

# CEEOT-MMS: A Macro Modeling System for Environmental Assessment

---

## Research Team

Edward Osei  
Asfaw Bekele  
Bing Du  
Ruth Rose  
Larry Hauck  
Ali Saleh  
Jim Houser  
Keith Keplinger  
Larry Beran

TR0303

March 2003

TIAER \* Texas Institute for Applied Environmental Research  
Tarleton State University \* Box T0410 \* Tarleton Station \* Stephenville TX 76402  
Phone 254-968-9567 \* Fax 254-968-9559 \* Internet <http://www.tarleton.tiaer.edu>

# Acknowledgments

The research on which this report is based was financed by the U.S. Environmental Protection Agency (USEPA) Office of Policy, cooperating agreement number CR 826807-01-1.

The authors acknowledge useful insights and critical review provided by Bruce Babcock, director of the Center for Agricultural and Rural Development (CARD), Iowa State University, and other CARD staff. The authors are also indebted to review comments provided by Clay Ogg, USEPA, and Ron Jones, director of the Texas Institute for Applied Environmental Research (TIAER). Editorial assistance and research support was provided by Lee Ann Huseman, Dianne Swanson, Don Gosdin, and Ruth Brannon. The findings of this report remain the sole responsibility of the authors.

## Authors

Asfaw Bekele, sr. research associate, TIAER, [abekele@tiaer.tarleton.edu](mailto:abekele@tiaer.tarleton.edu)

Larry Beran, assistant director for operations and programs, TIAER, [beran@tiaer.tarleton.edu](mailto:beran@tiaer.tarleton.edu)

Bing Du, sr. research associate, TIAER, [bdu@tiaer.tarleton.edu](mailto:bdu@tiaer.tarleton.edu)

Larry Hauck, assistant director of environmental sciences, TIAER, [hauck@tiaer.tarleton.edu](mailto:hauck@tiaer.tarleton.edu)

Jim Houser, research scientist, TIAER, [jhouser@tiaer.tarleton.edu](mailto:jhouser@tiaer.tarleton.edu)

Keith Keplinger, research economist, TIAER, [keplinger@tiaer.tarleton.edu](mailto:keplinger@tiaer.tarleton.edu)

Edward Osei, sr. research economist, TIAER, [osei@tiaer.tarleton.edu](mailto:osei@tiaer.tarleton.edu)

Ruth Rose, research assistant, TIAER, [rose@tiaer.tarleton.edu](mailto:rose@tiaer.tarleton.edu)

Ali Saleh, research scientist, TIAER, [saleh@tiaer.tarleton.edu](mailto:saleh@tiaer.tarleton.edu)

---

# Contents

<b>CHAPTER 1 Introduction</b> .....	9
<b>CHAPTER 2 Design of the MMS</b> .....	11
2.1 Components of the MMS.....	11
2.1.1 Central Interface.....	12
2.1.2 Data Repository.....	14
2.1.3 Environmental Modeling Component.....	14
2.1.4 Economic Modeling Component .....	14
2.2 Development and Operation of MMS.....	14
2.2.1 Development of MMS .....	14
2.2.2 Operation of the MMS .....	17
<b>CHAPTER 3 Development of Representative Farms</b> .....	19
3.1 Disaggregation Procedure.....	21
3.1.1 Selection of Primary View and Level of Aggregation.....	22
3.1.2 Generation of Variable Distributions .....	23
3.1.3 Allocation to Counties .....	24
3.1.4 Assignment to Farms .....	24
3.1.5 Augmenting Farm Definitions .....	24
3.2 Clustering Procedures .....	25
3.2.1 Data Preparation.....	26
3.2.2 Estimating Clusters with the Appropriate Method .....	26
3.2.3 Interpreting Cluster Output .....	27
3.2.4 Application of Statistical Clustering in CEEOT-MMS .....	28
<b>CHAPTER 4 National Database</b> .....	29
4.1 Data Types and Sources.....	29
4.2 Special Constructed Databases .....	32
4.2.1 Farm Definitions Database .....	32
4.2.2 Crop Operations Library.....	33
<b>CHAPTER 5 Indicators and Prominent Factors</b> .....	35
5.1 Environmental and Economic Indicators.....	35
5.2 Prominent Factors Impacting Indicators.....	36
5.3 Ecologically Uniform Regions (Ecoregions).....	37
5.4 Farm Types .....	37
<b>CHAPTER 6 Applications of CEEOT-MMS</b> .....	39
6.1 Static Scenarios.....	39
6.2 Spatial Comparative Analysis.....	41
6.2.1 Spatial Impacts.....	41

6.2.2 Vulnerability Indices.....	42
6.3 Use of Agro-Ecological Metamodels.....	42
6.4 Risk Analysis .....	43
<b>CHAPTER 7 Extensions of the MMS.....</b>	<b>45</b>
7.1 Advanced Scenario Options (Dynamics).....	45
7.2 Automated Calibration Routines.....	45
7.3 Routing of Environmental Indicators.....	45
<b>CHAPTER 8 Conclusions .....</b>	<b>47</b>
<b>REFERENCES .....</b>	<b>49</b>

---

# Tables

<b>Table 3.1</b>	Categories of Land in Farms and Corresponding Bounds in Acres.....	22
<b>Table 4.1</b>	Data Types and Sources Used in MMS .....	29
<b>Table 4.2</b>	Columns in the Crop Operations Library.....	33
<b>Table 6.1</b>	Categories of Static Scenarios with Illustrative Examples .....	39



---

# Figures

<b>Figure 2.1</b>	Schematic of the Macro Modeling System.....	11
<b>Figure 2.2</b>	Development of MMS .....	15
<b>Figure 2.3</b>	Steps Involved in a MMS Application.....	17
<b>Figure 3.1</b>	Procedure for Generating Representative Farms .....	20
<b>Figure 5.1</b>	Hierarchical Representation of Modeling Units .....	38



---

# Chapter 1

## Introduction

Nonpoint pollution from agricultural sources continues to receive attention in national agricultural and environmental policy. One of the most productive means to inform policy makers of the issues and options relating to agricultural nonpoint source issues is to investigate the efficacy of alternative practices and policies. To this end, several watershed studies have been conducted by various institutions (e.g., Pratt et al., 1997 on the Upper North Bosque River watershed in north central Texas; Heatwole et al., 1987 on the Lake Okeechobee watershed; McNitt et al., 1999 on the Lake Fork Reservoir watershed in northeast Texas; Keith et al., 2000 on the Upper Maquoketa River watershed in Iowa). These watershed studies provide insight into impacts of alternative water quality measures implemented across entire watersheds.

While watershed-based studies are useful in their own right, they are not directly applicable to evaluating broader policy efforts due to heterogeneity between watersheds. Other studies have attempted regional analyses of agricultural practices on water quality parameters (e.g., Babcock et al., 1997). To date, however, application of computer simulation models for national nonpoint source policy assessment is relatively rare, and applications for livestock agriculture are even rarer. The objective of this report is to outline the foundational structure and framework for a macro modeling system for national nonpoint source policy assessment using preexisting simulation models. This framework makes use of existing models used within the Comprehensive Economic and Environmental Optimization Tool (CEEOT) framework (Jones et al., 1993; Osei et al., 2000) for watershed policy assessment. Since the CEEOT suite of models has been used extensively for issues relating to livestock agriculture, the macro modeling system will enable very useful livestock and environmental policy assessments of broad scope.

Evaluation of environmental impacts of livestock practices at the national level requires special methodologies not needed for watershed analyses. Relevant environmental indicators differ from one region to another, and the factors responsible for their levels also differ considerably. Thus a standard methodology is needed to account for these differences in a systematic manner.

The methodology outlined in this report deals primarily with crop nutrients and manure, which have been the focus of previous watershed studies, and other nonpoint source issues relating to livestock agriculture. The modeling system uses a micro simulation approach. Representative farms are derived from aggregate data and these farms are simulated to help reflect variations in impacts across the entire nation or the region of choice. Appropriate statistical methods are then used to aggregate small-scale simulations to the national or regional level. The framework presented here is to be used to assess

impacts of various livestock practices within a national scope. Briefly, the purpose of the macro modeling system (MMS) is twofold.

1. To provide a computer modeling system for timely national/regional assessment of policies/practices related to livestock agriculture and the environment
  - a. Rapid assessment for large regions or the entire nation
  - b. Rapid assessment for small regions or large basins
2. To obtain an assessment of how and why economic and environmental impacts differ from one region to another.
  - a. Development of vulnerability indices for geographic areas and practices
  - b. Spatial impacts: efficiency and equity issues relating to various policy scenarios (e.g., best management practices).

Rapid simulation of large regions for broader policy assessment is attained by automation of two of the three CEEOT-component models Agricultural Policy Environmental eXtender model (APEX) and Farm Economic Model (FEM). Since the same farm-level models are used for small watershed studies, this process also enables application of the modeling system (CEEOT-MMS) for basin-scale studies. In addition, regional, basin, and nationwide studies present the challenge of spatial distributions of relevant indicators and factors. Consequently, CEEOT-MMS is being developed to account for and explain the relationship between the spatial distribution of these indicators and factors.

This report provides a brief description of the CEEOT-MMS system. The following chapter will present the design of the MMS, including the components as well as the methodology and operation of the system. Chapter 3 outlines the methodology used to derive representative farms for the simulations, a key component of the MMS. Chapter 4 presents the components of a national database being constructed to serve as input into the modeling system. Chapter 5 describes the prominent indicators and factors that will be the subject of interest in application of CEEOT-MMS. Chapter 6 outlines various applications of CEEOT-MMS as it is currently being developed. Then Chapter 7 provides suggestions for extension of the system for other applications. The concluding chapter summarizes the contents of the report.

---

# Chapter 2

## Design of the MMS

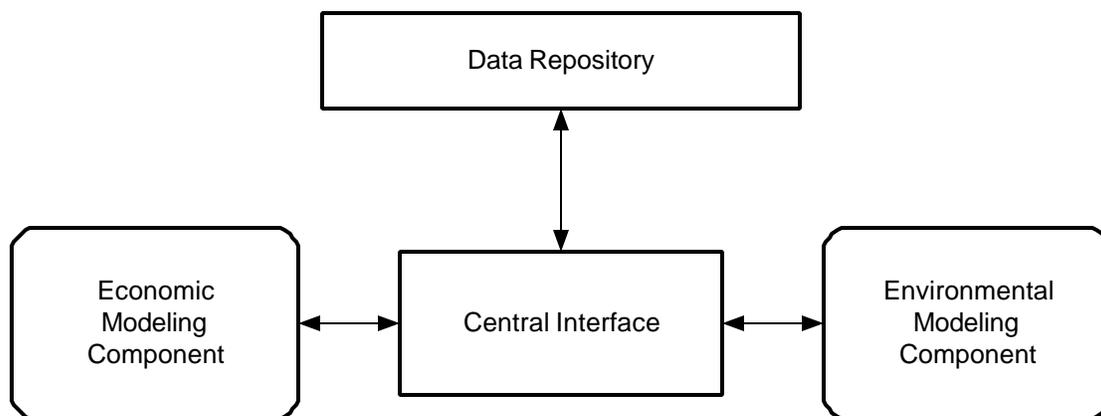
The objectives of the macro modeling system are achieved by a multidisciplinary system of models, linked together by a programming interface that allows for rapid and flexible simulations. This chapter discusses the components of the MMS as well as the methodology underlying its operation.

### 2.1 Components of the MMS

The macro modeling system consists of the following four components (Figure 2.1).

1. Central interface
2. Data repository
3. Environmental modeling component
4. Economic modeling component

The central interface links the other components and performs several automation functions. The data repository is the container for input and output data. The environmental modeling component performs environmental simulations and predicts the environmental response of livestock agricultural practices, providing estimates of edge-of-field nutrient losses. Finally, the economic modeling component performs economic simulations, yielding estimates of economic indicators. Each of these components is discussed below.



**Figure 2.1** Schematic of the Macro Modeling System.

## 2.1.1 Central Interface

The central interface (interface) embodies a major portion of the work involved in transforming the CEEOT suite of models into an effective macro analysis tool. To achieve this, the interface was developed with the following five functions in mind.

1. The interface provides a user interface for submission of automated tasks to evaluate multiple scenarios over a set of regions (a user-defined geographic study area, e.g., basin or a region comprising several states). The user interface consists of screens and supporting routines that allow computer modelers flexibility in specifying the following:
  - a. Geographic area for the simulations
  - b. Scenarios
  - c. Farm types to include in simulations
  - d. Other simulation options, such as time horizon
  - e. Output options
2. The interface provides for automated determination of modeling units to include in simulations.

The most important task in applying the computer simulation models is to determine the kinds of livestock agriculture operations in the study area. This assay involves an assessment of farm sizes, types, management practices, and a host of other endogenous and exogenous factors. Exogenous factors include weather, soil type, input and output prices, market conditions, and government policy. Endogenous factors include land uses, management practices, and producer responses to changes in policy scenarios. If the assessment of the operations is wrong, the results of the study would be inaccurate regardless of the simulation models used.

In the MMS, farm types are determined by applying statistical and mathematical programming tools to extant national databases and a wide variety of other state data sources. The major databases employed in this effort are the Agricultural Census, National Resources Inventory (NRI), crop production databases, cropping practices surveys, as well as other USDA data sources. Weather, land use, and other data needed for environmental model simulation are obtained from various sources including EPA's BASINS application system. Chapter 4 provides a list of the types of data used and the sources from which these data were obtained.

Using special disaggregation techniques and statistical clustering methods, representative farms are generated for economic and environmental model simulations depending on the factors (e.g., rainfall, land use) that have impact on the indicators of concern (e.g., phosphorus losses). Key indicators and prominent factors for agricultural nonpoint source issues are outlined in Chapter 5. Since environmental indicators are more heavily impacted by a wider array of factors, such as soil types, slopes, and weather, a different set of representative farms, usually including more farms, will typically be used for environmental model simulations than for economic

model simulations. Chapter 3 provides more detailed information on the procedure used to generate the representative farms.

3. The interface provides automated generation of input data for economic and environmental models. Once the farms to be included in the simulations have been selected, the input data files for the economic and environmental model simulations need to be generated. The interface generates farm and scenario definitions that will be read into the simulation models.
  - a. Various routines automatically generate input data files for the environmental modeling component.
  - b. Other routines develop input files for the economic model. The economic model then builds the farm and scenario definition files for batch simulations.
4. The interface provides for summary and analysis of output data and presentation in specified formats. Output data can be processed to provide more insight into the results from model simulations. Some of the uses of output processing include the following:
  - a. Metamodels: These functional relationships of output variables to input parameters enable us to predict or estimate the impacts of various scenarios on user-specified farm types without having to perform the MMS simulations for each of these situations.
  - b. Risk analysis: In many instances it is more useful to give interval estimates of scenario impacts in addition to point estimates. As examples: (i) there is a 95 percent probability that a phosphorus agronomic rate will lead to at least 35 percent decline in P losses relative to a nitrogen agronomic rate, while the expected (most likely or “point estimate”) impact is a decline of 45 percent; (ii) there is an 80 percent probability that the no-till scenario will lead to at least 40 percent decline in sediment losses and that the economic impact will be between  $-\$5/\text{acre}$  and  $-\$2/\text{acre}$ , while the expected (most likely or “point estimate”) impact is ‘55 percent decline in sediment losses and  $-\$4/\text{acre}$  change in profits’.
  - c. Development of vulnerability indices (and delineation of hot spots)
  - d. Graphical presentations: presentation of results in tables, charts, graphs, etc.
5. The interface provides for submission of output data to data repository. The interface will submit output data to be stored in MS Access® databases for future use and analysis.

## 2.1.2 Data Repository

The data repository serves as container for all input and output data. All output data and most of the input data are stored in MS Access database files. The types and sources of input data used within the MMS are discussed in the chapter on the national database.

## 2.1.3 Environmental Modeling Component

The Agricultural Policy Environmental eXtender (APEX) model will be used for farm level simulations within the MMS to predict edge-of-field nutrient losses. APEX has been widely used within watershed studies to generate environmental indicators of a wide range of scenarios. In order to facilitate rapid and automated simulations of large regions, a project-specific, modified version of I\_APEX, which is a user-friendly interface to the APEX model, will be used. I\_APEX was developed by the Center for Agricultural and Rural Development (CARD), Iowa State University, and allows for automated generation of multiple APEX simulations (Gassman, 2002). All input data are received in MS Access<sup>®</sup> format so that users do not have to generate APEX input files directly. Once the Access<sup>®</sup> database is developed the required number of simulations, often tens of thousands, can be performed in an automated manner within the I\_APEX environment.

## 2.1.4 Economic Modeling Component

Farm level economic indicators will be generated using the Farm Economic Model (FEM), which was developed jointly by the Texas Institute for Applied Environmental Research (TIAER) and CARD as part of the Environmental Protection Agency funded Livestock and the Environment: A National Pilot Project (NPP). FEM provides for flexible definition of farms and will be augmented to allow for more flexible scenario definition.

# 2.2 Development and Operation of MMS

## 2.2.1 Development of MMS

Development of MMS proceeds in two phases: a data generation phase and a development phase (Figure 2.2).

Data generation phase:

Two main activities are performed in the data generation phase.

1. Collection of input data and storing them in the data repository: Input data include environmental modeling data in GIS format and other databases that provide information on farm structure and operations.

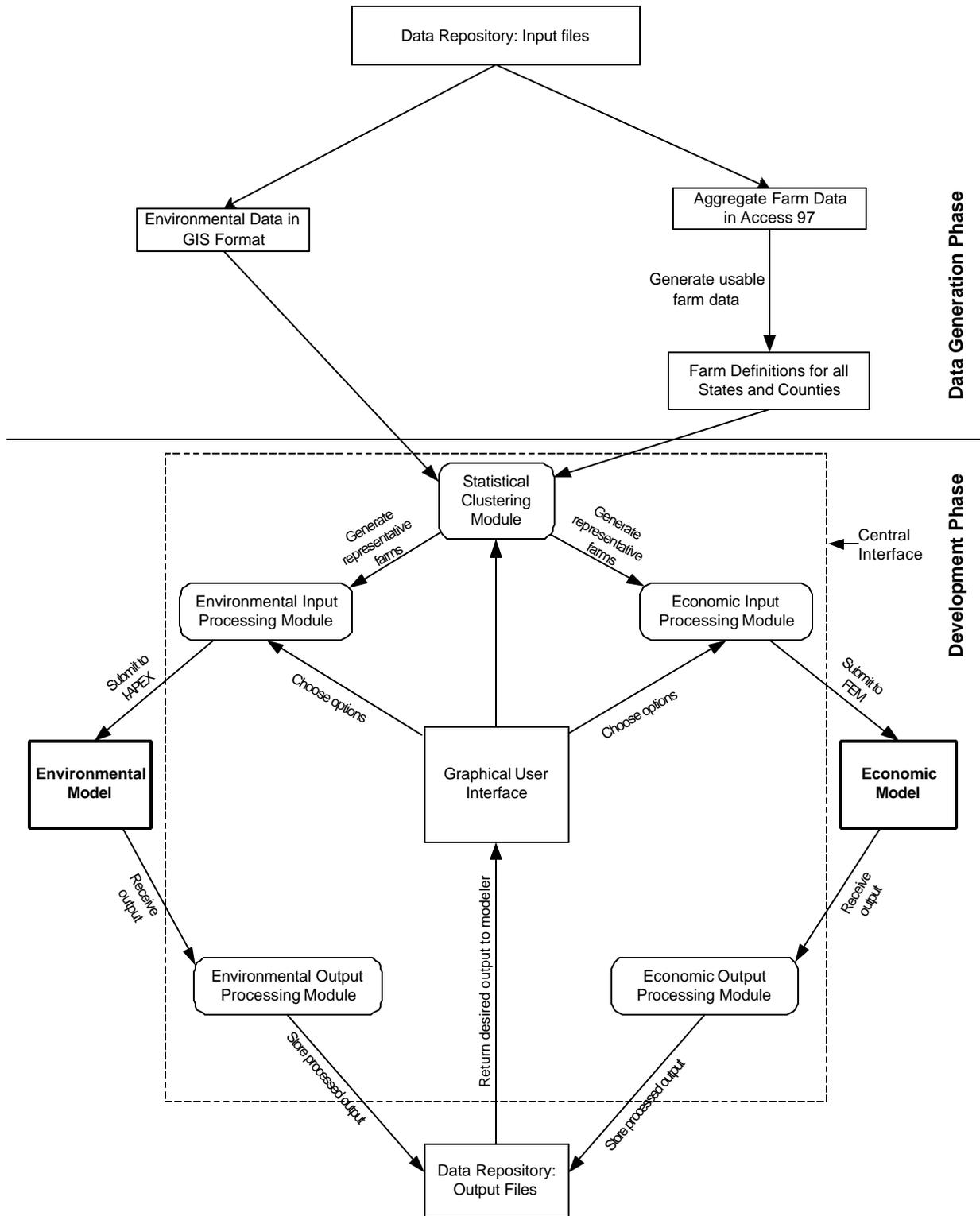


Figure 2.2 Development of MMS

2. Development of farm definitions: To facilitate model simulations using existing databases, a set of hypothetical farms are developed using a disaggregation procedure for all counties in every state. This provides a set of farm definitions for simulating any region of the country without the need to repeat the disaggregation process. Disaggregation needs to be performed again only when new data becomes available, such as when a new Agricultural Census is performed.

**Development Phase:**

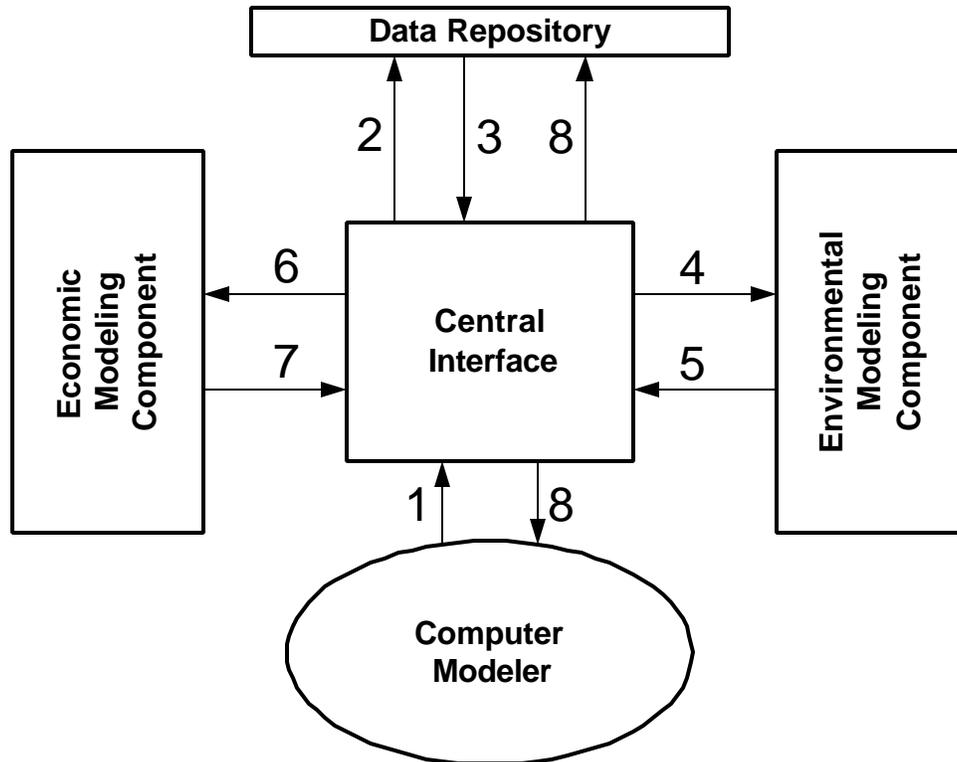
This entails development of several modules and a graphical user interface comprising the central interface.

Central interface components:

- ◆ Graphical user interface: This allows for the choice of simulation options – farm types, region, model-specific options, etc. It also provides for queries of database to view output from simulations.
- ◆ Statistical clustering module: This module clusters farms into groups and generates two sets of representative farms:
  - Environmental model representative farms – accounting for farm-specific and location-specific information (e.g., soil type, weather, etc.)
  - Economic model representative farms – accounting primarily for farm-specific information
- ◆ Environmental input processing module: This module obtains representative farm definitions from the statistical clustering module and transforms data into an I\_APEX database file ready for APEX simulations. This module also initiates batch runs of APEX simulations. The main tasks accomplished in this module are:
  - Generation of the I\_APEX database file including environmental data and farm-specific data as well as the required model parameter values.
  - Automated initiation of APEX simulations.
- ◆ Economic input processing module: This module obtains representative farm definitions from statistical clustering module and transforms the data into FEM input ready for economic model simulations. This module also initiates batch runs of FEM simulations.
- ◆ Environmental output processing module: This module processes I\_APEX output and submits them to the data repository (database).
- ◆ Economic output processing module: This module processes FEM output and submits them to the output data repository (database).

## 2.2.2 Operation of the MMS

The following figure displays the steps involved in a typical application of the MMS. The procedure is briefly described below.



**Figure 2.3** Steps involved in an MMS application

1. Computer modeler (user)
  - a. specifies geographic area to be used for analysis, e.g., entire U.S., corn belt, Southern Plains, Brazos River Basin of Texas, Cross-Timbers region of Texas.
  - b. specifies scenarios to be modeled
  - c. specifies output format/choices for results of the simulations.
- 2 & 3. Central interface receives user input and
  - a. reads in (queries) data from the database that is needed for simulations
  - b. performs statistical analyses to determine modeling units to be used [The basic modeling unit is the farm]. These analyses entail clustering to obtain farm types and representative farms within each farm type classification.
4. Central interface
  - a. generates input files [farm and scenario definitions and other data] for the environmental model(s) [APEX]

- b. calls (batch) execution of environmental model(s)
5. Environmental model(s)
    - a. reads input files and simulates scenarios.
    - b. sends output data to central interface
  6. Central interface
    - a. generates input files [farm and scenario definitions and other data] for the economic model
    - b. calls (batch) execution of economic model
  7. Economic model(s)
    - a. reads input files and simulates scenarios.
    - b. sends output data to central interface
  8. Central interface
    - a. stores relevant output data in data repository in specified format.
    - b. performs statistical analyses of output data and generates inferences; stores results of analyses in data repository, e.g., generates metamodels for indicators and stores these functions in data repository.
    - c. prints (portion of) output data to screen or file based on user's requests.

---

## Chapter 3

# Development of Representative Farms

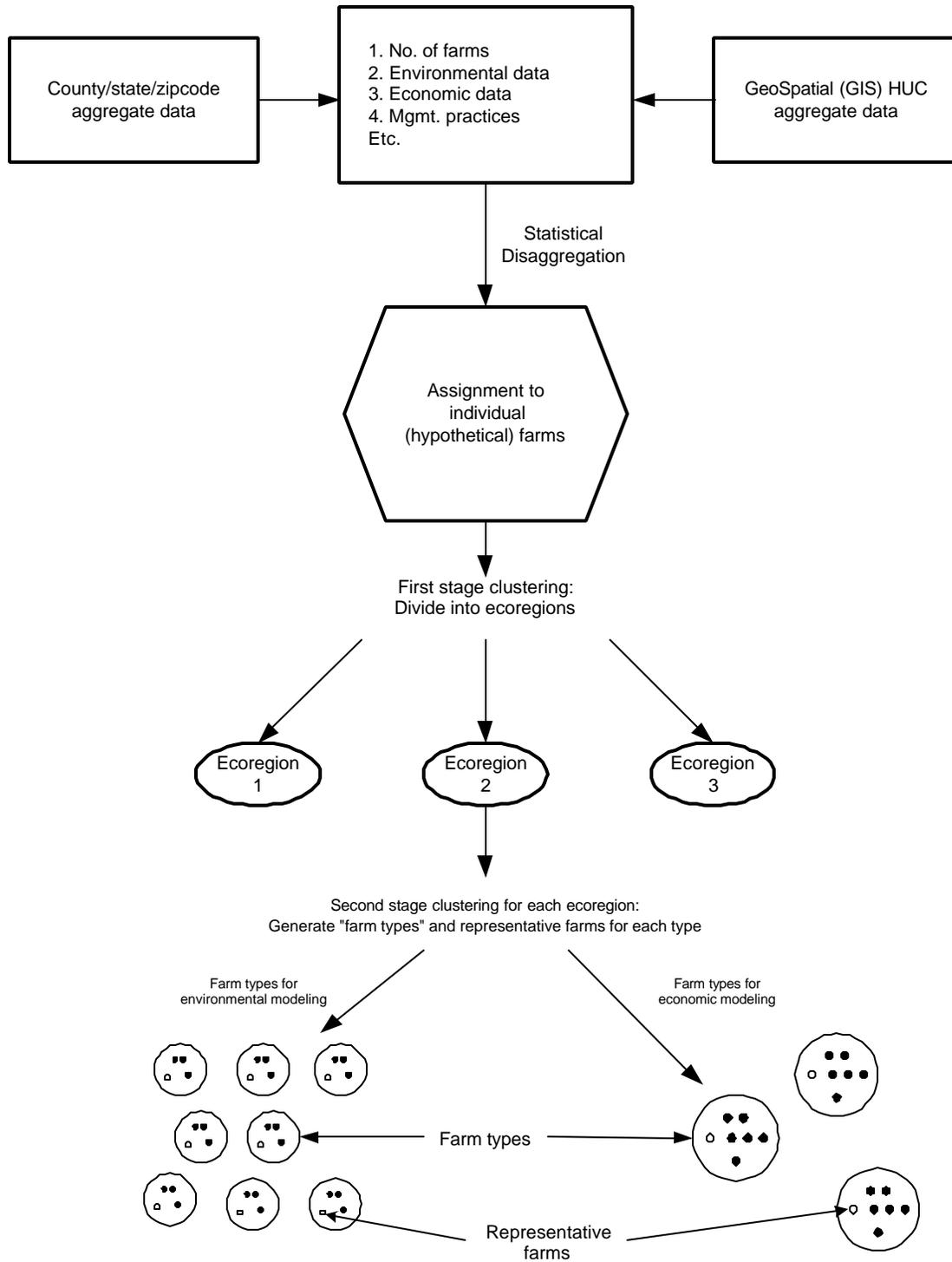
The design of the MMS is based on a micro simulation approach. This means that rather than performing aggregate estimations based on general features of farm types, the MMS simulates individual representative farms, which together are representative of the statistical distribution of farms in the study area. This approach has been employed in many large-scale studies and is particularly useful when it is important to account for wide variations in farm types across regions.

In determining farm types to be modeled, it is not important to know the exact sizes and operations of each individual farm. Due to the similarities among many farm operations, one can envision each farm as belonging to a cluster of operations by virtue of size, management practices, crop types, soil type, weather, and other key factors. Once these clusters have been determined, it is only necessary to simulate a weighted sample of the members of each cluster and then aggregate the results across all clusters within the study area.

Representative farms used in MMS simulations will be derived in a two-step process. First hypothetical farms will be generated for each state and county such that the aggregate distribution of hypothetical farms is approximately equal to that portrayed in the Agricultural Census. This is the data generation phase and is done only once. The data generation phase uses a disaggregation procedure, which is outlined in Section 3.1.

The second phase is a clustering process, which is performed once for each application of the MMS to a study area. Once the study area is chosen, the clustering procedure reads in the list of hypothetical farms generated for that study area, and generates clusters of farm types. Representative farms are members of each of these clusters. Typically, these are selected as the centroids of each cluster unless it is of interest to simulate more than one farm from each cluster. Section 3.2 outlines the clustering procedure.

Figure 3.1 summarizes the general procedure used to derive modeling units for simulations. It is important to note that the representative farms used for environmental model simulations are not necessarily the same as those used for economic model simulations.



**Figure 3.1** Procedure for generating Representative Farms

### 3.1 Disaggregation Procedure

Data disaggregation has been used in a number of disciplines to infer the underlying distribution of a variable based on measured aggregates. For instance, Koutsoyiannis and Onof (2001) disaggregated daily rainfall data into an hourly series using a Poisson cluster model. Temporal disaggregation has also been performed by Lin and Niemeier (1998) on travel demand, and Hodgess and Wei (2000) on U.S. Consumer Price Index and Total Personal Consumption Expenditures. Examples of spatial disaggregation applications include Tamis and Van't Zelfde (1998) on bird survey data and Howitt and Reynaud (2001) on agricultural land uses.

Lack of detailed information on agricultural production systems imposes some limitations on application of scientific tools and models to help solve problems relating to agriculture (Just and Pope, 1999). This lack of detail is especially true of environmental simulations of agricultural operations. As agriculture continues to receive attention in environmental policy and development of TMDLs continues to gain momentum, the need for more detailed and reliable data for model simulations will only increase. Accordingly there has been increased interest and effort towards obtaining more disaggregate data from aggregate measures. A recent example of such efforts is the work of Howitt and Reynaud (2001), who used a bayesian method to obtain spatial disaggregation of agricultural production (land use) data in California.

The need for data disaggregation is also quite crucial to successful implementation of the MMS. To make the most use of existing databases on livestock agriculture a disaggregation procedure was employed. The fundamental premise is that if a set of hypothetical farms is generated, whose combined distribution mimics all the distributions portrayed in the published aggregates, then that set of farms is consistent with the published data. Using the Agricultural Census as the main source of information, a set of hypothetical farm definitions will be generated for every state and county within the United States. While these farms do not represent actual existing operations, the distribution inherent in this set of farm definitions will be developed to mimic the information in the Agricultural Census rather closely. This set of farm definitions can be used for every application of the MMS until new information becomes available when the next (2002) Agricultural Census is published.

The data in the Agricultural Census is presented in tables at the national, state, county and zip-code levels. So far, for the MMS, only state and county data has been used. Each table can be considered a view or profile of the agricultural operations in a particular area. Each view is essentially a description of some features of the operations. For instance, Table 1 in the tabulations for each state in the Agricultural Census contains area summary highlights while cross-tab 49 for each state contains a cross-tabulation of livestock inventories, crop acreage, and other farm characteristics across various categories of farmland area. Since each view provides only a partial assessment of farm structure, a disaggregation procedure is necessitated in order to make use of as much information as possible from the Agricultural Census. The following procedure is being applied to Agricultural Census data to obtain a set of farm definitions consistent with the

published aggregates. Farm definitions will then be augmented with information from other data sources, including the NRI, cropping practices survey, and state budget information. The disaggregation procedure involves five steps:

1. Selection of primary view and level of aggregation
2. Generation of variable distributions
3. Allocation to counties
4. Assignment to farms
5. Augmenting farm definitions

### 3.1.1 Selection of Primary View and Level of Aggregation

The procedure begins with state-level data because it contains the most detailed information about farm structure (characteristics), though at a higher level of aggregation. In particular, the state cross-tab tables show a cross-tabulation of farms within each state by categories of land area, total value of sales, North American Industrial Classification System (NAICS) code, and other variables. This provides very rich information for derivation of the structure of representative farms.

In the disaggregation process, the primary view provides information that will serve as the starting point. Among the state cross-tabs, Table 49 of the Agricultural Census, dealing with tabulations across land area in farms was chosen as the primary view, since farmland area was deemed more important in describing farm structure than the other cross-tab categories. In addition to the state totals, there are 12 acreage categories in Table 49 of the Agricultural Census, as listed in Table 3.1.

**Table 3.1** Categories of land in farms and corresponding bounds in acres

Category	Lower Bound	Upper Bound
1	1	9
2	10	49
3	50	69
4	70	99
5	100	139
6	140	179
7	180	219
8	220	259
9	260	499
10	500	999
11	1000	1999
12	2000	—

### 3.1.2 Generation of Variable Distributions

The primary view [state cross-tabulation of farms by land in farms] and other cross-tabs contain data on over 500 items for each cross-tab category. These items are variables or subsets of variables such as land in farms, total value of sales, sales of specific crops, areas in various land uses and crops, livestock inventories and sales, and many other variables.

For most variables, data are given on the number of farms,  $N_x$ , with (positive) values for that variable, and the total value,  $T_x$ , that corresponds to the  $N_x$  farms for that variable. For instance, a pair of values  $N_x, T_x$  for  $x$ =corn acreage in the category of 10 to 49 acres, means there are  $N_x$  farms with land area between 10 and 49 acres, that have corn acreage in that state, and that the total corn acreage corresponding to these farms is  $T_x$ .

Given this pair of values, a set of  $N_x$  positive numbers are generated using a random number generator (RNG) such that  $\left| \sum_{i=1}^{N_x} a_i - T_x \right| \leq \frac{cT_x}{100}$  where  $c$  is a pre-defined tolerance value (usually less than 0.1),  $a_i$  is the  $i$ th corn acreage generated and  $a_i \in [a_{MIN}, a_{MAX}]$ ,  $a_{MIN}$  and  $a_{MAX}$  being prespecified or derived minimum and maximum bounds on corn acreage for that "land-in-farms" category. Minimum and maximum bounds, if not already indicated in the primary view, were derived from other census tables.

To ensure maximum performance of the disaggregation algorithms, a beta distribution will be used for the RNG process and the parameters for the distribution will be estimated from  $N_x$  and  $T_x$ . The beta distribution is highly flexible, with the specific form determined by the parameters. A number of additional constraints will be imposed on the generation process based on the nature of the variable and its relationships with other variables. To this effect, a highly structured array of variable definitions will be constructed and populated with the census aggregates from the primary view prior to generation of the distributions. The numbers thus generated are merely the initial seed for the disaggregation process.

The set of numbers generated for each variable will next be extensively revised using information from all the other state tables (views) pertaining to that variable. For this purpose, the structured array of census variables will be augmented in the algorithm and populated with pertinent data from all the other state views. The revision algorithm then proceeds in an iterative manner until the generated distribution is within a small pre-specified tolerance of all the census distributions for that variable.

### 3.1.3 Allocation to Counties

The next step of the disaggregation process is allocation of the variable distributions to the counties. The distributions will be allocated to the counties using search algorithms that meet all the constraints imposed by the county distributions and the already satisfied state distributions.

The county data essentially serves as an additional layer of information for the corresponding variables. Search algorithms allocate variable distributions to counties in such a way that census county distributions are preserved in the generated data and already satisfied state distributions are also maintained, all within a prespecified tolerance. Given the additional layer of information in county-level data, the search algorithms used for the allocation step will also revise the generated distributions further until all constraints are met.

Due to the complexity of the distributions portrayed in the Agricultural Census, the allocation step is necessary in order to reduce the size of the data to county-sized groups to achieve more effective assignment to farms.

### 3.1.4 Assignment to Farms

County-level generated data are mapped to individual farms to complete the disaggregation process. Farms will be initialized with specific land area values. The other variables will then be mapped to each farm using an assignment algorithm. The objective of this algorithm is to generate hypothetical farm data that reproduces the published Agricultural census aggregates within limits of a prespecified tolerance. This step also entails possible revision of the generated data until the resulting set of farm data is within tolerance of the Census aggregates.

### 3.1.5 Augmenting Farm Definitions

The Agricultural Census contains a wealth of information on farm structure, but is not sufficient for the purposes of environmental policy simulations such as required within the MMS. In order to provide all the data needed for simulations, information from other sources will be added to the set of farm definitions obtained from the disaggregation process. Major sources included the NRI (cropping system and practice information), state enterprise budgets (information on field operations), cropping practices surveys (field operations data), American Society of Agricultural Engineers (ASAE) Standards (manure production characteristics), United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) crop production database (crop yields), and USDA Natural Resources Conservation Service (NRCS) Agricultural Resource

Management Service (ARMS) data (farm management and manure handling practices). In addition, data on weather and other exogenous variables will be obtained from other sources and assigned to respective counties.

## 3.2 Clustering Procedure

The disaggregation procedure was developed to provide a set of farm definitions that can be used for various studies involving economic and environmental model simulation of agricultural operations. While much care is being taken to obtain a set of farm definitions consistent with the Agricultural Census and other published aggregates, the intent is not to exactly mimic real world operations. The *distributions* of the farms are intended to mimic real world farm *distributions* at the county and state levels. Thus studies conducted using these data at the multicounty-level and beyond, or at the scale of large watersheds, would not encounter problems with the statistical integrity of the data.

One of the most appropriate methods of using data for a given study area is to cluster the farms into groups. While one would not infer the characteristics of any specific farm from the generated data, correct general inferences can be made about groups (clusters) of farm types as long as these groups are large enough.

Cluster analysis is a multivariate data reduction technique used for solving classification problems. It is used to sort cases into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. The method has been successfully applied in various disciplines including ecology. For example, in the concept of “ecoregions” a region is subdivided into subregions based upon similarities of important measures of ecological conditions (Hargrove and Luxmoore, 1997). Such subdivision into similar regions has proved to be important to understanding and managing large geographic areas. Clustering particularly facilitates environmental and economic modeling and simulation of large areas since only a few simulations of representative entities, e.g., farms, can be made and the results applied to a subregion as identified by cluster analysis (Kobrich et al., In Press).

Cluster development is a multistep process; three major steps can be identified (Panel on Discriminant Analysis, Classification and Clustering, 1989): i) data preparation, ii) estimating clusters with the appropriate method, and iii) cluster output.

### 3.2.1 Data Preparation

The data preparation step involves the choice, transformation, and scaling of variables. The question of what variable(s) to use in clustering is a very important issue that involves statistical and expert knowledge of the subject matter. This is a crucial step because the result from cluster analysis is data dependent; that is, outcome from the analysis will depend upon the selection of variables. Transformation and scaling of variables is desirable when statistical assumptions of cluster analysis (e.g., normality) are violated, and the differences in scales of measurement lead to implicit weighting of variables.

There are few statistical procedures to assist in variable reduction and selection for clustering (Kobrich et al., In Press). One commonly used method and the one used for MMS is the principal component analysis (PCA). PCA aims at data reduction through linear combinations of the original variables. PCA removes correlations among the input variables, standardizes the mean and variance, and reduces the dimensionality of the data set. That is, in matrix terms, PCA decomposes a data matrix  $\mathbf{X}$  of rank  $h$ , as a sum of matrices of rank one. The rank indicates the number of linearly independent vectors of a matrix. The new rank one matrices are vector products of the score vectors,  $\mathbf{t}$ , and loading vectors,  $\mathbf{P}$ , as shown in Equation 3.1.

$$\mathbf{X} = \mathbf{t}_1 \mathbf{P}'_1 + \mathbf{t}_2 \mathbf{P}'_2 + \dots + \mathbf{t}_h \mathbf{P}'_h \quad (3.1)$$

These vectors can be calculated by a least squares procedure. The new coordinates of the system, the principal components (PCs), are mutually orthogonal, not correlated, and successively explain the maximum of the residual variation (Jobson, 1991). Usually, only the first few PCs account for the greatest amount of total data variance and can be utilized to represent the whole data set in a simpler manner.

PCA is beneficial in eliminating redundant variance, where the redundant variance represents the amount of irrelevant information present in the data due to correlation between variables. Further, the PCA is informative in its own right because it identifies the basic variables (dimensions) and their underlying factors that distinguish clusters.

### 3.2.2 Estimating Clusters with the Appropriate Method

Clustering methods can be broadly divided into two basic types: hierarchical and partitional clustering. Within each of the types there exists a wealth of subtypes and different algorithms for finding the clusters (SAS Institute, 1999).

Hierarchical clustering proceeds successively by either merging smaller clusters into larger ones or by splitting larger clusters. Hierarchical clustering methods differ in the rule by which it is decided which two small clusters are merged or which large cluster is

split. The end result of the algorithm is a tree of clusters called a dendrogram, which shows how the clusters are related. By cutting the dendrogram at a desired level a clustering of the data items into disjoint groups is obtained.

Partitional clustering, on the other hand, attempts to directly decompose the data set into a set of disjoint clusters. The criterion function that the clustering algorithm tries to minimize may emphasize the local structure of the data, as by assigning clusters to peaks in the probability density function, or the global structure. Typically global criteria involve minimizing some measure of dissimilarity in the samples within each cluster, while maximizing the dissimilarity between different clusters.

A commonly used partition clustering method is K-means clustering. In K-means clustering the criterion function is the average squared distance of the data items  $\mathbf{X}_k$  from their nearest cluster centroids,  $\mathbf{m}_{c(\mathbf{X}_k)}$ .

$$\mathbf{E}_k = \sum_k \left\| \mathbf{X}_k - \mathbf{m}_{c(\mathbf{X}_k)} \right\|^2 \quad (3.2)$$

where  $c(\mathbf{X}_k)$  is the index of the centroid that is closest to  $\mathbf{X}_k$ . One possible algorithm for minimizing the cost function  $\mathbf{E}_k$  begins by initializing a set of K cluster centroids denoted by  $\mathbf{m}_i$ ,  $i = 1, 2, 3, \dots, K$ . The positions of the  $\mathbf{m}_i$ 's are then adjusted iteratively by first assigning the data samples to the nearest clusters and then recomputing the centroids. The iteration is stopped when  $\mathbf{E}$  does not change markedly any more. In an alternative algorithm each randomly chosen sample is considered in succession, and the nearest centroid is updated. For this work, we employed the K-means clustering approach using the FASTCLUS<sup>®</sup> procedure of SAS (SAS Institute, 1999). The same clustering procedure is applied for the two stages of clustering illustrated in Figure 3.1.

### 3.2.3 Interpreting Cluster Output

The ability to identify the appropriate number of clusters is the most fundamental shortcoming of any clustering technique (SAS Institute, 1999). While this is an active area of research, there are numerous methods outlined in statistics literature detailing potential methods for detecting the appropriate number of clusters (Jobson, 1991; SAS Institute, 1999).

One of the most effective procedures for determining the number of clusters is the cubic clustering criterion (CCC) (SAS Institute, 1983; 1999). Graphical plots of CCC with respect to the number of clusters are commonly used to determine appropriate cluster groupings. Peaks greater than two or three on such plots suggest good clustering. Peaks between zero and two suggest potential clusters, but must be interpreted with caution (SAS Institute, 1999). Details on the computation of the CCC can be found in SAS Institute (1983). Another statistic that can be used in conjunction with the CCC criterion is the pseudo F (PSF). Large PSF values suggest a good stopping point. These criteria are appropriate for clusters that are roughly multivariate normal.

### 3.2.4 Application of Statistical Clustering in CEEOT-MMS

A two-stage clustering procedure using the methodology outlined above is employed in MMS applications (Figure 3.1). The entire simulation region is divided into ecoregions (clusters) in the first stage. Next, within each ecoregion, relevant farm types (sub-clusters) are determined and representative farms to be simulated in economic and environmental models are chosen for each farm type. Chapter 5 provides further rationale and information about the two-stage clustering process.

# Chapter 4

## National Database

The national database being constructed as part of the macro modeling system, contains data from various sources. With the exception of GIS data layers, all the data sets are being stored in MS Access® database format. Section 4.1 outlines the data types used within MMS and the sources for each of these data items. Section 4.2 presents some special databases being constructed as part of the data generation phase of MMS.

### 4.1 Data Types and Sources

The following table contains broad categories of data used within MMS.

**Table 4.1** Data types and sources used in MMS

Data Type	Model(s)	Sources
Cropping systems and operations	APEX, FEM	
Cropping systems	APEX, FEM	Ag Census, NRI
Cropping operations	APEX, FEM	CPS, Extension, NRI
Operations performed by producer	FEM	Extension
Equipment used	APEX, FEM	CPS, Ag Census
Supplies used (other than fertilizer or chemicals)	APEX, FEM	CPS
Discretionary fertilizer use	APEX, FEM	CPS, AAPFC data
Other chemical use	APEX, FEM	CPS
Operations custom performed	FEM	Extension, NRI
Custom rates	FEM	Extension
Crop (feed) yields	APEX, FEM	CPS, CPD, NASS
Agronomic requirements	APEX, FEM	Extension
Relationship between agronomic rates and crop yields	APEX, FEM	NRCS
Feed types and nutrient contents	FEM	
Feed types	FEM	Extension, NRCS
Feed nutrient contents	FEM	NRC
Feed prices	FEM	
Prices of purchased feed	FEM	Feedstuffs
Prices of feed sold	FEM	Feedstuffs
Rules/behavior governing use of raised feed	FEM	NRCS, Extension
Rules/behavior governing feed purchase	FEM	NRCS, Extension

**Table 4.1 Continued**

Livestock systems and nutrient requirements		
Livestock inventory	APEX, FEM	Ag Census
Livestock types	APEX, FEM	Ag Census
Characteristics of livestock types	APEX, FEM	NRCS
Livestock sales	FEM	NRCS, Ag Census
Livestock purchases	FEM	NRCS, Ag Census
Death losses	FEM	NRCS
Raised livestock	FEM	NRCS
Rules/behavior governing livestock inventory	FEM	NRCS
Livestock operations	APEX, FEM	NRCS
Rules/behavior governing livestock handling/operations	FEM	NRCS
Livestock nutrition	FEM	
Nutrient requirements	FEM	NRC
Bounds/restrictions on nutrient requirements	FEM	NRC
Bounds/restrictions on feed composition of rations	FEM	NRC
Manure nutrients and management	APEX, FEM	
Manure production characteristics	APEX, FEM	ASAE Standards
Manure types: solid, liquid, direct deposition	APEX, FEM	NRCS, Extension
Manure proportions for each type	APEX, FEM	NRCS, Extension
Characteristics of manure in each type	APEX, FEM	ASAE Standards
Manure management options	APEX, FEM	NRCS
Characteristics/impacts of manure management options	APEX, FEM	NRCS
Manure nutrient losses/changes during storage and handling	APEX, FEM	NRCS
Manure nutrient losses after land application	APEX, FEM	NRCS
Plant availability of manure nutrients	APEX, FEM	NRCS
Structures and facilities	APEX, FEM	NRCS
Buildings	FEM	NRCS
Livestock housing	FEM	NRCS
Characteristics of livestock housing	FEM	NRCS
Equipment housing	FEM	NRCS
Characteristics of equipment housing	FEM	NRCS
Commodity/grain/feed storage	FEM	NRCS
Characteristics of commodity storage	FEM	NRCS
Earthen structures	APEX, FEM	NRCS
Characteristics of earthen structures	APEX, FEM	NRCS
Other structures	APEX, FEM	NRCS
Characteristics of other structures	APEX, FEM	NRCS
Equipment inventory and characteristics	FEM	
Equipment inventory	FEM	Ag Census, NRCS
Equipment prices	FEM	Farmland, Extension
Equipment characteristics	FEM	ASAE Standards
Land uses and characteristics	APEX, FEM	
Land area owned	FEM	Ag Census
Land area leased	FEM	Ag Census

**Table 4.1 Continued**

Land types	APEX, FEM	Ag Census
Pasture	APEX, FEM	Ag Census
Cropland	APEX, FEM	Ag Census
Hay fields	APEX, FEM	Ag Census
Woodland	APEX, FEM	Ag Census
Buildings/Homestead/Barren	APEX, FEM	Ag Census, NRCS, Extension
Land use types	APEX, FEM	NRCS
Manure disposal: liquid and solid manure	APEX, FEM	NRCS
Land receiving no manure/waste	APEX, FEM	NRCS
Land used for grazing	APEX, FEM	NRCS, NRI, Ag Census
Acreage for each land use	APEX, FEM	Ag Census, NRCS, NRI
Rules/behavior governing land allocation to various uses	APEX, FEM	NRCS
Agronomic/manure rates governing land uses	APEX, FEM	NRCS
Cropping systems on each land use	APEX, FEM	Ag Census, NRI
Rules/behavior governing land allocation/use for cropping systems	APEX, FEM	NRCS
Other exogenous parameters	APEX, FEM	
Market conditions	FEM	
Government policy	APEX, FEM	
GIS maps	APEX, FEM	
Water use (consumption)	APEX	NRCS
Land use/cover	APEX	USGS
Soils	APEX, FEM	STATSGO, SSURGO
Digital elevation model (DEM)	APEX, FEM	USGS
Producer location	APEX	
Boundaries data (Streams, roads, etc.)	APEX	TIGER data
Weather information	APEX	
Rainfall	APEX	BASINS, NOAA
Temperature	APEX	BASINS, NOAA
Other	APEX	
Data for calibration/validation	APEX, FEM	
Soil loss and nutrient levels	APEX	NRI
Water quality monitoring data	APEX	
Water flow	APEX	
Sediment	APEX	
Nutrient concentrations	APEX	
Other water quality characteristics	APEX	
Farm budgets	FEM, APEX	NRCS, Extension

## Table 4.1 Continued

Notes on Data Sources:

CPS: USDA Cropping Practices Survey

CPD: USDA Crop Production Data, USDA-NASS

State EPA: State or Federal EPA offices

Watershed: Watershed sources, including producer surveys and information from area experts.

BASINS: USEPA BASINS application data

Extension: State extension offices.

ASAE: American Society of Agricultural Engineers.

NOAA: National Oceanic and Atmospheric Association.

USGS: United States Geological Survey

Model defaults: Default specifications in the model

Literature: Information from various academic publications on the subject of interest.

County Assessor: County Tax Assessor

NRC: National Research Council publications, e.g., NRC Tables on livestock nutrient requirements and feed compositions.

NASS: USDA-NASS: State or federal USDA National Agricultural Statistics Service.

NRCS: USDA-NRCS: Federal, state, or local USDA Natural Resources Conservation Service offices.

APEX Weather Generator: Weather generator routine in the APEX model.

TIGER: Topologically Integrated Geographic Encoding and Referencing system.

NRI: National Resources Inventory

STATSGO: State Soil Geographic Database: USDA NRCS

SSURGO: Soil Survey Geographic Data Base

AAPFC = Association of American Plant Food Control Officials ([www.aapfco.org](http://www.aapfco.org))

TFI = The Fertilizer Institute ([www.tfi.org](http://www.tfi.org))

Farmland: Hotline Farm Equipment Guide

Ag Census: United States Agricultural Census, USDA-NRCS.

## 4.2 Special Constructed Databases

As part of the data generation phase of the MMS, several databases are being developed from existing sources. Some of these are of particular importance in that they have wide applications beyond the scope of the MMS. This section outlines some of these databases that can be of importance elsewhere.

### 4.2.1 Farm Definitions Database

The farm definitions database will contain a set of farm definitions for every state and county within the United States. This is a product of the disaggregation procedure employed within the data generation phase of the MMS. For each farm this database will contain information on many of the variables represented in the Agricultural Census as well as other information pertinent to farm-level economic and environmental simulations. The set of farm definitions is being constructed in such a manner as to preserve the Agricultural Census distributions for farms in any given county or state.

## 4.2.2 Crop Operations Library

The crop operations library is being constructed to provide data on field operations performed on farms in any given state or county. These data are invaluable for environmental and economic model simulations and form the backbone of baseline scenario specifications. Data in the crop operations library are needed for any environmental simulation of agricultural operations. This library contains the field operations typically used by crop and livestock operations in various states and regions or counties within specified states. The structure of the crop operations library is shown in the following table.

**Table 4.2** Columns in the crop operations library

Column	Description	Data Type
Operation	Name of the operation, e.g. fertilizer spreading	Character
APEX operation code	If applicable, APEX numeric code corresponding to the operation	Numeric
Crop	Crop Name	Character
APEX crop code	APEX numeric code corresponding to the crop	Numeric
Crop definition	Specific Crop Definition, if any (e.g. high yield corn)	Character
Year in rotation	Year of this crop in cropping system rotation sequence	Numeric
Cropping system	Name of the cropping system or rotation	Character
Definition	Specific type of the cropping system, e.g., dryland sorghum forage	Character
Cropping system short	Unique identifier for the cropping system and definition	Character
Cropping system type	Code indicating whether operation description for cropping system is preceded by separate establishment operations	Character
CS establishment	Establishment operations to perform as indicated by type	Character
Rotation length	Length (in years) of cropping system sequence	Numeric
Applies to	Type(s) of crop fields to which this specific operation applies	Character
State	Name of the State	Character
Region	Name of the region/county or counties within the State where the operation is applicable	Char/Num
Year	Date of the operation: year	Numeric
Month	Date of the operation: month	Char/Num
Day	Date of the operation: day	Numeric
Frequency	Number of times this operation is performed within the production year	Numeric
Interval	Specific dates or time interval in which to perform multiple frequency operations	Char/Num
Proportion	Fraction of field on which this operation is performed	Numeric
Item1	First item used in or used to characterize this operation, e.g., 100 HP tractor, crop heat units, nitrogen fertilizer	Character

**Table 4.2 Continued**

Value1	Value of the first item	Char/Num
Unit1	Unit of the first item	Character
Item2	Second item used in or used to characterize this operation	Character
Value2	Value of the second item	Char/Num
Unit2	Unit of the second item	Character
Etc.	...	

---

# Chapter 5

## Indicators and Prominent Factors

Indicators are measures of the state of the subject of analysis. Environmental indicators are measures of environmental quality, e.g., sediment loss. Economic indicators are measures of economic performance, e.g., producer profitability or financial performance. Factors are the instruments that cause changes in the levels of the indicators. The indicators chosen for consideration determine the factors that will need to be included in a study. The following sections discuss the indicators and factors being considered in the MMS and their role in the development of the modeling system.

### 5.1 Environmental and Economic Indicators

Environmental and economic impacts of practices and policies relating to agriculture are presented as quantitative measures of relevant indicators. Bouzaher et al. (1993b) lists the key indicators considered in nonpoint source policy assessment. The choice of indicators depends on what is perceived to be socially relevant for the region. The key environmental and economic indicators considered within the CEEOT-MMS modeling framework include the following.

#### Environmental indicators:

- Nutrient (nitrogen, phosphorus) losses from agricultural land to surface water.
- Nutrient (nitrogen, phosphorus) losses from agricultural land to ground water.
- Sediment losses from agricultural land.

Most notably, due to an emphasis on livestock agriculture, the existing framework will not consider the role of pesticides in water quality, though these are also of prime importance in many areas. The key economic indicators are as follows.

#### Economic indicators

- Farm profitability
- County/regional/ecoregion economic performance
- Income distribution or enterprise relocation

## 5.2 Prominent Factors Impacting Indicators

A review of watershed and regional studies shows that numerous factors have varying degrees of influence on the levels of relevant indicators. In order to perform reasonable nationwide assessments, the list of factors needs to be reduced to manageable numbers. A priori determination of prominent factors is possible from a review of case studies in different regions of the country. Due to the nature of the problems addressed, most of the factors influencing economic indicators are subsumed under environmentally impacting factors.

All prominent factors can be categorized as *choice-based* factors or *location-specific* factors. Choice-based factors are those that can be altered by individual or collective choice. These include various managerial or structural practices, industry location and distribution, as well as government policy. Location-specific factors on the other hand, are typically impossible or impracticable to change. These include most biophysical parameters such as soil type, weather, and topography. Based on predominance in environmental and economic impact analyses, the following are the key factors considered for the present framework.

### **Prominent factors influencing environmental and economic indicators**

#### Choice-based factors

- ◆ Type of operation
  - Cropland acreage and yields
  - Size and types of livestock operations
  - Management practices
- ◆ Government policy
  - Policy scenarios

#### Location-specific factors

- ◆ Soil type
- ◆ Topography
- ◆ Weather
  - Rainfall
  - Temperature

The sole determinant of the levels of environmental and economic indicators is the spatial distribution of all relevant factors. More precisely, all the complex relations inherent in the levels of relevant indicators can be captured completely by the spatial distribution of the factors impacting those indicators. Data on the levels of prominent factors form part of a national database being developed to help address nonpoint source pollution issues related to agriculture.

The complex interactions among environmentally impacting factors must be addressed within the modeling framework in order to arrive at a reasonable assessment. In order to enhance the statistical validity of the simulations, the study region can be subdivided into areas of relative ecological uniformity using location-specific factors. Farm groups and

representative farms are then defined within each subregion, using both location-specific and choice-based factors.

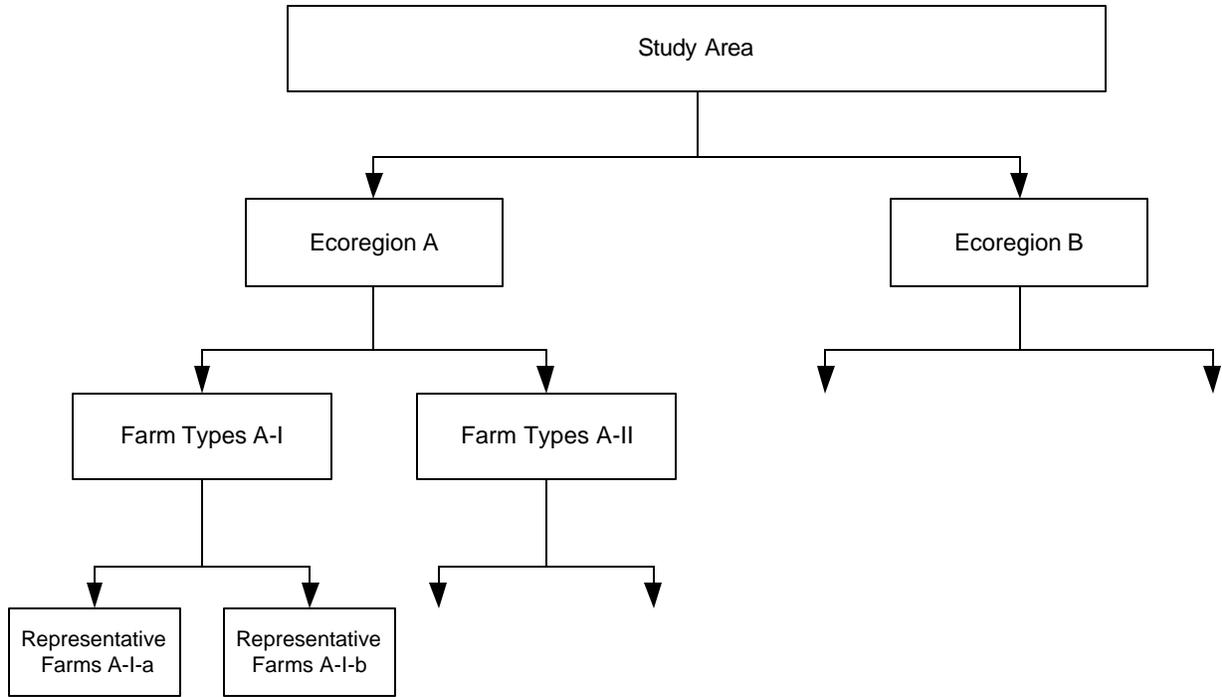
### 5.3 Ecologically Uniform Regions (Ecoregions)

Generally, the objective of national policy assessment is twofold. First, it is desired to obtain an estimate of the aggregate impacts of suggested policies or practices on environmental and economic indicators on a national scale. Secondly, it is also of interest to know how these impacts are distributed across the nation. To both estimate reliable aggregate impacts and provide inferences on distribution issues, the entire nation or study region will be divided into regions of relative ecological uniformity. These ecologically uniform regions or custom “ecoregions” will serve to enhance the validity of analyses by permitting us to capture the significant differences that exist across the nation without having to deal with each small watershed or local area individually.

Due to the complexities surrounding the distribution of ecologically impacting factors, the task of grouping a geographical area into ecoregions is not trivial. A clustering procedure will be used in MMS for categorizing all modeling units into ecologically “homogeneous” groups. The objective of clustering is to group individual items into groups such that members within the same group are more similar than members from different groups. Each resulting group contains a collection of watersheds or counties that share similar proportions of prominent factors. The number of custom ecoregions to be considered can be specified a priori or determined as the number that results in the statistically most distinct groupings of watersheds. Statistical clustering for custom ecoregions is based primarily on location-specific factors.

### 5.4 Farm Types

A second clustering procedure will identify farm types within each ecoregion, using choice-based factors as well. This procedure allows us to determine farm types and sizes based on existing data rather than subjective opinions. Figure 5.1 shows the hierarchical relationship between modeling units using this two-stage clustering procedure on the prominent factors.



**Figure 5.1** Hierarchical representation of modeling units

---

# Chapter 6

## Applications of CEEOT-MMS

CEEOT-MMS has a wide range of applications. In this chapter, we discuss four main areas of its use. Chapter 7 provides a number of additional applications that are being considered for future extensions of the modeling system.

### 6.1 Static Scenarios

Scenarios simulated in watershed studies have largely been of a static nature, though they do have some dynamic implications<sup>1</sup>. These scenarios can also be simulated within MMS for larger regions or the entire nation. Examples of these scenarios include the following taken from Chapter 5 of Osei et al., 2000 (Table 6.1). In past studies, static scenarios have also included combinations of the following practices. The letters beside scenario category headings and numeric codes in parentheses were used to represent scenario codes. For instance, scenarios dealing with changes in manure application rates fall under the M100 category.

**Table 6.1** Categories of static scenarios with illustrative examples

Manure Handling (On-farm) H
Collection [100]
Flushing
Scraping
Storage [200]
Manure pits
Waste storage ponds
Anaerobic lagoons
Stockpiling on open lots or fields
Processing [300]
On-farm composting
Liquid-solids separation
Manure application/deposition on-farm M
Application rate [100]
N rate
High P rate (Inorganic P)

---

<sup>1</sup> They are static in the sense that their specifications on the farms do not change over time. However, they have dynamic implications because the same management practices imposed by the scenario year after year cause dynamic changes in nutrient losses and other indicators. Dynamic scenarios are mentioned in Chapter 7.

**Table 6.1 Continued**

Low P rate (Total P)
Timing of applications [200]
Alternative timing of manure application
Mode of application [300]
Surface application of manure
Solid manure incorporation
Liquid manure injection
Pasture stocking density [400]
N-based
High P-based (Inorganic P)
Low P-based (Total P)
Manure disposal off-farm      D
Haul off [100]
Haul off manure/compost to other livestock farm in watershed
Haul off manure/compost to crop farm in watershed
Haul manure/compost outside watershed
Composting [200]
Haul off manure to composting facility
Manure trading [300]
Export of surplus manure
Fertilizer application      F
Rate [100]
Adjustments in fertilizer application rates: lower starter rates, etc.
Timing [200]
Changes in timing of fertilizer applications
Mode [300]
Surface application
Incorporation
Injection
Cropland tillage      T
Reduced till
No-till
Conventional tillage
Soil management      S
Terraces
Ration modifications      R
Phytase supplemented rations
Reduced P feed rations
Structural BMPs      B
Runoff control basins

**Table 6.1 Continued**

Filter strips	
Fencing	
Waterways development	
Lagoons: one-stage, two-stage	
Waste storage ponds	
Changes in livestock production systems	L
Livestock housing (and husbandry) [100]	
Hoop structures for swine operations	
Freestalls for dairies	
Open lots for dairies	
Grazing management [200]	
Intensive rotational grazing	
Seasonal grazing	
Grassed loafing lots	
Cropping systems and sequence	C
Rotations [100]	
Alternative crop rotations	
Spatial sequence [200]	
Strip-cropping	
Contouring	
Contour buffer strips	
Vegetative cover [300]	
Cover crops	
Permanent vegetative cover	
Other exogenous factors	X
Fiscal policy [100]	
Government price policy	
Taxes on feed prices	
Price subsidies/cost shares for approved BMPs	

## 6.2 Spatial Comparative Analysis

### 6.2.1 Spatial Impacts

The micro simulation approach used within CEEOT-MMS enables us to address a wide array of spatial issues. Simulation results can be used to infer relative impacts of a given scenario on different watersheds or regions, as well as spatial correlations with farm types, sizes or other factors.

## 6.2.2 Vulnerability Indices

CEEOT-MMS can be used to estimate the degree of vulnerability of various regions to environmental degradation. In this vein, two types of vulnerability may be assessed.

*Unconditional or absolute vulnerability:* This is the degree of vulnerability due largely to “unchangeable” or location-specific factors such as soil type or weather. A region is said to be absolutely vulnerable for a given indicator if it exhibits a high probability of environmental degradation under practicable levels of choice-based factors. In such cases only extreme levels of choice-based factors may result in acceptable levels of the indicator.

*Conditional or relative vulnerability:* Some regions become vulnerable only under certain conditions. We refer to this as conditional or relative vulnerability. It is based on location-specific as well as “changeable” or choice-based factors such as alternative scenarios or industry location (land use). A region is conditionally vulnerable for a given indicator if under practicable levels of choice-based factors, the probability of vulnerability is within the acceptable range.

A vulnerability index (VI) is defined for each indicator as the probability that the level of that indicator in the region of interest will exceed a predetermined threshold. The VI can be estimated as the proportion of simulation output resulting in higher levels of the indicator than the predetermined threshold. In order to obtain valid estimates of vulnerability, a broad range of factors must be used in the simulations.

Whereas the above discussion focuses on vulnerability of regions to environmental degradation, the concept can be applied to management practices or policies as well. It can also be used to assess vulnerability of regions or practices to adverse economic impacts.

## 6.3 Use of Agro-Ecological Metamodels

Metamodels are statistical functions that summarize the relationships between computer model inputs and outputs so that general inferences can be made without recourse to additional simulations. Metamodels have been used recently in regional studies to estimate the impacts of various policies on pesticide concentrations in surface and ground water (e.g., Bouzaher et al., 1993a) and nutrient and pesticide losses from cropland (Babcock et al., 1997). The use of metamodels in agroecological research is outlined in Bouzaher et al. (1993a). These statistical tools will also be employed in the MMS framework in the following manner.

Computer simulations on multiple representative farms in each ecoregion will generate model output based on input values representing the levels of prominent factors prevailing in those ecoregions. Standard regression procedures (e.g., Greene, 1995) can be used to estimate the relation

$$y_i = f^i(x_i, \theta_i) + u_i \quad (6.1)$$

for each ecoregion  $i$ , where  $y_i$  is an  $n \times k$  matrix of model output on the  $k$  indicators based on  $n$  simulations within the ecoregion,  $x_i$  is an  $n \times p$  matrix of specifications (values) for the  $p$  factors, and  $\theta_i$  is a  $k \times p$  matrix of parameter values that describe the relationship between the factors and the indicators. Depending on the underlying relations, the functional form  $f^i$ , could be linear or nonlinear and may include dummy variable representations, while the error term  $u_i$  either conforms to conventional regression assumptions or appropriate adjustments are made in the metamodel.

Estimation of the metamodel for each ecoregion yields the statistical relation

$$\hat{y}_i = f^i(x_i, \hat{\theta}_i) \quad (6.2)$$

for each ecoregion  $i$ . Given the estimated metamodel for each ecoregion, the predicted impact of the factors on the relevant indicators in farm type  $j$  in ecoregion  $i$  is given as

$$\hat{y}_j = f^i(x_j, \hat{\theta}_i), j \in i \quad (6.3)$$

The impacts on all farm types in the country can thus be approximated by the ecoregion-specific metamodels. National and regional impacts are obtained by aggregating across all relevant farm types and across all relevant ecoregions.

## 6.4 Risk Analysis

Automation of simulation procedures within MMS also enables us to perform risk analyses, which indicate a confidence interval in addition to the estimated impact of a scenario. By using historical or estimated distributions of various parameters, environmental and economic impact assessment can reflect the degree of uncertainty associated with the analysis. CEEOT-MMS can be used to perform risk analysis on the impacts of user-constructed scenarios.



---

# Chapter 7

## Extensions of the MMS

The following sections outline possible extensions to CEEOT-MMS, which will enable other kinds of applications, in addition to those discussed in Chapter 6. Additional applications include advanced scenarios, development of automated calibration routines, and routing of environmental indicators.

### 7.1 Advanced Scenario Options (Dynamics)

The key feature of the advanced scenario options being considered within MMS is dynamics. By allowing for scenario dynamics, one can investigate the impacts of scenarios, when the specifications of the scenarios themselves are contingent upon future producer behavior or other circumstances, which may be largely unforeseen. A typical example of this arises in situations where manure applications can be applied on land at the nitrogen uptake rate until such time that soil phosphorus levels exceed a given threshold, at which time applications would then be based on the phosphorus uptake rate. This extension of MMS is particularly useful for assessing the impacts of comprehensive nutrient management plans (CNMPs).

### 7.2 Automated Calibration Routines

Automated calibration routines will help ensure faster and more reliable assessments by providing built-in statistical routines that help modelers to select the optimal (input) parameter space for simulations.

### 7.3 Routing of Environmental Indicators

Current design of MMS uses edge-of-field modeling in APEX to estimate impacts of scenarios on environmental indicators. More extensive environmental assessment can be achieved by incorporating a model such as the Soil and Water Assessment Tool (SWAT), that routes edge-of-field indicators to a watershed outlet.



---

# Chapter 8

## Conclusions

Current trends in environmental and agricultural policies indicate a greater need for more readily available information on costs and environmental effectiveness of alternative agricultural practices. However, public and scientific concerns about use of computer model results in policy formation calls for improved reliability of model predictions. To address these legitimate concerns requires more detailed information than has been available to researchers in the past.

To help meet the growing need for more reliable and timely scientific analysis, a macro modeling system is being developed to help address issues relating to agriculture, particularly livestock farms, and the environment. CEEOT-MMS consists of two previously used simulation models (APEX and FEM), a data repository based on existing databases, and a central interface that provides the links between the other components. Substantial work accomplished as part of the data generation phase of MMS development enables use of the data for a much wider spectrum of applications than envisaged for the MMS. The central interface enables automation of tasks performed in previous watershed studies, but now at much coarser and larger scales as well. Since the modeling system is based on a micro-simulation approach, the MMS can also be used for large geographic and watershed studies provided the appropriate data are used.



---

# References

- Babcock, B.A., J. Wu, T. Campbell, P.W. Gassman, P.D. Mitchell, T. Otake, M. Siemers, and T.M. Hurley. 1997. RAPS 1997: Agriculture and environmental outlook. Center for Agricultural and Rural Development, Iowa State Univ., Ames, IA.
- Bouzaher, A., Lakshminarayan, P. G., Cabe, R., Carriquiry, A. L., Gassman, P., and Shogren, J. F. 1993a. Metamodels and Nonpoint Pollution Policy in Agriculture. *Water Resources Research*, 29:1579-1587.
- Bouzaher, A.; Lakshminarayan, P.G.; Johnson, S.R. Jones, T.; Jones, R., 1993b. "The economic and environmental indicators for evaluating the National Pilot Project on Livestock and the Environment: Livestock series report 1." Iowa-State-Univ. Ames, Iowa: Center for Agricultural and Rural Development. Oct 1993. (93-SR64).
- Gassman, P. W. 2002. Users Manual for the Interactive APEX (I\_APEX) modeling framework. Center for Agricultural and Rural Development. (Draft) Forthcoming.
- Greene, William H. 1993. *Econometric Analysis*. Second Edition. Macmillan Publishing Company. New York.
- Hargrove, W. W. and R. J. Luxmoore. 1997. A spatial clustering technique for the identification of customizable ecoregions. <http://research.esd.ornl.gov/~hnw/esri/>
- Heatwole, C. D., A. B. Bottcher, and L. B. Baldwin. 1987. Modeling cost-effectiveness of agricultural nonpoint pollution abatement programs on two Florida basins. *Water Resour. Bull.* 23: 127-131.
- Hodgess, Erin M., and William W. S. Wei. 2000. Temporal disaggregation of stationary bivariate time series. *Linear Algebra and its applications* 321 (2000) 175-196.
- Howitt, R. E., and A. Reynaud. 2001. Spatial disaggregation of agricultural production data by maximum of entropy. [http://www.inra.fr/Internet/Departements/ESR/vie/animations/jeuneschercheurs/pdf/2001/JJC\\_Reynaud.pdf](http://www.inra.fr/Internet/Departements/ESR/vie/animations/jeuneschercheurs/pdf/2001/JJC_Reynaud.pdf)
- Jobson, J. D. 1991. *Applied multivariate analysis*, Vol. 2: categorical and multivariate methods, Springer Verlag, New York, NY, 721 pp.
- Jones, R., L. Frarey, A. Bouzaher, S. R. Johnson, and S. Neibergs. 1993. Livestock and the environment: a national pilot project: Detailed problem statement. Texas Institute for Applied Environmental Research, Tarleton State Univ., Stephenville, Texas.
- Just, R., and R. Pope. 1999. Implications of heterogeneity for theory and practice in production economics. *American Journal of Agricultural Economics*. 81 (February): 711-718.

- Keith, G., S. Norvell, R. Jones, C. Maquire, E. Osei, A. Saleh, P. Gassman, and John Rodecap. 2000. Livestock and The Environment: A National Pilot Project: CEEOT-LP Modeling for the Upper Maquoketa River Watershed, Iowa: Final Report. Report No. PR0003. Stephenville, TX: Texas Institute for Applied Environmental Research, Tarleton State University.
- Kobrich, C., T. Rehman, and M. Khan. Typification of farming systems for constructing representative farm models: two illustrations of the application of multivariate analysis in Chile and Pakistan. *Agricultural Systems* (In press).
- Koutsoyiannis Demetris, and Christian Onof. 2001. Rainfall disaggregation using adjusting procedures on a Poisson cluster model. *Journal of Hydrology*. 246 (2001) 109-122.
- Lin, Kuo-Shian, and Debbie A. Niemeier. 1998. Temporal disaggregation of travel demand for high resolution emissions inventories. *Transportation Research Part D*, Vol. 3, No. 6, pp 375-387
- McNitt, J., R. Jones, E. Osei, L. Hauck, and H. Jones. 1999. Livestock and the Environment : precedents for runoff policy; policy option-CEEOT-LP. Report PR9909. Texas Institute for Applied Environmental Research, Tarleton State University, Stephenville, TX.
- Osei, E., P. Gassman, and A. Saleh. 2000. Livestock and the Environment: a national pilot project; Economic and environmental modeling using CEEOT. Report PR0002. Texas Institute for Applied Environmental Research, Tarleton State University, Stephenville, TX.
- Panel on Discriminant Analysis and Clustering. 1989. Discriminant analysis and clustering. *Statistical Science* 4:34-69.
- Pratt, Staci, Jones, Ron and Jones, C. Allan. 1997. Livestock and the environment: Expanding the focus: Policy options: CEEOT-LP, Texas Institute for Applied Environmental Research.
- SAS Institute Inc. 1983. SAS technical report A-108. Cubic Clustering Criterion, Cary, NC, SAS Institute Inc. 56 pp.
- SAS Institute. 1999. The SAS system for windows. Version 8. SAS Inst., Cary, NC.
- Tamis, W. L. M., and M. Van't Zelfde. 1998. An expert habitat suitability model for the disaggregation of bird survey data: Bird counts in the netherlands downscaled from atlas block to kilometre cell. *Landscape and Urban Planning* 40 (1998) 269-282.

A macro environment refers to the overall, broader economy and the forces affecting it versus a microenvironment, which focuses on a specific sector or region's economy. There are macroeconomic conditions or factors that affect how all businesses operate, which, in turn, affect the economy as a whole. In other words, a cyclical " which largely involve luxury goods and travel " that are more heavily affected by macroeconomic factors. Summary. A macro environment involves forces that affect the larger economy. A company should regularly perform a DEPEST analysis in order to best serve customers and the economy as a whole. Abstract. The interdisciplinary nature and increasing complexity of water- and environmental-resource problems require the use of modeling approaches that can incorporate knowledge from a broad range of scientific disciplines. The large number of distributed hydrological and ecosystem models currently available are composed of a variety of different conceptualizations of the associated processes they simulate. Assessment of the capabilities of these distributed models requires evaluation of the conceptualizations of the individual processes, and the identification of which conceptualizations are most appropriate for various combinations of criteria, such as problem objectives, data constraints, and spatial and temporal scales of application. Macro environment is the remote environment of the firm, i.e. the external environment in which it exists. As a rule this environment is not controllable by the firm, it is too huge and too unpredictable to control. Hence the success of the company, to a large extent will depend on the company's ability to adapt and react to the changes in the macro environment. Primarily the company has to closely monitor the various elements of macro environment. This will help them understand the dynamic nature of the macro environment. The modelling system used was the Comprehensive Economic and Environmental Optimization Tool (CEEOT), which was designed to evaluate the economic and environmental impacts of agricultural BMPs on water and soil quality at field and watershed scales. The CEEOT framework enabled interfacing among three separate computer models: Soil and Water Assessment Tool (SWAT), Agricultural Policy/Environmental eXtender (APEX), and Farm-level Economic Model (FEM). The modelling component consisted of two phases. In Phase 1 (2007 to 2011), CEEOT was applied to the two BMP Project watersheds, (Indianfarm Cree... The Macro Environment Analysis is the first step of a strategic analysis which in turn kicks off the traditional; strategic planning cycle; it is sometimes referred to as an external analysis, a pest analysis or a pestle analysis. The purpose of the Macro Environment Analysis is to identify possible opportunities and threats that will impact on your industry as a whole and that are outside the control of your industry. The goal of completing a macro environment analyses is to answer the following questions. What environmental factors will affect the growth of our industry as a whole? What is the likely cumulative impact of all of the things that affect the growth of our industry i.e. will our industry grow or shrink over the next 3-15 years?