International Congress on Environmental Modelling and Software

5th International Congress on Environmental Modelling and Software - Ottawa, Ontario, Canada - July 2010

Jul 1st, 12:00 AM

A Conservation Ontology and Knowledge Base to Support Delivery of Technical Assistance to Agricultural Producers in the United States

Jack R. Carlson

David B. Butler

Olaf David

Ken W. Rojas

James C. Ascough II

See next page for additional authors

Presenter/Author Information

Jack R. Carlson, David B. Butler, Olaf David, Ken W. Rojas, James C. Ascough II, George Leavesley, Wendall R. Oaks, Lane C. Price, and L.R. Ahuja

International Environmental Modelling and Software Society (iEMSs)
2010 International Congress on Environmental Modelling and Software
Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada
David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.)
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

A Conservation Ontology and Knowledge Base to Support Delivery of Technical Assistance to Agricultural Producers in the United States

Jack R. Carlson[a], David B. Butler[b], Olaf David[a], Ken W. Rojas[b], James C. Ascough II[c], George H. Leavesley[a], Wendall R. Oaks[b], Lane C. Price[d], and Lajpat R. Ahuja[c]
[a]Colorado State University Civil and Environmental Engineering, Fort Collins, CO USA (pspicata@gmail.com)
[b]USDA-NRCS Information Technology Center, Fort Collins, CO USA
[c]USDA-ARS Agricultural Systems Research Unit, Fort Collins, CO USA
[d]USDA-NRCS National Conservation Delivery Streamlining Team, Raleigh, NC USA

Abstract: Information systems supporting the delivery of conservation technical assistance by the United States Department of Agriculture (USDA) to agricultural producers on working lands have become increasingly complex over the past 25 years. They are constrained by inconsistent coordination of domain knowledge across databases, business applications, science models, and other repositories. The extent to which they interoperate is due to implicit understanding of core concepts across business interests. Domain knowledge has been embedded in policy and technical documents for more than 60 years, and with the advent of computing systems, some transformed into metadata of entity-relationship models and data dictionaries. However, these metadata usually are not transparent outside the particular business interests involved. A core conservation ontology and knowledge base (COKB) has been developed to work towards resolving these limitations. The COKB establishes core domain classes and their relationships for area of interest, assessment unit, management unit, response unit, management effect, conservation practice, management system, land use, land cover, management period, and management operation. It provides the foundation for a conservation delivery streamlining initiative.

Keywords: Agricultural natural resource conservation, ontology, knowledge base

1. INTRODUCTION

USDA delivers conservation programs legislated by Congress by providing technical and financial assistance to cooperating land owners and operators. Conservation assistance facilitates actions taken to sustain natural resources on agricultural working lands in the country. The planning and application of conservation addresses soil, water, plant, animal, air, and energy resource concerns, and also is influenced by social, economic, and cultural considerations. The programs are delivered using a conservation planning and application process that involves resource inventory, analysis of problems and opportunities, formulation and analysis of alternatives, application of a selected alternative, followed by monitoring, adjustment, and maintenance of the applied solution [USDA-NRCS 2006]. The data and information associated with the process are managed in a conservation plan developed and maintained with the land owner or operator. Currently there are more than 1 million active conservation plans covering more than 120 million hectares, managed in a central database that has been upgraded through our life cycles since 1988. To improve system efficiency, streamline business processes, integrate new natural resource science, and effectively assimilate knowledge from a variety of formats, this paper describes a Conservation Ontology and Knowledge Base (COKB).

The COKB underpins the data architecture of the USDA Natural Resources Conservation Service (NRCS) Conservation

Delivery Streamlining Initiative called CDSI [USDA-NRCS 2009a]. The initiative revamps the conservation delivery business model, with an emphasis on expanding and improving science-based technical assistance.

2. MATERIALS AND METHODS

Core conservation planning and application domain concepts are found in the USDA-NRCS National Planning Procedures Handbook [USDA-NRCS 2006] and technical guide policy [USDA-NRCS 2007]. Domain concepts also are expressed in entity-relationship models, data dictionaries, and database schemas for business application databases, such as the National Conservation Planning (NCP) database and the electronic Field Office Technical Guide (eFOTG). These sources were leveraged to create the domain classes and properties of the COKB. The classes and properties then were compared with concepts, parameters, and data definitions of agro-environmental models having priority to support conservation program delivery: (1) SWAT - Soil and Water Assessment Tool [Neitsch et al 2005]; (2) APEX - Agricultural Policy/Environmental Extender [Williams et al 2008]; (3) RUSLE2 - Revised Universal Soil Loss Equation [USDA-ARS 2008]; (4) AgES-Watershed [Ascough et al 2009]; (5) WEPS - Wind Erosion Prediction System [Wagner et al 1996]; (6) PRMS - Precipitation and Runoff Modeling System [Markstrom et al 2008], (7) SPUR2 - Simulation and Production of Rangelands [Baker and Hanson 2002], and (8) CENTURY carbon [Parton et al 2001]. These models collectively form a science foundation for the conservation planning and application process, and their integrated use requires the application of common core domain concepts. Although these models have been developed for different purposes at different times, common threads were evident, and the task involved adjusting COKB classes and properties to reconcile differences. The core domain concepts were entered into Protégé 4.0.2 and rendered in Web Ontology Language (OWL) / Resource Description Framework (RDF) format. The concepts were established as classes (subclasses), properties, and individuals. Some classes contain few individuals and therefore are maintained in OWL/RDF format. Other classes contain many to millions of individuals and reside in relational databases.

3. CORE CONSERVATION DOMAIN CLASSES

The following domain classes have been established to support CDSI and the integration of science model services. They are listed in alphabetical order, and within each section their relationships to other classes (bold font) is described. Figure 1 displays the core COKB classes and their relationships.

3.1 Area of Interest (AOI)

Definition: a geographical area encompassing the areas to be analyzed by the conservationist. An AOI is a bounding box, a rectangular or irregular polygon. With a business software application, the user zooms to an area on a digital map, delineates and sets the AOI. One or more **assessment areas** are found within an AOI. An AOI can overlap with one or more other AOIs. Although AOI may seem merely a convenience for geospatial software applications, it conveys business meaning: the area containing the entire resource problem space to be analyzed. None of the 8 models specifically include AOI as a geospatial entity. APEX contains the concept of a study containing areas called sites for resource analysis. Sites contain subareas, which correspond to the **management unit** domain described below. Geospatial interfaces to some models involve zooming to an area on a map where the analysis will be performed, and the COKB standardizes the concept across models. AOI and the classes Assessment Area, Management Unit, and Response Unit defined later are polygons, and in that context the terms “area” and “unit” are synonymous. Further standardizing terms in the COKB will occur in subsequent versions.

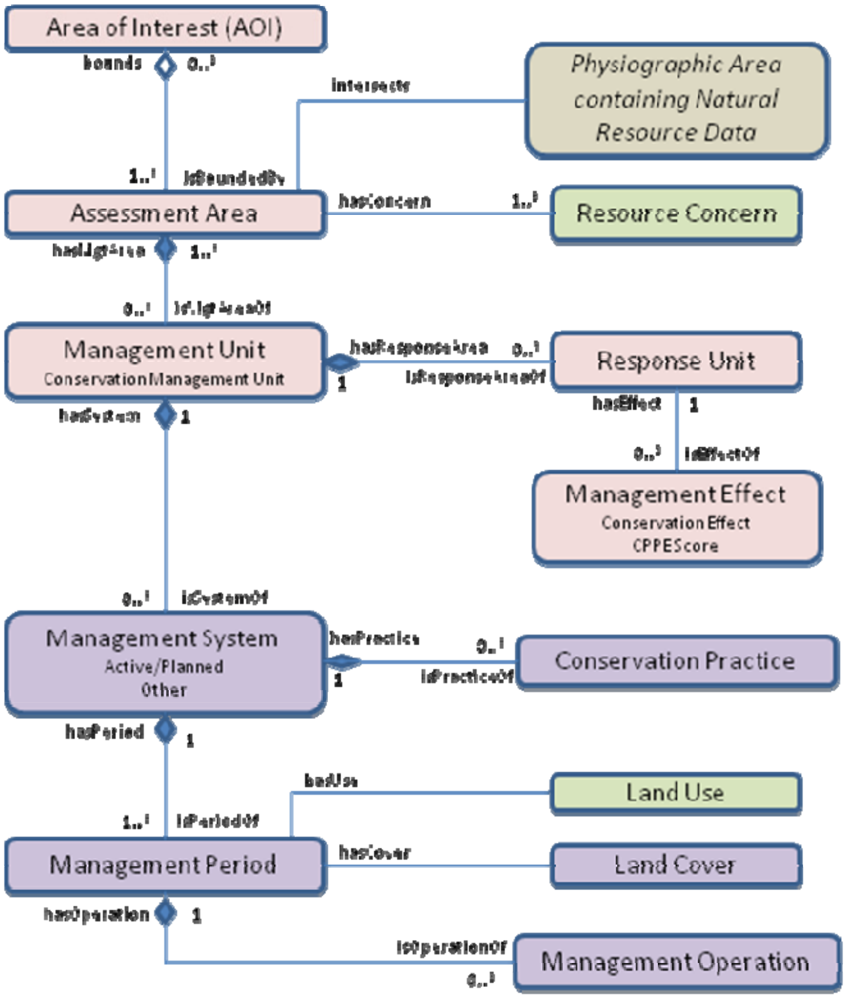

Figure 1. Conservation Ontology and Knowledge Base (COKB) core classes and their relationships.

3.2 Assessment Area

Definition: a geographical area inventoried and analyzed for the effects of management on one or more resource concerns. An assessment area is a polygon, for example a farm, part of a farm, a watershed, or a river basin. A farm split into two separate areas corresponds to two assessment areas. One of the first inventory steps is to identify one or more **resource concerns** for an assessment area. Another step involves delineating one or more **management units** within the assessment area. Various model terms are synonymous with assessment unit: site in APEX; basin or watershed in SWAT, AgES-Watershed (WS), PRMS, and SPUR2; and ranch in SPUR2. These models route water and pollutants through an area synonymous with assessment area. The RUSLE2, WEPS, and CENTURY models do not contain routing and therefore do not employ the assessment unit concept. However, model users usually apply output from these models to areas of land that aggregate to an area synonymous with assessment area. For example, the user may use RUSLE2 to estimate erosion rates in each of the fields of a farm, where the farm is the assessment area.

3.3 Conservation Practice

Definition: a physical structure (e.g. dam, terrace, well), vegetative measure (e.g. filter strip, cover crop), or set of **management operations** (e.g. residue management, pesticide management) that imparts a beneficial effect on one or more **resource concerns**. The number of nationally approved conservation practices [USDA-NRCS 2009b] has fluctuated between 150-175 through the years, maintained by NRCS national and state technical committees . Each practice conforms to a standard, and contains design criteria, specifications, and an expected life. All of the models except PRMS require land management inputs to calculate their outputs. Although most inputs are associated with **management operations** described below, some are found in the conservation practice domain. For example, APEX must know the depth and drainage rate from a subsurface drainage practice applied to a **management unit**. This data must be mapped to the appropriate model input file.

3.4 Land Cover

Definition: the biophysical cover on a **management unit** during a **management period**. Land cover parameters are required by the models that calculate biomass affecting infiltration, runoff, forage production, and other outputs. Land cover includes cultivated crops, seeded pasture, and natural plant communities. It also could include bare ground, asphalt, or other non-vegetative cover. APEX, SWAT, AgES-WS, and WEPS have adapted plant growth model components from the Erosion Productivity Index Calculator (EPIC) model [Williams et al 1984]. RUSLE2 applies other equations to adjust vegetative production to estimate canopy and residue effects on erosion. APEX uses RUSLE2 equations to estimate erosion and sedimentation, and APEX plant growth computations provide input to RUSLE2 calculation of canopy and residue effects. The plant growth components of SPUR2 and CENTURY are different from the others, conceptually rooted in ecosystem modeling. PRMS does not have a plant growth submodel, but requires plant cover type and density data for calculations that estimate effects on interception, infiltration, and runoff. Land cover will contain at least two, and probably three, domain subclasses in the short run to accommodate the requirements of the 8 models. The Unified Plant Growth Model (UPGM) attempts to reconcile and integrate the crop (EPIC-based) and rangeland (ecosystem-based) approaches, including an expanded plant/crop database [McMaster et al 2005].

3.5 Land Use

USDA designates 15 land uses for conservation planning and application [USDA-NRCS 2006]: Crop, Forest, Grazed Forest, Grazed Range, Hay, Headquarters, Mined, Native or Naturalized Pasture, Natural Area, Pasture, Recreation, Urban, Water, Watershed Protection, and Wildlife. Land use is associated with a **management period** of a **management unit**.

3.6 Management Effect

Definition: the effect of a **management system** applied by the land manager on a **resource concern**. Management effects of conservation systems are called conservation effects. The difference between a conservation effect and its corresponding benchmark condition effect (see section 3.9) is termed the conservation impact. All 8 models calculate management effects corresponding to resource concerns. For example, RUSLE2 calculates the effect of a management system on the sheet/rill erosion resource concern. Available models collectively do not estimate effects for all resource concerns (Table 1), leading to alternative methods. NRCS maintains conservation practice physical effects (CPPE) matrices, which scores effects of conservation practices on resource concerns reflecting the judgment of conservation experts in the region.

Table 1. Resource concerns addressed in the delivery of USDA conservation technical assistance and models that calculate effects of management on these concerns.

Resource Concern	Applicable Models
Soil Erosion	
Sheet/Rill, Wind, Irrigation	RUSLE2, WEPS, APEX, SWAT, AgES-WS
Gully	
Streambank, Shoreline	
Roadbank, Construction Site	RUSLE2
Soil Quality	
Organic Matter Depletion	APEX, SWAT, CENTURY
Organic Matter Oxidation	APEX, SWAT, CENTURY
Salinity / Contaminants	APEX
Nutrient Cycling	APEX, SWAT, CENTURY
Compaction	
Water Quantity	
Excess Water	APEX, SWAT, AgES-WS, PRMS, SPUR2
Insufficient Water	APEX, SWAT, AgES-WS, PRMS, SPUR2
Inefficient Use	APEX, SWAT, AgES-WS, SPUR2
Water Quality	
Sediment	APEX, SWAT, AgES-WS, RUSLE2
Nutrients	APEX, SWAT, AgES-WS
Pesticides	APEX, SWAT
Pathogens	APEX, SWAT
Salinity	APEX
Air Quality	
Airborne Soil Particulates	WEPS, APEX
Greenhouse gases and ozone	
Chemical Spray Drift	
Odors	
Plants	
Quantity, Diversity, Health, Vigor	APEX, SWAT, AgES-WS, SPUR2
Declining Populations (Threatened/Endangered)	
Animal	
Domestic Livestock – Food, Cover, Water	SPUR2, APEX
Terrestrial Wildlife – Food, Connectivity, Cover, Water	SPUR2
Aquatic Wildlife – Structure, Food, Water, Temperature	
Declining Populations (Threatened/Endangered)	
Energy	
Conservation	APEX

3.7 Management Operation

Definition: an operation during a **management period**, such as tillage, fertilizer application, irrigation, pesticide application, planting, harvesting, and grazing. Management operations contain detailed data important as inputs to most models. The CDSI effort involves developing a Land Management Operation Database (LMOD), which standardizes management operation definitions and attributes. Moving forward, model developers are expected to use LMOD directly, but this will require a transition period during which LMOD data are translated to model specific management operations. LMOD leverages definitions and terms from the 8 models, starting with RUSLE2 crop management zone records and the management operations of the multi-resource concern models APEX and SWAT.

3.8 Management Period

Definition: the period of time during which a set of **management operations** is applied. The period corresponds to producing and harvesting a crop, grazing livestock, etc. A management period is part of a **management system**. A management period is associated with one **land use** and one **land cover**. For example, a management system may consist of a corn and soybean crop rotation. The system can be divided into two management periods, one associated with growing the corn crop (land cover), and the other with growing the soybean crop. All models except PRMS embody the concept of management period, where management changes through the simulation period. The eight models run through daily time steps, and the periods during which the different land covers are managed must be known. PRMS currently sets land cover type and density constant through the simulation.

3.9 Management System

Definition: the **management operations**, **conservation practices**, and other measures applied on **management units**. A management system containing one or more conservation practices is called a conservation system. A management system has a status: active or non-active. An active management system is the one being applied by the land manager. A non-active system can be an alternative developed for the land manager's consideration, or a system that was active in the past. The COKB considers active and non-active systems to be subclasses of management system. If the land manager considers adopting a new management system, the existing active system becomes the benchmark condition. When a new alternative system is chosen, it becomes the active system, and the previous system is moved to non-active status, to the non-active subclass. A management system has one **land use**; one or more **management periods;** and may have one or more planned or applied conservation practices. Management system related data are fundamentally important inputs to all models, except PRMS. Model user data entry can be a time consuming process, and selecting from a database of pre-developed management systems is a viable option to reduce this burden.

3.10 Management Unit

Definition: an area of land designated having the same **resource concerns** and management. A management unit can be congruent with the boundary of a field, the outside boundary of a set of fields, the boundary of a hydrologic unit, or other configuration that meets the definition. Management units within an **assessment area** cannot overlap. All 8 models contain the concept of management unit. At the farm/field scale, a management unit may span ownership boundaries as long as it is operated the same by the same land manager. For data management reasons, conservationists usually attempt to make management units as large as possible without sacrificing technical integrity and usability for the land manager. At the watershed/basin scale, a management unit is congruent with a **response unit**. Land use and management are assumed to apply to the entire response unit.

3.11 Resource Concern

Definition: an expected degradation of the soil, water, air, plant, or animal resource base or energy efficiency to an extent that the sustainability or intended use of the resource is impaired. USDA has maintained a nationally approved list of resource concerns in the NRCS Technical Guide [USDA-NRCS 2007] for many years. CDSI currently is revising the list of resource concerns, similar to the list shown in Table 1. One or more resource concerns are associated with an **assessment area**. A **management unit** and associated

response units inherit the set of resource concerns from the assessment area. Model services (or other methods) are engaged to calculate a **management effect** for each resource concern.

3.12 Response Unit

Definition: an area of land to which one or more **management effects** are applied. At the farm/field scale, the response unit often is congruent with the management unit, but sometimes the management unit is large or physiographically complex and will contain more than one response unit. However, a response unit cannot contain more than one management unit. At the watershed/basin scale, the response unit is a relatively large physiographic area not associated with farms and fields. Land use and management is generalized for the response unit and in all cases the response unit is congruent with the management unit. Response unit corresponds to the following model entities: APEX subarea, SWAT sub-basin, AgES-WS hydrologic response unit, PRMS hydrologic response unit, WEPS region, SPUR2 sub-basin (watershed scale), and SPUR2 grazing unit (farm/ranch scale). RUSLE2 soil loss is calculated along a transect, which often represents a user-defined area synonymous with response unit. CENTURY output also is associated with a user-defined response unit.

4. DISCUSSION

The COKB provides a significant part of the foundation for the CDSI data architecture, providing a bridge to agro-ecosystem models containing functionality to quantify the effect of management on resource concerns. In the short term, these models are being mostly deployed as “black box” services, but in the longer term, they will be re-factored as automated services corresponding to specific resource concerns. In the short term, CDSI business application data entry will be translated to data files understood by “black box” model services. In the longer term, the re-factored model services will respond to business applications through the COKB in a more integrated manner. CDSI spans business domains beyond those of the COKB. Therefore the COKB is expected to be one of a family of ontologies and knowledge bases addressing the entire scope of conservation technical and financial assistance. The ontology provides the conceptual basis for CDSI logical and subsequent physical data models, as well as a quality gate to maintain conceptual consistency across business applications. The knowledge base contains or links to information not represented in databases, available to CDSI business applications.

Creating the COKB is somewhat similar to the SEAMLESS effort in Europe leveraging semantic tools to build a relational database system and model chains for integrated agro-environmental assessment [Athanasiadis et al 2009]. SEAMLESS is focused on integrating existing models and data stores into an integrated assessment system, whereas CDSI is focused on creating a new business model and must deal with transition of legacy system to the new model. In either case, expressing domain concepts in ontologies can facilitate synchronization and interoperability across organization information systems now and in the future. The CDSI model bases, as existing models are re-factored to science component services, also will standardize to leverage the controlled vocabularies of the CUAHSI Hydrologic Information System [Piasecki 2008] and to the extent feasible other accepted sources (http://aims.fao.org/website/Domain-Ontologies/sub).

To fully establish the knowledge base of COKB requires connecting to the NCP and other large conservation databases and information sources. This will occur as the CDSI data architecture is completed. The ontology currently resides at http://oms.javagforge.org. This paper describes the first version of COKB, but frequent updates are expected as the CDSI effort matures, and subsequent versions will be mediated through a formal change control process that includes opportunity for external review and comment.

REFERENCES

Ascough II, J., David, O., Krause, P., Green, T., Heathman, G., Kralisch, S., and L. Ahuja, A component-based distributed watershed model for the USDA CEAP Watershed Assessment Study, paper presented at Farming Systems Design 2009, Monterey, California USA, August 23-26, 2009.

Athanasiadis, I. N., Rizzoli, A-E, Janssen, S., Andersen, E. and F. Villa. Ontology for seamless integration of agricultural data and models. In Sartori, F., Sicilia, M. A., and Manouselis, N. (ed.) 3rd International Conference on Metadata and Semantics Research (MTSR'09), Springer-Verlag, 46, 282-293, 2009

Baker, B. and J. Hanson. Documentation and users guide for WinSPUR/SPUR2, USDA Agricultural Research Service, Great Plains System Research, 2002 revised.

Markstrom, S., Niswonger, R., Regan, R., Prudic, D., and P. Barlow, GSFLOW – coupled ground-water and surface-water flow, model based on the integration of the Precipitation-Runoff Modeling System (PRMS) and the modular ground-water flow model (MODFLOW-2005), USDI–USGS chapter 1 of section D, Ground-Water/Surface-Water, Book 6, Modeling Techniques, Techniques and Methods 6-D1, 2008.

McMaster, G., Ascough II, J., Edmunds, D., Andales, A., Wagner, L., and F. Fox, Multi-crop plant growth modeling for agricultural models and decision support systems, in Zerger, A. and Argent, R.M. (eds.) MODSIM 2005 International Congress on Modelling and Simulation, *Modelling and Simulation Society of Australia and New Zealand, December 2005*: 2138-2144, 2005. Retrieved 24 September 2009 from http://www.mssanz.org.au/modsim05/papers/mcmaster.pdf.

Neitsch, S., Arnold, J., Kiniry, J., and J. Williams, Soil and Water Assessment Tool – theoretical documentation, version 2005, USDA Agricultural Research Service and Texas Agricultural Experiment Station, 2005.

Parton, B., Ojima, D., Del Grosso, S., and C. Keough, CENTURY Tutorial – Supplement of CENTURY User's Manual, 2001.

Piasecki, M., HydroSeek functional description document (version 1.0), CUAHSI, March 2008. Retrieved 7 December 2009 from http://his.cuahsi.org/documents/HydroSeek_Functional_Description.pdf.

USDA-ARS, SPUR: simulation of production and utilization of rangelands, documentation and user guide, in J. Wight and J. Skiles, (eds.) USDA Agricultural Research Service ARS-63, 1982.

USDA-ARS, Draft science documentation, Revised Universal Soil Loss Equation Version 2, 2008. Retrieved 14 September 2009 from http://www.ars.usda.gov/sp2UserFiles/Place/64080510/RUSLE/RUSLE2_Science_Doc.pdf,

USDA-NRCS, National Planning Procedures Handbook, USDA Natural Resources Conservation Service Title 180, Part 600, 2006. Retrieved 24 September 2009 from http://directives.sc.egov.usda.gov/RollupViewer.aspx?hid=17088.

USDA-NRCS, Technical guides, USDA Natural Resources Conservation Service General Manual Title 450-Technology, Part 401, 2007. Retrieved 24 September 2009 from http://directives.sc.egov.usda.gov/RollupViewer.aspx?hid=17077.

USDA-NRCS, National conservation delivery streamlining initiative – overview of business model and requirements (draft), November, 2009a.

USDA-NRCS, National Handbook of Conservation Practices, 2009b. Retrieved 22 January 2010 from http://www.nrcs.usda.gov/technical/standards/nhcp.html.

Wagner, L, An overview of the wind erosion prediction system, USDA Agricultural Research Service and Kansas Agricultural Experiment Station Contribution No. 96-205-A, 1996.

Williams, J., Izaurralde, R., and E. Steglich, Agricultural Policy/Environmental Extender Model – theoretical documentation, version 0604, Texas Agricultural Research Experiment Station, AgriLIFE Reseach BREC Report #2008-17, 2008.

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