04-155

Environment Canada
Water Science and Technology Directorate

Direction générale des sciences et de la technologie, eau
Environnement Canada

Species at Risk Data and Knowledge Management within the WILDSPACE Decision Support System

By: I.W. Wong, R. Bloom, D.K. McNicol, P. Fong, R. Russell an X. Chen

NWRI Contribution # 04-155

National Water Research Institute - **Publication Form**

Database Locator No. 04-019
date assigned 10 March 2004
NWRI Contribution No. or Branch Approved No. (e.g., abstract, technical note) 04-155

A General Information

Date Submitted: March 10, 2004

Branch: WQMB

Project: Water Quality Information Management and Modelling

Senior NWRI Author: I. Wong

Additional NWRI Authors: P. Fong

Authors: I.W. Wong, R. Bloom, D.K. McNicol, P. Fong, R. Russell and X. Chen

Title: Species at Risk Data and Knowledge Management within the WILDSPACETM Decision Support System

Please specify the **program** or **EC Region** this work is in support of: SARA ______ or General ☐

Business Lines: Clean Environment ☐; Nature ☒; Weather and Environmental Predictions ☐.

B Scientific Manuscripts - To be Submitted to:

☐ **Journal** Name ______

☐ **Book or Chapter** Title ______
Editor(s) ______
Publishers ______ City, Country ______

☒ **Full Paper - Conference/Workshop Proceedings** - Please indicate above if Proceedings will be in Journal or Book
Conference/Workshop Name: International Symposium on Environmental Software Systems 2004
Location: Harrisonburg, Virginia, USA Date May 18-21, 2004

☐ **NWRI Printed Contribution** # of copies requested ______

C Other Types of Material

☐ **Branch Technical Note** (e.g., QA/QC report to clients, other technical reports for VNR clients)

☐ **Conference/Workshop Abstract** → Oral ☐ Poster ☐ Presenter's Name: ______
Conference/Workshop Name: ______
Location: ______ Date: ______

☐ **Other** (letter to editor, newspaper article, web page, etc.)
Provide full reference/details: ______

Additional Information: ______

D Review/Approval

Peer Reviewed by Dr. W.G. Booty Date March 10, 2004

Approved by Project Chief [signature] Date March 11, 2004

Approved by Branch Director [signature] Date [handwritten date] 25, 2004

Is a Briefing Note required (Branch Director to decide): ☐ Yes ☐ No

Does anyone else within the department need to be alerted? ☐ Yes ☐ No Names: ______

Does the report contain any confidential information? ☐ Yes ☐ No or sensitive information? ☐ Yes ☐ No

Action taken by Branch Director: ______

Comments and required changes ______

NOTE: After Branch Director's approval, Author to ensure Branch Coordinator has a final electronic copy of abstract, research summary and manuscript.

Species at Risk Data and Knowledge Management within the WILDSPACE™ Decision Support System

I.W. Wong, R. Bloom, D.K. McNicol, P. Fong, R. Russell and X. Chen

Abstract

The Species at Risk Act (SARA) in Canada provides legislation to prevent wildlife and other species from becoming extinct, to help in the recovery of threatened or endangered species, and to protect critical habitats for these species from further degradation. There is currently no single system to store, retrieve, analyse and interpret information on species and their critical habitats. To support the range of information requirements needed for SARA, it is highly desirable to develop network designs, infrastructures and applications that link distributed data sources into an integrated system that manages data and provides decision support. The system architecture described here will be built on the versatile WILDSPACE™ Decision Support System (hereafter referred to as 'WILDSPACE DSS') and will be web-based consisting of distributed servers (database servers, web servers and map servers) providing different kinds of information including species and habitat data, geo-spatial data, metadata, web services and decision support analyses. Its design takes into consideration the needs of different user groups (Intranet, Extranet and Internet) and data security. The complexity of Species at Risk data requires considerable "best practice" database design efforts. The main goal is to strike an optimum balance among storage, maintenance, and application performance. The WILDSPACE DSS provides an effective platform for the delivery of information and services to Species at Risk practitioners for better decision-making through its data mining, modelling and scenario gaming functionality.

Gestion des données et des connaissances sur les espèces en péril dans le système d'aide à la décision WILDSPACE^{MC}

I.W. Wong, R. Bloom, D.K. McNicol, P. Fong, R. Russell et X. Chen

Résumé

La *Loi sur les espèces en péril* (LEP) du Canada contient les dispositions visant à prévenir la disparition des espèces sauvages et autres, à favoriser le rétablissement des espèces menacées ou en voie de disparition, et à protéger les habitats critiques de ces espèces contre la dégradation. Il n'y a actuellement aucun système unique qui entrepose, extrait, analyse et interprète l'information sur les espèces et leurs habitats critiques. Pour combler les besoins en matière d'information de la LEP, il est grandement souhaitable de mettre au point des infrastructures et applications réseau reliant les diverses sources de données au sein d'un système intégré, qui gère les données et aide à la prise de décisions. L'architecture du système reposera sur le système d'aide à la décision WILDSPACE^{MC} et sera basée sur le Web. Le système consistera en divers serveurs (serveurs de bases de données, serveurs web et serveurs de cartes) fournissant différents types d'information, notamment des données sur les espèces et les habitats, des données géospatiales, des métadonnées, des services web et des analyses d'aide à la décision. Sa conception tient compte des besoins des différents groupes d'utilisateurs (Intranet, Extranet et Internet) et de la sécurité des données. La complexité des données sur les espèces en péril nécessite que des efforts considérables soient déployés pour la conception de bases de données optimales. L'objectif principal est d'atteindre un équilibre optimal entre stockage, entretien et performance des applications. Le système WILDSPACE^{MC} constitue une plate-forme efficace, qui fournit de l'information et des services aux spécialistes des espèces en péril pour les aider à prendre de meilleures décisions grâce à sa fonctionnalité de fouille de données, de modélisation et de simulation des scénarios.

NWRI RESEARCH SUMMARY

Plain language title

Species at Risk Data and Knowledge Management within the WILDSPACE Decision Support System.

What is the problem and what do sicentists already know about it?

Knowledge management is an important issue in science and research. Decision support systems (DSS) are useful analytical, planning and management tools that enable scientists, resource managers and decision-makers to carry out analyses and build a repository for knowledge management.

Why did NWRI do this study?

To bring forward an innovative approach that integrates Species at Risk data with other information such as water quality and spatial data to provide modelling, gaming scenarios and decision support for the research community and the decision makers.

What were the results?

The results were a set of advanced modelling and scenario gaming tools that provide the functionality in the areas of wildlife and aquatic ecosystems research.

How will these results be used?

This set of advanced tools is used in the research and in the decision making process of the Species at Risk program.

Who were our main partners in the study?

CWS-OR

Sommaire des recherches de l'INRE

Titre en langage clair

Gestion des données et des connaissances sur les espèces en péril dans le système d'aide à la décision WILDSPACE^{MC}.

Quel est le problème et que savent les chercheurs à ce sujet?

La gestion des connaissances est un élément important des sciences et des recherches. Les systèmes d'aide à la décision sont des outils d'analyse, de planification et de gestion qui permettent aux chercheurs, aux gestionnaires des ressources et aux décideurs de mener des analyses et de mettre sur pied un dépôt de gestion des connaissances.

Pourquoi l'INRE a-t-il effectué cette étude?

Il faut créer une approche innovatrice qui intègre les données sur les espèces en péril aux autres renseignements (par exemple les données sur la qualité de l'eau et les données spatiales) pour effectuer des modélisations, simuler des scénarios et aider la communauté scientifique et les décideurs à prendre des décisions.

Quels sont les résultats?

Les résultats forment un ensemble d'outils avancés de modélisation et de simulation de scénarios qui fournissent une fonctionnalité pour la recherche sur les espèces sauvages et les écosystèmes d'eau douce.

Comment ces résultats seront-ils utilisés?

Cet ensemble d'outils avancés sert à la recherche et à la prise de décisions du programme sur les espèces en péril.

Quels étaient nos principaux partenaires dans cette étude?

SCF-RO.

Species at Risk Data and Knowledge Management within the WILDSPACE™ Decision Support System

I.W. Wong[1], R. Bloom[2], D.K. McNicol[2], P. Fong[1], R. Russell[2] and X. Chen[3]

[1]National Water Research Institute, Environment Canada, 867 Lakeshore Road, Burlington, ON L7R 4A6, Canada
[2]Canadian Wildlife Service (Ontario Region), Environment Canada, 49 Camelot Drive, Ottawa, ON K1A 0H3, Canada
[3]University of Guelph, Guelph, ON N1G 2W1, Canada

Abstract

The Species at Risk Act (SARA) in Canada provides legislation to prevent wildlife and other species from becoming extinct, to help in the recovery of threatened or endangered species, and to protect critical habitats for these species from further degradation. There is currently no single system to store, retrieve, analyse and interpret information on species and their critical habitats. To support the range of information requirements needed for SARA, it is highly desirable to develop network designs, infrastructures and applications that link distributed data sources into an integrated system that manages data and provides decision support. The system architecture described here will be built on the versatile WILDSPACE™ Decision Support System (hereafter referred to as 'WILDSPACE DSS') and will be web-based consisting of distributed servers (database servers, web servers and map servers) providing different kinds of information including species and habitat data, geo-spatial data, metadata, web services and decision support analyses. Its design takes into consideration the needs of different user groups (Intranet, Extranet and Internet) and data security. The complexity of Species at Risk data requires considerable "best practice" database design efforts. The main goal is to strike an optimum balance among storage, maintenance, and application performance. The WILDSPACE DSS provides an effective platform for the delivery of information and services to Species at Risk practitioners for better decision-making through its data mining, modelling and scenario gaming functionality.

1. Introduction

Growing threats to biological diversity and ecosystem integrity call for innovative approaches and techniques to assess environmental risks that result from natural causes or deliberate actions by humans. Ever increasing demands on government agencies, land managers, and conservation practitioners, for timely, accurate, and spatially explicit environmental information, requires new thinking about the deployment of computer-oriented techniques and technologies (Manley et al. 2004). The growing need for real-time information dissemination and decision support is especially required to mitigate the effects of environmental threats or disasters, either natural or human-induced. Acute threats, such as floods, tornadoes, wildfires and spills, require immediate emergency response. Chronic threats posed by a variety of environmental hazards or conditions, such as pollution, pesticides, or disease, call for a range of management responses.

Increasingly, the search for clear courses of action for the protection and recovery of declining species of flora and fauna (Species at Risk) has become a complex challenge to environmental organizations. Cumulative effects of environmental stressors, such as climate change, invasive species, natural disturbance, habitat loss and fragmentation (e.g., urban sprawl, changing agricultural practices), result in unpredictable influences on individuals, populations, residences, critical habitats and landscapes. The Species at Risk Act (SARA) in Canada, proclaimed in June 2003, provides legislation to prevent wildlife and other species from becoming extirpated or extinct, to provide for the recovery of these threatened and endangered species, to protect residences and critical habitats for these species from further degradation, and to encourage sound management of other species to prevent them from becoming at risk (SARA 2002). Ecological information requirements to support SARA are considerable and broad-ranging, and include information or knowledge about: (1) species abundance and distribution; (2) species natural history; (3) habitats; (4) threats to species or habitats; (5) population viability; (6) recovery opportunities and stewardship; and (7) community or aboriginal traditional knowledge. The SARA requires that responsible agencies in both federal and provincial governments together make timely and informed decisions that serve to identify, protect, and recover both species at risk and the habitats they depend on. These new legal requirements are heavily dependant upon timely and accurate information (underlying data) for which recovery actions are based. Decisions associated with meeting the legal requirements of SARA have become a powerful new driver for the collection, storage, analysis, and reporting of information about species and habitats. Presently, there is no single authority for information on species and habitats; accordingly, there is a strong need for the development of enabling technologies that can link distributed systems and sources in a systematic and logical manner to efficiently store, retrieve, analyse and interpret this information.

Given the range of information requirements for SARA, it is necessary to design networks, databases, and applications to manage environmental and geo-spatial data. To be effective, these frameworks must be based on a distributed, inter-operable architecture which adheres to accepted standards and protocols. Here, we discuss the next generation of the WILDSPACE™ Decision Support System (hereafter referred to as 'WILDSPACE DSS'), originally introduced by Wong et al. (2003). The current generation of the WILDSPACE DSS is based on a desktop workstation with provision for connections to a Local Area Network (LAN), and is used to study complex wildlife conservation problems involving multiple projects, with data that are temporally and spatially heterogeneous. The next-generation system architecture will be web-based, consisting of distributed servers providing different kinds of information, including species and habitat data, geo-spatial data, metadata, web services and decision support analyses. The design and development of Species at Risk databases, and their supporting functions for data entry, facilitates mobile collection and transmission of data by recovery teams in the field, quality assurance and control, and data summarization through queries and analyses. In addition, the system supports advanced analytical, modelling and gaming functions, such as population viability and habitat suitability analysis, which allow for the simulation of effects of habitat change on population stability. Due to the sensitive nature of the data, a data access control and security protocol is embedded within the overall WILDSPACE DSS design. This system provides a more effective, flexible platform for delivery of information and services to Species at Risk practitioners, enabling them to make better decisions on habitat protection and stewardship options. The WILDSPACE DSS maintains the flexibility to support multi-species and multi-scale population and habitat queries required to assess acute or chronic environmental risks, and to assess potential actions to manage or mitigate effects.

2. Network Design and Configuration

The next generation for the WILDSPACE DSS must ensure that multiple users can access information, both locally and remotely. The system will use a Client/Server model (Reagan 2000) that can be accessed using web technologies. Client/Server is a computer network model for the interaction between concurrently executing software processes. The interaction between the client and server processes is a cooperative information exchange in which the client is proactive and the server is reactive. Like most Client/Server systems, the WILDSPACE DSS is based on a "many to one" design; that is, more than one client typically make requests to the server. Several SARA recovery team members will need to access information on the server simultaneously. These activities may include updating new data, accessing existing information and performing scenario gaming to gain knowledge about risk management in the Species at Risk program.

2.1. Client Requirements

In the WILDSPACE DSS wiring diagram (Figure 1), it is identified that there are three types of clients: the Intranet, the Extranet and the Internet. The Intranet refers to the implementation of a network using Internet technologies within the department or organization, rather than external connections to the global Internet. Advantages for an intranet-based approach to the WILDSPACE DSS include: (1) rapid prototyping and deployment of new applications; (2) effective scaling; (3) minimal training requirements, because applications and user interfaces are familiar; (4) open architecture, facilitating seamless addition of customized applications, such as Knowledge Templates; and (5) support for a range of media types, including audio, video and interactive applications. Though similar to the Intranet, the Extranet provides access to departmental resources for a different set of users, predominantly representing partner agencies (e.g., other government departments) or stakeholders. External access can be through the Internet or through other data communication networks (e.g., Virtual Private Network - VPN). The Extranet provides partners with more extensive access to data, in a fashion that enforces security policies, such as firewalls and user authentication. Advantages of an Extranet approach are: (1) information that must be shared is done in a highly automated fashion with minimal human involvement; (2) partners can be directly involved in the design process to ensure compatibility; and (3) partners can access the most up-to-date information. Finally, the Internet provides general access to summarized information, including tables, graphs and maps. The Client/Server model is still applicable, but with a different set of applications available.

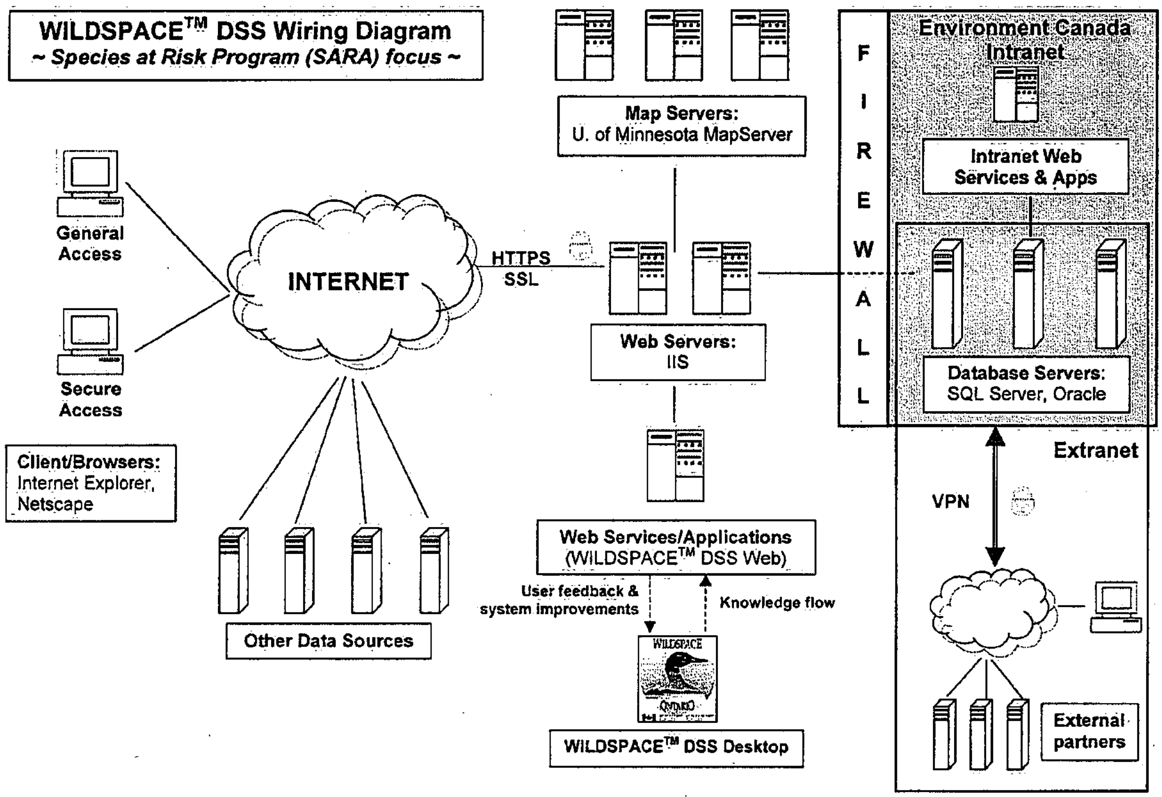

Figure 1. WILDSPACE™ DSS Wiring Diagram

2.2. Server Requirements

Figure 1 depicts a collection of servers required in the next-generation of the WILDSPACE DSS, comprised of web servers, database servers and map servers. These constitute the most common and successful implementations of the Client/Server network model. The web server provides the front-end or user interface to the WILDSPACE DSS in the form of web pages (using Hyper-text Markup Language - HTML), and is the primary server that users interact with, via their web browsers. The web server may make use of the services provided by the other servers to fulfill user requests. The database server provides data services to clients. The application client sends requests using Structured Query Language (SQL) to the database server. The server receives the SQL statements, validates them, executes them, and sends any results back to the client. The WILDSPACE DSS, using a distributed relational database server approach, has advantages over the more traditional local database approach. Administering and control of databases are made easier because these can be accomplished through a single user interface (the WILDSPACE DSS web site), which can be used from any location with an Internet connection. Another advantage is the ease of deployment of new database interfaces. The availability of web browsers across all platforms minimizes time required by application developers to implement graphical user interfaces across multiple operating systems. The map server, whose purpose is to publish maps on the web, is an integral component of the web-based WILDSPACE DSS. In particular, the WILDSPACE DSS will utilize both Web Map Service (WMS) and Web Feature Service (WFS), which are standards of the Open GIS Consortium (OGC 2001 and 2002). WMS provides services for generating visual representations (i.e., maps) from distributed heterogeneous geo-spatial data based on parameters such as the map layers to display, the bounding box of the viewing region and the map projection. The maps are usually returned in the form of a bitmap image. WFS allows clients to query and retrieve the actual geospatial data (features) from distributed heterogeneous data sources. Extensible Markup Language (XML) is used as the communication format; more specifically, geospatial data can be encoded in Geography Markup Language (GML), which facilitates editing of geographic feature geometry. Both WMS and WFS will be used prominently in the WILDSPACE DSS.

The reality of a Web-based system is that the Web is extremely vulnerable to a wide variety of compromises. As the WILDSPACE DSS is required to handle sensitive data, it will be necessary to use trusted security mechanisms. An increasingly popular solution is to implement security as a protocol that augments the underlying transport protocol (e.g., TCP). Security may be added to WILDSPACE DSS (or any web application) using Secure Sockets Layer (SSL), which is supported by most web browsers and servers. In the WILDSPACE DSS, SSL is part of the underlying protocol suite and therefore will be transparent to all applications. The SSL protocol starts with an initial "handshake" process whereby the client and server authenticate each other, and then establish an encryption algorithm and cryptographic keys to secure the connection. Afterwards, SSL encrypts all data being transmitted by either party. In addition, there must also be security at the application level to authenticate users logging into the WILDSPACE DSS.

2.3. Mobile Workers

Recovery team workers collect data in the field, often in remote locations, and are mobile by nature. Until recently, this has been accomplished using paper data sheets, or scannable forms. With Personal Digital Assistants (PDAs), it is now possible for mobile workers to enter data directly in electronic form. Ideally, the data should be updated immediately into the backend database using wireless communication links. However, current technologies are still not advanced enough and the wireless telecommunication networks do not cover all study areas. Although it is now possible to submit field data to a server in real time, it is still prohibitively expensive. A more economical approach is to collect and store data on a mobile computer, e.g., a personal digital assistant (PDA), and then upload it to the database server upon return to the office. When the cost of wireless networking declines, it may be feasible to deploy a real-time wireless network. Recent advancements in the wireless standards such as the Wi-Fi (IEEE 802.11a, b and g), WiMAX (IEEE 802.16a) and mobile wireless standards (IEEE 802.16e and 802.20) hold promise for the future.

The WILDSPACE DSS system will likely utilize third party middleware for the collection of field data by mobile workers. The basic purpose of the middleware is to rapidly develop and deploy customized data entry interfaces, specific to the requirements of each recovery team, without having to customize the code for each interface. WILDSPACE DSS requires different flavours of these data entry forms as a result of the amount and variety of data being collected by the Species at Risk recovery teams, ranging from tables describing and quantifying habitat characteristics, to tables characterizing reproductive status and productivity. The use of third party middleware is a cost-effective means of producing the diverse input interfaces required.

3. Database Design

Understanding the recovery needs and sensitivities of Species at Risk implies a need to understand various dimensions of their ecological niches (*sensu* Hutchinson 1957) and the biological and environmental factors that constrain populations (O'Connor 2000). Elucidation of such species-environment relationships however is known to be contingent on matching the spatial scales at which constraint variables are quantified, to spatial scales at which they influence the species (Huston 2000). Flexibility to match the spatial scale of predictor and response variables (environmental factors and species parameters, respectively) is facilitated in WILDSPACE DSS using a multi-scale spatial framework of digital map layers, which are used to join environmental attributes to occurrence data and habitat features. The resident framework comprises continental-, regional, and site-level digital map layers such as ecological land classifications, watersheds, landcover, bioclimatic envelopes, road networks and watercourses, and can be augmented by user-supplied data.

Mapping of Species at Risk occurrences and associated habitat is facilitated within the spatial framework by an integrated data management model that stores geo-referenced data pertaining to the life history of target species and quantitative descriptors of their habitat within the WILDSPACE DSS. This architecture provides swift access to spatial, temporal and biological data on species and their habitats for decision support and scenario gaming relative to recovery planning and emergency response. These components collectively provide a robust profile against which environmental stressors and threats can be assessed (Figure 2).

In the province of Ontario, Canada, there are 161 taxa (taxonomic categories or groups) legally listed as Extirpated, Endangered, Threatened or Special Concern (COSEWIC 2003). Since species are by definition biologically unique, efficient organization of species-specific information requires careful consideration. Tracking a variety of taxa with different characteristics through space and time in a quantitative framework requires deviations from straightforward approaches to database design. Complexity within data tables arises because of the need for unique field properties for many of the parameters which precludes use of a typical third (3rd) normal design (Date 1981). In addition, composite keys must be adopted in most tables because records are not identified uniquely on the basis of a single field. Such is the case in spatial and temporal data tables whose records are unique solely by combinations of species, sites and dates of observation. The challenge of storing data on many species is compounded by the fact that each species record is linked to sites (habitats) that are expected to be evaluated for species-specific suitability on a regular basis.

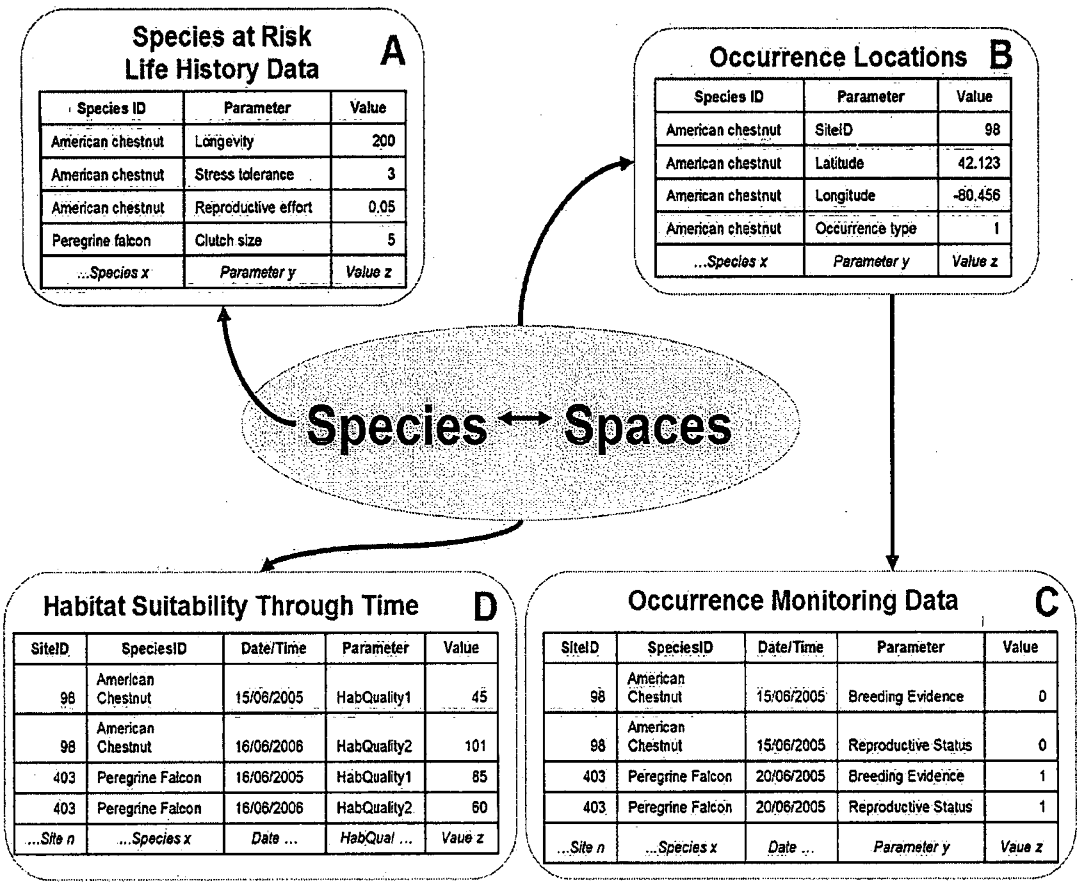

Figure 2. Components of the Species at Risk Database. Species and Spaces form the conceptual foundation for the data management system. Life history data tables provide biological data on the species (A). Species are placed in a spatial context by data tables that log the annual locations of occurrences (B). Occupied spaces are tracked within seasons to provide a temporal dimension to the status of occurrences (C). Similarly, the locations and characteristics of habitat patches for each species are tracked among years to monitor habitat supply in the landscape (D).

The multi-dimensionality of data on Species at Risk precludes the use of a non-relational model (flat file approach to data management), since such an approach produces excessively sparse tables and redundant data, which is inconsistent with best practices of data management (Fore 2003). The species-specific nature of life history, occurrence and habitat data precludes the use of a strict 3rd normal relational structure because unique tables for each aspect of each species biology are needed (potential of 483 data tables for 161 taxa). Due to the unwieldy nature of a database with this level of complexity, a more simplified approach to database design was pursued.

The list of legally listed Species at Risk comprises 10 taxonomic groups that share gross aspects of biological form and function (lichens, molluscs, invertebrates, vertebrates etc.). Species-level information are aggregated into tables at lower levels of differentiation as a means of simplifying the data structures. This lower level of organization corresponded to the "Class" and/or "Phylum" level in the taxonomic hierarchy. Adopting this organizational structure reduces the number of potential tables by more than 90%. Species-specific data fields are retained if they describe unique aspects of a species' biology deemed useful for risk and sensitivity assessment. As a result, biological tables have some fields that are not applicable to all species, but the sparseness of the data tables is not expected to be excessive (i.e., likely < 20%).

Similarly, a single habitat table with generic field names was designed to house information on all habitat sites, irrespective of the taxa. Records are indexed on species, site and date as a means of tracking species-specific habitat features through time in a single table. Fields in this generic table are dynamically customized based on the aspects of habitat quality relevant to each species based on published information and expert opinion. A secondary variable name table was designed to manage the identities of habitat variables on a species-by-species basis.

This organizational structure led to the creation of 31 tables that collectively capture biological, temporal and spatial dimensions of the available knowledge on Species at Risk and their habitats in Ontario and beyond.

With these database structures in place, a user interface was designed to provide efficient, reliable, and intuitive access to Species at Risk information. The data entry system consists of a series of steps that correspond to referential integrity rules among tables in the database. Species at Risk data records have dependencies on several tables that must be populated in a particular sequence in order for records to be accepted. For example, the data entry process guides the user through Sites and Observers tables to explicitly cross-reference SAR observations with a geographic location and the name of the observer for the purpose of accountability and quality assurance (Figure 3). Manual data entry is characterized primarily by a lookup framework in which manual character input is minimized but governed by validation rules and input masks where text entry is necessary. In the event that raw data are already in electronic format, batch mode record imports are also facilitated.

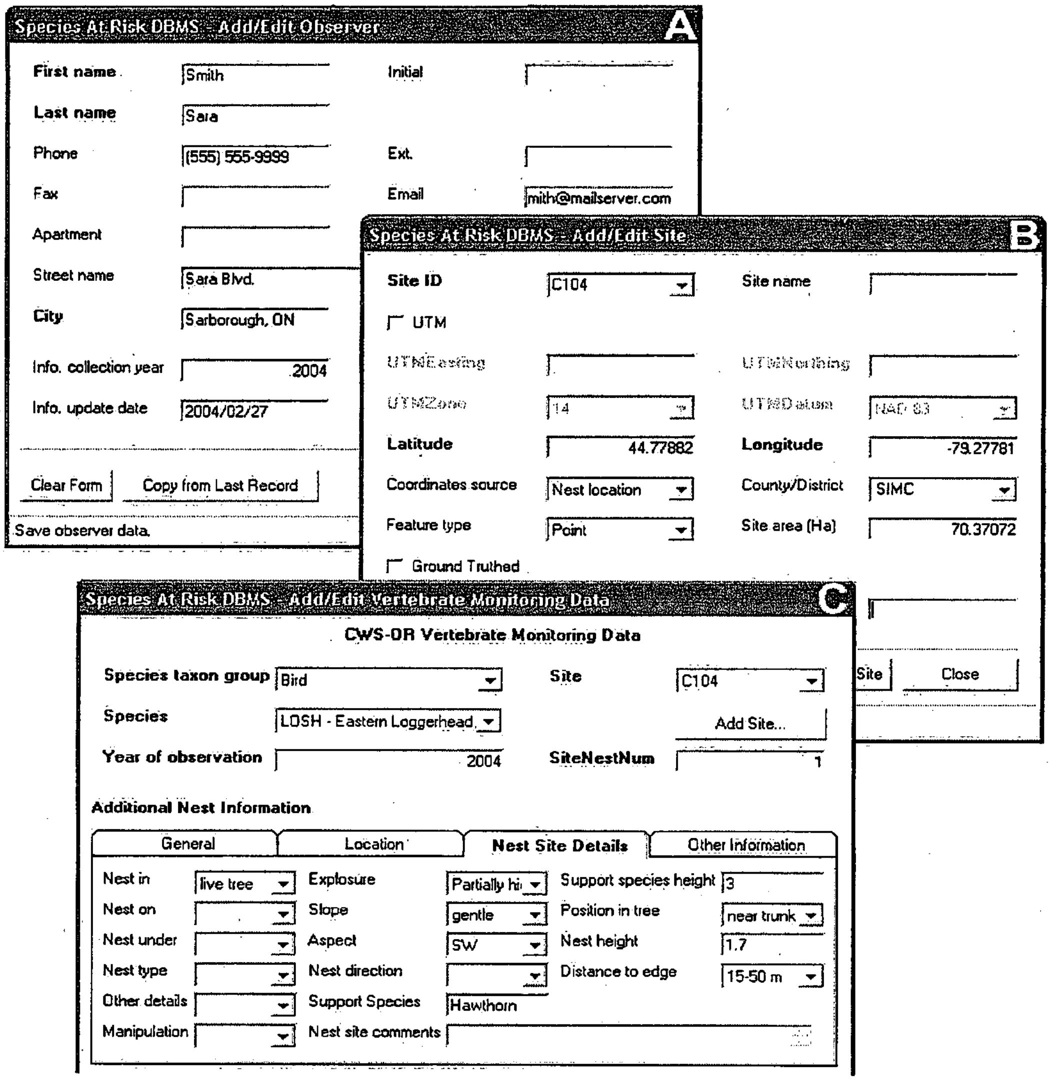

Figure 3. The Species at Risk Data Entry and Retrieval Interface. Addition of records begins by entering an observer profile to assign accountability (A). Next, details on geographic location are recorded (B), to georeference data to a known site. With forms A and B in place, the user completes the data entry process by specifying the biological parameters and selecting appropriate values for required fields from lookup tables (C).

4. Application Design

The current generation of the WILDSPACE DSS facilitates integration of diverse information, ranging from species life history data, to species productivity data, to map layers of species distribution and habitat. The application was designed to provide opportunities to query a variety of databases, to visualize spatial and/or temporal patterns, and to analyze data using standard and customized tools (Wong et al. 2003). Inherent to the WILDSPACE DSS architecture are relational databases, with common design structures for seamless integration, and a hierarchical spatial framework, for multi-scale analysis; this architecture facilitates multi-scale risk assessment using integrated data. The current-generation application also includes a rich portfolio of "Knowledge Templates" (KTs), which perform well-defined analytical procedures. The use of KTs further facilitates an integrated approach to exploring problems spanning multiple databases having temporal and spatial differences.

The next generation of the WILDSPACE DSS builds on this architecture with the addition of a "Scenario Gaming" module. This functionality facilitates hypothetical risk assessment analyses and predictive modeling, and has features tailored to support several diverse Environment Canada programs, including Species at Risk and Acid Rain Biomonitoring, among others. The use of scenario gaming will allow biologists and decision-makers to explore potential response to hypothetical situations. For example, "what if" several habitat patches of an endangered species are damaged by a natural or man-made disaster – how might the individuals respond to this abrupt change? The use of scenario gaming is ideal for exploring these types of hypothetical questions; answers to these questions can be used by decision-makers to take a pro-active approach to planning management actions. In addition, the scenario gaming feature could suggest solutions to improve current conditions, or to predict what future conditions might become, under different assumptions.

Consider the occurrence of an environmental threat – how does the design of the WILDSPACE DSS facilitate risk assessment, especially in the context of a Species at Risk? Generic steps required to carry out a risk assessment are illustrated in Figure 4. The initial step would involve the import of a map layer delineating the affected area into the spatial query module of the WILDSPACE DSS, 'SPACES' (Fig. 4; Step 1.). Spatial data on a variety of wildlife characteristics (range, distribution, occurrence, productivity, breeding, staging, dispersal, etc.) may then be loaded and compared for proximity with the threat (Fig. 4; Step 2.). The goal of this step is to identify the species, or group of species, that will potentially be at risk from the threat (several species may be shown to occur in the area). Sub-sets of survey sites or other locations where species have occurred, either historically or recently, can then be identified and tallied for use at subsequent steps. The variety of locational data queried is limited by the number of database server and map server connections available to the application; with the current-generation of the WILDSPACE DSS, only those species databases/map layers available internally on the desktop workstation or LAN-based server can be queried. Species' occurrence databases/map layers managed by partner agencies, or other sources, may be queried using the next-generation of the WILDSPACE DSS, via either an Extranet or Internet client, respectively. Having identified a collection of locations potentially at risk from the threat, further evidence must be gathered, in order to assess which, if any, species are likely to be in the vicinity of the threat (spatial evidence) at the time of the threat (temporal evidence). Step 3 (Fig. 4) illustrates a query to the module that holds the life history databases, 'SPECIES'; these databases include temporal characteristics of species including expected dates of breeding, migration, and over-wintering. The date of the threat can thus be compared to identify those species at least potentially affected simply based on seasonality. Again, the variety of data sources available for querying depends on the application generation and network configuration. Results from querying the SPACES and SPECIES modules can then be fed into the 'Data Analysis and Visualization' module. Database tables can be selected from a variety of data sources (again depending on connections within and/or beyond the local workstation or server), and filtered for the sub-sets of locations and species, as identified in steps 2 and 3. Resulting records can then be exported from the WILDSPACE DSS (Fig. 4; Step 4.) for interpretation in external applications, or further analyzed or visualized using any of a variety of standard or customized tools available in the system. For example, local population trend analysis may be performed for any species identified at Step 3, using the locations identified at Step 2; results can then become an influential component of the risk assessment for the purpose of planning emergency response and/or management actions. For example, if populations of a threatened species that occur within the vicinity of the threat have experienced statistically significant declines in recent years, and if it was thought that the threat would exacerbate that decline, then the risk would be heightened and appropriate and effective decisions could be then be taken to minimize the ecological damage posed by that threat. Another typical example would be to visualize the 'hot spots' of occurrence of multiple species potentially at risk from the threat – using the filtered records to calculate a richness index, and followed by a prediction surface generated and visualized in the 'SPACES' module. The resulting output at Step 4 could thus be a contoured map layer for use by decision-makers to prioritize areas according to how many species would be at risk from the threat.

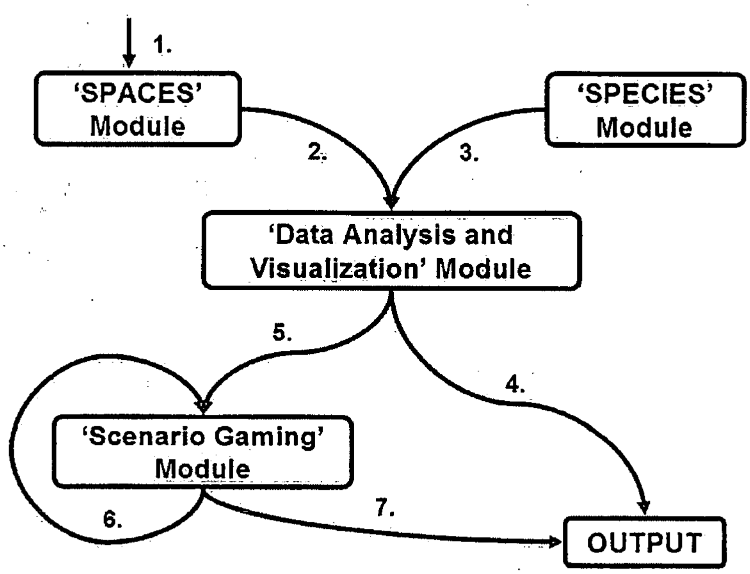

Figure 4. Steps to assess risk from threat using WILDSPACE DSS. Steps 1 through 4 can be accomplished using the current-generation WILDSPACE DSS; the next-generation application adds steps 5 through 7.

The formulation of meaningful scenarios requires the acquisition and synthesis of knowledge about situations and variables that play a major role in "what-if" questions. One of the strengths of the WILDSPACE DSS is its capacity for data mining and synthesis (Fig. 4; Steps 5-7), which facilitates the elucidation of situations and variables that influence both the scenario building process, and the type of modelappropriate in the gaming scenario. Modeling exercises may thus include qualitative and quantitative aspects, and stochastic, re-sampling, or statistical (e.g., Monte Carlo) approaches, in order to build and run the scenarios. Resulting output at Step 7 (Fig. 4) will thus illustrate empirically-based hypothetical situations, for consideration by biologists and decision-makers, during risk assessment and emergency response planning, and in evaluation and planning of long-term recovery actions.

5. Conclusion

The SARA serves as a catalyst and motivation for design of the next generation of the WILDSPACE DSS. The rationale behind the web-enabled approach is to serve multiple users who are dispersed in many locations. The network-wiring diagram (Figure 1) serves as a blue-print for the implementation of the next-generation WILDSPACE DSS. In order to serve the Intranet, Extranet and the Internet users, the Client/Server architecture is used, and the WILDSPACE DSS applications are written to communicate with the back-end database and map servers, and to provide information over the network via the Web server.

The database is an integral part of the WILDSPACE DSS. It stores the basic information and serves a portal to knowledge discovery. Through the best practise of database design and proper network configuration, the SAR data is protected against redundancy, inconsistency, inadvertent sharing, through enforcement of data structure standards, security, integrity, and balance between data storage and application efficiency.

The application design demonstrates the integrated philosophy of WILDSPACE DSS components. The data component, the map component, and the knowledge template component (data mining) are seamlessly integrated, to provide meaningful decision support. A new modeling component is being added to provide scenario gaming support, to answer hypothetical questions in the decision-making process for protecting species at risk, given the occurrence of natural or man-made environmental threats.

6. Acknowledgments

We thank the Species at Risk Program of Environment Canada (Ontario Region-OR) for supporting Project WILDSPACE™, especially S. Llewellyn, R. Pratt, M. Austen, R. Wenting, J. Robinson, L. Maynard, M. Cadman, L. Friesen, P. Ashley and J. Haggeman. We are grateful to A. Darwin, D. Fillman, K. Miller, and J. Neysmith for their contributions. Special thanks to D.C.L. Lam, G. Babish, B. Collins, W.G. Booty, K. Brown, R.C. McCrimmon, A. Storey and L. Robson for their invaluable input. We acknowledge contributions by ES Aquatic Inc. (D. Swayne), DM Solutions Group Inc. (D. McIlhagga), FieldWorker Products Ltd. (P. Neve), Dell Canada Inc. (I. D'Emilio) and ESRI Canada Inc. (M. Ballard) to WILDSPACE Research and

Development. We also thank T. Simmons and D. Wintjes (MSC-OR), L. Dunn, J. Pollock, I. Parsons, B. Hill and S. Painter (ECB-OR), S. Clement (EPB-OR), H. Auld and D. MacIver (MSC), and B. Young (NatureServe) for their continued support.

7. References

COSEWIC. 2003. COSEWIC Assessment Results, November 2003. Committee on the Status of Endangered Wildlife in Canada. 44 pp.

Date, C.J. 1981. An Introduction to Database Systems. Reading, Massachusetts, Addison-Wesley

Fore, L. S. 2003. Developing Biological Indicators: Lessons Learned from Mid-Atlantic Streams. Report prepared for EPA under Contract No. 50-CMAA-900065. EPA 903/R-003/003. U.S. Environmental Protection Agency, Office of Environmental Information and Mid-Atlantic Integrated Assessment Program, Region 3, Ft. Meade, MD. http://www.epa.gov/bioindicators/

Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22: 415-427.

Huston, M.A. 2002. Introductory essay: critical issues for improving predictions. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W. A. and Samson, F.B. (eds) Predicting Species Occurrences: Issues of accuracy and scale. Island Press, Washington, pp 7-21.

Manley, P.N., Zielinski, W.J., Schlesinger, M.D., Mori, S.R. 2004. Evaluation of a mulitple-species approach to monitoring species at the ecoregional scale. Ecological Applications 14(1): 296-310.

O'Connor, R. J. 2002. The conceptual basis of species distribution modeling: time for a paradigm shift? In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W. A. and Samson, F.B. (eds) Predicting Species Occurrences: Issues of accuracy and scale. Island Press, Washington, pp. 25-33.

OGC 2001. Web Map Service Implementation Specification. In: <http://www.opengis.org/docs/01-068r2.pdf>, 2004/03/02

OGC 2002. Web Feature Service Implementation Specification. In: <http://www.opengis.org/docs/02-058.pdf>, 2004/03/02

Reagan, P. 2000. Client/Server Network: Design, Operation and Management. Upper Saddle River, NJ: Prentice Hall.

Species at Risk Act. 2002. Bill C-5 An Act respecting the protection of wildlife species at risk in Canada. 1st Session, 37th Parliament, House of Commons of Canada. Ottawa, Canada. In:

<http://www.parl.gc.ca/37/1/parlbus/chambus/house/bills/government/C-5/C-5_3/C-5_cover-E.html>, 2004/03/02

Wong, I.W., McNicol, D.K., Fong, P., Fillman, D., Neysmith, J., Russell, R. 2003. The WILDSPACE™ Decision Support System. Environmental Modelling & Software 18 (2003): 521-530.

A General Information

Branch: WQM

Date Submitted: Mar. 10, 2004

Senior NWRI Author: Luis Leon

Project: Integrated Water Information Management and Modelling

Additional NWRI Authors: William Booty

Authors: L.F. Leon, G.S. Bowen and W.G. Booty

Title: Verification of the AGNPS Trace Source Contribution Algorithm for use in Southern Ontario Watersheds

B Scientific Manuscripts

To be Submitted to:

☒ **Journal** Name Environmental Software and Modelling

☐ **Book or Chapter**

Book Title

Editor(s)

Publishers City, Country

☐ **Conference/Workshop Proceedings**

Full Paper ☒ - Please indicate above if Proceedings will be in Journal or Book

Conference/Workshop Name: International Environmental Modelling and Software Society Conference

Location: Osnabruck, Germany Date June 14-17, 2004

☐ **NWRI Printed Contribution**

\# of copies Author NWRI Library Others (names)

Research/activity in support of program or Environment Canada regions (specify) Earth Observations Applications Development Program (EOADP) and Government Related Initiatives Program (GRIP)

or General ☐

Business Lines: Clean Environment ☒; Nature ☐; Weather and Environmental Predictions ☐.

Note: Author to submit Abstract AND Management Perspective with Manuscript

C Other Types of Material

☐ **Branch Technical Note** (e.g., QA/QC report to clients, other technical reports for VNR clients)

☐ **Conference/Workshop Abstract**

Short Abstract ☐ Extended Abstract (min. 2 pages) ☐

Oral ☐ Poster ☐ Presenter's Name:

Conference/Workshop Name:

Location: Date:

☐ **NWRI Web Page** or ☐ **Other Web Page**

Page Name

Site address http:\\

☐ **Other** (letter to editor, newspaper article, etc.)

Please specify type:

Title of journal, newspaper:

Additional Information:

D Review/Approval

Peer Reviewed by David Lam Date March, 2004

Approved by Project Chief *David Lam* Date *March 11, 2004*

Approved by Branch Director *J. Lawrence* Date *17 Mar '04*

Is a Briefing Note required (Branch Director to decide): ☐ Yes ☐ No

Does anyone else within the department need to be alerted? ☐ Yes ☐ No Names:

Verification of the AGNPS Trace Source Contribution Algorithm for Use in Southern Ontario Watersheds

L.F. León[a], G.S. Bowen[b] and W.G. Booty[c]

Abstract: The AGricultural Non-Point Source model (AGNPS), developed in the late 1980s by the US Department of Agriculture, is widely used around the globe as a decision support tool to evaluate opportunities for improved delivery of water quality best management programs (BMPs) in rural watersheds. This model has proven through numerous calibration studies to provide reasonable results. One component of the AGNPS model that has not been verified is the trace source contribution algorithm. This tool allows identification of priority grids in terms of their sediment and nutrient loads. In southern Ontario farming operation is often eligible for financial grant assistance to encourage implementation of BMPs. Usually these grant program are offered to all farm operation in the impaired watershed. Shrinking government funding for incentive programs and result based performance audits suggest that we could use the trace source contribution routine in the AGNPS model to identify priority grids where we would then provide, on a priority basis, BMP assistance grants. Before initiating this new funding program we needed to verify that the AGNPS model was identifying priority grids for sediment and nutrient loads. We have developed an approach that has combined standard field assessments with the interpretation of oblique low-level aerial photography. This methodology has been successfully applied on a subwatershed and watershed scale and has shown the AGNPS model did correctly identify the appropriate grids for targeting BMP programs. A secondary benefit of our study was the opportunity to verify model inputs established through GIS extraction with actual on the ground soil and land use conditions. This has lead to improvements in the GIS interface we have developed to link with the AGNPS model. The upgrade consisted mainly on porting the application to a 32bit platform and rewriting the extraction algorithms to make them more robust and accurate.

Vérification de l'algorithme d'établissement de la contribution des sources à l'état de traces du modèle AGNPS dans les bassins hydrographiques du sud de l'Ontario

L.F. León[a], G.S. Bowen[b] et W.G. Booty[c]

Résumé : Le modèle américain « Agricultural Non-Point Source » (AGNPS) (sources non ponctuelles agricoles), élaboré à la fin des années 1980 par le US Department of Agriculture, est largement utilisé dans le monde comme outil d'aide à la décision pour évaluer les moyens d'améliorer l'application des programmes de meilleures pratiques de gestion en matière de qualité de l'eau dans les bassins hydrographiques ruraux. De nombreuses études d'étalonnage ont permis d'assurer que ce modèle donne des résultats fiables. L'un des éléments du modèle AGNPS qui n'a pas été vérifié est l'algorithme d'établissement de la contribution des sources à l'état de traces. Cet outil permet d'établir des grilles de priorité en fonction des charges de sédiments et de nutriments. Dans le sud de l'Ontario, les exploitations agricoles sont souvent admissibles à une aide financière pour encourager l'implantation des meilleures pratiques de gestion. Ce programme d'aide est habituellement offert à toutes les exploitations agricoles d'un bassin hydrographique touché. La réduction des budgets du gouvernement consacrés à ces programmes d'encouragement et les audits de rendement basés sur les résultats nous incitent à envisager l'utilisation de l'algorithme d'établissement de la contribution des sources à l'état de traces du modèle AGNPS pour établir des grilles de priorité dans la prestation d'une aide financière pour les meilleures pratiques de gestion. Avant d'amorcer ce nouveau programme de financement, il fallait vérifier que le modèle AGNPS permet d'établir des grilles de priorité en fonction des charges de sédiments et de nutriments. Nous avons mis au point une méthode associant une évaluation classique sur le terrain et l'interprétation de photographies aériennes obliques prises à basse altitude. Cette méthode a été appliquée avec succès à l'échelle des sous-bassins et des bassins hydrographiques. Elle indique que le modèle AGNPS permet de déterminer correctement des grilles de priorité utilisables dans le cadre d'un programme de financement des meilleures pratiques de gestion. Notre étude a également permis de comparer les données du modèle obtenues à l'aide des SIG avec les valeurs réelles au sol et dans des conditions d'utilisation des terres. Nous avons pu ainsi améliorer l'interface SIG que nous avons mise au point pour le modèle AGNPS. L'amélioration consistait principalement à adapter l'application à une plateforme 32 bits et à réécrire les algorithmes d'extraction pour les rendre plus robustes et précis.

NWRI RESEARCH SUMMARY

Plain language title

Verification of the AGNPS Trace Source Contribution Algorithm for use in Southern Ontario Watersheds

What is the problem and what do sicentists already know about it?

The AGricultural Non-Point Source model (AGNPS) was developed in the late 1980's by the U.S. Department of Agriculture as a tool to evaluate rural watershed management plans. The model has been used around the globe and has been proven through numerous calibration studies. However, one component of the model that has not been verified is the trace source contribution algorithm that allows identification of priority watershed grid areas in terms of their contributions to sediment and nutrient loads. This tool is very important for Ontario watershed management programs such as those related to COA programs and Great Lakes Action Plans. Shrinking government funding for incentive programs and result-based performance audits makes the use of the source trace function very useful for optimizing best management practices in areas of the watersheds identified as major sources of sediment and nutrient loads.

Why did NWRI do this study?

The National Water Research Institute, in collaboration with the Great Lakes Sustainability Fund, Ontario Ministry of Environment and the Toronto and Region Conservation Authority, have been involved in the development and application of the AGNPS/RAISON decision support system for several years. This study is a continuation of this collaboration to provide technology support for the implementation of watershed management plans in support of the GL 2020, COA and RAP programs to reduce loads of contaminants to the Great Lakes.

What were the results?

An approach was developed to evaluate the trace source contribution function within the AGNPS model that combines standard field assesments with the interpretation of oblique low-level aerial photography. This methodology was successfully applied on a subwatershed and watershed scale and has shown that the AGNPS model correctly identified the most appropriate watershed areas for targeted best management practice (BMP) programs.

How will these results be used?

The technology developed during this study is now being used by the Toronto and Region Conservation Authority to direct loading reduction programs to priority areas within their watersheds. This technology has also been passed on to other Conservation Authorities within Ontario for similar applications.

Who were our main partners in the study?

The main partners in the study are NWRI, MOE, GLSF and TRCA.

Sommaire des recherches de l'INRE

Titre en langage clair

Vérification de l'algorithme d'établissement de la contribution des sources à l'état de traces du modèle AGNPS dans les bassins hydrographiques du sud de l'Ontario.

Quel est le problème et que savent les chercheurs à ce sujet?

Le modèle américain « Agricultural Non-Point Source » (AGNPS) a été mis au point par le U.S. Department of Agriculture, à la fin des années 1980. Le modèle a été utilisé partout dans le monde, et de nombreuses études d'étalonnage ont prouvé son efficacité. Toutefois, l'un des éléments de ce modèle n'a pas encore été vérifié; il s'agit de l'algorithme d'établissement de la contribution des sources à l'état de traces qui permet d'établir des grilles de priorité en fonction des charges de sédiments et de nutriments. Cet outil est très important pour les programmes de gestion des bassins hydrographiques de l'Ontario, tels les programmes de l'Accord Canada-Ontario et les Plans d'action des Grands Lacs. La réduction des budgets du gouvernement consacrés aux programmes d'encouragement et les audits de rendement fondés sur les résultats rendent la fonction de recherche des sources très utile pour optimiser les pratiques de gestion dans les régions des bassins hydrographiques qui contribuent le plus à la charge de sédiments et de nutriments.

Pourquoi l'INRE a-t-il effectué cette étude?

L'Institut national de recherche sur les eaux, en collaboration avec le Fonds de durabilité des Grands Lacs, le ministère de l'Environnement de l'Ontario et l'Office de protection de la nature de Toronto et de la région ont participé à l'élaboration et à l'application du système d'aide à la décision AGNPS/RAISON pendant plusieurs années. La présente étude poursuit cette collaboration à la recherche d'une technologie pour la mise en œuvre de plans de gestion des bassins hydrographiques à l'appui de GL 2020, de l'Accord Canada-Ontario et des programmes du PA dans le but de réduire les charges de contaminants dans les Grands Lacs.

Quels sont les résultats?

Nous avons mis au point une méthode d'évaluation de la fonction d'établissement de la contribution des sources à l'état de traces associant une évaluation classique sur le terrain et l'interprétation de photographies aériennes obliques prises à basse altitude. Cette méthode a été appliquée avec succès à l'échelle des sous-bassins et des bassins hydrographiques. Elle indique que le modèle AGNPS permet de déterminer correctement les régions des bassins hydrographiques où il est le plus urgent d'intervenir dans le cadre d'un programme de financement des meilleures pratiques de gestion.

Comment ces résultats seront-ils utilisés?

La technologie mise au point au cours de la présente étude est maintenant utilisée par l'Office de protection de la nature de Toronto et de la région pour orienter les programmes de réduction des charges vers les régions prioritaires des bassins hydrographiques. Cette technologie a également été transmise à d'autres organismes de protection de l'environnement en Ontario qui l'utilisent à des fins similaires.

Quels étaient nos principaux partenaires dans cette étude?

Les principaux partenaires de l'étude sont l'INRE, le ministère de l'Environnement de l'Ontario, le Fonds de durabilité des Grands Lacs et l'Office de protection de la nature de Toronto et de la région.

Verification of the AGNPS Trace Source Contribution Algorithm for Use in Southern Ontario Watersheds

L.F. León[a], G.S. Bowen[b] and W.G. Booty[c]

[a] University of Waterloo, Waterloo, Ontario, N2G3L1 (lfleonvi@uwaterloo.ca)
[b] Toronto Region Conservation Authority, Downsview Ontario, M3N1S4
[c] Environment Canada, National Water Research Institute, Burlington, Ontario, Canada.

Abstract: The AGricultural Non-Point Source model (AGNPS), developed in the late 1980s by the US Department of Agriculture, is widely used around the globe as a decision support tool to evaluate opportunities for improved delivery of water quality best management programs (BMPs) in rural watersheds. This model has proven through numerous calibration studies to provide reasonable results. One component of the AGNPS model that has not been verified is the trace source contribution algorithm. This tool allows identification of priority grids in terms of their sediment and nutrient loads. In southern Ontario farming operation is often eligible for financial grant assistance to encourage implementation of BMPs. Usually these grant program are offered to all farm operation in the impaired watershed. Shrinking government funding for incentive programs and result based performance audits suggest that we could use the trace source contribution routine in the AGNPS model to identify priority grids where we would then provide, on a priority basis, BMP assistance grants. Before initiating this new funding program we needed to verify that the AGNPS model was identifying priority grids for sediment and nutrient loads. We have developed an approach that has combined standard field assessments with the interpretation of oblique low-level aerial photography. This methodology has been successfully applied on a subwatershed and watershed scale and has shown the AGNPS model did correctly identify the appropriate grids for targeting BMP programs. A secondary benefit of our study was the opportunity to verify model inputs established through GIS extraction with actual on the ground soil and land use conditions. This has lead to improvements in the GIS interface we have developed to link with the AGNPS model. The upgrade consisted mainly on porting the application to a 32bit platform and rewriting the extraction algorithms to make them more robust and accurate.

Keywords: AGNPS; Model verification; GIS interface; Field study

1. INTRODUCTION

Rural Non point sources of pollution are recognized as a major source of pollution in the Great Lakes Basin and are a key focus of an International Water Quality Agreement between Canada and the United States [1972]. Following the Walkerton Drinking Water Tragedy in 2003 in which seven people died when agricultural runoff contaminated a municipal water supply, a renewed focus has been placed on drinking water source protection in Ontario [MAG 2004, MOE 2004]. While significant progress has made in developing agricultural best management practices (BMPs), there is a recognized need to target the application of BMPs in order to make the best use of available resources.

The main objective of our research was to gain an understanding of rural non point sources of pollution and to design more effective management strategies. The Agricultural Non Point Source Model (AGNPS) developed by the US Department of Agriculture in the late 1980s was determined to be the best modeling tool for our studies. This model is widely used around the globe as a decision support tool for planning agricultural BMPs. For the past eight years, we have been collaborating on applications of this model to watershed studies in southern Ontario. This paper describes our efforts to verify the application of the AGNPS model's trace source contribution algorithm in the Duffins Creek watershed and resulting improvements in a GIS interface developed for the model [Leon et al 2000].

Incentive programs were developed to encourage farming operations in Ontario to implement BMPs. For example, the Toronto and Region Conservation Authority (TRCA) rural water quality program provides grants to eligible farmers for water quality improvement projects, such as cattle fencing, buffer strips, no till cultivation and manure management practices across our jurisdiction [TRCA 2004].

Grant assistance programs offered by the TRCA are not targeted to specific locales within a watershed. Scarce funding and a movement towards result based management programs, suggest that it would be beneficial to target available resources at specific locales within a watershed that are known to be significant sources of pollution.

Identification of priority sources of NPS pollution within a watershed are normally undertaken by tributary monitoring programs. These programs are expensive to operate and can consume a large proportion of available water quality improvement budgets. In many cases the magnitude of water quality degradation is already understood and is often the rationale to undertake modeling studies. Through a previous study [León et al., 2004] we demonstrated that the AGNPS model provided a good understanding of water quality conditions in southern Ontario and that the model could be used to evaluate a variety of best management practices.

One of the potential advantages of distributed models like AGNPS is the ability to identify spatial factors contributing to water quality issues. Based upon an extensive literature search, we were not aware of any studies that have previously validated the AGNPS trace source contribution algorithm.

The main objective of this study was to determine whether the AGNPS model could accurately define-identify specific sources of pollution and to make the necessary improvements in the previous version of the interface developed to prepare input files for the model.

2. DUFFINS CREEK WATERSHED

The Duffins Creek watershed (283km2) is located approximately 20 km east of the City of Toronto on the north shoreline of Lake Ontario [TRCA 2002]. Over 50 % of the land in the watershed is in public ownership, as a result of expropriation by the Federal and Provincial governments in the 1970s for the purposes of building an International Airport and satellite urban community of 90,000. Only 7 % of the watershed is currently developed. After a 30-year hiatus, planning for a new airport and urban development are underway.

AGNPS modeling runs were used to develop effective surface water quality management actions for an integrated watershed plan developed by the Conservation Authority [TRCA 2003a]. The AGNPS model was employed in the watershed to evaluate water quality responses to changes in land use, identify priority sources of pollution and to investigate water quality implication of climate change [TRCA 2003b, Booty et. al. 2003].

3. MODEL DESCRIPTION

AGNPS is a distributed approach model designed to simulate agricultural watersheds for a storm event assuming uniform precipitation patterns [Young et al., 1989]. It simulates the sediment yield and the generation and transport of nutrients (nitrogen and phosphorus) in the watershed. The hydrologic component in the model is based in the Natural Resources Conservation Service method, known as the SCS curve number approach. In order to predict soil erosion, the universal soil loss equation (USLE) is used. Soil loss and sediment yield in the model is a two-step process. First, soil erosion is calculated and then compared to a sediment transport capacity of the flow. Finally the eroded sediment is then routed based on a steady-state continuity equation for sediment transport and deposition for each particle size described by Foster et al. [1980].

Among the factors in the USLE, the soil erodibility factor is a measure of potential erosion of the soil and is a function of the soil texture. A vegetative cover factor estimates the effect of ground cover conditions and accounts for the effect of vegetation and land management on erosion rates resulting from canopy protection (reduction of rainfall energy effect). The pollutant transport part of the model estimates transport of nitrogen, phosphorous and chemical oxygen demand (COD) throughout the watershed. It is divided into one part handling soluble pollutants and another part for sediment-based pollutants. The methods used to predict nitrogen and phosphorus yields were developed by Frere et al. [1980]. For nitrogen and phosphorus calculations, relationships between the chemical concentration, sediment yield and run-off volume are used. Soluble nitrogen and phosphorus in run-off waters represent the effects of rainfall, solid waste, fertilization and leaching from the soil in each cell. The contributions of soluble nutrients from each cell are calculated within the cell and then routed downstream.

Data needed for the model are classified into two categories: Watershed Data include information applying to the entire watershed, such as watershed size, number of cells, the storm type, duration and intensity. Cell Data includes information on the parameters based on soil type, land use, and management practices within the cell. An interface was developed for the AGNPS model using a decision support system with GIS algorithms that reduces time-consuming tasks by automatically extracting input data from digital maps of the watershed [León et al., 2000]. The interface tools provided easy data compiling for the model by using a digital elevation model (DEM), soil type, and land use maps in vector formats to extract the model input data.

4. MODEL SETUP AND RUNS

The main purpose of this study was to verify that the AGNPS model's trace source contribution sub routine accurately identified the priority sources of pollution in the watershed. Using the outlet cell for the watershed, to seed the AGNPS model's trace source routine, priority sources of sediment, nitrogen and dissolved phosphorus were identified for the watershed study (TRCA 2003b). For the purposes of this validation study we looked only at the priority grids identified by the AGNPS model as being key sources of sediment (Figure 1).

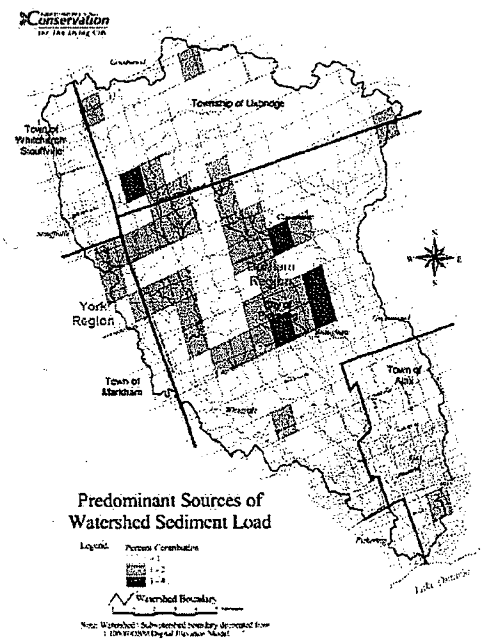

Figure 1. Cells identified as main sources

AGNPS setup for this study employed a 1x1km grid size. Two representative storm types were selected. The first was a 25mm storm with 12-hour duration. The second storm was a 15mm event spread over 9 hours. Both events used a US Soil Conservation Service, type II storm, which has a rainfall intensity and distribution typical for this area. Using the interface with the map extraction capabilities, the AGNPS model input parameters were automatically estimated for the base grid using 1:10,000 scale soils and land use mapping and a 100m resolution digital elevation model.

Fertilizer application rates and availability factors were based on a farm survey conducted for the Duffins Creek watershed in 1997 (JDE Ventures, 1998). This survey indicated that nitrogen and phosphorous applications rates varied with crop and soil types. As a result, fertilizer application rates and estimated availability factors were attached as attributes to a landuse layer to capture spatial distributions. In an earlier study, we calibrated the AGNPS model setup using the results from sampling 25 storm events over a three period study [Leon et. al. 2004].

The location of identified priority grids were overlaid on soils and land use mapping and on 30 m resolution color, orthographic photography. Site visits were then undertaken for 12 of the identified grids. Model input and outputs values for each of the priority grids and current agricultural practices were evaluated. Oblique photographs of the watershed (@500m elevation ASL) taken from a fixed wing aircraft were examined.

5. RESULTS AND IMPROVEMENTS

Characteristics of the 12 grids selected for field validation of sediment loads are summarized in Table 1. Field evaluations of these grids confirmed that significant sediments are being generated within all of the cells. However not all of the sediment generated within these cells is exiting and not all of these cells could as a result be considered to be significant sources of phosphorus. Four of the cells identified by the model however, were significant sources of sediment and phosphorus.

Table 1. Summary of cells for field validation

Cell	DA	Tributary	PSg	PSc	PPw
69	3409	West Duffins	Y	N	N
70	1136	Wixon	Y	N	N
81	1136	Major	Y	Y	Y
102	1847	Mitchell	Y	Y	Y
110	3977	Reesor	Y	Y	Y
111	5398	West Duffins	Y	Y	Y
115	142	Noatercourse	Y	N	N
125	142	Urfe	Y	N	N
126	426	Urfe	Y	N	N
128	142	Spring	Y	N	N
137	568	Urfe	Y	N	N
139	142	Brougham	Y	N	N

DA – Drainage area in hectares at the grid cell; PSg - Priority sediment generated within the cell; PSc – Priority sediment cell yield; PPw – Priority phosphorus in water.

To gain an understanding of why the observed conditions did not match model results, the input parameters (cell variables, land use and fertilizer availability) for the AGNPS were reviewed against conditions observed in the field (see Table 2). For each of the 12 grids, drainage conditions, cropping practices, soils and surrounding land use were evaluated, and specific locations were identified where BMP practices could be implemented. These locations were plotted on an ortho-photograph for follow up studies by TRCA's rural stewardship staff. Locales requiring BMPs were identified in all 12 grids.

Table 2. Results of field assessments of the model input data parameters.

Cell	Cell Variables					Land Cover, Fertilizer Use						
	Drain Area	Slope	K	C	Soil Texture	Urban	Forest	Field Crops	Hay Pasture	Wetland	Applied P	Avail. Fact.
69	Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok	Ok	Low	Ok	Ok
70	Ok	Ok	Ok	Ok	Ok	Ok	Ok	High	Err	Ok	Ok	Low
81	Ok	Ok	Ok	High	Ok	Ok	Ok	Low	Err	Ok	Ok	High
102	Ok	Low	Ok	Ok	Err	Ok	Ok	Ok	Err	Ok	Ok	Ok
110	Ok	Ok	Ok	High	Err	Ok	Ok	Err	Err	Ok	Ok	High
111	Ok	Ok	Ok	Ok	Err	Ok	Ok	Ok	Ok	Ok	Ok	Low
115	Ok	Ok	Ok	Ok	Err	Ok	Ok	Ok	Ok	Ok	High	Ok
125	Ok	Err	Ok	Low	Err	Ok	Ok	Ok	Ok	Ok	High	Ok
126	Ok	Ok	Ok	High	Err	Ok	Ok	Err	Err	Ok	Err	Err
128	Ok	Ok	Ok	Ok	Err	Ok	High	High	Err	Ok	Ok	Low
137	Ok	Ok	Ok	Ok	Err	Ok	Ok	Ok	Low	Ok	Ok	Ok
139	Ok	Ok	Ok	Low	Err	Ok	High	High	Err	Ok	High	Low

Table 3. Comparisson of the input parameters extracted with the upgraded version of the interface

Original 16 bit platform												
Variable	**10000**	**31000**	**35000**	**38000**	**39000**	**40000**	**41000**	**42000**	**43000**	**46000**	**47000**	**49000**
SCS_No	72	93	74	83	83	72	84	84	73	72	74	81
Mannings_n	0.31	0.053	0.198	0.174	0.156	0.245	0.134	0.134	0.22	0.308	0.187	0.225
K_Factor	0.328	0.354	0.31	0.271	0.26	0.309	0.298	0.249	0.278	0.3	0.231	0.263
C_Factor	0.3385	0.0098	0.4795	0.4929	0.5358	0.4209	0.555	0.563	0.4319	0.2513	0.4759	0.4224
Upgrade to a 32 bit Platform												
Variable	**10000**	**31000**	**35000**	**38000**	**39000**	**40000**	**41000**	**42000**	**43000**	**46000**	**47000**	**49000**
SCS_No	71	81	73	79	72	69	80	80	66	71	69	69
Mannings_n	0.29	0.057	0.174	0.199	0.167	0.272	0.117	0.122	0.202	0.344	0.162	0.263
K_Factor	0.365	0.372	0.308	0.284	0.286	0.314	0.288	0.257	0.311	0.333	0.201	0.297
C_Factor	0.3779	0.3587	0.5427	0.4843	0.5271	0.4046	0.58	0.5742	0.4835	0.2929	0.4111	0.4123
Differences in %	**10000**	**31000**	**35000**	**38000**	**39000**	**40000**	**41000**	**42000**	**43000**	**46000**	**47000**	**49000**
SCS_No	1.4%	12.9%	1.4%	4.8%	13.3%	4.2%	4.8%	4.8%	9.6%	1.4%	6.8%	14.8%
Mannings_n	6.5%	7.5%	12.1%	14.4%	7.1%	11.0%	12.7%	9.0%	8.2%	11.7%	13.4%	16.9%
K_Factor	11.3%	5.1%	0.6%	4.8%	10.0%	1.6%	3.4%	3.2%	11.9%	11.0%	13.0%	12.9%
C_Factor	11.6%	C<<	13.2%	1.7%	1.6%	3.9%	4.5%	2.0%	11.9%	16.6%	13.6%	2.4%

Parallel to this work, improvements were made to the Interface. It was ported to a 32bit platform from the initial 16bit application and re-designed with the main focus of being 100% backward compatible with its predecessor. Furthermore it was engineered to include the capabilities of the more powerful object oriented module ROS (Raison Object System) and integrated with its own MapViewer, based on a prototype from the NWRI (National Water Research Institute). The integrated system no longer requires launching two different applications as with the prior version and the model itself has being integrated in a way that additional manual installation of the AGNPS model is not needed.

The map data and DEM extract procedures were completely re-written so that fertilizer propagation and land use extraction allow for individual land class definitions. The extraction methods are much faster, more robust and accurate. Several tests were performed using previous running times with the 16bit system and it confidently reproduces the required data in less than a fifth of the time. No limitation is imposed in the 32bit version and the accuracy of the area calculations is down to 0.001 margin of error. One of the limitations of the original interface was imposed by the 16 bit platform on the number of polygons that the data extraction process can handle. To sort this out, the land classes in the raw map data were aggregated in groups (i.e. 6, 9, 10 classes).

During the field validation of the priority cells, a problem was identified when comparing land use percentages of areas. This was an artificial result of the landuse aggregation; the model was set up with the 6 class scheme option, so there was no room to handle map attributes such as hay/pasture, golf courses or grassed areas as individual land covers. With the improvements in the interface, this no longer posses a problem.

The new interface was used to re-calculate the model parameters for each land class in the original map. Table 3 compares, for the selected cells, the land use related variables. To evaluate the impact on the model results, the sensitivity analysis is quite useful, where it can be seen that K and C will have a lesser impact than that of precipitation or curve number for sediment yield results.

6. CONCLUSIONS

The AGNPS model's Trace Source Contribution (TSC) analysis algorithm has been shown to be a useful tool that can assist watershed managers by directing their attention to management efforts to specific areas. This capability allows water quality improvement efforts to be focused on design and implementation of BMPs. Previously significant proportions of available water quality management resources were spent identifying rural pollutant sources through monitoring or detailed agricultural surveys. Improvements to the model interface allow the manipulation of individual land classes. This, together with the more robust extraction methods for input data, increases the usability and flexibility of the model as a decision support tool.

7. ACKNOWLEDGEMENTS

We would like to thank Brian Howieson for assistance with the fieldwork, Tim Van Seters for his help with the watershed AGNPS modeling and Patricia Lowe for her assistance with the design of the field study. The Toronto Region Conservation Authority, Environment Canada and watershed municipalities provided funding for this research project.

REFERENCES

Canada and the United States 1972, Water Quality Agreement (amended 1978). www.epa.gov.glnpo/glwqa/1978/

Foster, G.R., Lane, L.J., Nowlin, J.D., Laflen, J.M., Young, R.A., 1980. Model to estimate sediment yield from field-sized areas: development of the model. In: CREAMS: A Field Scale Model for Chemicals, Run-off, and Erosion from Agricultural Management Systems. Research Report 26, USDA, pp. 36–63.

Frere, M.H., Ross, J.D., Lane, L.J., 1980. The nutrient submodel. In: CREAMS:AField Scale Model for Chemicals, Run-off, and Erosion from Agricultural Management Systems. Research Report 26, USDA, pp. 65–85.

JDE Ventures, 1998. 1998 Survey of Farming Practices and Fertilizer use in the Stouffville and Reesor Creeks Watershed for AGNPS Model Application, A final report to the Ontario Ministry of the Environment, Standards Development Branch and the Toronto and Region Conservation Authority, 15pp.

León, L.F., Lam, D.C., Swayne, D.A., Farquhar, G.J., Soulis, E.D., 2000. Integration of a nonpoint source pollution model with a decision support system. *Environ. Model. Software*, 15, 249–255.

León, L.F., Booty, W.G. , Bowen, G.S. and Lam, D.C.L., 2004. Validation of an agricultural non-point source model in a watershed in southern Ontario. *Agricultural Water Management*, Vol. 65 pp. 59–75.

Ministry Attorney General 2004., Walkerton Commission of Inquiry Reports www.attorneygeneral.jus.gov.on.ca/english/about/pubs/walkerton/

MOE 2004. White Paper on Watershed Based Source Drinking Water Protection www.ene.gov.on.ca/programs/3585e01.pdf

Toronto Region Conservation Authority 2002 Duffins Creek State of the Watershed Report. www.trca.on.ca

Toronto Region Conservation Authority 2004. Rural Clean Water Program. www.trca.on.ca

Toronto Region Conservation Authority 2003a. A Watershed Plan for Duffins Creek and Carruthers Creek Watersheds. A Report of the Duffins Creek and Carruthers Creek Task Forces. www.trca.on.ca

Toronto Region Conservation Authority 2003b. Agricultural Non Point Source (AGNPS) modeling of the Duffins Creek and Carruthers Creek Watersheds in Technical Reports In Support of the Duffins Creek and Carruthers Creek Watershed Plan 63 pgs.

Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P., 1989. Agricultural non-point source pollution model for evaluating agricultural watersheds. *J. Soil Water Conserv.* 44, 168–173.

Environment Canada Environnement Canada

Canada

Canada Centre for Inland Waters
P.O. Box 5050
867 Lakeshore Road
Burlington, Ontario
L7R 4A6 Canada

National Hydrology Research Centre
11 Innovation Boulevard
Saskatoon, Saskatchewan
S7N 3H5 Canada

St. Lawrence Centre
105 McGill Street
Montreal, Quebec
H2Y 2E7 Canada

Place Vincent Massey
351 St. Joseph Boulevard
Gatineau, Quebec
K1A 0H3 Canada

Centre canadien des eaux intérieures
Case postale 5050
867, chemin Lakeshore
Burlington (Ontario)
L7R 4A6 Canada

Centre national de recherche en hydrologie
11, boul. Innovation
Saskatoon (Saskatchewan)
S7N 3H5 Canada

Centre Saint-Laurent
105, rue McGill
Montréal (Québec)
H2Y 2E7 Canada

Place Vincent-Massey
351 boul. St-Joseph
Gatineau (Québec)
K1A 0H3 Canada