. LAKSHMINARAYAN

Center for Agricultural and Rural Development, Iowa State University, Ames

RICHARD CABE

Department of Economics, New Mexico State University, Las Cruces

ALICIA CARRIQUIRY, PHILIP W. GASSMAN, AND JASON F. SHOGREN

Center for Agricultural and Rural Development, Iowa State University, Ames

Complex mathematical simulation models are generally used for quantitative measurement of the fate of agricultural chemicals in soil. But it is less efficient to use them directly for regional water quality assessments because of the large number of simulations required to cover the entire region and because the entire set of simulation runs must be repeated for each new policy. To make regional water quality impact assessment on a timely basis, a simplified technique called metamodeling is suggested. A metamodel summarizes the input-output relationships in a complex simulation model designed to mimic actual processes such as groundwater leaching. Metamodels are constructed and validated to predict groundwater and surface water concentrations of major corn and sorghum herbicides in the Corn Belt and Lake States regions of the United States. The usefulness of metamodeling in the evaluation of agricultural nonpoint pollution policies is illustrated using an integrated environmental economic modeling system. For the baseline scenario, we estimate that 1.2% of the regional soils will lead to groundwater detection of atrazine exceeding 0.12 μg/L, which compares well with the findings of an Environmental Protection Agency monitoring survey. The results suggest no-till practices could significantly reduce surface water concentration and a water quality policy, such as an atrazine ban, could increase soil erosion despite the conservation compliance provisions.

1. INTRODUCTION

Control of nonpoint pollution from agricultural practices and source reduction of agricultural pollutants for water quality protection are increasingly debated policy goals. These debates must be based on informed evaluation of agricultural nonpoint pollution in relation to policies, management practices, and hydrogeological factors. Ideally, water quality monitoring should provide policy analysts with the needed information. But due to high monitoring costs, mathematical models are generally used to simulate the physical processes that describe the agricultural chemical movement in soil and predict their concentrations in groundwater and surface water [*Walton*, 1984; *Wagenet and Hutson*, 1991]. Use of these process models is economical and practical for site-specific problems only [*Evans and Myers*, 1990]. To use these field-scale models for regional water quality assessments we have to simulate them for the area-wide distribution of soil and weather parameters. But it is costly and time consuming to do area-wide simulation for all combinations of crop, chemical, management practice, and technology. Nearly 75,000 simulation runs are required to cover a study area comprising the Corn Belt and Lake States regions of the United States. Furthermore, to evaluate a new policy within a regional integrated modeling system, we have to repeat the simulation runs for all combinations of factors used in the baseline evaluation. For instance, a policy scenario in an integrated modeling system requires a mutually consistent combination of policy, environmental, chemical, management, and technological parameters and behavioral equations. Integrated systems analysis requires both timely integration of diverse process models and integration of outcomes over a distribution of diverse input sets. Therefore a simplified technique to ease the computational burden while abstracting the key process characteristics is needed. Metamodels are simple, but statistically validated, analytical tools capable of addressing both of these difficulties.

Metamodeling is a statistical method to abstract away from unneeded detail for regional analysis by approximating outcomes of a complex process model through statistically validated parametric forms. The simplification provided by metamodels allows us to evaluate the consequences of alternative policies without the need for additional simulations. If the complex simulation model is a tool to approximate the underlying real-life system, the analytic metamodel attempts to approximate and aid in the interpretation of the simulation model and ultimately the real-life system. *Blanning* [1975] and *Kleijnen* [1979] recommend analytic metamodels for simulation experiments; *Lawless et al.* [1971] propose their use for sensitivity analysis. Empirical application of metamodels in industrial, computer, and management fields is documented by *Kleijnen* [1987]. To our knowledge, use of metamodels in agriecological systems simulation and, particularly, the simulation of real processes describing the fate of agricultural chemicals, is fairly new [see *Bouzaher*, 1991].

This paper discusses metamodeling in an agriecological economic system with specific reference to evaluate agricultural nonpoint pollution policy. We identify, estimate, and

Paper number 93WR00286.
0043-1397/93/93WR-00286$05.00

validate regression metamodels for concentrations of chemical in groundwater and surface water. We generated these concentrations from process model simulations calibrated on a sample of soils in a study area comprising the Corn Belt and Lake States. A representative, stratified, self-weighted random sample of soil was drawn for the simulation experiment. We find simple nonlinear exponential functions to adequately explain and predict the simulation model responses. We use the estimated metamodels to predict the groundwater and surface water chemical concentrations and their distributions for the entire set of soils in the study area for the baseline regime of herbicide use. This baseline is determined by the agricultural decision model in the Comprehensive Environmental Economic Policy Evaluation System (CEEPES). CEEPES is an integrated agriecological economic system designed to evaluate the trade-offs of alternative policies restricting the use of herbicides, particularly atrazine, in corn and sorghum production [*Cabe et al.*, 1991].

We compared our estimate for the spatial distribution of groundwater concentration of atrazine with that of the Environmental Protection Agency's (EPA) groundwater monitoring survey estimate [*EPA*, 1990]. Our estimate of 1.2% of the soils in the region contributing to an atrazine detection level exceeding the survey's minimum reporting limit of 0.12 μg/L (parts per billion (ppb)), is bounded by the monitoring estimate of 0.7% in the rural wells and 1.7% in the community water systems. We also derived cumulative spatial probability distributions for groundwater and surface water concentrations of atrazine under conventional- and no-till practices.

Some of our results are as follows: (1) The probability of exceeding the toxicity-weighted benchmark for human exposure from atrazine, as suggested by the EPA, is relatively larger for surface water than groundwater. (2) No-till practices significantly reduce the surface water concentrations of atrazine and other herbicides relative to conventional tillage. (3) A water quality policy that bans atrazine could increase soil erosion, even with the conservation compliance provisions fully incorporated.

2. Metamodeling in an Agriecological Economic System

A metamodel is a regression model explaining the input-output relationship of a complex simulation model, which is a mathematical model structured to mimic the underlying real-life process. Let Φ be the unknown function which characterizes the underlying real phenomena relating the response y to the input vector $\mathbf{v}$:

$$y = \Phi(\mathbf{v}). \tag{1}$$

Most simulation models mimic outcomes for a variety of possible response variables, and specification of the response of interest may not be a trivial matter.

A simulation experiment is a set of executions of the simulation models intended to approximate the values of y associated with a specified set of input vectors. The output of a simulation experiment is a data set consisting of specified input vectors and their associated responses, as determined by the simulation model. Choice of the number and values of input vectors for which the simulation model will be executed is the subject of experimental design. For

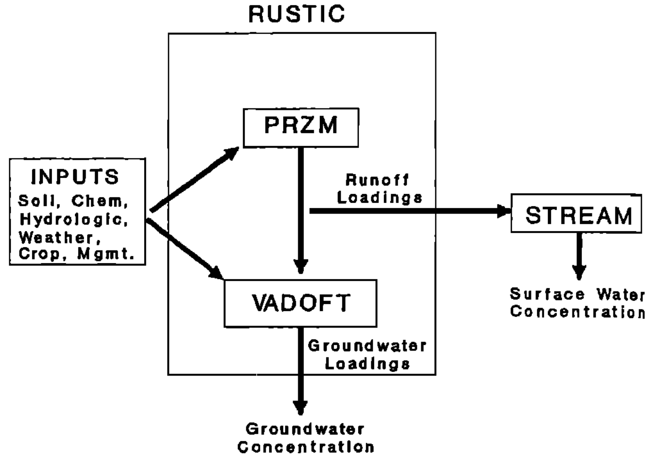


Fig. 1. Schematic representation of the multimedia (groundwater and surface water) simulation models.

statistical purposes, it would be preferable to experiment with the real-life system rather than a simulation model of the system. In that case we would have a statistical model of the system rather than a metamodel. This approach is not adopted because it would mean incurring the cost and delay of waiting, in this case for 30 years of weather to present itself to the real-life system.

Given the output of a simulation experiment, we can specify an analytic metamodel with relatively few inputs, x_1 through x_k. Let the metamodel explaining the simulated outcome be represented as

$$y = \phi(x_1, x_2, \cdots, x_k, u), \tag{2}$$

where u is the stochastic disturbance term. We can use statistical procedures to identify and estimate the function ϕ describing the metamodel. Because of their simple and precise representation of the complex mathematical model, simulation practitioners are favoring metamodels for purposes such as validation, sensitivity analysis, estimation of interactions among inputs, control, and optimization, without the need for additional simulation runs [*Kleijnen*, 1987].

The groundwater and surface water process models we use were configured to simulate the environmental fate of herbicides in the major corn- and sorghum-growing regions of the United States. This regional application is part of an overall CEEPES configuration to evaluate the set of herbicide strategies applicable to corn and sorghum production. Figure 1 illustrates that the core of multimedia (groundwater and surface water media) fate and transport component is the Risk of Unsaturated/Saturated Transport and Transformation of Chemical Concentrations (RUSTIC) system developed by *Dean et al.* [1989]. RUSTIC is the system that links the Vadose Zone Flow and Transport (VADOFT) model and the Pesticide Root Zone Model (PRZM) to trace pesticide movement in soil. PRZM is a one-dimensional, dynamic, and compartmental model that can simulate chemical movement in the unsaturated root zone. Chemical, soil, and plant characteristics, tillage and management practices, and local hydrometeorological conditions are the major parameters for PRZM. VADOFT performs one-dimensional transient or steady state simulations of water flow and solute transport in the variably saturated vadose zone.

Soil parameters for PRZM and VADOFT were automati-

TABLE 1. Summary Statistics of the Herbicide Concentrations Simulated by the Process Models

Summary Statistics	Average Herbicide Concentration at 1.2 m	Average Herbicide Concentration at 15 m	Peak Herbicide Concentration in Surface Water (Peak Stream)
Sample mean, ppb	3.25	0.087	242
Standard deviation	11.8	0.5	269
Skewness	5.1	8.5	2.9
Range	0–110	0–7.3	2–2114
Percent zeros	20	48	0

cally generated with the Data Base Analyzer and Estimator (DBAPE) soil data base [*Imhoff et al.*, 1990]. Given edge-of-field loadings from RUSTIC, the Surface Transport and Agricultural Runoff of Pesticides for Exposure Assessment (STREAM) model [*Donigian et al.*, 1986] is used to simulate chemical concentrations in surface water. Because these concentrations are simulated from edge-of-field loadings, they are considered to be accurate within an order of magnitude and typically overestimate actual concentrations [*Donigian and Mulkey*, 1992]. Hence these should be considered as estimates from the worst case scenario. The basic RUSTIC and STREAM configurations for this simulation experiment are described by *Gassman et al.* [1991].

3. Experimental Design and Procedure

Soils selected for the RUSTIC simulations were chosen from a total of 2076 P1 (prime agricultural land) and P4 (irrigated agricultural land) soils. A representative, stratified, self-weighted random sample of soils was drawn; soils were randomly chosen within each stratum, with sampling probability proportional to the percentage use. The soil selection was also based on their ability to support corn and sorghum. In all, 180 soils, representing four hydrologic groups (A–D), from 16 states (strata) were chosen for the RUSTIC simulations. Soils were classified into four hydrologic groups based on their minimum infiltration rate, ranging from A soils, which have the highest infiltration rate, to D soils, which have the lowest [*Soil Conservation Service*, 1972]. Sixteen herbicides used in corn and sorghum production were selected. Assuming that chemical use is independent of soil type, each of the 16 chemicals was applied to the 180 soils. The simulations were performed separately for conventional-, reduced-, and no-till cultivation practices. Simulated timings of application of herbicide were early preplant, preplant incorporated, preemerge, and postemerge. Early preplant herbicides are applied before the crop is planted and may be soil incorporated as in preplant incorporated. Preemergence (postemergence) herbicides are applied after planting and before (after) the crop and weed emerges. A total of 7518 simulation runs were performed. The number of simulation runs do not match the number of combinations of crop, soil, chemical, tillage, and application timings because (1) some herbicides are recommended for specific timings only (for instance, butylate and EPTC are only preplant incorporated herbicides) and (2) some application timings are not defined for certain tillage practices (for instance, no-till planting does not allow preplant incorporation of herbicides) [*Hartzler and Owen*, 1990].

Many groundwater studies have indicated an inverse relationship between pesticide concentration and well depth. The groundwater table at 15 m below the soil surface is the most vulnerable to chemical contamination [*Detroy et al.*, 1988]. Therefore, the pesticide concentrations in the solute phase were estimated for 1.2 and 15 m (the assumed water table depth) for each RUSTIC simulation. Adsorptive properties of the chemicals under study are such that the sediment phase is negligible and was therefore not simulated. The simulation was performed dynamically for each day over a 30-year period. Historical weather data (1950–1979) used for one weather station in each state of the study region was taken from the RUSTIC weather data base [*Imhoff et al.*, 1990]. From each simulation run, the average (chronic) groundwater concentration at 1.2 and 15 m and the runoff loadings were recorded. The runoff loadings were used to estimate the peak (acute) surface water concentrations of herbicides using STREAM.

4. Regression Metamodeling

4.1. *The Data*

The concentrations in groundwater and surface water recorded from the simulation experiment comprise the data for the dependent variables in the regression metamodels. Table 1 presents the descriptive statistics and distributional characteristics of the data. Preliminary analysis of the data showed large variability in concentrations from one soil to another, highlighting the need for a spatial dimension, and from one management practice to another within a soil. In 90% of the observations, herbicide concentrations in groundwater were less than 1 ppb. Twenty percent of the concentrations at 1.2 m and nearly one half of the concentrations at 15 m were zero. The distributions, in general, were nonnormal and positively skewed (to the right). The sample mean of surface water concentrations was 242 ppb with a standard deviation of 269. The data for the regressors were mostly represented by the simulation inputs. Soil properties (organic matter, water retention capacity, bulk density, sand and clay proportions, and soil depth) were obtained from DBAPE. Pesticide characteristics (decay rate, Henry's law constant, and organic carbon partition (soil sorption) coefficient (K_{oc})) were obtained from *Wauchop and Goss* [1990]. See *Carsel and Jones* [1990] for a description of these data bases and their applicability to regional studies.

4.2. *The Models*

In the metamodeling literature, the most commonly used models are the general linear and nonlinear models, often referred to as "regression metamodels." We first fitted a simple linear model using an ordinary least squares (OLS) procedure. Let **Y** be an $n \times 1$ vector of observations of the simulated response; $\mathbf{X}$ be a known, full-rank $n \times p$ matrix of observations on the explanatory variables; and $\boldsymbol{\beta}$ be a $p \times 1$ vector of unknown, fixed parameters. The simple linear regression model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad E(u_i) = 0 \qquad E(u_i)^2 = \sigma_i^2, \qquad \operatorname{Cov}(u_i, u_j) = 0. \tag{3}$$

Given that the response variable is nonnormal with heterogeneous (nonconstant) variance, the parameter vector $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$, and the corresponding predictions $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ are inefficient (in the minimum variance sense). We examined the residual plots for any patterns indicating heterogeneity of variance. These plots for the linear model (3) exhibited a clear wedge-shaped pattern violating the classical assumption of homogeneity of variance.

We used a standard variance stabilizing transformation on the data, and fitted the linear model in the transformed space. An estimated regression of the transformed data should have an error structure that is normally distributed with constant variance. A variance-stabilizing transformation for y_i (ith element of $\mathbf{Y}$) can be found by using either the general form for a power transformation, y_i^λ, proposed by *Box and Cox* [1964], or a procedure similar to the one proposed by *Lin and Vonesh* [1989]. Although the regression with the transformed data gave a higher R^2 and well behaved residuals compared to the regression with the untransformed data, the predictions were poor. We also fitted a weighted least squares (WLS) model, using appropriately derived weights. Given the variance-stabilizing transformation $\mathbf{Y}^\lambda$, we used *Bartlett*'s [1947] procedure to relate the variance of the response variable in the original and transformed spaces and get an approximate weight. Although residual diagnostics were greatly improved when using WLS, predictions were still poor. The failure of these linear models to adequately predict the response variable naturally led us to fit a nonlinear model using nonlinear least squares (NLS).

Sometimes variance heterogeneity may be introduced by specifying a linear model when the actual underlying structure is a nonlinear one. Such instances are common in models for chemical, biological, and kinetic processes [*Box and Hill*, 1974]. Therefore, we fitted a nonlinear model of the form

$$z_i = g(x_i; \Theta) + \varepsilon_i \qquad E(\varepsilon_i) = 0 \qquad E(\varepsilon_i)^2 = \sigma^2 \qquad \operatorname{Cov}(\varepsilon_i, \varepsilon_j) = 0, \tag{4}$$

where g is the nonlinear expectation function, ε is the random disturbance term, and Θ is the unknown parameter vector to be estimated. A desirable estimate of Θ, denoted by $\hat{\Theta}$, has optimal large sample properties (i.e., $\Theta \sim AN\{\Theta, \sigma^2[\Sigma(\partial g(x_i;\ \Theta)/\partial\Theta)(\partial g(x_i;\ \Theta)/\partial\Theta)^T]^{-1}\}$). Because our objective is to find a model with theoretical as well as empirical justification and better predictive ability we chose the simple exponential model, $g(x_i; \Theta) = \exp(\mathbf{X}\Theta)$. The exponential model is a satisfactory representation for several reasons: (1) the optimal power transformation parameter λ^* was small; (2) the original (untransformed) data have a positively skewed distribution; and (3) other studies that evaluated the groundwater pollution potential of pesticides [*Jury et al.*, 1987; *Khan and Liang*, 1989] used an exponential model.

5. THE RESULTS

Table 2 summarizes the results from the nonlinear fit. The parameter estimates of the nonlinear model for (transformed) average groundwater concentrations at 1.2 and 15 m and peak concentration in surface runoff are shown. The high degree of skewness in the groundwater data justified the use of transformation even with the nonlinear model. We used SAS's Gauss-Newton algorithm to solve for the optimal parameter vector. We relied on past studies in identifying a parsimonious specification. Care was taken to avoid significant multicollinearity among the regressors. Collinearity between linear and quadratic regressors was reduced by centering the variables. That is, we defined these variables as deviation from the respective mean values. This transformation does not change the meaning or fit of the model, but by reducing collinearity it tends to stabilize the sampling variance of the estimates. The adjusted R^2 was more than 80% in all three fitted equations. The correlations between the actual and the predicted concentrations in groundwater and surface water were between 70 and 95%. Figure 2 shows the distribution of actual and predicted concentrations.

The coefficients of the continuous regressors (other than the 0,1 type dummy variables) were all different from zero at the 5% level of significance, and their signs were consistent with theory. The interaction term between bulk density and sorption coefficient ((BD)(K_{oc})) generally referred to as the retardation factor [*Khan and Liang*, 1989], is expected to have a negative impact on chemical concentration. The estimated coefficient of this regressor is negative and significant. The estimated coefficient for decay is significant, with a negative sign for groundwater and a positive sign for surface water because fast decay implies less leaching and more runoff potential. Organic matter enhances the soil sorption capacity and the microbial activity, both resulting in reduced leaching [*Stevenson*, 1982]. The negative impact of organic matter on groundwater leaching is consistent with the theory. The higher the sand percentage, the greater the seepage, implying a positive impact on groundwater, which is what our result shows. The herbicide concentrations in groundwater increased with available water and decreased with soil depth.

Qualitative variables were represented in the nonlinear model by a 0, 1 dummy variable. The dummy variables for tillage practice were all different from zero at the 5% level of significance. These coefficients measure the difference in leaching/runoff potential of reduced- and no-till practices relative to conventional tillage. Intuitively, conservation tillage practices, which allow greater infiltration of moisture and less surface runoff, should increase pesticide leaching losses [*Baker*, 1992]. The coefficient on no-till, which has a positive impact on groundwater and a negative impact on surface water, clearly supports this. The estimated equation also captured the differences between hydrologic groups and timing of application through 0, 1 regressors. Fifteen dummy variables were included to represent the 16 different weather stations covering the study area. Most of these coefficients were significantly different from zero, highlighting the importance of climate in determining chemical concentration levels. A dummy variable to capture the difference in the leaching/runoff potential of sorghum was included. This coefficient was significant, with a positive sign for ground-

TABLE 2. Coefficients of the Estimated Metamodels Explaining the Response Variable (Simulated Herbicide Concentrations in Groundwater at 1.2 and 15 m and Surface Water)

	Dependent Variable		
	Eighth Root of 1.2-m Average	Eighth Root of 15-m Average	Peak Stream
Adjusted R^2	0.84	0.84	0.83
Root-mean-square error	0.19	0.11	112
$\hat{\rho}$	0.78	0.73	0.91
	Regression Coefficients		
Intercept	−0.892	−1.239	7.258
Predicted concentration for 1.2-m average	···	0.374	···
(OM) × (Henry's constant)	123.317	···	−1685.510
(BD) × (K_{oc})	−0.002	−0.001	−0.006
(OM) × (Decay)	−8.359	−67.379	−0.859
Decay	−19.051	−20.333	6.484
$(\text{Decay})^2$	142.391	···	−197.149
Organic matter	−1.070	−0.496	···
$(\text{Organic matter})^2$	0.222	···	···
Percent sand	0.003	0.008	···
WRC	0.529	1.458	···
Soil depth	−0.002	−0.001	−0.0004
D-sorghum	0.199	0.453	−0.054
D–reduced till	0.071	0.045	−0.005
D–no till	0.101	0.126	−0.341

Sample size, $N = 7518$. All the coefficients are significant at 5% level of confidence interval; $\hat{\rho}$ is the coefficient of correlation between the actual (simulated) and predicted data. OM denotes organic matter; BD, bulk density; and WRC, water retention capacity (available water). (OM) × (Henry's constant), (BD) × (K_{oc}), and (OM) × (Decay) are the respective interaction terms. D indicates the dummy variable. The linear and quadratic variables representing decay and organic matter are centered.

water, implying that leaching of herbicides is more severe in sorghum than in corn.

6. Validating the Metamodel

Metamodel "validation" refers to testing the robustness and predictive ability of the estimated models. Since the metamodel is built with simulated data, validation of the metamodel differs from the usual sense of validation, in which statistical and process models are compared with actual (observed) data. Validating the metamodels is important because they are two steps away from the underlying real processes. We will have greater confidence in the regression metamodels and their estimated parameters and predictions when they are statistically validated before being integrated into the unified modeling system. The practical statistical validation methods [*Snee*, 1977] include (1) validation with new data; (2) cross validation (split-half validation) in which the original data set is randomly split into two halves, a model is fitted for each half separately, and the fitted models are used to predict the other half of the data; and (3) comparison of empirical results with those from simulations and monitoring surveys. Validation with new data and cross validation are widely used methods in the literature [*Snee*, 1977; *McCarthy*, 1976; *Friedman and Friedman*, 1985].

In the absence of any limitations to obtaining new data, model validation with new data is the best method. In some cases, however, it is either impossible or too costly to obtain new data. *Snee* [1977] regards data splitting using either a random split half or splitting the data based on the underlying structural makeup as alternative procedures when the preferred method of evaluation on new data is not feasible. Models explaining time series data can use a natural time split and those explaining cross-sectional data can use a subset of sample points as validation samples [*Berk*, 1984]. Because of time and cost constraints we chose to validate the estimated metamodels using the random split-half validation (cross validation) technique.

6.1. *Cross Validation of the Metamodels*

Stone [1974] and *Snee* [1977] offer a good review and discussion of cross validation and alternative data-splitting methods. According to Snee, cross validation by data splitting is a method to test the in-use prediction accuracy of the model and simulate the complete or partial replication of the study. For purposes of cross validation, we split the data randomly into two approximately equal halves. The first subset, ss1, was used to estimate the model, while the second subset, ss2, was used to measure the predictive ability of the model, and vice versa. The cross-validation results shown in Table 3 demonstrate robustness and predictive power of the estimated metamodels. We also compared the sign and magnitude of the estimated coefficients from the two split-half models. In the groundwater metamodel, the signs of all the coefficients were the same in both samples, and the estimated coefficients were comparable in their magnitude. In the surface water metamodel, only two out of 31 coefficients had unmatching signs. These two coefficients, however, were not significantly different from

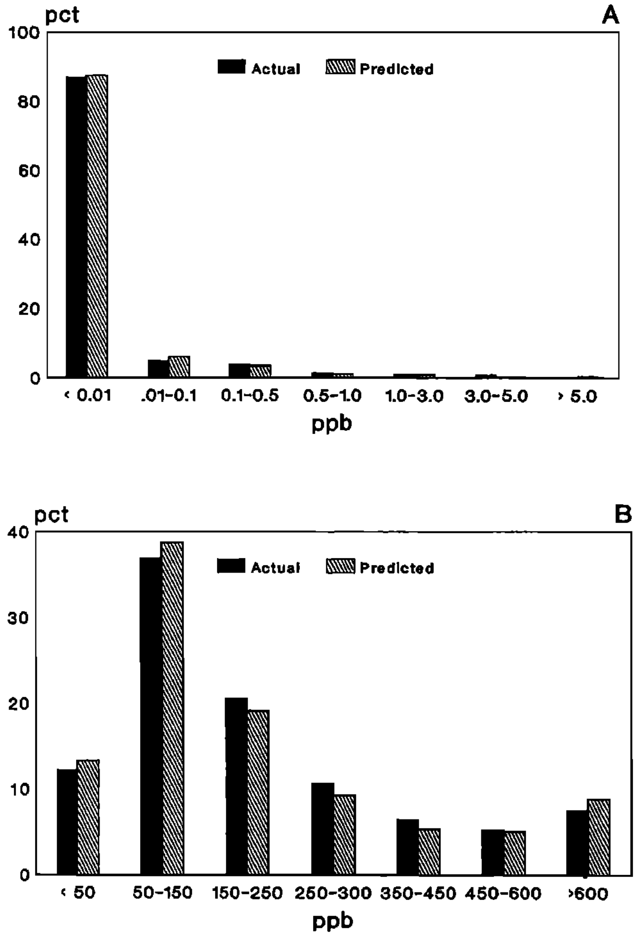


Fig. 2. Frequency distribution of the actual (simulated) and predicted observations for herbicide concentrations in (*a*) groundwater (15 m) and (*b*) surface water.

zero in both models. The cross-validation results for 1.2 m indicated same trend as that of 15 m.

6.2. *Validating the Metamodels With Monitoring Survey Estimates*

Validating the metamodels with monitoring data would be the ideal method of validation provided we had adequate monitoring data and the process models were adequately validated. Given the limited information on groundwater and surface water monitoring in a wide geographical area, we elected to perform approximate (crude) validation tests with the EPA's groundwater monitoring survey estimates [*EPA*, 1990]. Table 4 shows some of these results. Atrazine and simazine are the two herbicides detected at reasonably high percentage rates in the survey. The estimates predicted by using the metamodels indicate the same trend. The EPA estimates that atrazine is present, at or above 0.12 μg/L (survey minimum reporting limit), in about 1.7% of community water systems and 0.7% of rural domestic wells. Our estimation indicates that 1.2% of the soils in the region contribute to the groundwater detection limit of atrazine at or above 0.12 μg/L, which is clearly bounded by the EPA's estimates. At a minimum, we can state that the trends from our results are consistent with actual monitoring data.

7. HERBICIDE POLICY APPLICATION

Statistically validated metamodels for predicting regional agricultural nonpoint pollution enhance the evaluation of alternative agricultural chemical policies. By integrating the metamodels with an agricultural economic decision-making model that allows for substitution between herbicides and between weed control management strategies, chemical and nonchemical, we can evaluate the consequences of water quality policies regulating or restricting the use of herbicides. In this section we briefly discuss the decision-making model and its integration with the metamodels and the results from a herbicide policy of banning atrazine in corn and sorghum production in the study area. Atrazine is the most commonly detected herbicide in groundwater and surface water, forcing the EPA to reevaluate its ecological-economic trade-offs [*EPA*, 1990].

Assume agricultural production is represented by a joint production process where the two outputs, crop and pollution, are separable. The agricultural production and the nonpoint pollution process can be represented by the following expressions:

$$\mathbf{q} = f(\mathbf{x}), \tag{5}$$

$$\mathbf{z} = h(\Omega_x, \Psi, \delta). \tag{6}$$

Expression (5) represents farm outputs ($\mathbf{q}$) as a function of inputs ($\mathbf{x}$). The production technology f is assumed to follow the standard regularity conditions, including strict concavity. The function h translates the level of polluting inputs and practices employed in the production process into the amount of chemical concentrations in groundwater and surface water ($\mathbf{z}$), via the application rates and the physical and chemical characteristics of the polluting inputs (Ω_x), the soil characteristics (Ψ), and the meteorologic conditions (δ). A notable feature of our damage function is the inclusion of spatial characteristics and climatic conditions.

For the empirical analysis we used the agricultural economic decision-making model Resource Adjustment Modeling System (RAMS) [see *Bouzaher et al.*, 1990], which is an optimization model specified for a representative farm defined at the watershed (producing area) level. RAMS is a short-run profit-maximizing model that assumes a risk-neutral and competitive producer managing a multioutput-farm firm. A major feature of RAMS is that it has a weed control subsector, which defines the weed control and herbicide application activities and provides the important link to the chemical policy space.

The information on yield loss and cost trade-off from

TABLE 3. Cross Validation of the Metamodels for Herbicide Concentrations in Groundwater (15 m) and Surface Water

	15-m Average		Peak Stream	
Validation Statistics	ss1 (Pre-ss2)	ss2 (Pre-ss1)	ss1 (Pre-ss2)	ss2 (Pre-ss1)
R^2	0.83	0.85	0.83	0.83
	(0.87)	(0.88)	(0.82)	(0.82)
MSE^v/MSE^o	0.70	0.84	1.02	1.04

Here, ss denotes split-half sample. N_{ss1} = 3748; N_{ss2} = 3770. MSE^v and MSE^o are the validation and the original mean squared errors.

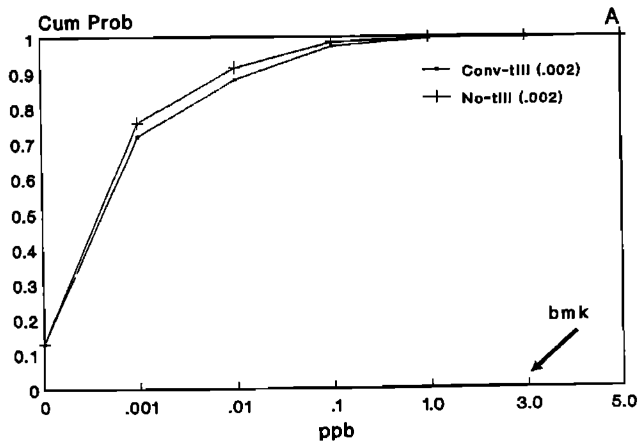
TABLE 4. Comparison of the Estimated Spatial Probability Value of Groundwater Concentrations of Selected Herbicides With the Groundwater Monitoring Estimates

Nonpoint Pollutant	Percent Occurrence of Pollutant*		Estimated Spatial Probability of Exceeding Minimum Reporting Limit, %	Survey Minimum Reporting Limit, μg/L
	Rural Wells	CWS†		
Atrazine	0.7	1.7	1.2	0.12
Alachlor	<0.1	0	0	0.50
Bentazon	0.1	0	0	0.25
Simazine	0.2	1.1	1.2	0.38

*Occurrence is defined as concentrations in excess of the *EPA*'s [1990] minimum reporting limit.
†Community water systems.

alternative weed control strategies and the relative herbicide substitution is inputed into RAMS through the WISH (weather impact simulation of herbicide) simulator [*Bouzaher et al.*, 1992]. A weed control strategy captures both the management and the technological aspects of weed control. Its structure is assumed to be made up of a primary herbicide treatment and a secondary herbicide treatment that will be applied only if the primary treatment fails for weather-related reasons. The choice of alternative weed control strategies determines the rate of substitution between herbicides and also the substitution between chemical and mechanical weed control. The estimated metamodels, which are proxies for social damage functions, and the RAMS model were exogenously linked to determine the concentration of atrazine and other herbicides used in corn and sorghum production under different tillage practices.

Given the baseline estimates of RAMS, we determined the chemical concentrations for the complete distribution of soils in the study area and computed spatial probability distributions. The spatial probability is measured as the proportion of soils for which a particular chemical under a given technology exceeds the toxicity-weighted benchmark (maximum contaminant level (MCL)). This measure, "probability that a soil is at risk," is more intuitively interpreted as a measure of the "spatial distribution of risk," and its usefulness is to target vulnerable soils and areas. Figure 3 illustrates the cumulative spatial probability distribution of atrazine under conventional- and no-till practices. Comparing our estimates with the MCL for chronic and acute exposure levels of atrazine in drinking water, 3 ppb and 100 ppb, we find that the probability of exceeding the benchmark is higher for surface runoff than for groundwater. The probability that the average concentration in groundwater will exceed the chronic benchmark value of 3 ppb is only 0.2%, regardless of tillage. The probability that the peak concentration in surface runoff will exceed the acute bench-



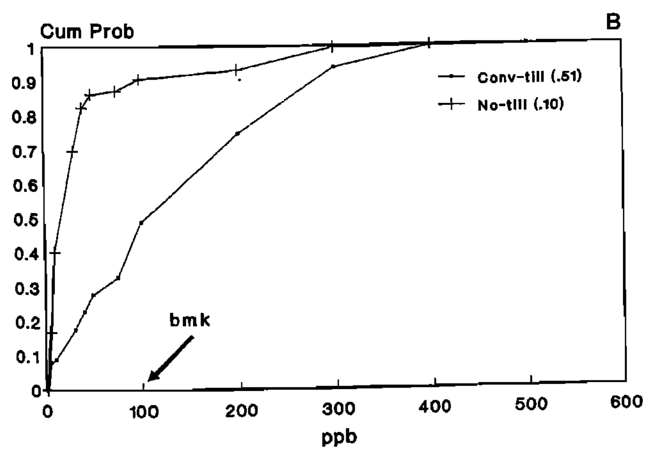


Fig. 3. Cumulative spatial probability distribution of atrazine concentration by tillage practice in (*a*) groundwater (15 m) and (*b*) surface water. Values in parentheses are the probability that the concentration will exceed the benchmark.

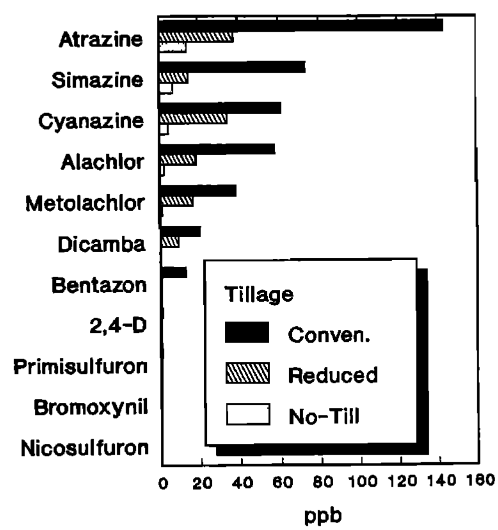


Fig. 4. Acute (peak) concentrations of selected corn and sorghum herbicides in surface water by tillage practice.

TABLE 5. Total Soil Erosion in the Corn Belt for the Baseline and Atrazine Ban Scenarios

	Soil Erosion, million tons	
Tillage	Baseline	Atrazine Ban
Conventional	429.1	452.3 (+5.4)
Conservation*	131.9	128.9 (−2.3)
All	561.0	581.2 (+3.6)

1 ton equals 907.2 kg. Values in parentheses are the percentage change from the baseline.
*Conservation tillage includes reduced till and no-till.

mark value of 100 ppb is reduced from 51% under conventional tillage to 10% under no-till practice. In general, a similar result holds for other herbicides (see Figure 4).

A major implication of these results is that groundwater quality is unimpaired by the conservation compliance policy. This suggests that the implementation of soil conservation policy will not lead to any unfavorable trade-offs between soil conservation and groundwater quality goals. But this is not the case for the water quality policy of banning atrazine. Our preliminary investigation suggests increased soil erosion in the Corn Belt because of shifts in cultivation practices from conservation tillage to conventional tillage. Table 5 shows the changes in the total soil erosion caused by an atrazine ban policy relative to the baseline for conventional and conservation tillage practices in the Corn Belt. Overall, soil erosion increased by 3.6% despite the conservation compliance provisions. By relaxing those provisions, we expect a more significant increase in soil erosion. These results are interesting in light of the recent debate on compatibility of soil conservation and water quality policies. Water quality is determined by a multiattribute vector, comprising such elements as sediment, nutrient, chemicals, and biotoxicants. Therefore a comprehensive analysis using a multiobjective framework is required to analyze trade-offs between these two policies [*Lakshminarayan et al.*, 1991].

8. CONCLUSION

Informed debate on agricultural nonpoint pollution policy requires evaluation of water quality on a regional basis in relation to management practices and hydrogeological conditions. Metamodeling has enormous potential for use in integrated agriecological economic systems designed for policy evaluation on regional levels. The overall implication of this study is that the metamodeling strategy can support integrated multimedia policy analysis in an environment of existing policy interventions with agents who respond to policy changes. The present illustration incorporates groundwater and surface water media, models relevant to existing policy interventions such as conservation compliance, and allows agents to respond to policy changes by altering weed control strategies. Without the method of metamodels, policy analysis would necessarily be less comprehensive, and consequently, less adequate to deal with the difficult but important task at hand.

Acknowledgments. Journal paper J-14921 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project 2872. The authors acknowledge the valuable suggestions of Wayne Fuller, appreciate the programming and modeling efforts of Todd Campbell and Derald Holtkamp, and thank the associate editor and the two anonymous reviewers for their helpful comments and suggestions.

REFERENCES

Baker, J. L., Effects of tillage and crop residue on field losses of soil-applied pesticides, in *Fate of Pesticides and Chemicals in the Environment*, edited by J. L. Schnoor, pp. 175–187, John Wiley, New York, 1992.

Bartlett, M. S., The use of transformations, *Biometrics*, *3*(1), 39–52, 1947.

Berk, K. N., Validating regression procedures with new data, *Technometrics*, *26*(4), 331–338, 1984.

Blanning, R. W., The construction and implementation of metamodels, *Simulation*, *24*(6), 177–184, 1975.

Bouzaher, A., Metamodeling in large-scale environmental economic systems, paper presented at the Operations Research Society of America/The Institute of Management Sciences Conference, Anaheim, Calif., Nov. 1991.

Bouzaher, A., D. Holtkamp, and J. F. Shogren, The Resource Adjustment Modeling System (RAMS), research report, Center for Agric. and Rural Dev., Iowa State Univ., Ames, Dec. 1990.

Bouzaher, A., D. Archer, R. Cabe, A. Carriquiry, and J. F. Shogren, Effects of environmental policy on trade-offs in weed control management, *J. Environ. Manage.*, *36*(1), 69–80, 1992.

Box, G. E. P., and D. R. Cox, An analysis of transformations, *J. R. Stat. Soc., Ser. B*, *26*(2), 211–252, 1964.

Box, G. E. P., and W. J. Hill, Correcting inhomogeneity of variance with power transformation weighting, *Technometrics*, *16*(3), 385–389, 1974.

Cabe, R., P. J. Kuch, and J. F. Shogren, Integrating economic and environmental process models: An application of CEEPES to atrazine, *Staff Rep. 91-SR54*, Center for Agric. Rural Dev., Iowa State Univ., Ames, May 1991.

Carsel, R. R., and R. L. Jones, Using soil and meteorologic data bases in unsaturated zone modeling of pesticides, *Ground Water Monit. Rev.*, *10*(4), 74–79, 1990.

Dean, J. D., P. S. Huyakorn, A. S. Donigian, Jr., K. A. Voos, R. W. Schanz, and R. F. Carsel, Risk of unsaturated/saturated transport and transformation of chemical concentrations (RUSTIC), vol. II, User's guide, *Rep. U.S. EPA/600/3-89/048b*, Environ. Prot. Agency, Washington, D. C., 1989.

Detroy, M. G., P. R. B. Hunt, and M. A. Holub, Groundwater-quality monitoring program in Iowa: Nitrate and pesticides in drinking water wells, in *Proceedings of the Agricultural Impacts on Groundwater*, pp. 75–88, National Well Water Association, Dublin, Ohio, 1988.

Donigian, A. S., Jr., and L. A. Mulkey, STREAM: An exposure assessment methodology for agricultural pesticide runoff, in *Fate of Pesticides and Chemicals in the Environment*, edited by J. L. Schnoor, pp. 297–330, John Wiley, New York, 1992.

Donigian, A. S., Jr., D. W. Meier, and P. P. Jowise, Stream transport and agricultural runoff of pesticides for exposure: A methodology, *Rep. U.S. EPA/600/3-86/011a*, Environ. Prot. Agency, Washington, D. C., 1986.

Environmental Protection Agency (EPA), National Pesticide Survey: Summary results, report, Washington, D. C., 1990.

Evans, B. M., and W. L. Myers, A GIS-based approach to evaluating regional groundwater pollution potential with DRASTIC, *J. Soil Water Conserv.*, *45*(2), 242–245, 1990.

Friedman, L. W., and H. H. Friedman, Validating the simulation metamodel: Some practical approaches, *Simulation*, *45*(3), 144–146, 1985.

Gassman, P., R. Cabe, A. Carriquiry, and J. F. Shogren, Multimedia assessment of atrazine fate and transport, paper presented at the 1991 International Summer Meeting, Am. Soc. of Agric. Eng., Albuquerque, N. M., June 1991.

Hartzler, G., and M. D. K. Owen, Weed management guide for 1991, report, Coop. Ext. Serv., Iowa State Univ., Ames, Dec. 1990.

Imhoff, J. C., R. R. Carsel, J. L. Kittle, and A. S. Donigian, Data base analyzer and parameter estimate (DBAPE) interactive computer program user's manual, *Rep. U.S. EPA-600/3-84-066*, Environ. Prot. Agency, Washington, D. C., 1990.

Jury, W. A., D. D. Focht, and W. J. Farmer, Evaluation of pesticide groundwater pollution potential from standard indices of soil-chemical adsorption and biodegradation, *J. Environ. Qual.*, *16*(4), 422–428, 1987.

Khan, M. A., and T. Liang, Mapping pesticide contamination potential, *J. Environ. Manage.*, *13*(2), 233–242, 1989.

Kleijnen, J. P. C., Regression metamodels for generalizing simulation results, *IEEE Trans. Syst. Man Cybern.*, *SMC-9*(2), 93–96, 1979.

Kleijnen, J. P. C., *Statistical Tools for Simulation Practitioners*, Marcel Dekker, New York, 1987.

Lakshminarayan, P. G., J. D. Atwood, S. R. Johnson, and V. A. Sposito, Compromise solution for economic-environmental decisions in agriculture, *J. Environ. Manage.*, *33*(1), 51–64, 1991.

Lawless, R. W., L. H. Williams, and C. O. Richie, A sensitivity analysis tool for simulation with application to disaster planning, *Simulation*, *17*(6), 217–223, 1971.

Lin, L., and E. F. Vonesh, An empirical nonlinear data-fitting approach for transforming data to normality, *Am. Stat.*, *43*(4), 237–243, 1989.

McCarthy, P. J., The use of balanced half-sample replication in cross-validation studies, *J. Am. Stat. Assoc.*, *71*(355), 596–604, 1976.

Snee, R. D., Validation of regression models: Methods and examples, *Technometrics*, *19*(4), 415–428, 1977.

Soil Conservation Service (SCS), National engineering handbook, section 4, Hydrology, report, U.S. Dep. of Agric., Washington, D. C., 1972.

Stevenson, F. J., *Humus Chemistry: Genesis, Composition, and Reaction*, pp. 402–419, John Wiley, New York, 1982.

Stone, M., Cross-validatory choice and assessment of statistical predictions, *J. R. Stat. Soc., Ser. B*, *36*(2), 111–147, 1974.

Wagenet, R. J., and J. L. Hutson, Modeling pesticide transport and transformation in the soil-plant system, in *Pesticide Chemistry*, edited by H. Freshe, pp. 347–357, VCH Publishers, New York, 1991.

Walton, W. C., *Practical Aspects of Groundwater Modeling*, National Well Water Association, Dublin, Ohio, 1984.

Wauchop, R. D., and D. Goss, Pesticide properties database, report, Coastal Plains Res. Sta., U.S. Dep. of Agric., Tifton, Ga., 1990.

A. Bouzaher, A. Carriquiry, P. Gassman, P. G. Lakshminarayan (corresponding author), and J. Shogren, Center for Agricultural and Rural Development, Department of Economics, Iowa State University, Ames, IA 50011.

R. Cabe, Department of Economics, New Mexico State University, Las Cruces, NM 88003.

(Received May 12, 1992;
revised January 9, 1993;
accepted January 29, 1993.)