From Landscapes to Waterscapes:
A PSE for Landuse Change Analysis

R. Dymond[†], V. Lohani[‡], D. Kibler[†], D. Bosch[§],
E.J. Rubin[*], R. Dietz[†], J. Chanat[†], C. Speir[§],
C.A. Shaffer[*], N. Ramakrishnan[*], and L.T. Watson[*]
Virginia Polytechnic Institute and State University,
Blacksburg, VA 24061

Abstract

We describe the design and implementation of L2W — a problem solving environment (PSE) for landuse change analysis. L2W organizes and unifies the diverse collection of software typically associated with ecosystem models (hydrological, economic, and biological). It provides a web-based interface for potential watershed managers and other users to explore meaningful alternative land development and management scenarios and view their hydrological, ecological, and economic impacts. A prototype implementation for the Upper Roanoke River Watershed in Southwest Virginia, USA is described.

Keywords: Problem solving environments (PSEs), landuse change analysis, watershed management, multidisciplinary modeling, decision support systems, web-accessible software.

Corresponding Author:
Naren Ramakrishnan
Department of Computer Science
Virginia Tech, VA 24061
Tel: (540) 231-8451
Fax: (540) 231-6075
Email: naren@cs.vt.edu

---

[*] Department of Computer Science.

[†] Department of Civil and Environmental Engineering.

[‡] Department of Engineering Fundamentals.

[§] Department of Agricultural and Applied Economics.

1 Introduction

Watershed management is a broad concept whose scope entails 'the plans, policies, and activities used to control water and related resources and processes in a given watershed' [1]. Effective watershed management requires that decision-makers receive input about, and balance consideration of, a number of competing factors. The fundamental drivers of change are modifications to landuse and settlement patterns. These changes affect surface and ground waterflows, water quality, wildlife habitat, economic value of the land and infrastructure (directly due to the change itself such as building a housing development, and indirectly due to the effects of the change, such as increased flooding), and cause economic effects on municipalities (taxes raised versus services provided).

To model the effects of landuse and settlement changes properly requires, at a minimum, the ability to model and integrate codes/procedures related to surface and subsurface hydrology, economics, and biology. In a recent study defining new strategies for America's watersheds, the National Research Council recommended that watershed researchers emphasize the integration of environmental, economic, and social perspectives, with more attention to linkages, their implications for management, and overcoming barriers to implementation. It was further recommended that the scientific communities develop better, more user-friendly decision support systems to help decision-makers understand and evaluate alternative approaches [1]. The emerging discipline of problem solving environments (PSEs) [2, 3] seeks to attain precisely this goal by combining discipline-specific software tools into integrated systems for decision-making and problem solving. PSEs free the computational scientist from managing individual software components and enable the specification of parameters of the problem at a high level (in the vernacular of the domain), rather than in terms of low-level modeling subsystems or software. PSEs then integrate results of the submodels into coherent, visual feedback suitable for high-level comprehension. Finally, PSEs are meant to be used by people who have diverse backgrounds and levels of expertise, and who are certain not to be experts in all of the domains that are modeled.

This paper presents the design and implementation of L2W (From Landscapes to Waterscapes) — a PSE for landuse change analysis. L2W organizes and unifies the diverse collection of software typically associated with ecosystem models (hydrological, economic, and biological). It provides a web-based interface for potential watershed managers and other users to explore meaningful alternative land development and management scenarios and view their hydrological, ecological, and economic impacts.

Organization of the Paper

Section 2 outlines various design principles of PSEs, with specific reference to watershed assessment. The current state-of-the-art in implementation technologies for PSEs is also presented here. Section 3 presents the design architecture of the L2W PSE. It outlines the various models considered in this study. Usability and performance considerations are outlined, and comparisons to other systems are drawn. Section 4 describes experimental results from a prototype implementation of the L2W PSE for the Upper Roanoke River Watershed in Southwest Virginia, USA. It also emphasizes the configuration and tuning of the models presented earlier. Section 5 outlines various avenues for extending the capability of L2W.

2 Problem Solving Environments

PSEs were originally introduced in domains such as partial differential equations (PDEs) [4, 5] and linear algebra [6] where they provided high-level programmatic interfaces to widely used software libraries [3, 7, 8]. With rapid advances in high performance computing, GIS, software interfaces, computational intelligence, and networking, interest in PSEs has expanded to diverse application domains such as wood-based composite design [9] and aircraft design [10]. While these projects concentrate on developing domain-specific PSEs, considerable attention has also been devoted to developing generic tools for building PSEs. The software engineering of customizable architectures, leveraging the Web, supporting distributed, collaborative problem solving, and providing *middleware* constitute some of the enabling technologies.

The focus of this paper is a PSE for landuse change analysis; while there is no doubt that the need exists for better models for all aspects of watershed assessment, including hydrology (flooding and erosion effects), biology (effects of contaminants and population changes), and economics (valuations resulting from landuse changes and surrounding environment, economic effects on governments), the synergy resulting from integrating them in a PSE will help leverage them in ways that best benefit planners and other observers. We identify a number of distinct aspects that should be part of a full-fledged PSE for watershed management, along with rationale for the desirability of each point.

2.1 Design Features

Internet Access to Legacy Codes

Linking models (e.g., from hydrology, economics, and biology) together, possibly via the Internet, is a primary aspect addressed by the integration of existing codes in a PSE. Such an approach avoids platform dependency issues and users are not required to install the system on a compatible platform. Perhaps more importantly, by using a network-based approach, it is not even necessary that all of the models reside/run on the same platform and PSEs can be envisioned as providing network-based 'software services' [11, 12].

Interactive Visualization

Users of a PSE typically wish to visualize the output, rather than process numeric results from the models. Such visualization processes should be integrated seamlessly with the computational pipeline by the PSE. An important aspect of such integration relates to inlined simulation and visualization tasks. It can be argued that if one can identify specific processes (and/or subdomains) that are interesting, then computational resources could be steered towards these processes, while supporting other simulation tasks only in so far as to maintain the fidelity of the interesting phenomena. This concept of *computational steering* [13] plays an important role in involving stakeholders in the management decision process.

Scenario and Experiment Management

PSEs should encourage users to experiment with various management options or scenarios. Such scenarios should be at a cognitive level relevant to the user, *i.e.* typically higher than the raw input demanded by the model. As each scenario is evaluated, the results can be recorded in a database for later retrieval, and for automated comparison to other scenarios. It is not uncommon for a typical user to run a model several times, with various combinations of input parameters, to generate output that meets some performance criteria. In some cases, users may conduct hundreds of experiments. Recording scenarios can thus aid in experiment management [14], parameter tuning, and automated optimization. In the context of watershed assessment, scenario management in PSEs is intricately coupled to GIS support for physically-based models.

Multidisciplinary Support and Usage Documentation

Since the collection of models comprising a watershed assessment system are multidisciplinary in nature, a PSE must provide support to users who will not be expert in every (or any) aspect of the domain. This requires alternate interfaces to different aspects of the modeling subsystems to reflect various levels of expertise. Typically, expert users desire more detailed control of models while novice users will wish to control only the coarse details, and need the maximum amount of guidance on reasonable setting(s) for models. The simulation interface could provide recommendations on reasonable interactions of parameters, or on which submodels to use in particular circumstances. Such advisory support regarding parameters is an integral aspect for the practical utility of PSEs.

Recommender Systems

A full-fledged PSE will likely provide a rich collection of simulations for modeling various aspects of the problem. Unfortunately, the multitude of choices available can bewilder novice users. Recommender systems for PSEs [15] serve as intelligent front-ends and guide the user from a high level description of the problem through every stage of the solution process, providing recommendations at each step.

Collaboration Support

Decision makers often would like to either communicate their rationale to others, or work collaboratively with others during the planning process. While the ability to save and restore prior results can be used to provide asynchronous collaboration, ideally a PSE would allow multiple users at multiple sites to work together collaboratively and interactively. For instance, one user can create a scenario and display the results to others who can perform further analyses. Alternatively, two or more users (e.g., a hydrologist, an economist, and a biologist) can jointly set up scenarios. Together with component-based architectures, collaborative systems help realize the paradigm of ‘programming-in-the-large,’ where powerful programming abstractions harness widespread computing resources in an intuitive and transparent manner.

Optimization

Selecting a ‘best’ configuration to balance competing goals within a watershed can be cast as an optimization problem. A given run of a model is typically an evaluation at a single point in a multi-dimensional space. In essence, the goal is to supply to the model that vector of parameters that yields the best result under some figure of merit. As such, decision-making processes can often be improved by applying automated optimization techniques, rather than have someone manually try a large number of parameter sets. Automated optimization techniques are quite sophisticated today, and are woefully underutilized by decision support systems in many disciplines, including watershed management.

High Performance Computing

Many of the models used in watershed assessment (e.g., for simulation of hourly flood hydrographs for a period extending over multiple years) require significant computing resources, such as a parallel supercomputer or an 'information grid.' PSEs can incorporate a computing resource management subsystem [16] such as Globus [17] or Legion [18], and hide the details of accessing the necessary computational resources from the user.

Preservation of Expert Knowledge

Like books in libraries, programs codify and preserve expert knowledge about the application domain. By using and preserving legacy code, the expert knowledge embodied in the legacy codes is (indirectly) employed by the PSE. Yet, state-of-the-art codes in their native form are nearly impossible for nonexperts to use productively. By providing advice, either from knowledge culled from experts or by automatic inference and mining, PSEs can make legacy codes and knowledge more usable by nonexperts.

Pedagogical Uses

PSEs in domains such as watershed assessment can also help to improve education in all of the related disciplines. Students in environmental and civil engineering can more easily be made aware of biological and economic issues, and likewise biologists and economists can acquire sensitivity regarding issues in the other disciplines. In addition, the general public gets heavily involved in controversial zoning and planning decisions. Using PSEs, citizens could go online and learn various aspects involved in resource management decisions. For example, the EPA TMDL (Total Maximum Daily Load) development process [19] requires the involvement of stakeholders who could evaluate for themselves the rationale for planning choices made in particular projects. Ultimately, a better understanding of the complex issues involved will benefit all parties.

2.2 Other Issues Specific to Watershed Management

There are many types of constituencies which reside in, or have an impact on, a watershed. Individual residents, manufacturing companies, governmental agencies, chambers of commerce, and transportation corridors have differing (and contradictory) goals that influence the environment in myriad ways. Watersheds do not form political boundaries, forest boundaries, or transportation boundaries. Only within the last 30 years or so have some larger watersheds had an organization that manages resources (which predominantly is water). These factors imply that (i) modeling a watershed as a closed entity is unrealistic, especially with regards to economics and development, (ii) multiple perspectives on how resources should be protected, preserved, or utilized exist, (iii) optimization is influenced heavily by the quantification of a negotiated set of evaluation criteria as created by a multidisciplinary team of stakeholders, (iv) substantial assumptions need to be made and clarified in order to create a watershed model, and an infinite variety of models are possible. In terms of a PSE, this is a significant deviation from more quantifiable (and structured) problems based purely on scientific and engineering principles. Furthermore, the spatial, temporal scales, and uncertainty in data call for a systematic integration of GIS services into the PSE design [20].

The level of refinement of input data, modeling processes, and output data also influence PSE design. For example, in a hydrologic sense, models are generally either lumped or distributed parameter models. Lumped models describe runoff, infiltration, evapotranspiration, and other parameters for 'similar' areas being clustered for a sub-watershed, while distributed models attempt to handle a finer spatial resolution of parameter characterization and modeling. In economic modeling terms, spatial scale is dependent on location of infrastructure, existing tax base, and many other parameters. The base spatial unit for economics may be a parcel, or a municipality, while the base spatial unit for a hydrologist is often a sub-watershed. Besides this difference in spatial unit of interest by discipline, the modeler must determine, for each system, the desired outcome scale as a function of the available scale of input data. Similar issues arise on a temporal level of detail. While hydrologic modeling can occur for an individual storm of an hour's duration, economic models are typically a multi-year event. PSEs are thus required to support site-specific predictions of hydrological and economic variables, by incorporating background knowledge when available.

2.3 Available Technologies

We now describe the present state of various software technologies that help to realize these different aspects. The key need is the ability to link together multiple models, and provide access to the aggregate via the Internet. Fortunately, the techniques for doing this are becoming well understood. 'Middleware' refers to software that mediates between a user interface (usually provided via a Web browser) and back-end database(s) and/or simulation(s), routes queries, performs information integration, and supports distributed problem solving. Many systems in use everyday by millions of people are based on the middleware model. Typically, scripting languages such as Perl are used to access the models and visualization tools, wherein a Web server accepts commands from the user interface to drive the scripts. Custom Java applets can be used for the front-end interface(s).

One tool that we have found to be particularly useful for developing a watershed management PSE is MapObjects from the Environmental Systems Research Institute (ESRI) [21]. The purpose of MapObjects is to provide a Web-based interface to ESRI's ARC/INFO product, which is already familiar to many watershed planners. It provides the ability to call user-defined functions, which in turn can access Perl scripts to drive outside models and visualization tools.

Another alternative is to develop component-based software using, for example, Sun's JavaBeans technology. The goal of JavaBeans is to allow developers to make reusable software components to simplify program development. However, JavaBeans can also be used to develop systems where the 'beans' are surrogates for various distributed tools that can be linked together in various ways. Thus, we can envision a system that allows the user to select one or more modeling tools, link them together, and then in turn link the output to the user's choice of visualization tool. Once again, middleware acts as the intermediary between the various components, addressing data formatting and transfer issues.

The technologies just described for linking together distributed components are now well understood, and currently being used in various PSEs. Somewhat more speculative is technology for supporting synchronous collaboration. The success of Microsoft's NetMeeting demonstrates that collaborative systems are now reaching the level of limited commercial success. NetMeeting is rather limited in its capabilities, but it is the first practical collaborative system that is widely used by typical users. The research field known as computer supported cooperative work (CSCW) is pushing forward on more advanced collaborative systems. Once again, Sun's Java technology provides reasonable possibilities for practical collaborative systems in the near future.

Large-scale simulations can require massive amounts of computing power. A plausible alternative to making super-computer class equipment available to local government planners is to harness the computing power that normally goes untapped in desktop computers. Recently, the SETI@home project (Search for Extraterrestrial Intelligence [22]) gained prominence due, in large part, to its ingenious approach to harnessing the large computational resources of the Internet to search for patterns and anomalies indicating extraterrestrial intelligence. A number of efforts are underway to create a computing 'power grid.' The Information Power Grid [23] (IPG) being envisioned by NASA and the national laboratories is a general, all-encompassing PSE. While some of the requisite technologies are in place (e.g., Globus [17] for distributed resource management, and PETSc [24] for a scientific software library), it is unclear how the remaining components can be built and integrated. At this time, IPG is a vision rather than a working prototype.

As the number of algorithms and models made available to the computational scientist increases, there is a concomitant need to support the knowledge-based selection of solution components. This requirement is addressed by recommender systems, introduced earlier. Recommender systems are typically designed by organizing a battery of benchmark problems and algorithm/model executions, and subsequently generalizing the results to obtain high-level rules that can form the basis of a recommendation. Such generalizations are typically obtained by data mining software [25, 26] that seek to find higher order patterns hidden in data. The reader will be familiar with the beers-diapers discovery in commercial market basket data ('People who buy diapers in the afternoon are more likely to buy beer too') [27], but the role of data mining in computational science is a larger and more complicated application. Data mining thus constitutes a key computational technology, supporting traditional analysis, visualization, and design tasks [28].

Like most of PSE work, recommender systems research has concentrated on both (i) creating reusable knowledge-bases for specific domains, and (ii) designing software architectures for the rapid prototyping of recommender systems. The PYTHIA kernel, described in [29], provides a database infrastructure for problem and method definition, experiment management, performance data analysis, and automatic mining of recommendation spaces. Its generic design permits applications to structured domains such as PDEs, numerical quadrature as well as to more amorphous domains, such as watershed management. PYTHIA is built using the Postgres object-relational database system (for storage, retrieval, and management), Tcl/Tk (for interfaces and scripting), statistical software in C (for performance analysis), PROGOL (an induction package for data mining), and CLIPS (a production system shell for making recommendations).

Recommender systems thus contribute directly to automated decision making and also have pedagogical uses in providing phenomenological explanations of their choices and selections. The recently concluded NSF SIDEKIC Workshop on PSEs underscores the importance of recommender systems in several key applications [30].

Once recommendations for models are configured, such choices and selections can be optimized to achieve user-defined objectives. Multidisciplinary and multiple-objective optimization is a well-understood area of technology, and can thus be deployed immediately in the context of watershed management. In multidisciplinary optimization [31, 32, 33], a large system comprising several disciplinary components (e.g., hydrology, hydraulics, economics, biology) is optimized in parallel, by optimizing the subsystems concurrently using for each subsystem a detailed model in one discipline and approximate models for the other disciplines. There are several known successful strategies for managing the parallel optimizations and ensuring convergence [34].

Multicriteria optimization is typical in landuse management, where there are contradictory goals for the involved stakeholders. The approach is to find pareto optima, similar to game equilibria, where no one participant can unilaterally improve their position. Giving planners and managers a family of such optima permits them to consider a range of tradeoffs. Again, well understood theory and algorithms exist for multicriteria optimization that could be immediately deployed in landuse management systems.

Those parts of optimization theory best known outside the mathematical sciences—linear programming and derivative based algorithms—are perhaps the least useful in this context. There are direct search [35] and simplicial pattern search [36] algorithms that only require candidate points to be ranked; these methods coupled with statistical response surface methodology [37] can be very effective for the type of problems with sparse and noisy data encountered in landuse models. There is certainly research work to be done on improved optimization techniques, but standard tools could be integrated with existing models quite quickly.

3 Design of the L2W System

3.1 Related Research

PSEs for watershed management are typically centered on physically-based conceptual models which delineate a watershed into multiple classifications based on landuse and drainage connectivity. One of the primary systems available for hydrological modeling is the commandline program HSPF (Hydrological Simulation Program in FORTRAN) [38]. A scenario generator called GenScn (GENeration and analysis of model simulation SCeNarios) [39] that implements a graphical user interface over HSPF is also available, making it easier to enter necessary data and parameters to drive HSPF. GenScn is meant to help the user in analyzing various what-if scenarios in a watershed involving landuse change, landuse management practices, and water management operations. Such scenarios involve analyzing and managing high volumes of input and output data and hence follow a difficult process. GenScn helps in this process by creating simulation scenarios, analyzing results of the scenarios, and comparing scenarios. The GUI uses standard Windows 9x/NT components and MapObjects LT from ESRI. The model outputs include interactive and batch graphical and tabular displays of both observed and simulated data.

An example of an integrated system is the LUCAS (Land Use Change Analysis System)

PSE [40], designed on a Markov probabilistic model that attempts to capture the influence of market economics (ownership characteristics), transportation networks (access and routing costs), human institutions (population density), and ecological behavior on landscape properties. The primary motivation is socioeconomic modeling; LUCAS uses a transition matrix to assess random spatial variations in landuse which, in turn, are used for assessing the expected impact of a given set of factors. LUCAS has an advanced GUI for displaying landuse scenarios and habitat changes, based on the public domain Geographic Resources Analysis Support System (GRASS) GIS from the U. S. Army Construction Engineering Research Laboratories [41]. The Markov models used are derived from time series data and expert opinion, and thus predictions must come from averaging many simulation runs. Currently only a small number of economic factors are considered, and biological effects are only inferred from (probabilistic) habitat changes.

The modeling philosophy of L2W is quite different from that of LUCAS. L2W uses a well-established physics based hydrology model to simulate surface water runoff, subsurface flow, stream flow, stream bank erosion, sedimentation, and pollutant transport. It has a comprehensive economic model including roads, taxes, water and sewer infrastructure, and numerous zoning and developmental considerations. Furthermore, L2W has the potential to predict effects of landuse change (residential or industrial development) on biological indicators, e.g. fish diversity and health. L2W is site-specific in its predictions (e.g., flooding or the disappearance of a particular species at a given location), rather than global and probabilistic as LUCAS, which is based solely on Markov transition matrices for landscape changes. Currently, LUCAS is superior in its GIS based display of predictions, and LUCAS can more easily incorporate expert opinion and known isolated facts than the physics based L2W. Systems such as L2W (akin to climate modeling) have the advantage of high resolution and detailed prediction, but also the matching burden of obtaining initial and boundary conditions, and physical constants (e.g., soil permeability for subsurface flow).

Various other PSE-like products have been proposed in the water resources and geographical engineering communities. WISE [42] is a PSE for simulating complete ecosystem models and has goals similar to L2W: link models from multiple domains and subdisciplines. AQUATOOL [43], a system for water resources planning and operational management, is composed of modules linked through geographically referenced databases and knowledge bases. These modules are designed to model water resources schemes optimization, carry out simulation of management of water resources systems including conjunctive use of surface and ground water, and preprocess a groundwater model designed to include distributed aquifer submodels in the simulation model. BASINS [44], released by the EPA, supports environmental and ecological analysis on a watershed basis through use of models and a GIS. Osmand et al. developed a decision support system (DSS) called WATERSHEDSS [45] to aid watershed managers in handling water quality problems in agricultural watersheds. The key objectives of this DSS are to transfer information to watershed managers for making appropriate land management decisions, to assess nonpoint-source pollution in a watershed based on user supplied information and decisions, and to evaluate water quality effects of alternative land treatment scenarios. Lal et al. [46] and Negahban et al. [47] describe a DSS named LOADSS that is designed to evaluate phosphorus loading and control in the Lake Okeechobee basin through the use of GIS linked modules. The WISE environment [48] lets researchers link models of ecosystems from various subdisciplines. Chen et al. [49] present the design of the watershed analysis risk management framework (WARMF) for calculation of the total maximum daily loads (TMDLs) of various pollutants within a river basin. WARMF contains five integrated modules—Engineering, TMDL, Consensus, Data, and Knowledge. A GUI that provides menus for the user to issue commands, store, and display the output in the forms of GIS maps, bar charts, and spreadsheets helps to integrate these modules.

A number of European organizations jointly developed a DSS called Waterware [50] to assist government agencies and river basin commissions in decision making for the efficient qualitative and quantitative management of water resources. Waterware consists of a GIS and a database management system (DBMS) coupled to a large number of analytical components including demand forecasting, water resources planning, ground water pollution control, surface water pollution control, and hydrological processes. The authors have applied the DSS to two river basins to address the problems of water resource assessment, reservoir site selection, decontamination of groundwater, estimation of sustainable irrigation abstractions, and derivation of required effluent quality standards.

Dunn et al. [51] have described the hydrology component of the NERC-ESRC Land-Use Programme (NELUP) — a DSS with the objective of analyzing the implications and predicting the impact of agricultural landuse change at the river basin scale. The components of NELUP include models representing agricultural economics, ecology, and hydrological regimes of the basin. Due to the complexity of various landuse change problems, the authors opine that the NELUP DSS cannot be used by non-specialists without assistance.

Leavesley et al. [52] describe the development of a DSS for water and power management called the modular modeling system (MMS). MMS uses a master library that contains compatible modules for simulating a variety of water, energy, and biogeochemical processes. It provides a common framework in which to develop and apply models that are designed for basin and problem specific needs. The GIS interface of MMS is developed to facilitate model development, parameterization, application, and analysis. Typical applications of the MMS are in the management of a multi-reservoir river system within the constraints of competing water users and selected environmental constraints such as water temperature limits or fisheries habitat needs.

Simonovic and Bender [53] discuss the concept of a collaborative planning support system (CPSS) in water resources planning. CPSS is intended to be less a full-fledged PSE than a systematic framework to empower participants by identifying areas of common understanding, then encouraging them to explore solutions and reach a consensus.

In order to examine the effects of potential landuse and land management policies on water quality and the resulting costs in South Australia, Davis et al. [54] developed a DSS consisting of three modules, namely a policy module, catchment module, and query module. The policy module allows the user to construct a suite of policies, and the catchment module estimates the effects of these policies on total phosphorus, total nitrogen, and turbidity levels in a catchment under consideration. The query module allows the user to see the results of the simulation.

While most of these systems provide sophisticated models and link appropriate simulation codes using a GIS, none are web-accessible to the best of our knowledge. Also, the availability of a PSE explicitly focusing on evaluating hydrologic and economic impacts of residential settlement patterns is limited. Most systems are somewhat restrictive in their scope and do not provide a truly multidisciplinary assessment of management changes in watersheds.

The Market Manager (MM) model of Carpenter et al. [55] takes an agent-based and dynamical system approach to modeling socioecological systems. The entire (dynamical)

	WARMF	WATERSHEDSS	BASINS	AQUATOOL	LUCAS	MM	L2W
Internet Access to Legacy Codes							●
Interactive Visualization	●	●	●	●	●	○	●
Scenario Management	●	●	●	●		●	●
Multidisciplinary Support			○		○	○	●
Recommender Systems[‖]							
Collaboration Support[‖]							
Optimization[‖]				●			
High Performance Computing					●		●
GIS Support	●	●	●	●	●		●
Site-Specific Prediction	●	●	●	●			●
Incorporates Prior Knowledge?					●		○

Table 1: Comparative tabulation of features of some PSEs and decision support systems for watershed assessment. The features are distinguished as aspects common to all PSEs (top eight) and those specific to watershed assessment (bottom three). Caption guideline: • – feature present; ○ – partial support for feature available.

system can have stable or unstable equilibria, and the actions of the various stakeholders (agents) drive the system toward an equilibrium, a periodic solution, or even toward chaos. The agents have only incomplete, local information, and no small group of agents can learn and control the total system. The authors consciously avoid cost-benefit optimization, fully intending the model to be metaphorical, i.e., illustrating general patterns of system behavior rather than making specific predictions. A notable observation from the work is that stable ecological systems can have intrinsic oscillations, and intervention failing to recognize this can be worse (drive the total socioecological system away from desirable solutions) than doing nothing.

Thus, Market Manager is quite dissimilar from L2W, being metaphorical, dynamical system (ordinary differential equation) based, and only mildly multidisciplinary, rather than (as L2W) predictive, physics based, and strongly multidisciplinary. See Table 1 for a detailed comparison of some of these systems along the PSE aspects introduced in Section 2.

3.2 System Architecture and Implementation

The architecture of the L2W PSE is based on leveraging existing software tools for hydrology, economic, and biological models into one integrated system. Geographic information system (GIS) data and techniques merge both the hydrologic and economic models with an intuitive web-based user interface. Incorporation of the GIS techniques into the PSE produces a more realistic, site-specific application where a user can create a landuse change scenario based on local spatial characteristics. Design of the PSE/GIS follows the model developed by Fedra [56] and Goodchild [57] in which one user interface interacts with the GIS and the models employed by the application. Another advantage of using a GIS with the PSE, as described by [58], is that the GIS can obtain necessary parameters for hydrologic and other modeling

---

[‖] These features are supported in the system design for L2W but were not implemented at the time of writing of this manuscript.

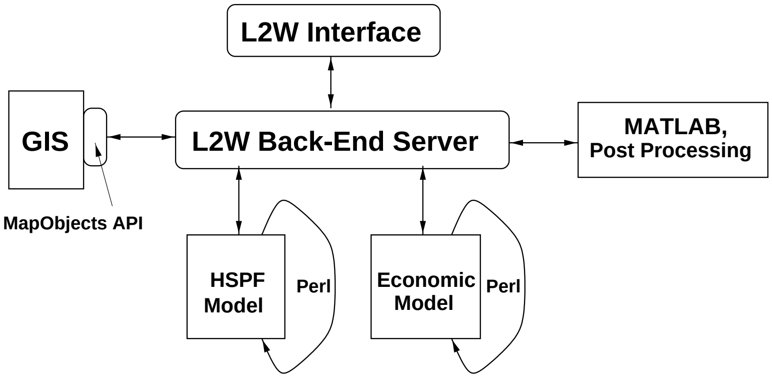


Figure 1: Software Architecture of L2W.

processes through analysis of terrain, land cover, and other features.

As described earlier, the surface hydrology model used is the HSPF V11.0 system [38] that incorporates a watershed scale ARM (Agricultural Runoff Management Model) and NPS (Nonpoint Source Pollutant) loading models into a basin-scale framework. HSPF models hydrological processes mathematically as flows and storages and uses a spatially lumped model for each *subarea* for a watershed (referred to as a subwatershed). In contrast, fully distributed, physically based models use a gridded rectangular cell as the building block and attempt to provide greater resolution in the modeling process. However, this enhancement in modeling power is not accompanied by corresponding spatial detail in the various input data sources (e.g. precipitation) and hence does not necessarily translate into improved hydrological forecasts. Furthermore, HSPF poses no topographic limits on the size of the subareas, is capable of modeling the hydrological processes on a continuous basis, and supports the analysis of various scenarios where the user changes landuse.

The hydrologist's interface to HSPF in L2W allows users to specify the percentage of basic landuse types (forested, herbaceous, and disturbed) to be applied within specified subwatersheds, which are selected from a map. These percentage figures reflect introduction of various land settlement patterns in a subwatershed. Landuse changes are also provided to the economic model for analysis of economic impacts. The back-end prototype is written as a Visual BASIC application (chosen because it supports the MapObjects system) and the simulations for watershed runoff are accessed via Perl scripts wrapped around HSPF. Postprocessing tools for statistics and visualization are provided by Matlab and operating system utilities. MapObjects's programming interfaces that allow implementors to add map features and other GIS functions quickly without writing a lot of code in-house aids in the specification of spatial input. By combining HSPF, Matlab, and MapObjects into one integrated system, we provide a way for the user to experiment with various land development scenarios within the watershed. Fig. 1 depicts the software architecture of L2W.

The economic model estimates the effects of residential developments on water and sewer costs, property values, property tax base, and property tax revenues. Maintenance requirements are determined according to the layout of each development and its location relative to existing water and sewer lines. These infrastructure requirements are used in conjunction with unit cost data from generally accepted industry sources to calculate total costs. Property values and estimated tax base are based on lot size and other characteristics. We now describe these models in more detail.

3.3 Models, Codes, and Software

HSPF: Model Structure

HSPF was developed in the late 1970's as a union between the Stanford Watershed Model [59] and several water quality models developed by the USEPA. The USEPA and USGS agencies have since been involved in the development and maintenance of HSPF, which has witnessed over 150 applications in the country and abroad [60]. The model contains three application modules and five utility modules. The application modules, representing the hydrologic/hydraulic processes, are referred to as PERLND, IMPLND, and RCHRES. The PERLND module simulates runoff and water quality constituents from pervious land areas in the watershed and is the most frequently used part of the model. The IMPLND module simulates impervious land area runoff and water quality. The movement of runoff water and its associated water quality constituents in stream channels and mixed reservoirs are modeled by the RCHRES module. The utility modules perform operations involving time series which are essentially auxiliary to application modules, e.g., input time series data from ASCII files to a watershed data management (WDM) file using COPY, multiplying two time series etc.

HSPF: PERLND

The application modules are divided into several distinct sections, each of which may be selectively activated in a given simulation by the user. The PERLND module contains 12 sections, the first for correcting air temperature for elevation difference (ATEMP) and the last for simulating the movement of a tracer (TRACER). The key section of the PERLND module is called PWATER which is used to calculate the water budget components resulting from precipitation on the pervious land segments. PWATER models processes such as evapotranspiration, surface detention, surface runoff, infiltration, interflow, baseflow, and percolation to deep groundwater using both physical and empirical formulations.

The PWATER section requires precipitation and potential evapotranspiration time series for performing water balance computations. When snow accumulation and melt are considered, additional information on air temperature, snow cover, ice content of the snowpack etc. are required. The time series of precipitation representing moisture supplied to the land segment is first subjected to interception losses. Typically on pervious areas, the interception capacity represents storage on grass blades, leaves, branches, trunks, and stems of vegetation. It can either be supplied on a monthly basis or as one single value. Water held in interception storage is removed by evaporation. Moisture exceeding the interception capacity overflows the storage and becomes available for either infiltration or runoff. The infiltration rate is modeled as a function of time and is related to the soil moisture content based on the work of Philip [61].

Spatial variation in infiltration rate is considered using a linear probability distribution. For each time step, the available depth of water is divided between infiltrated depth and potential direct runoff (PDRO). The PDRO either enters the upper zone storage or becomes available for either interflow or overland flow. The fraction of PDRO that goes to the upper zone storage is dictated by the ratio of storage in upper zone and its nominal capacity. The overland flow is simulated using the Chezy-Manning equation and an empirical expression that relates outflow depth to detention storage. The overland flow computations require Manning's roughness, slope, and length of flow plane. The Manning's roughness can be input on monthly basis to allow for surface roughness variations over the year. The rate of interflow is assumed as a linear function of interflow storage. An interflow recession parameter is used in interflow computation that is taken as ratio of present rate of interflow outflow to the value 24 hours earlier. This parameter can be given monthly values to allow for variation in soil properties. The inflow computed for the upper zone storage gets added to the existing storage and depending on the status of storages in upper and lower zones, percolation of water takes place from upper storage to the lower storage. An empirical relationship is used to compute the fraction of infiltration and percolation entering the lower zone storage. The amount entering the lower zone storage is dictated by the ratio of lower zone storage and the nominal capacity of lower zone that is one of the model parameters and can be input on monthly basis to allow for annual variation. The fraction of the moisture supply remaining after the surface, upper zone, and lower zone components are subtracted is added to the groundwater storages. The flow to groundwater is split between active and inactive groundwater storage. This split is based on a user supplied parameter. The groundwater outflow akes place from the active storage based on a relationship that involves cross sectional area and energy gradient of the flow.

The model requires time series of potential evapotranspiration (PET). This can be developed using data from a Class A pan or by using various empirical relationships for estimating PET. The input time series of PET is compared to the available water on the watershed during each time step and the flux of actual evapotranspiration (ET) is calculated from five sources in the following order. The first source of meeting ET demand is the baseflow or groundwater outflow. The fraction of total PET met by this source is dictated by a user supplied parameter. The remaining PET comes from interception storage which is depleted until the PET is met or until there is no more water in interception storage. The next source of meeting PET is the upper zone storage. The contribution of this storage is controlled by the ratio of upper zone storage to the nominal value of upper zone storage. PET not satisfied from the above storages is met from active ground water storage and is controlled by a user supplied parameter. The lower zone is the last storage from which ET is drawn and the amount withdrawn is based on a user supplied parameter that can have monthly values to reflect vegetation density, rooting depth, density of vegetation, and stage of plant growth.

HSPF: IMPLND and RCHRES

In a land segment modeled as IMPLND, no infiltration occurs and only land surface processes are modeled. Many of the sections of the IMPLND module are similar to corresponding sections in the PERLND module. In fact, IMPLND sections are simpler because infiltration and sub-surface flows are not considered. The flow in a RCHRES segment is assumed unidirectional. Inflow to a RCHRES comes from upstream RCHRESs, overland flow, diversions and enter through a single gate. The volume of RCHRES is updated and the downstream discharge is computed from the volume-discharge relationship specified at the downstream end. Tables of volume-discharge relationships for each RCHRES thus form part of the input file. Outflows may leave the RCHRES through one or several gates or exits.

Economic Model

The economic model estimates the effects of residential developments on property values, property tax base, property tax revenues, and water and sewer costs. The user can place any combination of four development tract forms within the subwatershed—low density, mid density conventional, mid density cluster, and high density [62, 63, 64]. Property values are estimated as the sum of bare land values and estimated construction costs for housing and infrastructure. In reality, the value of a development is jointly determined by the supply of housing and the demand by potential home buyers in an area. However, developers should expect housing sales revenues to cover their costs over the long term, otherwise they would not invest in developments. If sales revenues exceed costs by a large margin, more developers will invest in housing developments causing housing supplies to increase and driving housing prices down.

Bare land values are statistically estimated using land values based on land transactions from sources such as the Roanoke County Division of Planning and Division of Tax and Assessment's database [65]. Housing construction costs are estimated from secondary sources [66, 67]. Assessed values of developed land equal the sum of bare land values plus housing construction hosts. Assessed values of undeveloped land can be based on use value or market value. Tax revenues from land equal the assessed value of land times the tax rate. Costs to link sewer and water systems from the edge of the development to the central water or sewage treatment system are assumed to be borne by the local government. The unit cost of water transmission mains is determined by the sum of the costs per meter of pipe (materials, labor and equipment), excavation, trench bedding, fire hydrants, and valves [68].

4 Experimental Studies

User Interface Description

An initial prototype of our system is available at the URL `http://landscapes.ce.vt.edu` and covers the 57 square-mile Back Creek subwatershed of the Upper Roanoke River watershed (see Fig. 2) in Southwest Virginia, USA. Typically, the user invokes the thin-client Java applet (see Fig. 3) depicting the Back Creek subwatershed and uses the cursor to specify landuse distributions for individual land segments within Back Creek subwatershed. By selecting the "hydrology expert interface" (see Fig. 4) over the "decision maker" interface in Fig. 3, hydrologists can use an HSPF input file that they have created, allowing more control when greater expertise is available. The cursor locations are converted and communicated via messages to a server, where each individual message contains details of the coordinates on the map (where clicked), parameters for running a simulation, or a command to indicate a particular simulation. Using MapObjects on the $600m^2$ per pixel grid helps us provide map layer functions, automatic drawing of the map on the server, and transmission of maps across the internet. In particular, MapObjects provides primitives for intercepting coordinates of clicks on the map in the applet. Based on the user input, L2W calculates the new distribution of landuses, suitable for input to HSPF, which is then run on one "base" rainfall pattern for a pre-selected duration.

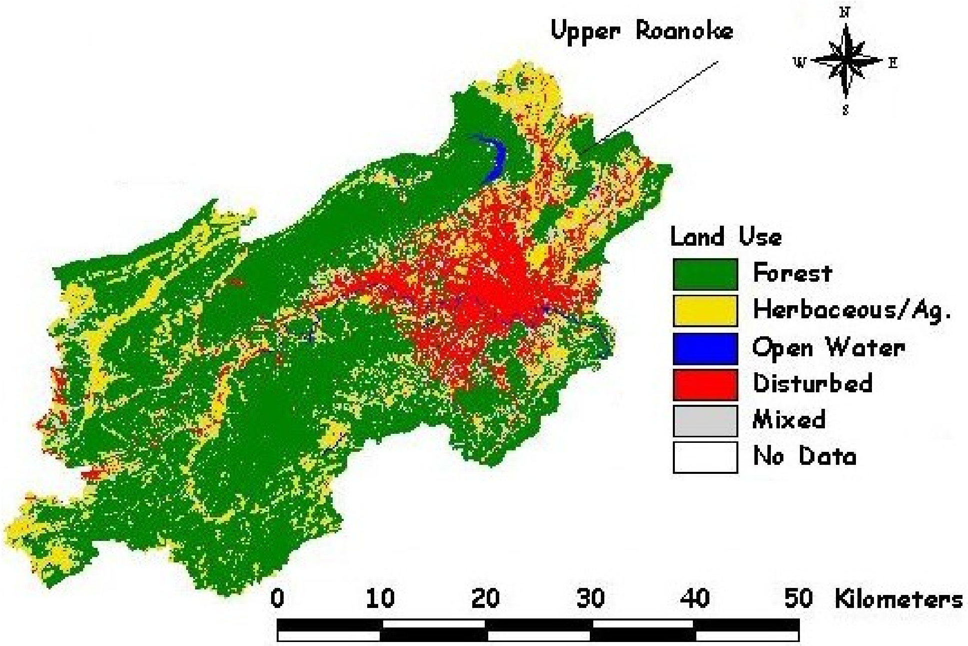
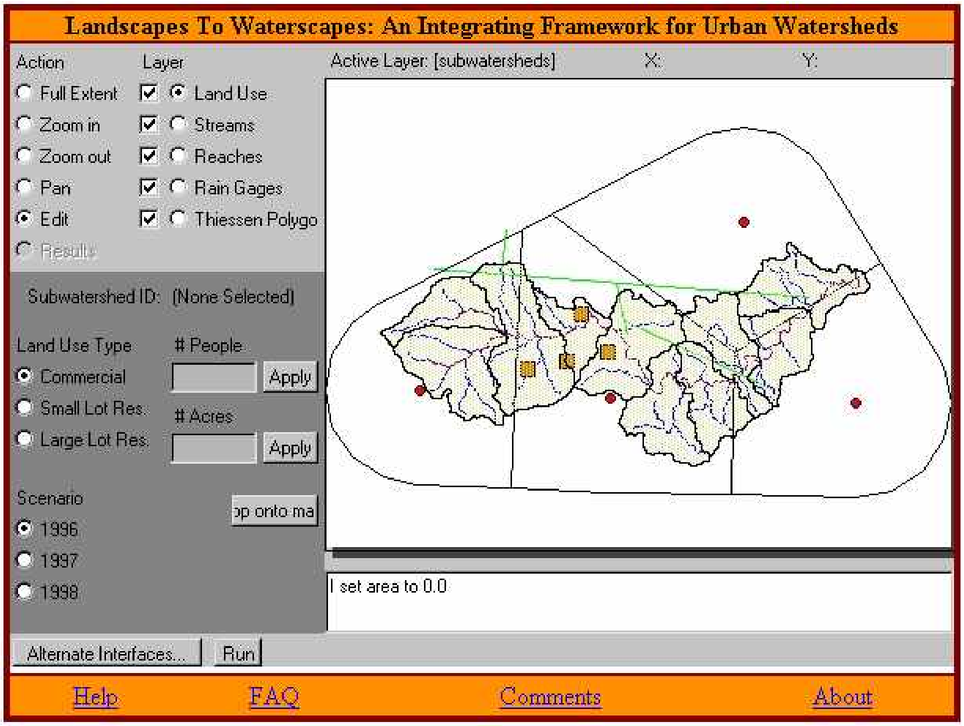


Figure 2: Landuse segmentation of the Upper Roanoke River Watershed in Southwest Virginia, USA.



Landscapes To Waterscapes: An Integrating Framework for Urban Watersheds

Active Layer: [subwatersheds] X: Y:

Action: ● Edit

Layer: ● Land Use

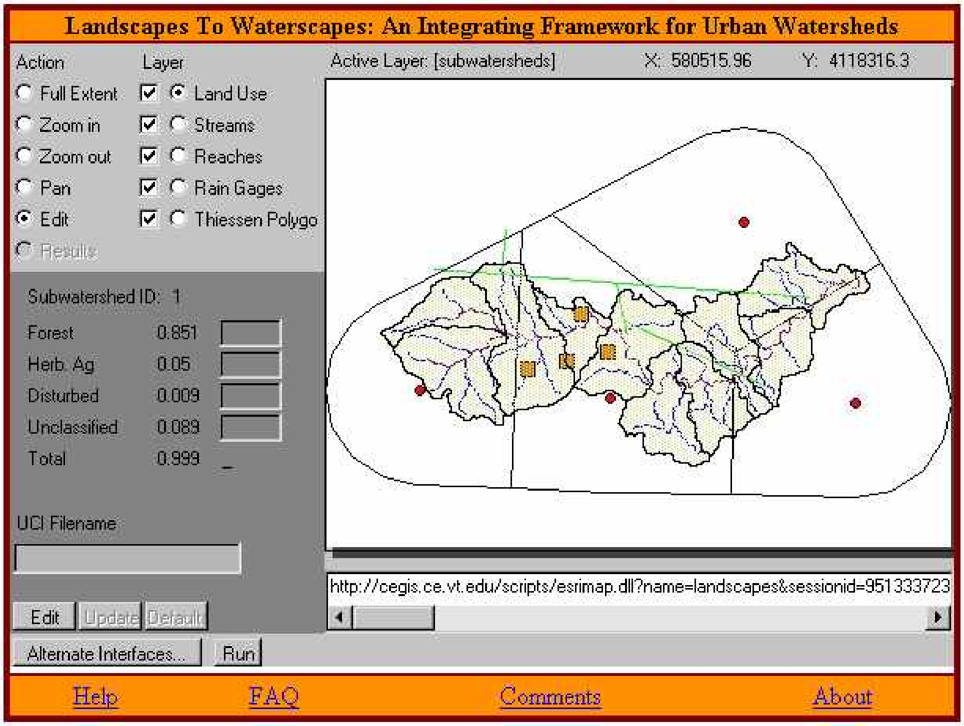
Subwatershed ID: (None Selected)

Land Use Type: ● Commercial

Scenario: ● 1996

I set area to 0.0

Figure 3: Front-end decision maker interface to the L2W PSE.



Active Layer: [subwatersheds] X: 580515.96 Y: 4118316.3

Subwatershed ID: 1

Forest	0.851
Herb. Ag	0.05
Disturbed	0.009
Unclassified	0.089
Total	0.999

http://cegis.ce.vt.edu/scripts/esrimap.dll?name=landscapes&sessionid=951333723

Figure 4: Front-end hydrology expert interface to the L2W PSE.

HSPF Model Parameters and Calibration

The HSPF model requires input data on rainfall, stream flow, evaporation, soil, and landuse information. Hourly records of rainfall data were obtained from the Blacksburg office of the National Weather Service. Flow data for Dundee stream gage on Back Creek was obtained from the USGS office in Richmond, VA. Potential evapotranspiration values were calculated on a monthly basis using the Thornthwaite method [69]. Physical watershed data were obtained from USGS 30-meter DEMS, USGS stream reach overlays, and Virginia Gap landuse data. Reach cross-section data was collected in a field visit and from the Roanoke Valley Regional Stormwater Management Plan [70]. Based on the distribution of landuse and stream reaches, the watershed is divided into ten segments drained by ten stream reaches (see Fig. 5).

HSPF is a heavily parameterized model and uses both conceptual and physical parameters to represent hydrologic processes occurring within a watershed. The initial estimate of parameters was made based on published studies including the Upper James River study, conducted as a part of the EPA's Chesapeake Bay model [71]. In general, parameters associated with the upper soil zone varied with landuse, while the watershed slope varied among the ten physical land segments. Forest, herbaceous/agriculture, mixed, and disturbed lands were modeled as PERLND segments while impervious land was represented as an IMPLND segment.

The model was calibrated for water years 1995, 1996, and 1997 using the USGS/EPA HSPEXP expert system shell. Calibration consisted of matching simulated and observed results for annual flow volume, high and low flow volumes, storm peaks, and seasonal volume differences. Parameter changes were made by varying the parameter by a fixed percentage for all landuses in all areas, while maintaining the relative differences in parameters between landuses. Calibration was considered complete when expert system advice did not improve model performance. The performance of the calibrated model was validated on water year 1998 and the final model was incorporated into the PSE design. Results of simulation runs taken with PSE version of HSPF for examining various 'what if' scenarios were satisfactorily compared to the results of similar runs taken by running the model in a standalone setting.

Land Value Model Estimation

A total of 1,844 transactions of vacant and nonvacant land parcels (in and around Roanoke county) for the period of 1996 to 1997 were used to estimate bare land values, which equal the value of the parcel minus the value of structures on the land. The assessed values of structures located on parcels was deducted from the parcel transaction prices—a procedure used by Bockstael and Bell [72]. Estimation was performed using traditional linear least squares approximations. Further work is being done to evaluate alternative statistical procedures. The resulting estimated model is

$$
\begin{aligned}
\log(Price) \;=\; & -17.87 \\
& -0.53[\log(Size)] - 0.02[\log(Size)]^2 \\
& +0.41[\log(Elevation)] - 0.13[\log(Elevation)]^2 \\
& -0.05(Soil1) - 0.10(Soil2) \\
& +0.0037(Population) - 0.0005(Population)^2 \\
& +1.60[\log(Mall)] - 0.25[\log(Mall)]^2 \\
& +2.47\log[\log(City)] + 0.13(Developed) \\
& -0.07(Road) + 0.05(Year) \\
& +4.09[\log(X)] + 3.72[\log(Y)] \\
& -0.91[\log(X)\log(Y)],
\end{aligned}
\tag{1}
$$

where $Price$ is the price of the parcel per square meter, $Size$ is the area of the parcel in square meters, $Elevation$ is the average elevation of the parcel in meters, $Soil1$ and $Soil2$ are dummy variables for soil permeability with $Soil1$ being least permeable and $Soil2$ intermediate in permeability, $Population$ is the population density (persons/hectare) in the U.S. Census block containing the parcel, $Mall$ is the minimum distance to an existing mall, $City$ is the minimum distance to the closest city (Roanoke or Blacksburg depending on parcel location), $Developed$ indicates whether the parcel is vacant or contains a commercial or a residential structure, $Road$ reveals whether the parcel is adjacent to a major Road, the variable $Year$ shows if the parcel was sold in 1996 or 1997, and the coordinates X and Y determine the exact location of the parcel [65].

Determination of Developable Land

In order to analyze effects of alternate land development scenarios, the extent of developable land within the watershed is first determined. For this purpose, a raster overlay of four

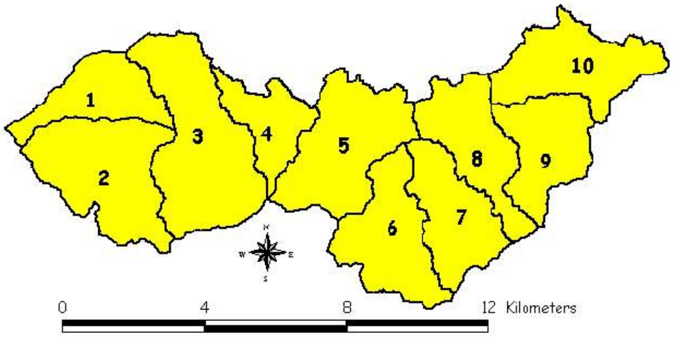
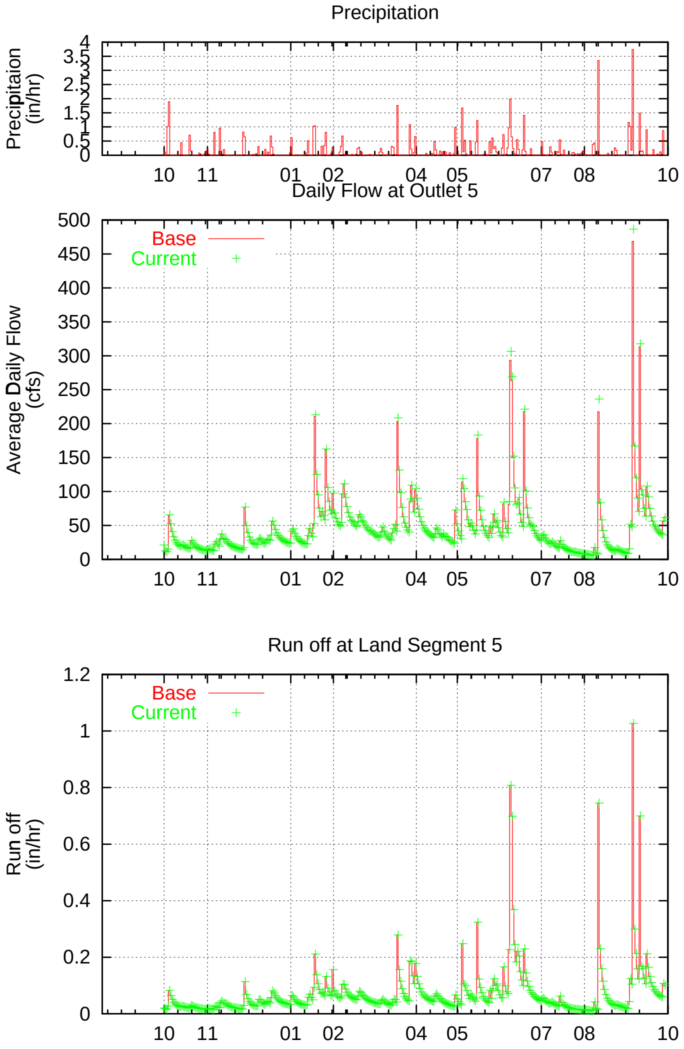


Figure 5: Land segments in the Back Creek subwatershed.

GIS Layer	*Criteria*	*Value*
Slope	>20	1
	<20	0
Landuse	Disturbed and Water	1
	Forest and Herb/Agr	0
Preservation Status	Preserved	1
	Unpreserved	0
Flood Plain Location	Inside Flood Plain	1
	Outside Flood Plain	0
Raster Overlay	Sum of Values	0 = Developable
		>0 = Undevelopable

Table 2: Method of raster overlay for determining developability.

spatial data layers — slope, landuse, preservation status, and flood plain location — was designed. In the overlay, the pixels in each of the four layers are reclassified with a value of either 0 or 1. The value assigned depends on whether the original value meets the criteria for developability. For example, pixels with average slopes of less than 20% are developable and are assigned a 0, while those over 20% are not developable and are assigned a 1. Each of the other layers is reclassified in a similar method (see Table 2). If any one of the four layers for a pixel is equal to one, then the pixel is not developable. Overall developability for a particular pixel, therefore, is achieved by summing the values for each of the four layers. If they sum to zero, then the pixel is developable; if the sum exceeds zero, then it is not developable. Within Back Creek subwatershed, land segments 3, 4, 5, and 10 (see Fig. 5) have significant portions of developable lands. These land segments, therefore, provide prime sites for adding new development tracts. The larger problem of misclassified and unlabeled data caused by out-of-date field measurements and lack of knowledge of precise commercial and vegetation boundaries is endemic to this domain; in the future, we plan to make use of machine learning techniques [73] to aid in automatic landuse segmentation. This is related to the broader task of map analysis using GIS data, a problem that has received much attention in areas such as identifying clusters of wild life behavior in forests [74], modeling population dynamics in ecosystems [75], and socioeconomic modeling [40].


Precipitation

Daily Flow at Outlet 5

Run off at Land Segment 5

Figure 6: Typical graph output indicating runoff impact resulting from altering landuse values in Segment 5.

	Development Scenario		
Landuse Summary (Acres)	*Baseline*	*Low Density*	*High Density*
Forest	25,077	22,216	24,373
Herbaceous/Agriculture (including lawns)	3,157	5,248	3,888
Mixed Forest	5,504	4,438	4,963
Disturbed Pervious	1,643	1,643	1,643
Disturbed Impervious	337	2,173	851
Total Land	35,718	35,718	35,718

Table 3: Back Creek landuses for pre-development baseline and after development.

Description of Test Scenarios

Two test scenarios are presented to discuss the hydrologic and economics results generated by the PSE. The first scenario is referred to as the low density scenario in which case all available land in the Back creek watershed (18,377 acres) was developed using a low density pattern. This scenario represented the conventional planning approach in the Roanoke basin. The low density development tract was assumed to be 150 acres to accommodate 125 people and included 10% impervious land, among other landuses. This scenario resulted in importing 15,300 new people in the watershed. The second scenario, referred to as the high density scenario, involved importing the same number of people near Roanoke city (i.e., land segment 5 in Fig. 5) using the high density pattern. This scenario represented likely outward expansion of already developed land that is part of Roanoke City currently. The high density development tract was assumed to be 12 acres to accommodate 125 people and included 35% impervious land, among other landuses. The proportions of various landuses were computed for both scenarios. Table 3 shows landuse for the pre-development baseline and the two development scenarios.

Interpretation of Results from Economic Model

Total land area and land devoted to housing lots and infrastructure are shown in Table 4. Average lot size for low density housing is 2.76 acres compared to 0.2 of an acre for high density. Estimated bare land values are based on average housing lot sizes as shown in Eq.(1). For the purposes of this demonstration, all variables in Eq.(1) except lot size are set at their average values for Roanoke County. The categorical variable *Soil Quality* has a setting of 1 for *Soil2* (and 0 for *Soil1*), while *Road* is set at 0. These settings describe the majority of land tracts in Roanoke County. *Year* is set at 1 indicating that land values are based on sales for 1997.

Bare land value for low density development is over twice as high compared to high density because more land is occupied by housing lots. New house construction cost is higher with low density because larger, more expensive homes are assumed to be built in low density settlement compared to high density. Total tract development cost and total tract values are almost 50% higher with low density.

Developed land shifts from use value to market value assessment meaning that the estimated use value assessment of land prior to development is subtracted from its estimated

Development Tract Form	*Low Density*	*High Density*
Development Landuses		
Land Occupied by Housing (ac)	16,892	1,218
Land Occupied by Infrastructure (ac)	1,468	250
Total Land (ac)	18,360	1,468
Total Number of Housing Lots	6,120	6,120
Dollar Change Relative to Predevelopment Baseline		
A. Bare Land Value	$276,213,627	$122,559,010
B. Tract Development Cost	$1,464,732,555	$1,030,619,534
C. Estimated Total Value (A+B)	$1,740,946,182	$1,153,178,544
D. Assessed Value	$1,628,092,618	$1,113,603,195
E. Tax Revenue	$18,397,447	$12,583,716
F. Annualized Sewer and Water Cost to Localities	$0	$105,708

Table 4: Estimated tax revenues and fiscal costs by development tract form. Well and septic system costs are considered for low density development. A tax rate of 0.0113 was assumed.

market value after development. Twenty six percent of developed land was assumed to be assessed as agricultural land ($543/acre) and 74 percent as forest land ($296/acre) prior to development. The market value assessment was calculated as the sum of bare land value plus new house construction cost. The tax revenue equals the sum of use value plus market value assessment times the property tax rate (1.13%). Tax revenues from the low density scenario are about 50 percent higher than from high density because of higher market value assessment.

The total sewer and water connection costs equal the cost per foot of water and sewer connecting mains times the assumed distance between the development and the connection point with the established sewer or water system. In this example, the assumed distance was 15,840 feet. Annualized costs are calculated by multiplying the capital costs by the annualization factor for a 30-year investment life and 7% interest rate (.0806). In this example, the low density scenario has more desirable fiscal impacts to localities because of higher tax revenues and lower sewer and water connection costs.

Interpretation of Results from Hydrological Model

The HSPF model incorporated into the PSE was run for water year 1996 using the interface shown in Fig. 4 for simulating the hydrologic effects of the two scenarios described earlier. Note that in the model set up, the land segments run downstream from 1 to 10, with 10 being the closest to the outlet for Back Creek (see Fig. 5). For each land segment, there is an associated river reach. The hydrologic model simulation generates runoff volumes from each of the ten segments and reaches. The runoff volume at the outlet of each land segment represents the onsite effects and the runoff volumes at the reaches indicate cumulative effects of land development scenarios. The effects of changes on the flood hydrograph for the entire period of simulation can also be viewed graphically (see Fig. 6 for an example).

As can be seen in Table 5, in the low density scenario there is on average about 8% increase in the annual runoff at the outlets of the land segments and river reaches. There is slight

Land Segment	Baseline (inches)	Low Density (% increase)	High Density (% increase)	River Reach	Baseline (cfs-hours)	Low Density (% increase)	High Density (% increase)
1	15.7	8.85	0.00	1	39,743	8.85	0.00
2	15.7	8.17	0.00	2	74,079	8.17	0.00
3	19.2	8.58	0.00	3	224,774	8.49	0.00
4	23.9	10.5	0.00	4	272,903	8.84	0.00
5	25.0	6.29	16.20	5	383,717	8.11	4.69
6	23.6	7.95	0.00	6	468,221	8.08	3.85
7	23.6	8.46	0.00	7	540,513	8.13	3.33
8	21.5	7.46	0.00	8	608,954	8.06	2.96
9	15.9	8.71	0.00	9	656,115	8.11	2.75
10	22.8	9.32	0.00	10	740,120	8.25	2.43

Table 5: Hydrologic simulation results - annual runoff (1996 water year).

variation among land segments that can be explained by the extent of development in various segments. In the high density scenario, as described earlier, it was only land segment 5 that was developed to accommodate 15,300 people. Therefore, it is only land segment 5 that is showing an increase of 16.2% in the annual runoff at its outlet. The percent increase in runoff at the outlet of river reach 5 is 4.69% and this figure keeps reducing along reaches 6 through 10 owing to river routing effects with an increase of 2.43% at the outlet of Back Creek (i.e. reach 10 outlet). In both cases, the number of people imported was kept constant at $15,300$. However, the low density scenario resulted in almost 3.5 times larger impact on annual runoff at the watershed outlet than the high density scenario, which is counter intuitive. However, a reasonable explanation can be drawn using the landuse numbers given in Table 3. It is seen that in the low density scenario the disturbed impervious fraction of land increased to a level of 2,173 acres from 337 acres in the baseline condition, representing about 545% increase in impervious land, while in the high density scenario the new level of disturbed land is of the order of 851 acres representing only a 153% increase. Therefore, due to the higher value of per capita impervious land, the low density scenario resulted in a greater impact on annual runoff volume as compared to the high density scenario. Interestingly, the ratio of percent increase in disturbed impervious land in low density and high density scenarios (i.e., 545/153) closely matches with the ratio of increase in annual runoff at the watershed outlet in these scenarios (i.e., 8.25/2.43).

5 Concluding Remarks

The long-term goal of our project is to provide a holistic approach to watershed management by an integrated assessment of the alternative landscape scenarios that occur during the urbanization/suburbanization process. On the PSE front, we plan to explore various additional aspects, as outlined in Table. 1. The operational strength of watershed management PSEs will increasingly rely on an integration of methodologies for storage, retrieval, and postprocessing of scenarios and experiments. The importance of support for such data intensive operations is increasingly underscored in scientific circles [30, 76, 77, 7]. One of the emerging areas in database research is to provide native support for domain specific analyses. This is the approach taken by the multi year, multi institution Sequoia earth science project

[78]. In the L2W context, we plan to extend this methodology to provide storage for scenario populations in a structured way, and enable management of the execution environment (e.g., HSPF) by keeping track of constraints implied by the physical characteristics of the application. This will be achieved by a one-to-one correspondence between the entities in the scenario description to, say, tables in a relational database system (RDBMS). In addition, scenario evaluation can be efficiently formulated as query answering. For example, the SQL query

```
SELECT RunOff(*)
FROM Roanoke
WHERE slope < 12 AND landuse = 'Preston Forest';
```

can be used to evaluate the runoff arising in subwatersheds that satisfy the desired conditions. Powerful query optimization algorithms have been developed [79, 80] that selectively 'push' costly GIS operations into the computational pipeline. In addition, useful conceptual abstractions for reasoning about the watershed domain and supporting the problem solving process need to be developed. The ZOO desktop experiment management system [14] has taken the first steps towards this goal by providing a compositional modeling environment for data collection, pre-processing, and management of experiments. However, ZOO lacks decision support capabilities and will require fairly detailed domain modeling before application to watershed management. The connections to GIS based services also need to be strengthened in PSE design methodology. Wildlife and fisheries biologists were involved in the L2W project, but their data and models were not completed as of this writing. The intent of L2W is to integrate hydrologic, economic, and biological models. Finally, as mentioned in Table 1, we intend to explore the incorporation of collaboration support, optimization, and recommender systems (for selecting among various choices of simulation models) within the L2W framework.

Acknowledgements: The authors express appreciation to Kurth Stephenson, Ioannis Kaltsas, Brendon Lockhart, and Binaya Regmi for research assistance.

References

[1] National Research Council (NRC). New Strategies for America's Watersheds. Technical report, National Academy Press, Washington, DC, 1999.

[2] E.N. Houstis, J.R. Rice, E. Gallopoulos, and R. Bramley, editors. *Enabling Technologies for Computational Science: Frameworks, Middleware, and Environments.* Kluwer Academic Publishers, Boston, March 2000.

[3] E. Gallopoulos, E.N. Houstis, and J.R. Rice. Computer as Thinker/Doer: Problem-Solving Environments for Computational Science. *IEEE Computational Science and Engineering*, Vol. 1(2):pp. 11–23, 1994.

[4] E.N. Houstis, J.R. Rice, S. Weerawarana, A.C. Catlin, P. Papchiou, K.Y. Wang, and M. Gaitatzes. Parallel ELLPACK: A Problem Solving Environment for PDE Based Applications on Multicomputer Platforms. *ACM Transactions on Mathematical Software*, Vol. 24(1):pp. 30–73, March 1998.

[5] J.R. Rice and R.F. Boisvert. *Solving Elliptic Problems Using ELLPACK.* Springer Verlag, New York, 1985.

[6] D. Gannon, B. Bramley, T. Stuckey, J. Villacis, J. Balasubramanian, E. Akman, F. Breg, S. Diwan, and M. Govindaraju. Component Architectures for Distributed Scientific Problem Solving. *IEEE Computational Science and Engineering*, Vol. 5(2):pp. 50–63, May/June 1998.

[7] J.R. Rice and R.F. Boisvert. From Scientific Software Libraries to Problem-Solving Environments. *IEEE Computational Science & Engineering*, Vol. 3(3):pp. 44–53, Fall 1996.

[8] R.F. Boisvert and P.T.P. Tang, editors. *The Architecture of Scientific Software.* Kluwer Academic Publishers, Norwell, MA, 2001.

[9] A. Goel, C. Phanouriou, F. Kamke, C.J. Ribbens, C.A. Shaffer, and L.T. Watson. WBCSim: A Prototype Problem Solving Environment for Wood-Based Composites Simulations. *Engineering with Computers*, Vol. 15(2):pp. 198–210, 1999.

[10] A. Goel, C. Baker, C.A. Shaffer, B. Grossman, R.T. Haftka, W.H. Mason, and L.T. Watson. VizCraft: A Problem Solving Environment for Configuration Design of a High Speed Civil Transport. *Computing in Science and Engineering*, Vol. 3(1):pages 56–66, Jan/Feb 2001.

[11] H. Casanova and J. Dongarra. NetSolve: A Network Server for Solving Computational Science Problems. *Int. J. Supercomp. Appl. High Perf. Comput.*, Vol. 11(3):pp. 212–223, 1997.

[12] S. Markus, S. Weerawarana, E.N. Houstis, and J.R. Rice. Scientific Computing via the Web: The Net Pellpack PSE Server. *IEEE Computational Science and Engineering*, Vol. 4(3):pp. 43–51, 1997.

[13] S.G. Parker, C.R. Johnson, and D. Beazley. Computational Steering Software Systems and Strategies. *IEEE Computational Science and Engineering*, pages 50–59, October–December 1997.

[14] Y. Ioannidis, M. Livny, S. Gupta, and N. Ponnekanti. ZOO: A Desktop Experiment Management Environment. In *Proc. 22nd International VLDB Conference*, pages 274–285, 1996.

[15] N. Ramakrishnan, E.N. Houstis, and J.R. Rice. Recommender Systems for Problem Solving Environments. In H. Kautz, editor, *Technical Report WS-98-08 (Working Notes of the AAAI-98 Workshop on Recommender Systems)*, pages 91–95. AAAI/MIT Press, 1998.

[16] A.S. Grimshaw, A. Ferrari, G. Lindahl, and K. Holcomb. MetaSystems. *Communications of the ACM*, Vol. 41(11):pp. 46–55, November 1998.

[17] I. Foster and C. Kesselman. Globus: A Metacomputing Infrastructure Toolkit. *Internat. J. Supercomput. Appl.*, Vol. 11(2):pp. 115–128, 1997.

[18] A.S. Grimshaw, J.B. Weismann, E.A. West, and Ed. C. Loyot Jr. MetaSystems: An Approach Combining Parallel Processing and Heterogeneous Distributed Computing Systems. *Journal of Parallel and Distributed Computing*, Vol. 21(3), June 1994.

[19] USEPA. Overview of Current Total Maximum Daily Load — TMDL — Program and Regulations. Technical Report EPA 841-F-00-009, UOffice of Water, USEPA, Washington D.C., October 2000.

[20] P.A. DeBarry and R.G. Quimpo. GIS Modules and Distributed Models of the Watershed. American Society of Civil Engineers, 1999.

[21] ESRI. MapObjects Internet Map Server User Guide. Technical report, Environmental Systems Research Institute, 1998. URL: `http://www.esri.com/software/mapobjects`.

[22] D.P. Anderson et al. SETI@home: The Search for Extraterrestrial Intelligence. Technical report, Space Sciences Laboratory, University of California at Berkeley, 1999. URL: `http://setiathome.ssl.berkeley.edu/`.

[23] I. Foster and C. Kesselman. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Publishers, July 1998.

[24] S. Balay, W.D. Gropp, L.C. McInnes, and B.F. Smith. Efficient Management of Parallelism in Object-Oriented Numerical Software Libraries. In E. Arge, A.M. Bruaset, and H.P. Langtangen, editors, *Modern Software Tools in Scientific Computing*. Birkhauser Press, 1997.

[25] N. Ramakrishnan and A.Y. Grama. Data Mining: From Serendipity to Science (Guest Editors' Introduction to the Special Issue on Data Mining). *IEEE Computer*, Vol. 32(8):pp. 34–37, August 1999.

[26] N. Ramakrishnan and A.Y. Grama. Mining Scientific Data. *Advances in Computers*, Vol. 54, 2001. to appear.

[27] R. Agrawal, T. Imielinski, and A. Swami. Mining Associations between Sets of Items in Large Databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 207–216. ACM Press, May 1993.

[28] N. Ramakrishnan and C.J. Ribbens. Mining and Visualizing Recommendation Spaces for Elliptic PDEs with Continuous Attributes. *ACM Transactions on Mathematical Software*, Vol. 26(2):pp. 254–273, June 2000.

[29] E.N. Houstis, V.S. Verykios, A.C. Catlin, N. Ramakrishnan, and J.R. Rice. PYTHIA II: A Knowledge/Database System for Recommending/Testing Scientific Sofware. *ACM Transactions on Mathematical Software*, Vol. 26(2):pp. 227–253, June 2000.

[30] K.M. Chandy, R. Bramley, B.W. Char, and J.V.W. Reynders. Report of the NSF Workshop on Problem Solving Environments and Scientific IDEs for Knowledge, Information and Computing (SIDEKIC'98). Technical report, Los Alamos National Laboratory, 1998.

[31] S. Burgee, A.A. Giunta, V. Balabanov, B. Grossman, W.H. Mason, R. Narducci, R.T. Haftka, and L.T. Watson. A Coarse Grained Parallel Variable-Complexity Multidisciplinary Optimization Paradigm. *Internat. J. Supercomputer Appl. High Performance Comput.*, Vol. 10:pp. 269–299, 1996.

[32] J.F. Rodríguez, J.E. Renaud, and L.T. Watson. Convergence of Trust Region Augmented Lagrangian Methods using Variable Fidelity Approximation Data. *Structural Optimization*, Vol. 15:pp. 141–156, 1998.

[33] J.F. Rodríguez, J.E. Renaud, and L.T. Watson. Trust Region Augmented Lagrangian Methods for Sequential Response Surface Approximation and Optimization. *ASME J. Mech. Design*, Vol. 120:pp. 58–66, 1998.

[34] J. Rakowska, R.T. Haftka, and L.T. Watson. Multi-Objective Control-Structure Optimization Via Homotopy Methods. *SIAM Journal of Optimization*, Vol. 3:pp. 654–667, 1993.

[35] D.R. Jones, C.D. Perttunen, and B.E. Stuckman. Lipschitzian Optimization without the Lipschitz Constant. *Journal of Optimization Theory and Application*, Vol. 79(1):pp. 157–181, 1993.

[36] R.M. Lewis, V. Torczon, and M.W. Trosset. Direct Search Methods: Then and Now. *J. Computational and Applied Mathematics*, 2000. to appear.

[37] A.A. Giunta, V. Balabanov, D. Haim, B. Grossman, W.H. Mason, L.T. Watson, and R.T. Haftka. Multidisciplinary Optimization of a Supersonic Transport using Design of Experiments Theory and Response Surface Modeling. *Aero J.*, Vol. 101:pp. 347–356, 1997.

[38] B.R. Bicknell, J.C. Imhoff, J.L. Kittle, A.S. Donigian Jr., and R.C. Johnson. Hydrological Simulation Program - FORTRAN (HSPF). User's Manual for Version 11.0. Technical report, National Exposure Research Laboratory, Research Triangle Park, US EPA, NC 27711, USA, August 1997.

[39] J.L. Kittle Jr., A.M. Lumb, P.R. Hummel, P.B. Duda, and M.H. Gray. A Tool for the Generation and Analysis of Model Simulation Scenarios for Watersheds (GenScn). Technical Report 98-4134, U.S. Geological Survey Water-Resources Investigations, 1998. 152 pages.

[40] M.W. Berry, R.O. Flamm, B.C. Hazen, and R.L. MacIntyre. The Land-Use Change Analysis System (LUCAS) for Evaluating Landscape Management Decisions. *IEEE Computational Science and Engineering*, Vol. 3(1):pp. 24–25, 1996.

[41] M. Shapiro et al. GRASS Version 4.1 Programmer's Manual. Technical report, U.S. Army Corp. of Engineers' Construction Engineering Research Laboratory, Champaign, IL, March 1994.

[42] R.G. Knox, V.L. Kalb, and E.R. Levine. A Problem-Solving Workbench for Interactive Simulation of Ecosystems. In E.N. Houstis, J.R. Rice, E. Gallopoulos, and R. Bramley, editors, *Enabling Technologies for Computational Science: Frameworks, Middleware,*

and Environments, pages 159–170. Kluwer Academic Publishers, Boston, MA, March 2000.

[43] J. Andrew, J. Capilla, and E. Sanchis. AQUATOOL, A Generalized Decision-Support System for Water-Resources Planning and Operational Management. *Journal of Hydrology*, Vol. 177:pp. 269–291, 1996.

[44] USEPA. Better Assessment Science Integrating Point and Nonpoint Sources - Users Manual, Version 1.0. Technical Report EPA-823-R-96-001, Environmental Protection Agency, District Court, Washington, DC, 1966.

[45] D.L. Osmond, R.W. Gannon, J.A. Gale, D.E. Line, C.B. Knott, K.A. Phillips, M.H. Turner, M.A. Foster, D.E. Lehning, S.W. Coffey, and J. Spooner. WATERSHEDSS: A Decision Support System for Watershed Scale NonPoint Source Water Quality Problems. *Journal of the American Water Resources Association*, Vol. 33(2):pp. 327–341, 1997.

[46] H. Lal, C. Fonyo, B. Negahban, J.W. Jones, W.B. Boggess, G.A. Kiker, and K.L. Campbell. Lake Okeechobee Agricultural Decision Support System (LOADSS). ASAE Paper No. 91-2623, American Society of Agricultural Engineers, St. Joseph, Michigan, 1991.

[47] Negahban, B. and Moss, C.B. and Jones, J.W. and Zhang, J. and Boggess, W.D. and Campbell, K.L. Optimal Field Measurement for Regional Water Quality Planning. ASAE Paper No. 94-353, American Society of Agricultural Engineers, St. Joseph, Michigan, 1994.

[48] R.G. Knox, V.L. Kalb, E.R. Levine, and D.J. Kendig. A Problem-Solving Workbench for Interactive Simulation of Ecosystems. *IEEE Computational Science and Engineering*, Vol. 4(3):pp. 52–60, 1997.

[49] C.W. Chen, J. Herr, L. Ziemelis, R.A. Goldstein, and L. Olmsted. Decision Support System for Total Maximum Daily Load. *Journal of Environmental Engineering*, Vol. 125(7):pp. 653–659, 1999.

[50] D.G. Jamieson and K. Fedra. The Waterware Decision Support System for River-Basin Planning: 3. Example Applications. *Journal of Hydrology*, Vol. 177:pp. 199–211, 1996.

[51] S.M. Dunn, R. Mackay, R. Adams, and D.R. Oglethorpe. The Hydrological Component of the NELUP Decision Support System: An Appraisal. *Journal of Hydrology*, Vol. 177:pp. 213–235, 1996.

[52] G.H. Leavesley, S.L. Markstron, M.S. Brewer, and R.J. Viger. The Modular Modeling System (MMS) - The Physical Process Modeling Component of a Database-Centered Decision Support System for Water and Power Management. *Water, Air, and Soil Pollution*, Vol. 90:pp. 303–311, 1996.

[53] S.P. Simonovic and M.J. Bender. Collaborative Planning Support System — An Approach for Determining Evaluation Criteria. *Journal of Hydrology*, Vol. 177:pp. 237–251, 1996.

[54] J.R. Davis, P.M. Nanninga, J. Biggins, and P. Nad Laut. Prototype Decision Support for Analyzing Impact of Catchment Policies. *Journal of Water Resources Planning and Management*, Vol. 117:pp. 339–414, 1991.

[55] S. Carpenter, W. Brock, and P. Hanson. Ecological and Social Dynamics in Simple Models of Ecosystem Management. *Conservation Ecology*, Vol. 3(2), 1999.

[56] K. Fedra. GIS and Environmental Modeling. In M. Goodchild, B. Parks, and L. Steyaert, editors, *Environmental Modeling with GIS*, pages 35–50. Oxford University Press, 1993.

[57] M.F. Goodchild. The State of GIS for Environmental Problem-Solving. In M. Goodchild, B. Parks, and L. Steyaert, editors, *Environmental Modeling with GIS*, pages 8–15. Oxford University Press, 1993.

[58] D.R. Maidment. GIS and Hydrological Modeling. In M. Goodchild, B. Parks, and L. Steyaert, editors, *Environmental Modeling with GIS*, pages 147–167. Oxford University Press, 1993.

[59] N.H. Crawford and R.K. Linsley. Digital Simulation in Hydrology: Stanford Watershed Model IV. Technical Report 39, Department of Civil Engineering, Stanford University, July 1966.

[60] A.S. Donigian Jr., J.C. Imhoff, and J.L. Kittle Jr. HSPFParm, An Interactive Database of HSPF Model Parameters Version 1.0. AQUA TERRA Consultants, CA 94043, 1999.

[61] J.R. Philip. The Theory of Infiltration. *Soil Science*, Vol. 84:pp. 257–264, 1957.

[62] R. Arendt. *Designing Open Space Subdivisions: A Practical Step-by-Step Approach*. Natural Lands Trust, Media, PA., 1994. 96 pages.

[63] T. Schuler et al. The Benefits of Better Site Design in Residential Subdivisions. *Watershed Protection Techniques*, Vol. 3(2):pp. 633–646, 2000.

[64] K. Stephenson. Proposed Development Tracts for Development Scenarios. Unpublished Manuscript, Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, VA, 2000.

[65] I.K. Kaltsas and D.J. Bosch. Land Value Modeling in Roanoke County. Unpublished Manuscript, Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, VA, April 2000.

[66] R.S. Means Company Inc. Site Work and Landscape Cost Data 1999: 18th Annual Edition. Kingston, MA, 1999.

[67] C. Speir. Two Cost Analyses in Resource Economics: The Public Service Costs of Alternative Land Settlement Patterns and Effluent Allowance Trading in Long Island Sound. Unpublished M.S. Thesis, Virginia Tech, Blacksburg, VA, 2000.

[68] R.S. Means Company Inc. Residential Cost Data 1999. Kingston, MA, 1999.

[69] C.W. Thornthwaite. An Approach Toward a Rational Classification of Climate. *Geograph. Rev.*, Vol. 38:pp. 55–94, 1948.

[70] Dewberry and Davis. Roanoke Valley Regional Stormwater Management Plan, Study Report. Fifth Planning District Commission, Roanoke, VA, December 1996.

[71] A.S. Donigian Jr., B.R. Bicknell, and J.L. Kittle Jr. Conversion of the Chesapeake Bay Basin Model to HSPF Operation. Prepared by AQUA TERRA Consultants for the Computer Sciences Corporation, Annapolis, MD and U.S. EPA Chesapeake Bay Porgram, Annapolis, MD, 1986.

[72] N.E. Bockstael and K. Bell. *Land Use Patterns and Water Quality: The Effect of Differential Land Management Controls.* Kluwer Academic Publishers, Boston, 1998.

[73] C.E. Brodley and M.A. Friedl. Identifying Mislabeled Training Data. *Journal of Artificial Intelligence Research*, Vol. 11:pp. 131–167, 1999.

[74] M.W. Berry, J. Comiskey, and K. Minser. Parallel Analysis of Clusters in Landscape Ecology. *IEEE Computational Science and Engineering*, Vol. 1(2):pp. 24–38, Summer 1994.

[75] C.A. Abbott, M.W. Berry, J. Comiskey, L.J. Gross, and H.-K. Luh. Parallel Individual-Based Modeling of Everglades Deer Ecology. *IEEE Computational Science and Engineering*, Vol. 4(4), October-December 1997.

[76] R.W. Moore, C. Baru, R. Marciano, A. Rajasekar, and M. Wan. Data-Intensive Computing. In C. Kesselman and I. Foster, editors, *The Grid: Blueprint for a New Computing Infrastructure.* Morgan Kaufmann, 1998.

[77] R.W. Moore, T.A. Prince, and M. Ellisman. Data-Intensive Computing and Digital Libraries. *Communications of the ACM*, Vol. 41(11):pp. 56–62, November 1998.

[78] M. Stonebraker. Sequoia 2000: A Reflection on the First Three Years. *IEEE Computational Science and Engineering*, pages 63–72, Winter 1994.

[79] S. Chaudhuri and K. Shim. Optimization of Queries with User-Defined Predicates. *ACM Transactions on Database Systems*, Vol. 24(2):pp. 177–228, June 1999.

[80] J.M. Hellerstein. Optimization Techniques for Queries with Expensive Methods. *ACM Transactions on Database Systems*, Vol. 23(2):pp. 113–157, September 1998.