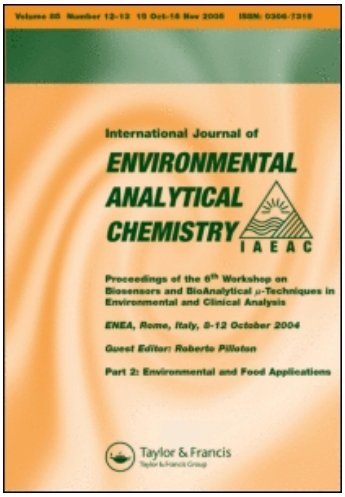

International Journal of Environmental Analytical Chemistry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713640455

Conducting Field Studies for Testing Pesticide Leaching Models

Charles N. Smith[a]; Rudolph S. Parrish[b]; David S. Brown[a]
[a] U.S. Environmental Protection Agency, Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, GA, U.S.A. [b] Computer Sciences Corporation, Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, GA, U.S.A.

To cite this Article Smith, Charles N. , Parrish, Rudolph S. and Brown, David S.(1990) 'Conducting Field Studies for Testing Pesticide Leaching Models', International Journal of Environmental Analytical Chemistry, 39: 1, 3 — 21
To link to this Article: DOI: 10.1080/03067319008027678
URL: http://dx.doi.org/10.1080/03067319008027678

Intern. J. Environ. Anal. Chem., Vol. 39, pp. 3–21
Reprints available directly from the publisher
Photocopying permitted by license only

CONDUCTING FIELD STUDIES FOR TESTING PESTICIDE LEACHING MODELS*

CHARLES N. SMITH

U.S. Environmental Protection Agency

RUDOLPH S. PARRISH

Computer Sciences Corporation

DAVID S. BROWN

U.S. Environmental Protection Agency

(*Received 17 February 1989; in final form 12 May 1989*)

A variety of predictive models are being applied to evaluate the transport and transformation of pesticides in the environment. These include well known models such as the Pesticide Root Zone Model (PRZM), the Risk of Unsaturated-Saturated Transport and Transformation Interactions for Chemical Concentrations Model (RUSTIC) and the Groundwater Loading Effects of Agricultural Management Systems Model (GLEAMS). The potentially large impacts of using these models as tools for developing pesticide management strategies and regulatory decisions necessitates development of sound model validation protocols. This paper offers guidance on many of the theoretical and practical problems encountered in the design and implementation of field-scale model validation studies. Recommendations are provided for site selection and characterization, test compound selection, data needs, measurement techniques, statistical design considerations and sampling techniques. A strategy is provided for quantitatively testing models using field measurements.

KEY WORDS: Field design, ground water, leaching, model testing, pesticides, validation.

INTRODUCTION

Pollutant transport and transformation models are used widely as predictive tools for assessing the impact on groundwater quality of chemicals released to the environment. Such models aid in developing water quality management strategies partly because they enable investigators to examine many combinations of pesticides, soils, crops, management systems, hydrogeologic settings, and meteorological conditions. The U.S. Environmental Protection Agency's Offices of Pesticide Programs and Groundwater Protection need field-validated models of pesticide fate for use in regulatory actions. Some models are designed to predict concentrations of chemical compounds within the soil profile, having the general objective of estimating the probability of groundwater contamination or the mass

*Correspondence: Environmental Research Laboratory, U.S. Environmental Protection Agency, Athens, GA 30613-7799, U.S.A.

that is likely to reach groundwater and subsequently be transported to domestic water supplies (i.e., wellheads). Accurate predictions from these kinds of models are essential if the process of risk-based management is to succeed. In addition, many of these models also are used to design cost-effective stategies for trends, early warning, and post-action monitoring.

In a sometimes ill-termed analysis, a model usually is examined for "validity" prior to release or endorsement for general use. There can be many aspects of such a validation exercise, but comparisons of model predictions with values observed in real-world situations are the most revealing. Several workers have addressed the need for the generalized framework to conduct model validation studies.[1–3] As part of an overall model-testing strategy, field studies sometimes are conducted to obtain data that can be compared statistically with model predictions. Such comparisons are considered essential to the proper testing of chemical transport and transformation models. In agricultural applications, interest centers on both unsaturated and saturated zones within the soil profile. A specific objective usually involves characterization of the distribution of chemical compounds within the (three-dimensional) study area, thus requiring statistical estimation of related parameters such as means and standard errors of various variables. Aside from statistical design and methodology considerations, there also are analytical and practical aspects of conducting field studies that ultimately affect the efficacy of the model-testing effort as well as the scope of inference based on the experimental data.

This work treats various aspects of conducting field studies, particularly in relation to testing chemical transport and transformation models, such as the Pesticide Root Zone Model (PRZM),[4] the Groundwater Loading Effects of Agricultural Management Systems (GLEAMS)[5] model, and the Risk of Unsaturated/Saturated Transport and Transformation Interactions for Chemical Concentrations (RUSTIC)[6] model. Although it is not possible in this forum to provide either a comprehensive treatment of or a generic design for field-based testing, the topics discussed are each considered relevant from either a statistical, analytical, or practical point of view. In any given field effort, study design should be driven by specific study objectives that relate to methodology for conducting statistically (or otherwise) a well-defined test of model performance.

The authors' work on a project for testing PRZM at a field site (Dougherty Plain) in southwest Georgia, as well as other work in the past and in progress, has largely influenced the recommendations that are made herein. One approach to field testing is presented, but other approaches exist. In particular, the quantitative methodology described by Parrish and Smith[7] outlines specific data objectives that must be achieved in order to conduct a rigorous test. Most existing databases, however, were not developed in a manner that permits testing along the lines suggested by Parrish and Smith. The emphasis here is on inferring the state of the field site at selected times, with the idea of comparing model predictions with inferred values representing the site as a whole.

No single validation study, regardless of the scientific or statistical rigors involved, would be considered convincing evidence of model acceptability for a given purpose by all observers. For example, any single (but elaborate and

statistically sound) field study is sure to be criticized as being inadequate to substantiate model performance at other geographic locations. Yet, attempting to implement several less elaborate studies with the same limited funds would likely result in shallow data sets inadequate for establishing any statistically defensible conclusions. A similar controversy exists regarding field versus laboratory testing. Some will argue that validation studies are most effective and more efficient when performed under highly controlled laboratory conditions. The counter argument is that all the complexity of the real world can never be duplicated in the laboratory and that such tests are inconclusive. Ultimately, several professional peer groups usually must be convinced using a variety of field and laboratory evidence before a complex predictive model can be said to have gained acceptance as an adequate predictive tool for a particular purpose. The focus of this work is on the design and implementation of statistically sound field validation studies as part of the overall model validation stategy.

SITE SELECTION AND PREPARATION

One of the first and most important steps in getting a field study underway is selecting a site that is appropriate for the model-testing effort. Generally, a large amount of resources will be involved over the course of the study (typically lasting 3 to 5 years), and the data generated from the effort likely will serve researchers for some time afterward, perhaps having usefulness in testing other models and developing new modeling approaches. More specifically, it is essential that the selection of a site be consistent with the scope and requirements of the model test intended. Monitoring equipment, wells, and other instrumentation become semi-permanent fixtures of the site that usually cannot be moved easily to another location.

Sandy soils, such as those found in coastal zones, have higher potential for infiltration and water–chemical movement than upland soils. In some areas, for example, sandy soil is present down to the level of the water table. Naturally, such sites generate relatively low variability in water and chemical movement, and they favor the formation of textbook style concentration profiles and contamination plumes as illustrated in Figure 1. Indeed, many models are likely to perform reasonably well in this situation, due to basic model formulation and simplifying assumptions. Other areas, including highly farmed lands, may have very complicated structures that produce much higher variabilities. Sandy soils with aquatards (i.e., clay lenses) may present obstacles to flow and transport, whereas the presence of nonhomogeneous preferential flow paths may enhance, in a relative sense, chemical movement and increase overall variability. The sampling requirements for these different types of soil structures are affected greatly by inherent variability. In addition, it is probably safe to assume that the more complex the site, the less likely it is that any existing process-based or empirical model will be able to provide acceptable predictions.

An investigation is needed early on to characterize the soil substructure at the site and to determine whether the site is within the scope of the intended testing

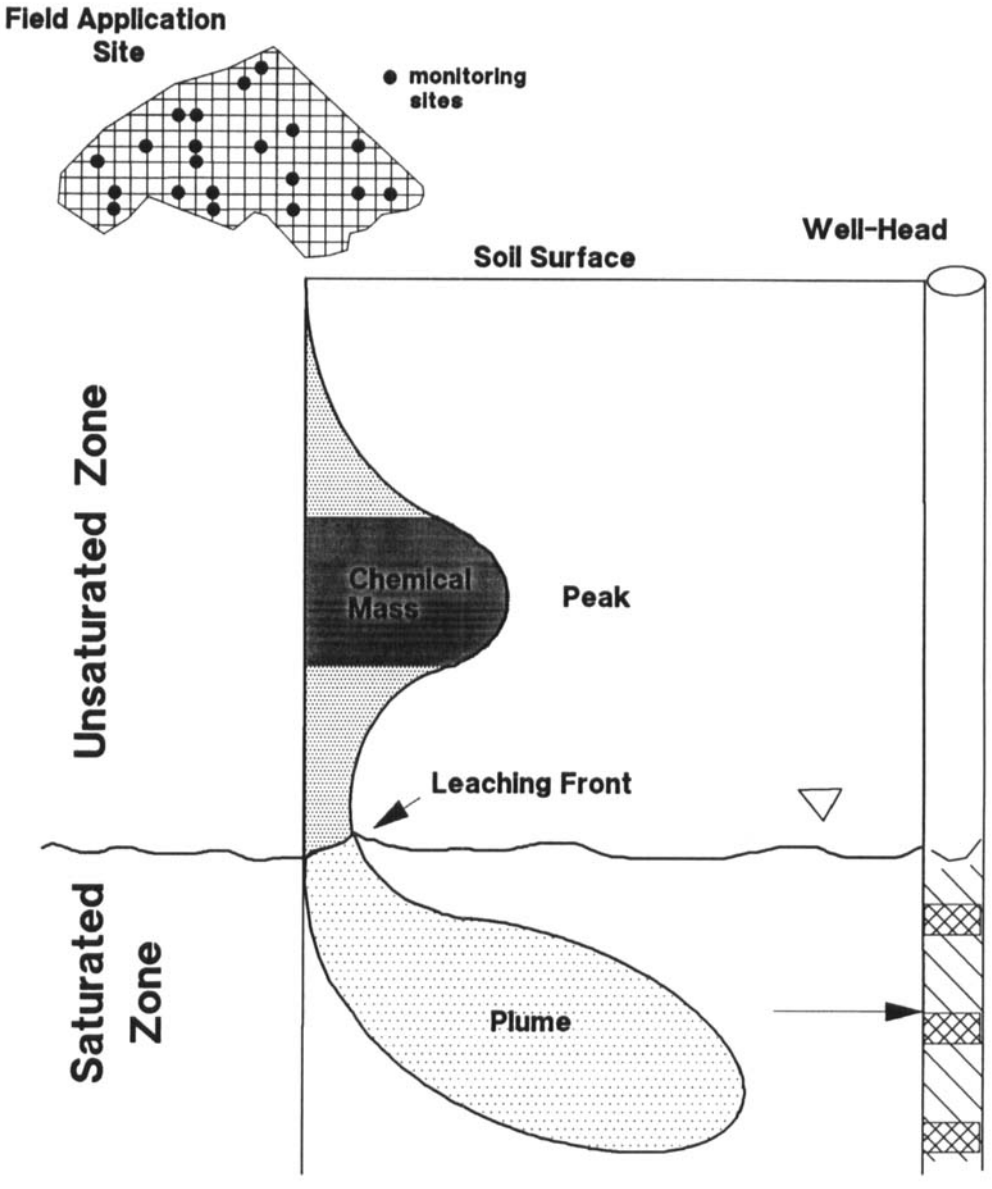

Figure 1 Idealized concentration profile and plume.

effort. Observation pits usually can be developed along the site perimeter with a backhoe, and soil samples for various characteristics can be obtained from the excavated side walls. These sampling areas can be restored before the study gets underway. From these pits, horizon depths can be determined, samples can be collected for obtaining preliminary estimates of chemical residues, and the presence or absence of significant preferential flow paths or lenses can be verified. High levels of pesticide residues or an overly complex flow structure might serve to

disqualify the site from further consideration, or the information can be used to advantage for developing sampling strategies.

The scope of inferences to be made from the study should be taken into account when determining the appropriateness of the site. Site characteristics that qualify the site as "representative" of a larger population of sites may be large in number, but scenarios in which the model is to be applied usually will dictate a set of characteristics or conditions that make some sites acceptable and others unacceptable. For example, testing a model in sandy coastal zone soils alone may not constitute a test of whether the model will perform adequately when applied in agricultural regions having vastly different soil and meteorological conditions. In an ideal statistical sense, many different field sites should be studied, just as many samples are selected randomly from a population; unfortunately, the realities of field work generally preclude using more than one or two study areas.

Several points of view could be considered when choosing a study site. First, testing the model under near-ideal conditions might identify gross model deficiencies that also would prevent adequate performance in more complex situations. Second, in a distributional sense, an "average" site might be selected as the most representative of the population of sites. Third, a site having more complex characteristics, while representing a relatively small part of the total population, could be chosen as a "worst case". This case might be used more for the purpose of assessing the limits of a model that already is known to perform reasonably well. A fourth case might be to choose a site that tests only a few well-defined aspects of the model. Whatever the criteria applied, however, the manner of choosing the site will determine to what extent the resulting inferences will apply.

Specific factors that should be considered in the site selection process include location within a major agricultural production area, complexity of soil structures and soil series, meteorological conditions, size of field, cropping practices, pesticide use, topography, runoff potential, accessibility, irrigation, proximity to domestic wells, management practices, depth to the water table, and others. The size of the field is of particular interest insofar as sampling requirements are concerned, which in turn depend on the overall level of effort required to meet study objectives. A field that is too small ultimately may become sampled so intensely as to destroy the integrity of the site. Consequently, it is prudent to select a site that is large enough to withstand the planned level of sampling activity while permitting reasonable distances to be maintained between all soil cores, wells and instrumentation. For pesticide leaching studies, one hectare is perhaps a minimal field size.

SITE CHARACTERIZATION AND SELECTION OF TEST COMPOUNDS

Attendant to each field site is a specific set of physical, chemical, biological, and crop management factors that influence hydrologic processes for pesticide movement and transformation. These characteristics are critical to the overall study because they provide input and parameterization for subsequent modeling. Characterization of the site involves collecting information concerning weather conditions, soil and hydrogeologic characteristics, crop management practices, pesti-

Table 1 General site characterization and pesticide data for model testing

Weather records
– precipitation
– evaporation
– max/min air temperature
– relative humidity
– solar radiation
Soil characteristics (by depth)
– identification of soil series
– horizon depths
– infiltration rate (or percolation) for surface
– hydraulic conductivity
– water content (moisture release curves)
– bulk density
– texture
– porosity
– organic carbon content
– depth to water table
– runoff potential
– pH and temperature
– microbial populations
Pesticide application and other field-determined parameters
– application method
liquid
wettable powder
granular formulation
date
rate
incorporation depth
– distribution coefficient of plant-soil application
– foliar washoff for foliar applied pesticides
– transformation rate with depth for soil and foliage
– sorption partition coefficients (by depth)
– volatilization
– plant uptake
– pesticide concentration profile for each sampling time
Crop management
– tillage practices
– other cultural practices

cide application, and processes. Specific data requirements may vary depending on the selected model. A summary of some important site characteristic data is presented in Table 1.

Weather conditions drive the hydrologic process, thereby affecting the movement and/or transformation of chemicals. In particular, the frequency, intensity, and magnitude of precipitation are major factors in movement. Because below-average rainfall conditions often result in limited chemical movement, supplemental irrigation systems should be considered as a means to ensure desirable soil moisture conditions (i.e., greater than or less than field capacity). If such a system is used, it should be designed so as to minimize variability in distribution patterns.

To provide appropriate input for the model, applied irrigation water and natural rainfall should be measured. Detailed information on establishing an on-site weather station is discussed by Smith *et al.*[8]

Soil characteristics affect the extent of chemical advection and dispersion. The potential for chemical movement may vary significantly among locations within a given study site, due largely to varying soil-related parameters. Detailed surveys can be used to collect soil physical data for developing soil maps. Generally, if different soil series exist within the field, it will be of interest to identify them for possible stratification uses. Other soil characterization data, as shown in Table 1, should be collected for specific models. Information on measurement of many of these parameters is provided by Smith *et al.*[8] These data are needed for various depths down to the aquifer. Within the United States, general soils data for the root zone (0 to 150 cm) are available from the United States Department of Agriculture Soil Conservation Service.[9]

A tracer compound, such as bromide, can be used to map the hydrologic response patterns in the field and to calibrate the hydrologic component of the leaching model. Such an application is recommended highly.

Crop management should be conducted in a manner that represents normal farming practice for the particular crop and area (e.g., spring plowing, planting, pesticide application, and cultivation).

Pesticides are available in liquid, wettable powder, and granular formulations. The formulation and the degree of soil incorporation, if any, will affect runoff, leaching and volatilization losses. Incorporation of pesticides into the upper few centimeters of soil has been shown to reduce runoff losses by as much as 50%, and it would be expected to affect the mass that potentially could leach to groundwater.[10] The amount of chemical applied to the site is important for defining source terms. Detailed sampling designs are recommended to quantify the total mass applied and to characterize the assoicated variability. If a foliar application is used, the distribution between the amount reaching the soil (throughfall) and the amount intercepted by the plants must be determined. Chemical and biological degradation process information (e.g., pathways and rate constants) is required for determining transformation rates in soil, pore water, and/or foliage. Similarly, sorption partition coefficients are required. Determination of partition coefficients as a function of soil depth requires collection of soil samples down through the profile for use in laboratory analyses.

Transformation rates, on the other hand, can be determined from soil samples collected after application. It is difficult to estimate field-based rates for each depth by using soil core sampling because the mass flux is affected by movement of the chemical into and out of the zone of interest. In addition, transformations can result from a combination of chemical reactions and microbiological degradation. In some cases, it might be possible to relate chemical reactions to specific chemical characteristics of the soil. An overall transformation rate can be determined over a given depth range based on soil cores, although this might have limited value for the modeling exercise. Another alternative involves estimating a rate coefficient in the surface zone that is adjusted in some manner for depth-related factors. Unfortunately, transformation rates, for some models, are sensitive parameters that

are difficult to estimate in the field. Thus, substantial effort is recommended for determining the field values of these rate constants.

Many factors must be considered when selecting pesticides to be used in a field study. The first question that usually arises involves the available resources. Often, compromises must be made concerning projected sample numbers based upon available instrumentation and availability of sensitive, efficient analytical methods. These factors must be balanced against the technical goals of the model testing project relative to the chemical properties and degree of usage of the candidate compounds. Another constraint is imposed on the experiment by pesticide registration restrictions, label application rates, and formulation types. Finally, it is often necessary to consider multiple-compound extraction and analysis compatibilities, particularly where measurement of both parent and daughter products is deemed necessary.

Analytical methods are required that will determine pesticide and daughter product residues with sensitivities in the low parts-per-billion (ppb) range. Methods must be adaptable for sampling various media including filter disks, soil, water, sediment, and plant tissue. Often, such large numbers of samples are involved that production-line analysis approaches are mandated. Protocols are required for careful handling of samples from the time collected in the field and throughout the analysis process. Quality assurance activities are required to ensure that the resulting data are of sufficiently high quality. Clear data-quality objectives (DQO's) based upon the model testing scheme to be used must be established prior to final design and costing of the field sampling and analysis protocols. Additional information is available in Sherma[11] and U.S. EPA.[12]

Analytical costs usually consume a large part of the project resources, especially for model testing studies involving large numbers of samples for chemical analysis. The planned number of samples is a question that usually surfaces initially from statistical considerations of the study design. Careful planning must be done to adequately address this crucial issue.

MEASUREMENT CONSIDERATIONS

When any chemical compound is applied to soil, it nearly always is distributed in a spatially variable manner despite efforts to maintain uniformity. The degree of variability will differ according to many factors, and, for sampling purposes, this variability directly affects the sample sizes required to achieve prescribed levels of precision in associated estimates. Whatever the variability present for the parameter of interest, it is important to avoid increasing that variability unnecessarily through other more controllable factors. In particular, the way samples are handled, extracted, and analyzed in the laboratory often affects overall variability to a significant degree.

Soil sampling can be viewed on the day of pesticide application as a surface-zone sampling activity, whereas on post-application days, it is a subsurface-zone effort. These types of sampling differ in that surface-zone measurements usually are based on mass per unit area, whereas subsurface samples are interpreted as mass per unit volume. Because some models predict on a compartment or horizon basis,

where a compartment covers a well-defined depth range, it is convenient to express measured values in terms of a volume determined laterally by a unit area and vertically by the depth dimensions of the horizon of interest. Model predictions must be properly integrated over compartments to match the physically integrated measurements associated with subsurface sampling. In other words, if soil cores are used to obtain subsurface samples within a specific depth range, model predictions on a compartmental basis for the same range ordinarily should be combined so as to mimic the physical sampling.

For liquid-applied compounds, there sometimes is a choice of methods for monitoring the application. One technique involves using filter paper positioned on the soil surface to intercept the spray as it is applied to the field; another technique involves analyzing soil samples collected after application.[8] Data developed during the Dougherty Plain project using the pesticide metolachlor indicated that soil samples, on average, accounted for 85% of the mass indicated by filter paper values; the coefficients of variation were approximately 37% and 45% for filter paper and soil samples, respectively. Relative to the mass actually distributed by the application equipment, the filter disk values accounted for approximately 83%, indicating about 17% was lost to drift. In that experiment, soil samples could not be collected as quickly as filter disks, so that volatilization at the soil surface was suspected as causing additional losses. By contrast, depending on soil moisture conditions, pesticide bound in granular formulations may not be released at all for some period of time after application. This supposition is supported by data collected during a drought-plagued season at Dougherty Plain.

Statistical Techniques in the Laboratory

Randomization of analyses within the laboratory is considered essential in order to avoid inadvertent trending and bias problems. An easy way to accomplish this is to assign sample numbers randomly prior to actually obtaining samples, then having the laboratory analyze them sequentially according to those numbers. In order to maintain analytical sensitivity and resolution, randomization plans should account for chemical-horizon characteristics. It is often advantageous to randomize and analyze such sample sets separately. More elaborate methods could be used for controlling other potential sources of bias. This helps to ensure that machine variations, operator differences, storage effects, or other chronological factors do not influence the measured outcomes inappropriately.

Quality control methods can be applied in the laboratory in certain circumstances. Control charts can be used for assessing consistency of data and for graphically presenting abnormalities. If samples have been properly randomized, control charts often can point out failing instrumentation, trending, or other problems.

Quality Assurance Procedures

The success or failure of any major field testing project is strongly dependent on

Table 2 Quality assurance planning

Field sampling
– decontamination of equipment
– sample collection protocols
– sample labeling, tracking and custody
– statistical randomization
– sample storage procedures and integrity checks
– adherence to statistical design
Extraction and analysis
– instrument calibration and maintenance
– internal standards
– calibration checks
– storage stability checks
– detection and evaluation of systematic and nonsystematic error sources
– personnel training procedures
– error correction
– supervisory analyst oversight
Data collation and processing
– database documentation (format, units)
– error checks
– calculation procedures
– data storage (database management)

the availability of dedicated analytical and field capability that incorporates rigorous quality assurance procedures. Because of the spatial variation associated with field soils, very large sample loads frequently are required to support meaningful conclusions. Sampling events usually continue through most or all of the growing season and frequently are concentrated most heavily during the early post-application period and influenced by major rainfall events. Sample storage procedures often must be established to circumvent temporary sample backlogs. The quality assurance plan should cover all phases of sample collection, handling and storage, extraction and analysis, and data collection. Exact details of the procedures should be tailored to the particular project. Some areas of consideration are given in Table 2.

Subsample Versus Whole-sample Extraction

When samples are collected from the field and brought to a laboratory for analysis, the objective usually is to determine the mass or concentration of specific compounds in the sample. Quite often the sample material will be blended into a nearly homogeneous mixture from which one or more subsamples are taken. This approach may be adopted for any of several reasons. Unfortunately, because the sample material often cannot be blended perfectly, this technique introduces an additional source of variation. That is, the mass or concentration for the whole sample that is inferred from the subsample measurements will vary from subsample to subsample. This source of variability is especially pronounced for granular-formulated pesticides that are relatively sparsely distributed within the whole sample. This situation is most likely to be encountered for sampling on the

day of application where moisture levels may be insufficient to release the chemical from its carrier. In some cases, even after thorough mixing, the numbers of granules that occur in individual subsamples can vary markedly, thus causing higher intrinsic variability that can be compensated for only by using more samples from the field.

One solution to this problem is to perform whole-sample extractions, which avoids the introduction of variation due to the mixing-subsampling process. Moreover, the degree of variability that is at risk should be investigated early in the study to determine whether alternatives to subsampling should be considered. This can be accomplished by using analysis of variance techniques to estimate the variance components of interest. For a granular formulation of aldicarb, based on preliminary work for the Dougherty Plain study for day-of-application monitoring, it was found that as much as 40% to 75% of total variation within sites could be attributed to subsampling. In that study, whole samples consisted of about 500 g of soil with 50 g subsamples. Consequently, whole-sample extraction techniques were selected.

Sampling from subsurface zones via soil cores requires a different approach because the mass of material that is removed from the bore hole, as it intersects the horizon of interest, generally is too voluminous to retain as a whole sample. While in the field, the soil mass usually is blended by hand and a subsample is taken. Even though this method may be subject to the same kinds of variability as discussed previously, the net effect is considered to be minimal because of the manner in which the chemical was transported (i.e., by water) into the sampling zone, because of the associated mixing phenomena, and because of other factors that contribute, perhaps more strongly, to overall variability.

Lysimeter Samples Versus Soil Samples

Different methods are used for sampling the unsaturated zone soil and soil pore water to determine concentrations of the compound under study. At the termination of the Dougherty Plain field study in which soil-solution samplers (i.e., suction lysimeters) had been installed,[13] pits were excavated in order to recover the samplers. Immediatiely prior to the excavation, samples were drawn from the lysimeters. Pits were dug adjacent to each of 20 sites corresponding to lysimeter installations, and 3 soil samples per depth were obtained from the depths corresponding to lysimeter locations. Both the soil and lysimeter water samples were analyzed for residues of bromide, a conservative tracer that had been applied 3 years earlier for flow calibration. For the soil samples, the total mass of bromide extracted was presumed to exist in the soil water; therefore, concentrations were calculated on the basis of moisture content.

The results revealed rather high variability among soil samples within individual pits. Correlation between the lysimeter samples and the soil samples was low to moderate (on the order of 0.4). Means were computed for each depth using data from the individual sampling pits, and simple linear regression was applied. This provided a small data set for analysis, but the coefficient of determination was very high ($R^2 = 0.999$). Interestingly, the equation for soil (ppb) versus lysimeter (ppb)

was: [Bromide in soil] = −3599 + 9.56 [Bromide in lysimeter water]. From an interpretive viewpoint, it is unclear what conclusions can be drawn from this relationship. The negative intercept might be due to the irreversible loss of bromide ion by ion exchange or sorption into soil particles. Doubt arises as to whether the lysimeters actually are providing measurements on the same parameter as are the soil samples.

For at least three reasons, lower bromide concentrations could be expected in lysimeter samples than in soil samples from the surrounding area. First, it is possible that the silica flour envelope installed around the lysimeter cups during installation may have affected the transport of electrolytes. Assuming the isoelectric pH of the flour was on the order of 2, it would have been characterized by a substantial net negative surface charge at pH values consistent with those of the test soils. This being the case, a significant ion exclusion effect on the bromide ion would be expected. Second, the flour was noticeably moister than the surrounding soils at the time excavations were made for removing the lysimeters; thus, a significant dilution effect may have occurred owing to the higher moisture content in the immediate area of the lysimeter. This is consistent with the conclusions of Litaor.[14] Third, the ceramic cup itself may have acted as a sorbent for the bromide ion. Any or all of these effects could have contributed to the substantially lower observed bromide concentrations for the lysimeter samples; observed soil means were larger than the lysimeter means by factors of 3 to 5. It is unclear from these studies whether the disparity between soil and lysimeter samples can be considered constant in time and space, but the strong correlation suggests that there may be some potential labor-saving utility in lysimeter measurements.

If such a strong relationship could be shown to hold definitively, lysimeter sampling could be used rather than soil sampling to measure concentrations in the soil profile, with adjustments made appropriately. On a field-mean basis, a strong linear relationship indicates that lysimeters could be used to locate regions of relatively higher concentrations within the profile. If so, lysimeter measurements may be used at least for determining when it is appropriate to sample the soil and at what depths. If the lysimeter samples could be used directly, or after some mathematical adjustment, as reliable measures of concentrations, there would be a distinct advantage realized because soil sampling via cores is destructive to the site and it is resource intensive, whereas lysimeter samples are easy to obtain, easy to analyze, and nondestructive to the site. They also can be obtained more frequently than soil core samples.

For model testing, however, it is critical that the method of sampling represent a bona fide means for measuring the concentrations that the model predicts. The current state of knowledge surrounding lysimeter sampling should be improved; that is, a better understanding is needed before such related data can be used in model testing situations. For a detailed review of soil-solution samplers, see Litaor.[14]

In planning field measurements of any sort, it is crucial that the intended use of the data be thoroughly considered at the outset. A common error is to inadvertently omit a subsidiary measurement needed to facilitate use of the primary measurements. For example, measuring the concentration of a tracer ion

in water samples obtained from lysimeters is of no use in estimating the concentration of tracer ion in the bulk soil unless the soil moisture content, porosity, and bulk density also are known. To avoid oversights of this kind, it is prudent to trace through all intended data applications prior to initiating any measurements.

Prior to collecting site-specific data and installing on-site monitoring equipment, model runs can be made using best available estimates of associated parameters to predict potential leaching characteristics. Such results can be useful for planning and design of the full-scale study with respect to placement of monitoring equipment and sampling frequencies.

DESIGN

General Testing Strategy

Once a field site has been selected, samples should be collected at various points in space and time thoughout the course of the study to determine concentrations of the test compound (including reaction products and metabolites), usually on a field-level or horizon basis. Corresponding model predictions should be made so that, ultimately, estimated means or other observed values can be compared with model predictions. Often, only field-level values (for example, the concentration in a given soil horizon at some point after application) are of interest, but this depends on the kinds of predictions that the model is capable of producing and the type of testing that is planned. In general, the practice of pairing specific model predictions (at a given place and time) with individual observations is insufficient for model testing because of high natural variability. Most deterministic models are not capable of accounting for such variances. It seems much more appropriate to consider model predictions in relation to more global parameters, such as the average concentration in a given horizon of the field or mass flux of a chemical leaving a horizon or entering the saturated zone.

Parrish and Smith[7] proposed a testing approach and an associated model prediction capability index for the case in which model predictions of concentration are to be tested for accuracy within a prescribed factor of true values. The method is based on joint confidence statements for horizon means at various times after application of pesticides. In a typical field situation, soil horizons (i.e., depths) are identified for which model performance is considered critical and, therefore, of interest in subsequent testing. For example, it may be desired to test whether a model is capable of predicting adequately at three distinct depths within the soil profile and at, say, three dates after application. Also, sample sizes are influenced strongly by the desired power of the statistical tests. Guidance for sample size determination is given by Smith *et al.*[15] and in other references as well.

In short, the experimental data must provide information on mean levels of concentration and the associated uncertainties of estimates at all points where comparisons are deemed to be of interest. In addition, samples must be collected for various other parameters to provide estimates for model parameterization. Thus, statistical considerations largely are centered on issues of sampling to derive

statistically valid estimates of unknown values. The essence of performing an acceptable test of the model, therefore, lies first in obtaining appropriate estimates for concentrations or other desirable measurement parameters.

Statistical test objectives are achievable only if adequate resources are allocated so as to obtain precise estimates of concentration means. The power of the final model test can be controlled by appropriate choices of sample sizes. For prescribed levels of power, variability, and statistical significance, required sample sizes usually can be determined. In the practical situation where resources are limited, however, there will be a choice of whether to reduce sample sizes or change other features of the test design. If sample sizes are curtailed too much, uncertainty in estimates of means will increase to the point where subsequent statistical tests will have low power, and, consequently, the results of the field-testing effort will be inconclusive and questionable. The alternative is to reduce the number of test points, either by using fewer depths or fewer dates. If the number of test points is reduced, sample sizes are more likely to be appropriate and the model test will be more powerful, albeit with less information on breadth of model capability.

Sampling Designs

In most situations where parameters are to be estimated from sample observations, various statistical sampling designs can be utilized, and several references on sampling are available. In much field work, simple random sampling or stratified random sampling with proportional allocation will provide good estimates with appropriate control and assessment of associated variability. Other approaches, of course, might be applicable in some situations. The emphasis here is on estimating field-level values, such as a mean concentration. In general, the locations of wells, instrument clusters, soil cores, and surface samples all should be randomized with respect to the field or strata involved. There normally is a practical advantage to locating instrument clusters and wells together, and it is best from a farming standpoint to keep these on row centers.

Whenever sample sizes are to be projected, consideration must be given to several factors simultaneously. These include the level of precision required in the parameter to be estimated, the level of confidence desired, and the level of variability in the observations. When estimating means, these values generally determine minimal sample sizes. The sample sizes, in turn, can be controlled by modifying these parameters. If, however, precision and confidence requirements are fixed, only the variability can be addressed. For specific sampling protocols, the level of variability is an implicit characteristic of the parameter of interest. Some situations, though, permit the variability to be reduced by changing the sampling approach. The overall objective of estimating a given field-level parameter can be achieved by selecting a large enough sample, holding as fixed the precision, confidence, and variability. If, on the other hand, variability can be reduced, a smaller sample size can be used.

Many considerations will influence the selection of test depths and dates that are to be used in testing the model, but once chosen, these "test points" drive the

subsequent sampling designs and strategies. The combined characteristics of the field site and the compounds under study, together with meteorological conditions, require that sampling dates be considered carefully. In some situations, both movement of the compound through zones of interest and compound degradation can occur rapidly. For this reason, post-application sampling should begin relatively soon after pesticide application.

Grids

Grid-based networks commonly are implemented in conjunction with random designs. Usually, a relatively large number of candidate sampling sites are identified by placing an imaginary rectangular grid at regular spacings over the site to be sampled (see Figure 1). A set of points, as determined by grid intersections, is selected at random. The selection itself should be based on tables of random digits or on a programmed algorithm. Grid spacing usually is related to a requirement or desire to maintain some minimal distance between all pairs of sample locations. Ideally, such a minimal distance should equal or exceed the so-called "range of influence" for the parameter of interest. That is, in order to achieve independence among sample observations and, therefore, maximize the information in the sample, the threat of spatial correlation should be removed insofar as possible. Sometimes it is possible to assess spatial variability for a particular parameter, but this generally requires a relatively heavy sampling effort. The fundamentals of spatial analysis are discussed in Clark.[16]

Pesticide Application Monitoring (Surface Zone)

Granular-formulated and liquid-formulated pesticides present different sampling problems during application. In either case, it is better to monitor the application rates through sampling rather than by calculating mass distributed. In addition, uniformity assumptions can be examined. Granular formulations, as discussed above, may involve non-uniform coverage, and often the application is in banded form. The size and location of the sampling unit, relative to the level of coverage and the band of application, must be considered. Generally, a simple randomized design is adequate for determining application rates of such pesticides. Information on variability and distributional form can be obtained from such data, and possible trending can be examined provided that samples are properly identified as to location in the field. When trends are being explored, characterization in terms of the distance traveled by the application equipment often is revealing, thereby preserving the order associated with the application itself. Usually, multiple rows are covered in each pass of the farming equipment during application. If the sample sizes are small, care should be exercised to ensure that sampling is not biased toward individual hoppers or other distribution channels.

For liquid-applied pesticides, spray nozzles normally produce overlapped applications. If sampling is restricted to row centers in this case, areas of higher concentration may be left unsampled. Whether such nonhomogeneity will have an impact on model testing depends on many factors. Nonetheless, it is desirable to

characterize variability across the spray boom and to sample accordingly. When using filter disks to intercept spray applications, it may be rather important to retrieve the filter papers as quickly as possible to minimize volatilization losses. For this reason and other practical considerations, systematic sampling designs generally are advantageous. Workers usually can retrieve the filter disks almost immediately after the spray equipment has passed over them. Other calibration-level efforts are geared toward obtaining samples directly from the nozzles prior to application.

Post-application Monitoring

Sampling at times after application may involve the root zone, the unsaturated zone, and the saturated zone. In each case, some form of randomized sampling is recommended. The saturated zone usually will be sampled via wells, perhaps at multiple levels, at various points in time. Information on design and construction of monitoring wells was provided by Barcelona *et al.*;[17,18] see also ASTM.[19] The initial location of the well sites (and instrument clusters) should be determined carefully with regard to randomization and stratification factors. Soil core locations, on the other hand, should be selected by random design before each sampling event, and the set of locations used in each event should be different. Because of spatial variability, it usually is not fruitful to attempt to produce profiles at individual locations in the field. Thus, it is recommended that mean profiles be developed for the field site as a whole or for strata within the site, based on randomly obtained samples.

Composite Sampling

Composite sampling techniques, as generally used, do not provide estimates of among-sample variability, which is needed for producing estimates of standard errors of means. Standard errors are essential for conducting quantitative model tests. In essence, composite sampling is an averaging process that physically combines separately obtained samples. When perfectly blended and perfectly analyzed, the result is equivalent to a mathematical average based on perfectly analyzed individual samples. A single composite sample, however, masks the variability among the samples, so that an estimate of the standard error of the mean is not obtainable. Also, variability among replicate composite samples is not the same as variability among individual samples, so the standard errors cannot be estimated from them. Although some specialized designs can be used with composite sampling to recover this type of information, the in-field requirements are substantial. There is a natural tradeoff between the field effort and laboratory analysis effort when composite sampling approaches are used. Because of this, it often is better to adopt a random sampling design.

Size of Sampling Unit

In most sampling efforts, it is important to obtain samples that physically are

representative of the material under study. With soil sampling, there can be substantial variability in the nature of the sampled material. Emphasis, however, must be placed specifically on variability associated with the parameter being measured. In addition, the levels of mass contained in samples may be important from an analytical viewpoint. The mass of sample needed usually can be projected on the basis of the expected mass of compound expressed on an area or volume basis. Variability in concentration level generally tends to increase with depth and time after application, and it may become more pronounced when concentration levels are low.

In agricultural applications where pesticides in various formulations are introduced to the field, granular formulations particularly can be more difficult to measure accurately. This is especially true at or near the time of application before natural mixing forces are operable. In most cases, there is a significant variability factor to consider when granular formulations are used. If the application were nearly uniform, the size of sample would be relatively unimportant. But where variation is significant, samples that are too small may have much higher inherent variability than larger samples, thus resulting in larger required sample sizes to achieve prescribed levels of precision in estimates. Of course, this translates to increased analytical effort in the laboratory.

Auxilliary Studies for Transformation Rates

Some aspects of the modeling effort might need to be resolved using relatively small scale independent studies. For example, field degradation rates are likely to differ from laboratory projections, so rates might need to be determined from separately managed experiments in the field. Results from laboratory transformation studies for aldicarb using soil from the Dougherty Plain field site showed an average half-life of 48 days,[20] while corresponding field-determined values were determined to be on the order of two to five times faster. Higher field temperatures might have been partly responsible for the difference. The differences could be explained partly by identifying and relating chemical characteristics of soils to the kinetic rate constants. Certainly, if model predictions were made on the basis of laboratory values alone, the model testing results would be affected adversely in this case. Similar transformation studies were conducted for the pesticide metolachlor, wherein no difference was found between field-determined and laboratory-determined rates.

MODEL TESTING

In all of the above, it has been assumed that the primary objective of the field study is to address the question of how well a given model can predict behavior of compounds in a real-world environment. Much of the field effort is directed toward simply measuring what is happening in the test site, but doing so in a statistically rigorous manner. The importance of this part of the effort lies ultimately in making comparisons with model predictions so that an objective

criterion can be applied to answer the model performance question. Unless the state of the field site is accurately assessed for parameters of interest, comparisons with model predictions cannot be conducted properly. As with any estimate of an unknown parameter, there always will be uncertainty surrounding the estimate, and comparisons with other values are meaningful only when made in relation to that uncertainty.

Confidence statements can be formed for each of the means that are estimated at each date and depth that are of interest in the testing of the model. If such statements are to be made jointly, the overall level of confidence can be controlled, and indices can be formed that measure how well the model predictions agree. Parrish and Smith[7] have described this method in detail. That work also discusses a test for determining whether model predictions fall within a prescribed factor of true values.

Given an adequate data base and an objective test criterion, any number of testing scenarios could be contrived to improve or better understand the model. This applies in the sense that different model parameterizations or various model assumptions can be tried, comparatively, to determine model performance. It is to this end that estimation of field means (i.e., concentrations or mass) are desired. Thus, the emphasis on appropriate statistical sampling designs pays off by enabling objective model tests to be carried out.

Acknowledgments

The authors express appreciation to W. R. Payne, Jr. of the U.S. EPA; to M. J. Brady, P. A. Churchman, and T. A. Simmons of Technology Applications, Inc., for analytical support and field sampling assistance; and to J. M. Cheplick of Computer Sciences Corporation for data processing and graphics support.

References

1. L. A. Burns, Validation methods for chemical exposure and hazard assessment models. In: *Environmental Modelling for Priority Setting Among Existing Chemicals*, Workshop 11–13 Nov. 1985, Munchen-Neuherberg. Munich, Gesellschaft fur Strahlen und Umweltforschung mbH (1986).
2. A. S. Donigian, Jr. and P. S. C. Rao, *Selection, Application and Validation of Environmental Models.* Presented at International Symposium on Water Quality Modeling of Agricultural Nonpoint Sources. Logan, UT, June 19–23 (1988).
3. U.S. Environmental Protection Agency. Work Plan: Model validation for predictive exposure assessment. Risk Assessment Forum (1988).
4. R. F. Carsel, C. N. Smith, L. A. Mulkey, J. D. Dean and P. P. Jowise, User's Manual for the Pesticide Root Zone Model (PRZM): Release I. U.S. Environmental Protection Agency, Athens, GA, EPA-600/3-84-109 (1984).
5. R. A. Leonard, W. G. Knisel and D. A. Still, GLEAMS: Groundwater loading effects of agricultural management systems. *Transactions of the ASAE* **30** (5), 1403–1418 (1987).
6. J. D. Dean and R. F. Carsel. A linked modeling system for evaluating impacts of agricultural chemicals. In: *Proceedings of the Agricultural Impacts on Ground Water—A Conference.* March 21–23, 1988. National Water Well Association, Dublin, OH, pp. 477–502 (1988).
7. R. S. Parrish and C. N. Smith. A method for testing whether model predictions fall within a prescribed factor of true values, with an application to pesticide leaching. Submitted for publication (1989).

8. C. N. Smith, D. S. Brown, J. D. Dean, R. S. Parrish, R. F. Carsel and A. S. Donigian, Jr., *Field Agricultural Runoff Monitoring (FARM) Manual.* U. S. Environmental Protection Agency, Athens, GA, EPA/600/3-85/043 (1985).
9. W. D. Goran, *An Interactive Soils Information System Users Manual.* U.S. Army Corps of Engineers, Construction Engineering Research Laboratory and U.S. Department of Agriculture, Technical Report N-163, 25 pp. (1983).
10. T. E. Waddell and B. T. Bower. *Managing Agricultural Chemicals in the Environment: The Case of a Multimedia Approach.* The Conservation Foundation, Washington, DC, 78 pp. (1988).
11. J. Sherma, *Manual of Analytical Quality Control of Pesticides and Related Compounds in Human and Environmental Samples.* U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-600/2-81059 (1981).
12. U.S. Environmental Protection Agency. *Handbook for Analytical Quality Control in Water and Wastewater Laboratories.* U.S. Environmental Protection Agency, Cincinnati, OH, EPA-600/4-79-019 (1979).
13. C. N. Smith and R. F. Carsel. A stainless-steel soil solution sampler for monitoring pesticides in the vadose zone. *Soil Sci. Soc. Am. J.* **50** (1), 263–265 (1986).
14. M. I. Litaor. Review of soil solution samples. *Water Resources Research* **24** (5), 727–733 (1988).
15. C. N. Smith, R. S. Parrish and R. F. Carsel, Estimating sample size requirements for field evaluations of pesticide leaching. *Environmental Toxicology and Chemistry* **6**, 343–357 (1987).
16. I. Clark, *Practical Geostatistics* (Applied Sciences Publishers, London, 1979), 129 pp.
17. M. J. Barcelona, J. P. Gibb, J. A. Helfrich and E. E. Garske. *Practical Guide for Ground-water Sampling.* (Illinois State Water Survey, Champaign, Illinois, 1985), Report no. 374, 94 pp.
18. M. J. Barcelona, J. F. Keely, W. A. Pettyjohn and A. Wehrmann, *Handbook of Ground Water.* U.S. Environmental Protection Agency, Ada, OK, EPA/625/6-87/016 (1987).
19. ASTM. *Groundwater-Contamination: Field Methods.* A. G. Collins and A. I. Johnson (eds.). Symposium sponsored by ASTM Committee D19 on water and D18 on soil and rocks. (Cocoa Beach, FL, Feb. 2–7 1986). ASTM Special Technical Publication 963, 491 pp. (1988).
20. P. S. C. Rao, K. S. V. Evardsson, L. T. Ou, R. E. Jessup, P. Nkedi-Kizza and A. G. Hornsby. Spatial variability of pesticide sorption and degradation parameters. In: *Evaluation of Pesticides in Groundwater* (W. Y Garner, R. C. Honeycutt and H. N. Nigg, eds.) (ACS Symposium Series No. 315, American Chemical Society, Washington, DC, 1986), pp. 100–115.