Data Collection for the 3MRA Modeling System

Robert S. Truesdale
RTI International

Presentation Overview

- Objectives, scope, and approach
- Methodology overview (by approach and data type)
  - National data
  - Regional data
  - Site-based data
- Results and conclusions

Objectives and Scope

- Assemble dataset for 3MRA Model (over 900 inputs) that is complete and appropriate for a national analysis
- Site-based, regional data collected for 201 sites representing industrial non-hazardous waste management across the U.S.
- Data collected nationally for variables where site-based or regional values are not appropriate or readily available
- Data collected to represent nationwide variability NOT site-specific conditions

Site Selection and WMU Settings

- EPA's 1985 Industrial Subtitle D Survey used to represent the nationwide management of non-hazardous industrial wastes
- 201 sites randomly selected from over 2,800 facilities in survey with onsite waste management units (WMUs)
- Survey data used to place landfills, waste piles, land application units, and surface impoundments at each site
- Aerated tanks placed at each surface impoundment site (tanks not in survey)
- Multiple WMU types at 201 sites = 419 WMU settings

Data Collection Approach

- Data collected using one of 3 approaches:
  - *Site-based*: collected around each site using geographic information systems (GIS) technology
  - *Regional*: characterize regional conditions, assign sites to regions
  - *National:* characterize nationwide conditions
- Approach determined by data availability and variability; site-based generally preferred

Data Quality

- 5-site pilot study to determine data availability, collection approach
- Data collection plan with sources, methods, and QA/QC
- Detailed system specifications define data type, amount, and quantity (and are enforced by system for each variable)
- Data quality objectives (for extant data) defined by data type (see Vol. 2)
- Iterative quality process:
  - Senor technical direction and review of methodology
  - Automated to ensure precision, consistency, and completeness (over 8 million records in final databases)
  - QA/QC checks of both manual and automated data entry and processing

National Data

- Variables characterized by single value or distribution
- Variables for which site-based or regional data are not available or not appropriate
- Every site uses the same data

Example National Data

- *Chemical-specific data*: collected from literature or estimated with models (SPARC, MINTEQ); 3MRA system adjusts for site-specific temperature, pH
- *Human exposure factors:* national distributions developed by age cohort from data in *Exposure Factors Handbook*
- *Human health benchmarks:* collected from a hierarchy of established data sources (IRIS, HEAST, etc.)
- *Ecological benchmarks:* collected from extant sources considering weight of evidence and focusing on relevance to population sustainability, community structure and function.
- *Aerated tank data:* developed dataset of model tanks from EPA databases and engineering analysis

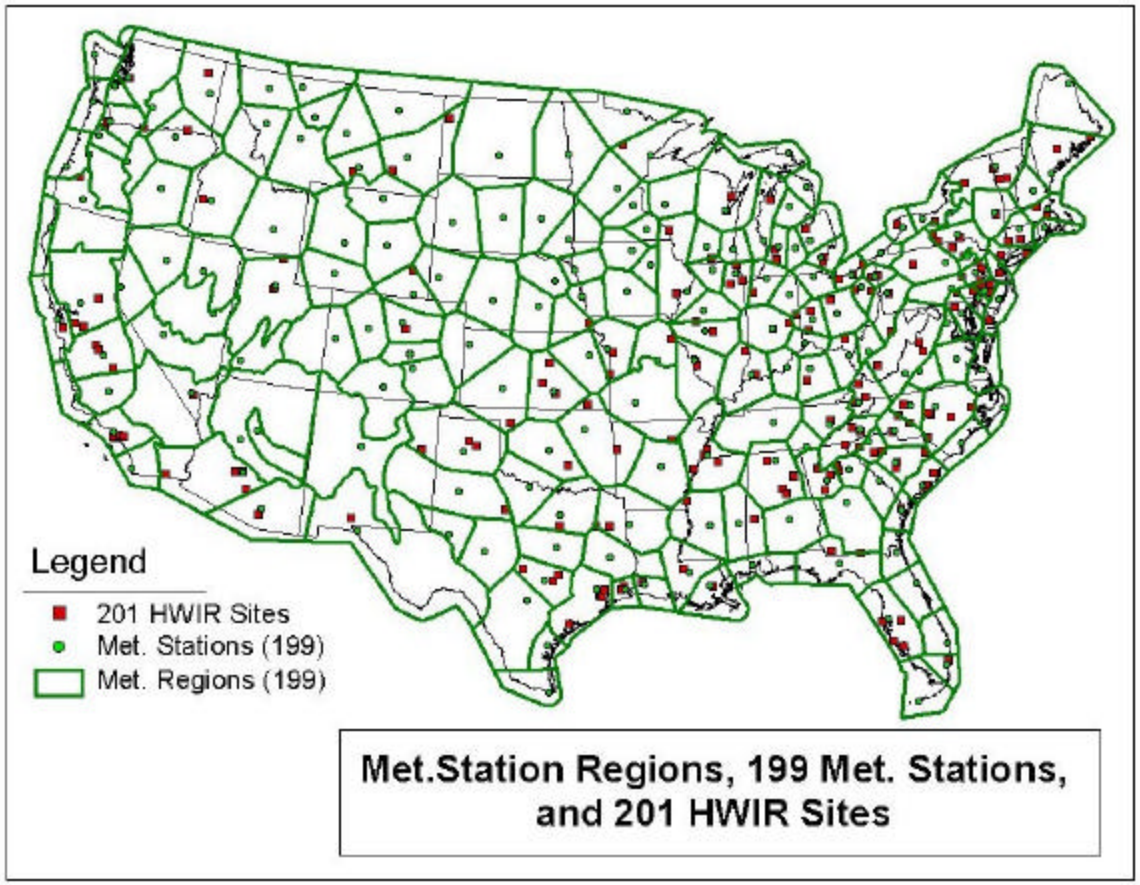
Regional Data

Sites assigned to regions or settings; variables collected to characterize each region:

- *Fixed regions:* meteorological data (hourly, daily, monthly, annual, long-term); stream base flow data
- *Variable regions*: water quality data (cataloging unit, accounting unit, subregion, region)
- *Hydrogeologic settings:* correlated dataset from API's Hydrogeologic Database (HGDB)

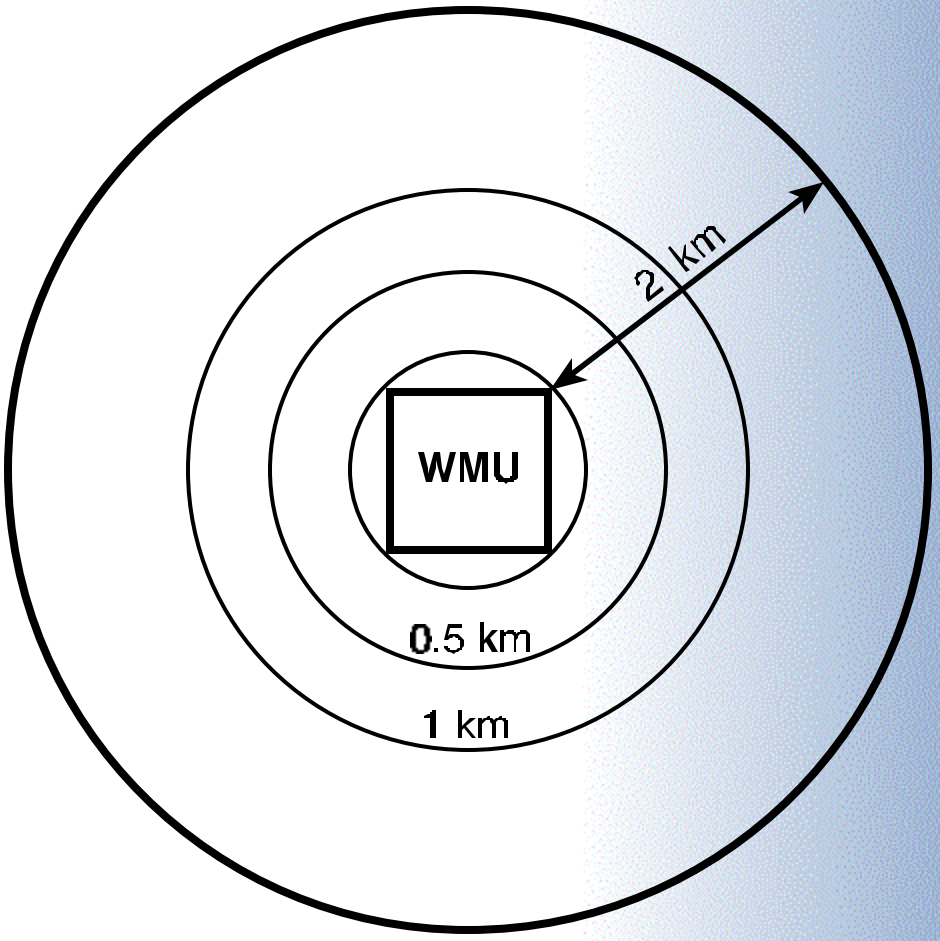
Meteorological Data

- Met. regions defined by Thiessen polygons, conflated to climatic boundaries
- SAMSON, Cooperative Summary, and Radiosonde Data data processed for hourly, daily, monthly, annual, long-term data files
- QC checks and data replacement routines ensured adequate record length (10 yr)



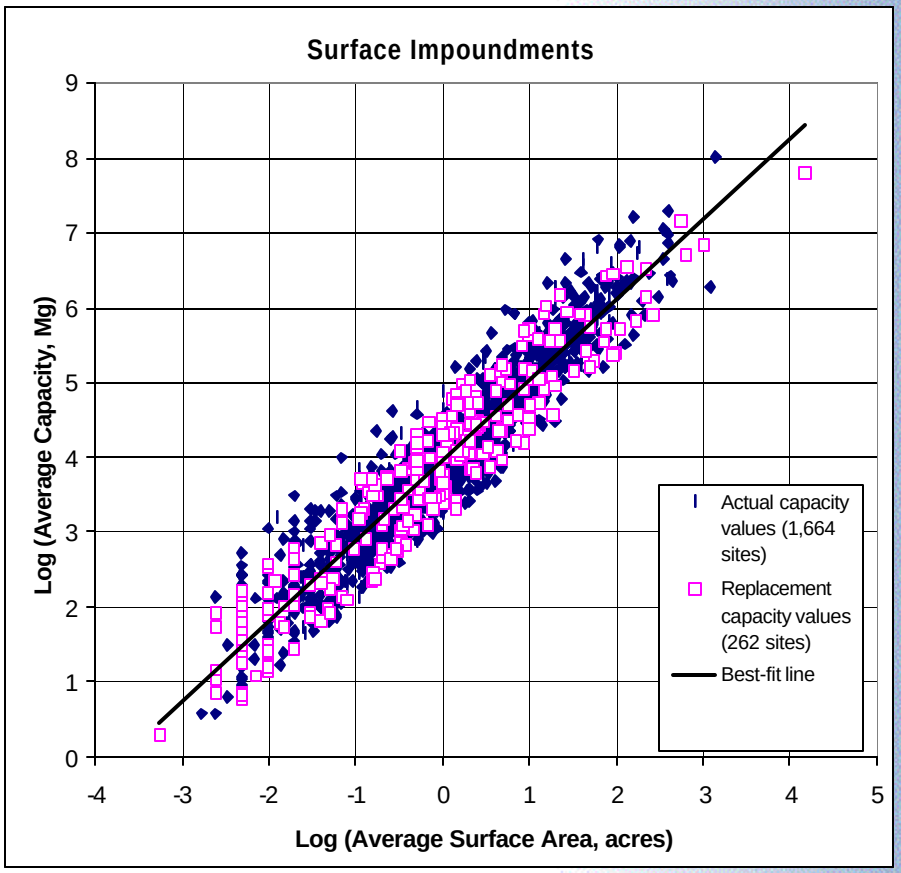
Site-based Data

- Data collected for each of the 201 Industrial D sites
- Data collected within 2 km radius from square WMU
- 3 risk distance rings (2 for eco)
- Sources modeled separately; different WMU = separate area of interest (AOI) per 419 settings
- Used GIS data to define waterbodies and watersheds and place human and ecological receptors at site



Waste Management Unit Data

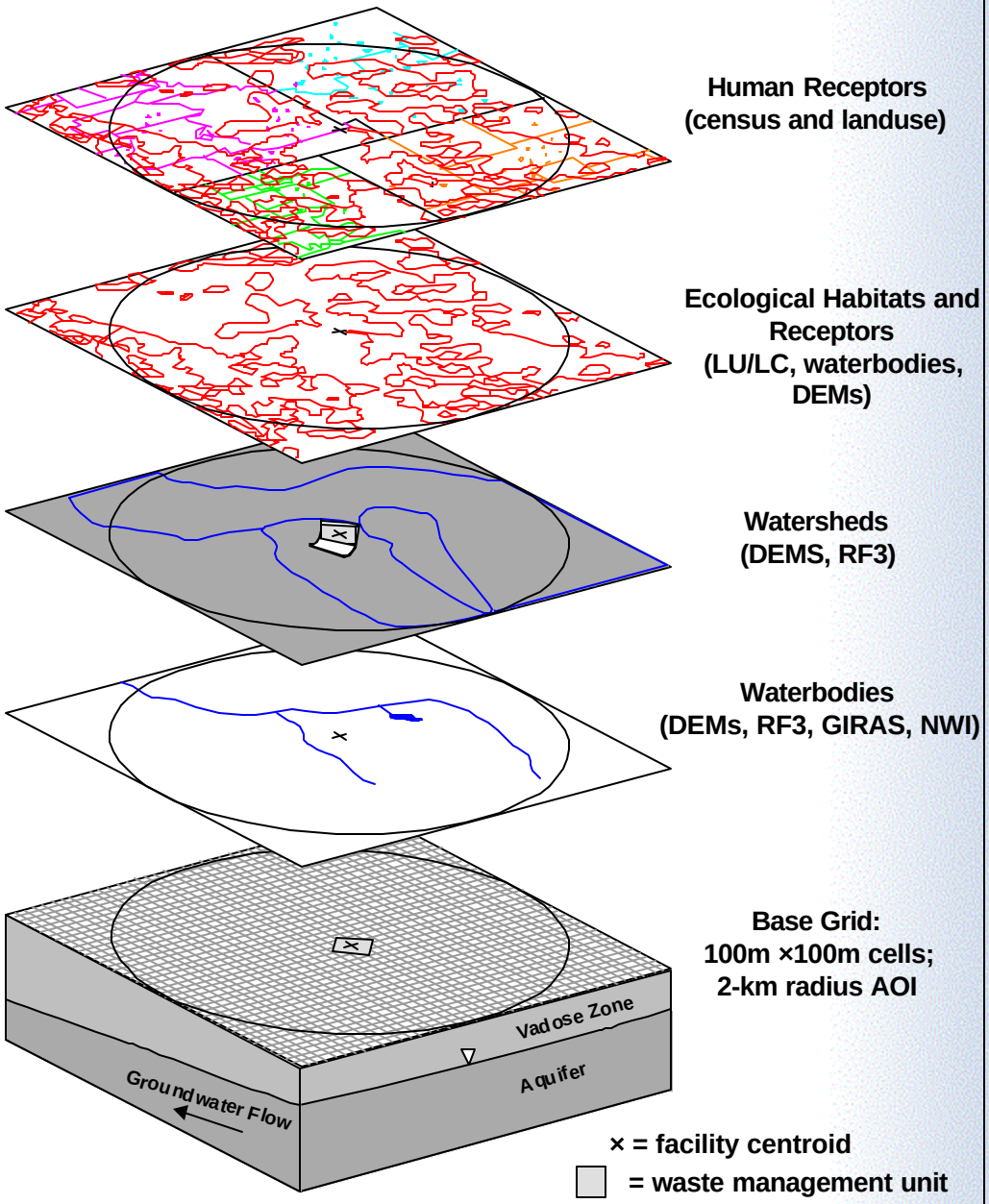
- Industrial D data: WMU area, capacity, annual waste volume
- Data screened and randomly replaced using correlations
- Other WMU variables scaled to Ind. D size data
- Aerated tanks randomly selected and placed at every site with an SI



Spatial Data Sources (GIS)

- *Land use*: Geographic Information Retrieval and Analysis System (GIRAS)
- *Hydrography*: Reach Files 3 (RF3), National Wetlands Inventory (NWI), GIRAS
- *Topography*: 1:250,000-scale Digital Elevation Model (DEM)
- *Soils*: State Soil Geographic (STATSGO) database
- *Population*: U.S. Census (block, block group) and County Agricultural Census

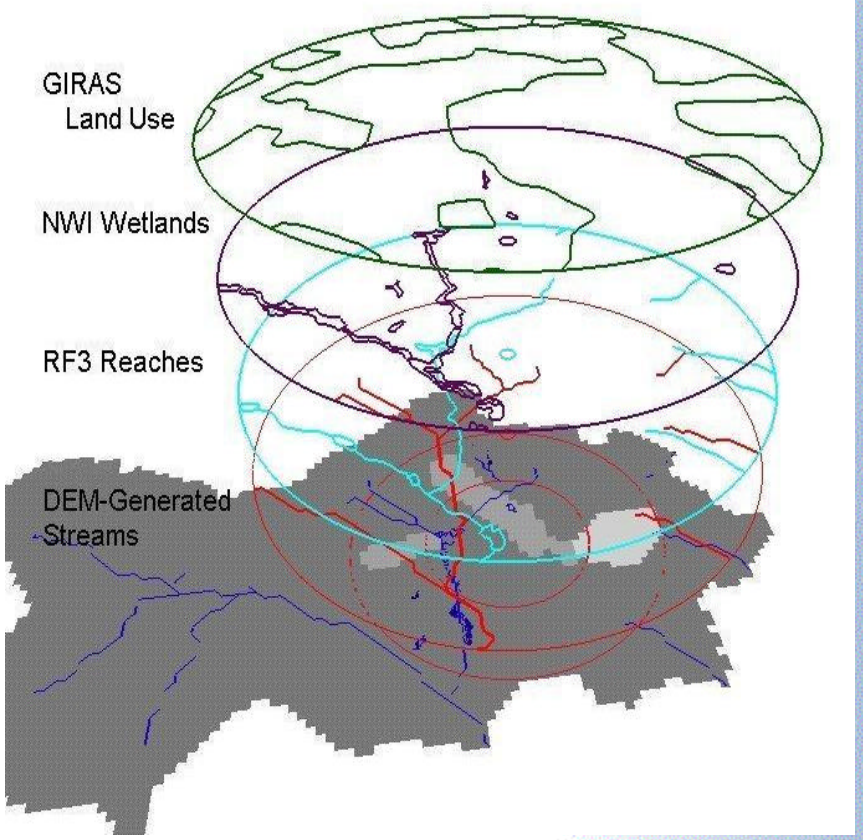
3MRA Spatial Data Layers



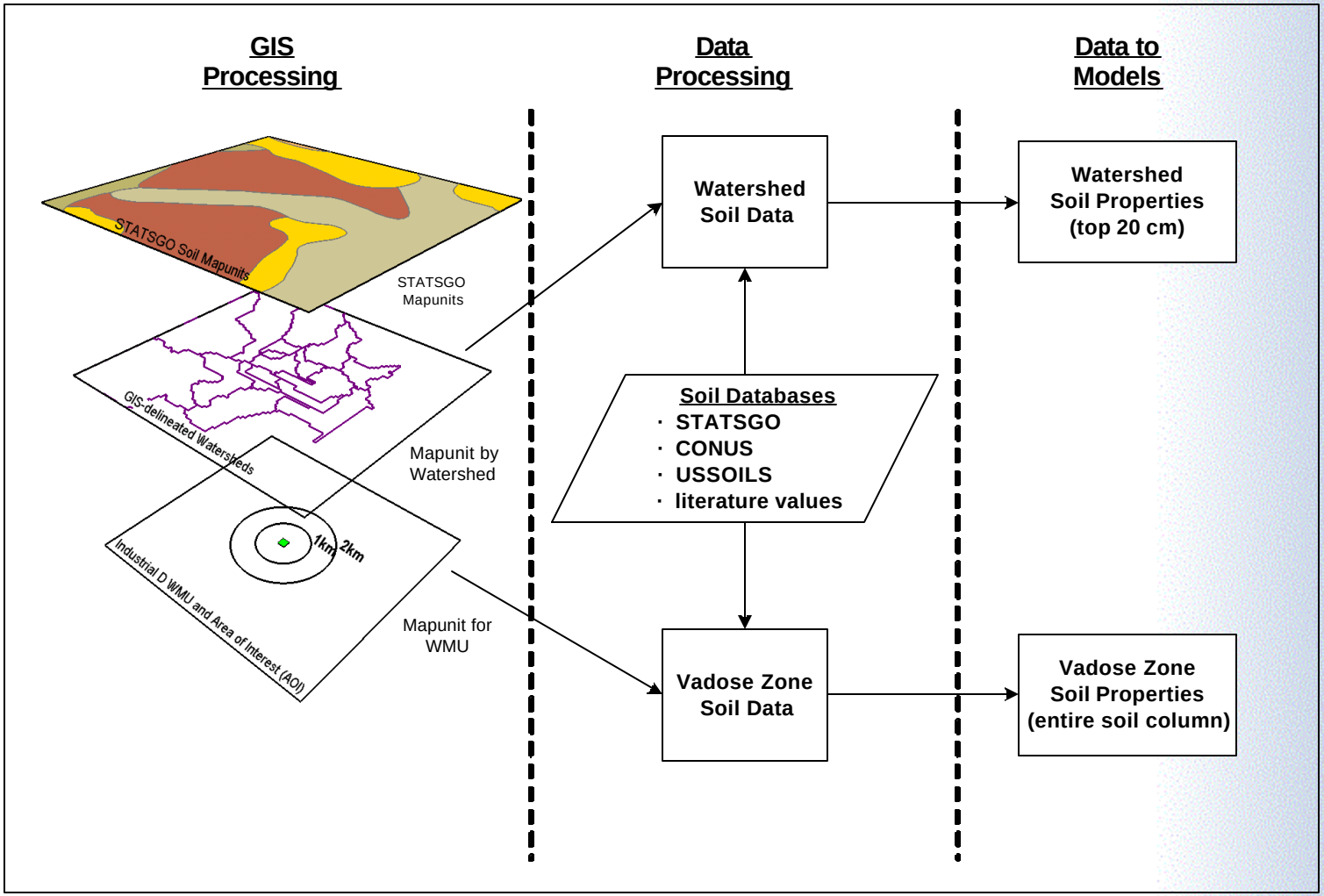
Waterbody/Watershed Delineation

- Reach files (RF3) used as basis for waterbody network
- Wetlands added from NWI or landuse data
- Watersheds delineated manually or automatically, depending on relief (extend outside AOI for reach order 5 and less)
- All delineations and connectivity checked for consistency, correctness, and system compatibility

Data Layers

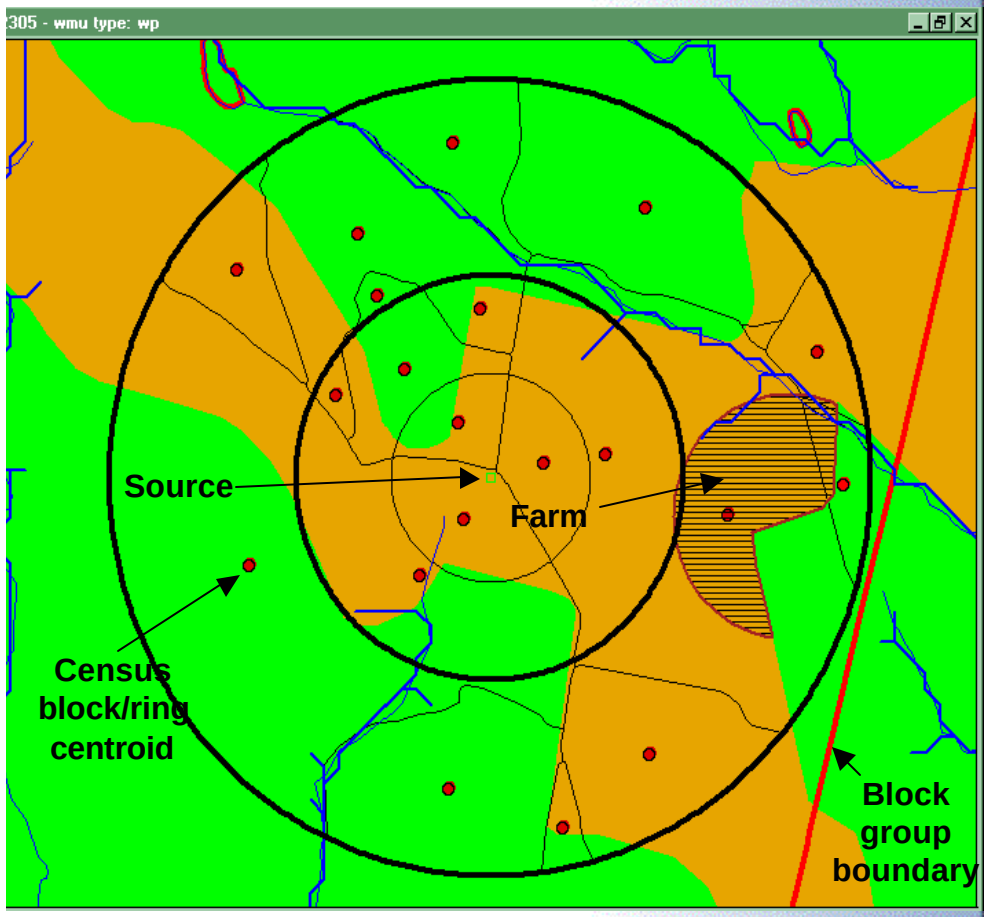


Soil Data Collection and Processing



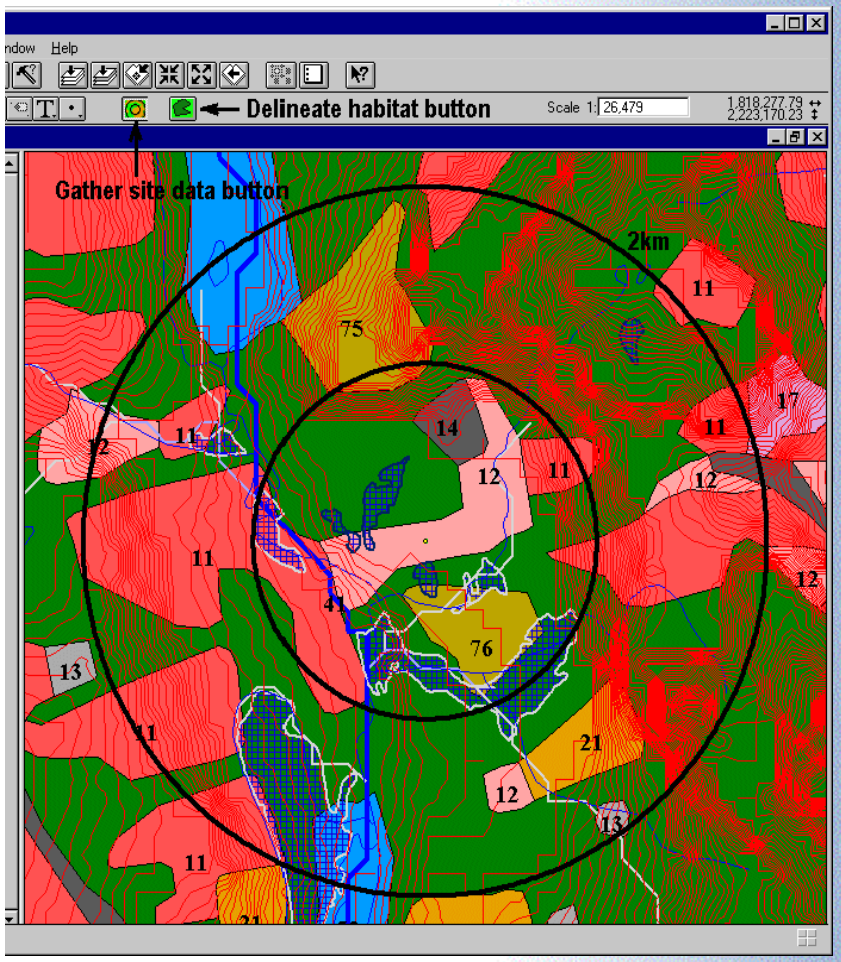
Human Receptor Placement

- Receptor point placed at each census block/ring centroid
- Farms placed within agricultural landuse / block group polygons
- U.S. Census data: population by age group (block), wells (block group)
- Agricultural census data: farm type, median size (county-level)



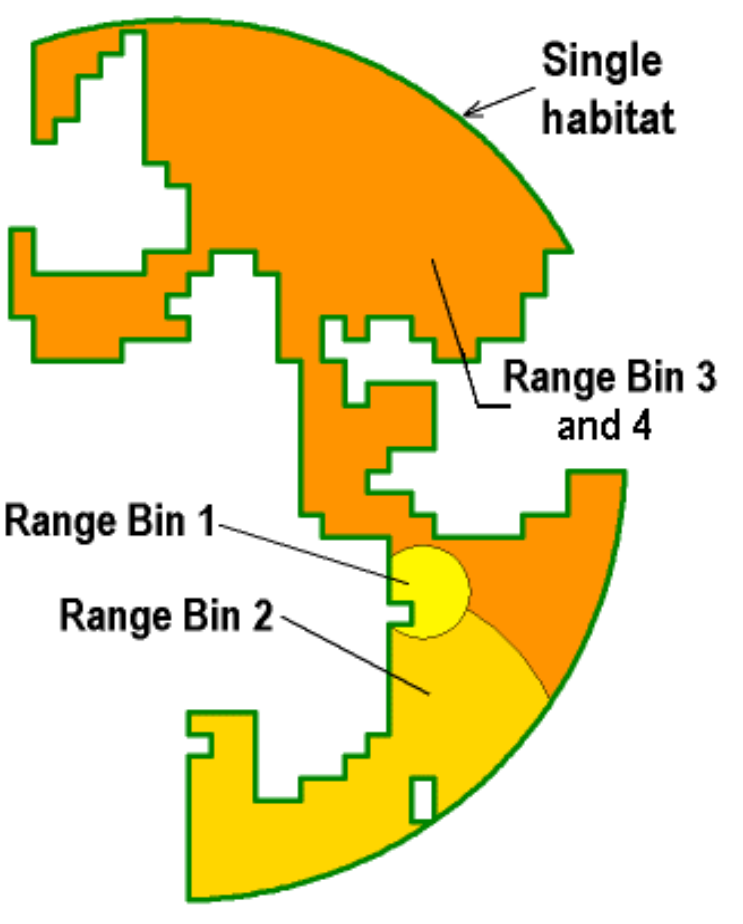
Ecological Habitat Delineation

- 14 representative terrestrial, wetland, and margin habitats
- GIS habitat delineation tool: WMU, landuse, waterbodies, watersheds, topography
- Habitats delineated by assigning grid cells to habitats (“no habitat” for highly developed, impervious areas)
- Elevation contours used to help define stream corridors, lake margins



Ecological Home Range Delineation

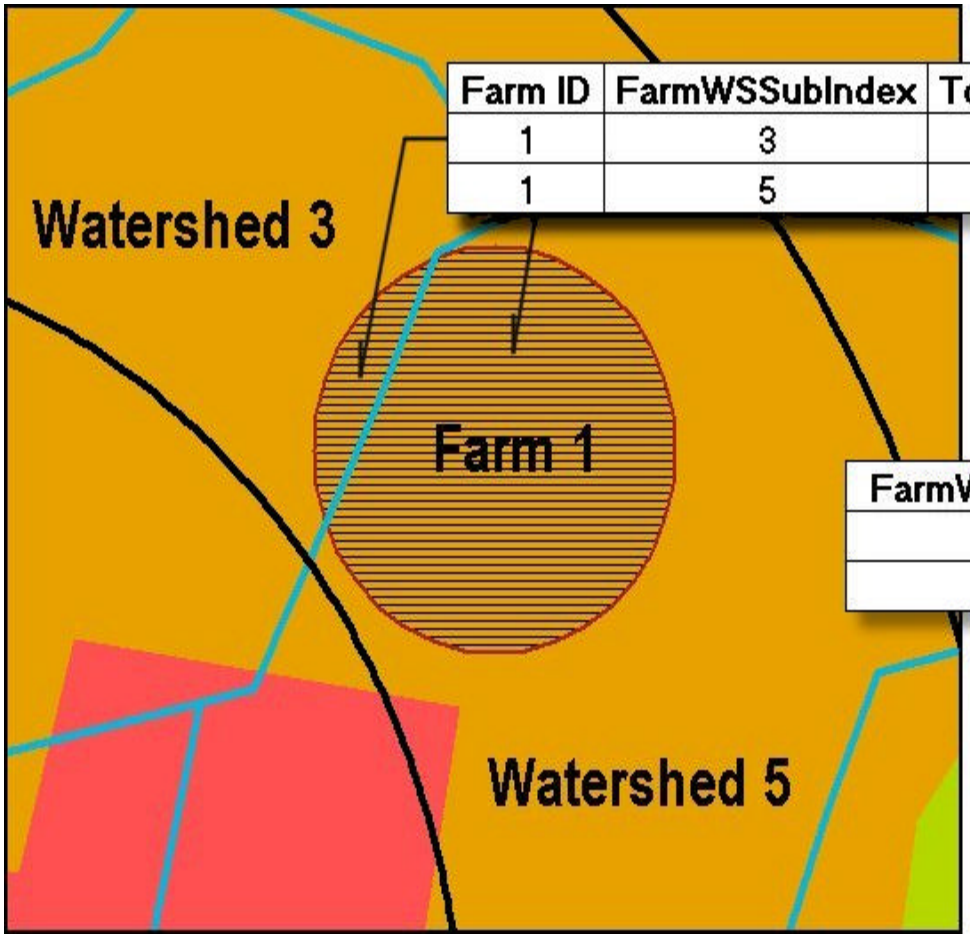
- 4 home range bins used to simplify delineation
- Bins randomly placed within each habitat using GIS program
- Bin ranges overlap; largest bins often cover entire habitat
- Receptors assigned to bins based by habitat and home range size



Data Processing

- Database programs processed GIS and other data to 3MRA model specifications
- Database includes references by variable
- Grid database for all spatial features (used to place air modeling points)
- Meteorological and chemical data processed separately to create ASCII input data files
- Quality control checks programmed into data processing (iterative process)
- 3MRA system checks specifications, completeness, min-max range in final database

Spatial Data Conversion



Farm ID	FarmWSSubIndex	Total WSSub Area	Intersection
1	3	4999875	33084
1	5	5999850	300241

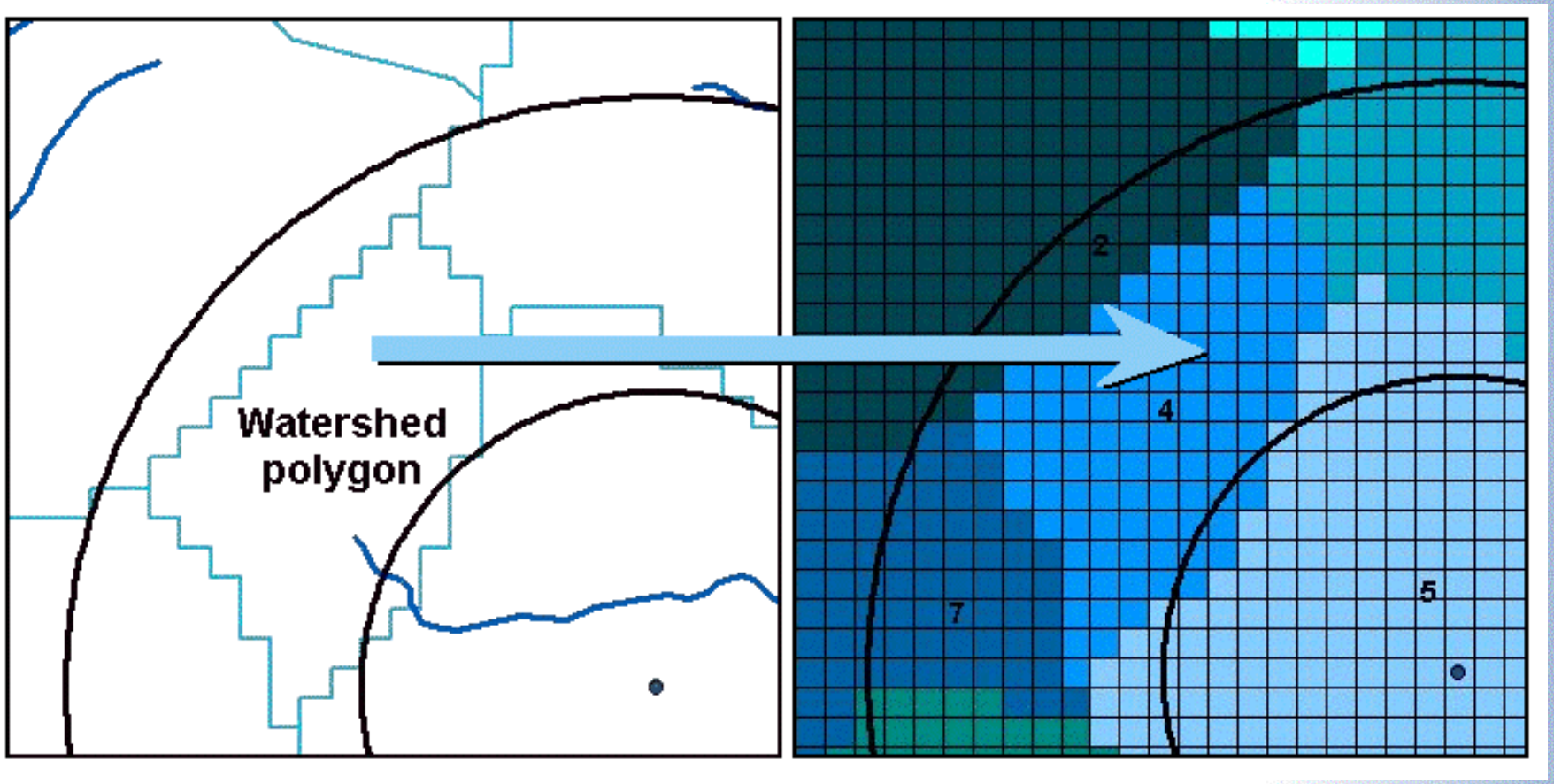
Passed by GIS

FarmWSSubIndex	FarmWSSubFrac
3	0.0066
5	0.0500

Processed in database

Fractions and indices calculated for overlapping areas

Spatial Data Conversion



Spatial features converted to 100 m x 100 m grid cells

Results

- GIS and data processing programs produced 3MRA database currently used for testing model
- 3MRA variable distribution database
  - National data table (446 variables; 35,507 records)
  - Regional data table (7 variables; 1,629 records)
  - Site-based data table (264 variables; 3,531,901 records)
- Meteorological datafiles (99 stations x 5 files; 45 variables)
- Chemical specific datafiles (41 chemicals, 19 files, 145 variables)

Conclusions

- Data for national multimedia modeling can be obtained from existing secondary data.
- GIS coupled with conventional database programs can effectively and accurately collect, process, and analyze large volumes of site-based spatial data for a national analysis.
- Site-based, regional, national approach is flexible, allows best use of data, and minimizes correlation problems.
- 201-site 3MRA dataset has been useful for model testing and can serve as a data source for national risk analyses.