Collaborative Research: CyberSEES: Type 2: A New Framework for Crowd-Sourced Green Infrastructure Design

Overview: This project will develop a novel computational green infrastructure (GI) design framework that integrates interactive, neighborhood-scale, collaborative design by multiple stakeholders ("crowd-sourced" design) with multi-scale models of ecosystem and human impacts. The following **research questions** will be addressed**: (1) How well does coupling of site-scale ecohydrology with catchment-scale hydraulic routing improve predictions of nutrient dynamics and GI performance? (2) How well can stakeholder preferences in GI design be predicted using design image feature extraction and machine learning? (3) What interactive optimization and visualization techniques lead to the most rapid and complete consensus among diverse stakeholders involved in urban GI design? (4) Do stakeholders using interactive cyberinfrastructure tools consider more options and explore more of the GI design space?**

A "crowd-sourced" design framework will be developed to enable stakeholders to interactively create and evaluate potential GI designs that reflect consideration of the full breadth of social, economic, and environmental criteria. The following specific research tasks will be undertaken: (1) create integrated models to predict hydrologic, human, and ecosystem impacts of green infra-structure designs from site to catchment scales (Research Questions 1 and 2); (2) develop interactive methods for crowd-sourcing green infrastructure design (Research Question 3); and (3) implement modeling and crowd-sourced design methods in a cyberinfrastructure (CI) framework (Research Question 4).

The research questions will be evaluated in diverse neighborhoods within three urban catchments in the Baltimore Ecosystem Study, which have extensive existing data on pretreatment stormwater and nutrient conditions, and planned or ongoing GI implementation. These data will be used to calibrate and validate the hydrologic and ecosystem models. Environmental non-governmental organizations (NGOs) in Baltimore will provide access and interface with communities that are currently implementing GI. Their input will be used to evaluate and improve predictions of human GI preferences, the efficacy of the crowd-sourced design framework, and improvements in stakeholder engagement in GI design through interactive CI.

Intellectual Merit: The models developed in this project will be the first to integrate criteria for human and ecosystem wellbeing with site- and watershed-scale hydrologic processes, a key advance for improving understanding and implementation of GI design. Advancing interactive optimization approaches and model parameterization into a crowd sourcing method is novel and will have applications in many other types of design where diverse input early in the design process is important for acceptance. Map and image visualization will identify which visualization approaches are most supportive in achieving consensus in collaborative design. The project will also provide the first evaluation of interactive CI for improving stakeholder engagement in collaborative design.

Broader Impacts: The project team will work closely with community partners, ongoing studies, and GI implementation efforts in Baltimore, ensuring that the results will provide significant benefits to community stakeholders. By coordinated, collaborative planning for GI, this project will increase not only water quality but also the greening of urban spaces, with benefits to human and ecosystem health and wellbeing. Making the models accessible through simple Web interfaces can help educate and persuade people of all ages of the benefits of GI. Finally, graduate and post-doctoral students will be introduced to the valuable skills needed to integrate technology with community needs and public policy.

Collaborative Research: CyberSEES: Type 2: A New Framework for Crowd-Sourced Green Infrastructure Design

Vision Statement.

This project develops a novel computational green infrastructure (GI) design framework that integrates storm water management requirements with criteria for ecosystem health and human wellbeing. The framework synthesizes multi-scale modeling, interactive and collaborative visualization and optimization, and cyberinfrastructure (CI) to promote active community input to the planning and design of catchment restoration and management efforts.

Urbanization over the last century has contributed to increasing load from stormwater runoff and pollutants, reducing ecosystem nutrient retention, and creating poor water quality and ecosystem health downstream (NRC, 2008; Wendel et al., 2011). The loss of tree canopy and expansion of impervious area and storm sewer systems have significantly decreased infiltration and evapotranspiration, increased streamflow velocities, and increased flood risk. These problems have brought increasing attention to catchment-wide implementation of green infrastructure (e.g., decentralized green storm water management practices such as bioswales, rain gardens, permeable pavements, tree box filters, cisterns, urban wetlands, urban forests, stream buffers, and green roofs) to replace or supplement conventional storm water management practices and create more sustainable urban water systems (Dietz, 2007; Roy et al., 2008; Sullivan et al., 2010). Current GI practice has the goal of mitigating the negative effects of urbanization by maintaining or restoring pre-development hydrology (Dietz, 2007) and ultimately restoring aquatic ecosystems and addressing water quality issues at the catchment scale (Walsh et al., 2005; Filoso and Palmer, 2011; Burns et al., 2012).

Despite increasing attention to GI, currently available urban GI design methodologies cannot adequately integrate site-scale design decisions with catchment-scale impacts. Municipal stormwater managers and homeowners, for example, make decisions at the site scale (i.e., patch or parcel scales, where patches are land areas with relatively uniform physical and biological characteristics and parcels are legal or management areas that may have multiple patches) and receive credits for expected pollutant reductions, but it is difficult to estimate or verify the impacts of particular GI installations at the catchment scales that concern regulators.

Furthermore, the benefits of green infrastructure extend well beyond local storm water control, as urban green spaces (e.g., lakes, parks, and community gardens) are also major contributors both to the quality of the urban ecosystem and to human health (Morris, 2003; NRC, 2008; and Wendel et al., 2011). Quality green spaces encourage people to walk, run, cycle, play, and engage in recreation that provides healthy physical activity and reduces mental stress (Mass et al., 2006; Morris, 2003). Green spaces also improve air quality, reduce noise pollution, filter out air-borne dust and contaminants, and can partially offset greenhouse gas emissions (Dunn, 2010; Pataki et al., 2011; Pincetl, 2007). A wide range of water and nutrient capture activity by natural and quasi-natural green environments contribute to human and ecosystem wellbeing. However, these multiple benefits (or bundled ecosystem services) are not yet formally considered in GI design frameworks.

The computational design framework developed in this project will address these issues through the following **research questions: (1) How well does coupling of site-scale ecohydrology with catchment-scale hydraulic routing improve predictions of nutrient dynamics and GI performance? (2) How well can stakeholder preferences in GI design be predicted using design image feature extraction and machine learning? (3) What interactive optimization and visualization techniques lead to the most rapid and complete consensus among diverse stakeholders involved in urban GI design? (4) Do stakeholders**

using interactive cyberinfrastructure tools consider more options and explore more of the GI design space?

A novel "crowd-sourced" design framework, shown in Figure 1, will enable engineers, landscape designers, and community members (collectively called "stakeholders" hereafter) to interactively create and evaluate potential GI designs that reflect consideration of the full breadth of social, economic, and environmental criteria. The computational elements of the framework combine innovative models and optimization algorithms and a Web application user interface. The models and algorithms will be implemented in a scientific workflow system that utilizes computing resources in the Cloud; the Web interface will enable stakeholders to view, create, and rate GI designs using advanced visualization methods.

By developing this framework, our multidisciplinary team will produce tools that: (1) perform analysis of landscape patches (uniform areas of land parcels) for feasible GI practices and landscape design at site to neighborhood to catchment scales; (2) perform multicriteria optimization to select the best practices to achieve catchment-scale response targets; (3) capture stakeholder values and preferences to obtain a diverse range of solutions; and (4) allow data updates and what-if scenarios that build confidence in the model and support collaborative design of GI. Particular attention will be paid to developing and linking a new machine learning model of human preferences for green space designs, as described in Subtask 1.2 below. Human psychological impacts of GI have not been rigorously modeled before; therefore this project would create the first GI design model that considers both human and ecosystem impacts.

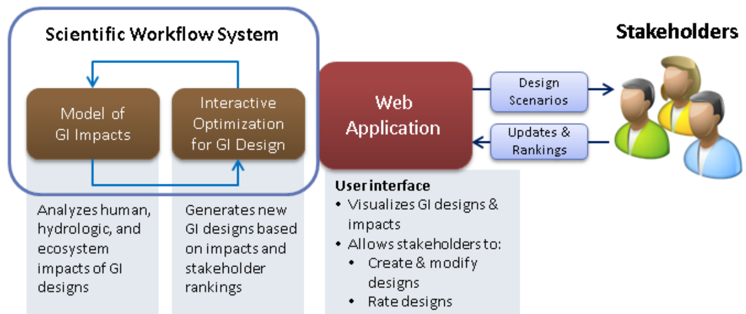


Figure 1. Computational crowd-sourced GI design framework.

Background and Significance.

Developing an effective crowd-sourced GI design system will require advances in modeling, information visualization, interactive optimization, and cyberinfrastructure. Background and significance of the proposed work in each of these areas are given below.

Green Infrastructure Modeling. Green infrastructure design guidelines provide site-specific (patch or parcel) design criteria with only qualitative discussion of catchment-scale impacts of multiple GI installations (e.g, CalTrans, 2010; City of Portland, 2008; Harper and Baker, 2007; MDE, 2009; NCDWQ, 2007). Practitioners typically use either these types of site-scale design tools or catchment-scale lumped-parameter stormwater models (e.g., MARC 2008; Vassilios, et al. 1997; and tools such as HSPF, SWMM and HEC–HMS) that do not represent detailed site-specific hydrology or GI processes.

In catchment-scale models, both traditional ("grey") and green infrastructure have typically been modeled as "edge-of-field" or "in-line" filters and sinks for stormwater runoff received from source catchment areas. Attenuation of stormwater volumes and pollutants are often included as fixed reduction percentages or first-order decay reactions based on limited input and output water quality measurements (e.g., Lee et al., 2012, in the SUSTAIN modeling framework, and Wong et al.,2001, in the MUSIC framework). Except for the use of fixed retention rates for each GI practice, or first order retention parameters to adjust for flows and temperatures, these models do little to incorporate the ecosystem processes representing the continuous cycling and storage of water, carbon, and nutrients with time-varying hydroclimate conditions over a range of local ecosystem and landscape conditions.

Furthermore, most of these models address surface water loading only and consider infiltration to be a sink, or loss from the system, without adequate coupling with groundwater. The arrangement and drainage sequence of flowpath features (e.g. from roofs to lawns, streets, and GI) has been modeled by "correcting" composite curve numbers (CN) within the SLAMM model (Pitt and Vorhees, 2002, 2011). However, this approach does not adequately incorporate dynamics or kinematics of flows, biogeochemical processes, or subsurface flow response. SLAMM has been used to simulate urban areas and export estimated stormwater nutrient loads to SWAT for larger basin nutrient loads (NRC 2008). However, SLAMM simulates event-based hydrology and does not treat subsurface flows between storms. To rigorously consider ecosystem services in GI design, including carbon sequestration and nutrient retention, requires an integrated ecosystem process approach with a continuous distributed hydrologic representation.

Research over the past decade as part of the Baltimore Ecosystem Study suggests that significant carbon sequestration and nitrogen retention can occur in a range of urban ecosystem features, including lawns, gardens, and stormwater detention structures, but that these processes are sensitive to specific characteristics of the integrated drainage system, including contributing areas, flow regimes, soils and structure design (e.g. Raciti et al., 2011a,b; Bettez and Groffman, 2012). Living components of green infrastructure will grow and adjust to prevailing water, climate, and nutrient conditions, and there may be a long, transient development of ecosystem cycling and retention capacity following development. Design of sustainable green infrastructure as either edge-of-field or at-source treatment should incorporate transient development as the ecosystem develops in response to local climate, soil, and drainage position (e.g. location within a flow field). It is critical that GI modeling extend to encompass the full catchment as a continuum beyond the discrete GI sites, including runoff source areas in addition to edge-of-field or in-line treatment systems.

This project will address these limitations by coupling the Illinois Urban Hydrologic Model (IUHM, Cantone and Schmidt, 2011), which builds a catchment-scale hydraulic routing network and performs probabilistic analysis of multiple flow paths, with RHESSys (Tague and Band, 2004), which builds a fine-scale continuous model of ecosystem patch dynamics. The underlying representation will link the ecosystem dynamics of source area patches along a drainage sequence with catchment scale routing to better simulate the site- and catchment-scale physical, chemical, and biological response of specific GI and land management practices.

Human Impacts of Green Infrastructure. There is mounting evidence that exposure to urban green spaces have profoundly positive impacts on individuals and communities. Forty years of research has established the powerful and consistent effects of the presence of natural elements in increasing preference for urban landscapes (for review, see Kaplan & Kaplan, 1989). These elements, in turn, are now associated with a variety of health benefits including faster recover from stressful experiences, reduced physiological symptoms of stress (Thompson, et al., 2012; Chang & Chen, 2005), and increased life expectancy after controlling for a host of features associated with mortality (Mitchell & Popham, 2008; Takano, et al., 2002). Exposure to these natural features has been shown to increase a person's capacity to pay attention (Berman, Jonides, & Kaplan, 2008, Taylor & Kuo, 2009) and deal with life's challenges (Kuo, 2001). Urban green spaces also seem to promote a calmness in neighbors that leads to lower levels of aggression and violence (Kuo & Sullivan, 2001a) and fewer property and violent crimes (Kuo & Sullivan 2001b).

Many dozens of studies confirm the beneficial impacts on human wellbeing of having everyday exposure to urban green spaces. But it is not only the content of green settings that predict wellbeing. The arrangement of green elements within the setting also matters (Kaplan & Kaplan, 1989). Green settings that are well organized (coherent) and display distinctive features (legible) attract people and engage them longer than less coherent or legible green settings. Green settings that have some complexity and mystery (e.g., when a path is partially concealed by foliage) are highly preferred by urban residents (Sullivan, Anderson, & Lovell, 2004).

These types of stakeholder preferences cannot be easily reduced to engineering requirements, thus visual depictions are critical to evaluating potential GI designs. Our aim is to map landscape features that are correlated with human wellbeing to features that can be extracted from design images via a novel machine learning model. This will create the first GI design tool to predict human benefits of green infrastructure.

Interactive Optimization. Interactive and collaborative design, in which multiple stakeholders are engaged in evaluating candidate design proposals, addresses recent concerns that environmental design for efficiency alone can lead to unsustainable solutions and stakeholder resistance (Ostrom, 2007; Ostrom et al., 2007; Brock and Carpenter, 2007). Interactive optimization methods have had a wide variety of applications, including shuttle scheduling (Chien et al., 199), vehicle routing (Waters 1984; ; Baker and Carreto, 2003), constraint-based graph drawings (Do Nascimiento and Eades, 2002), oceanographic campaign planning (Ibarbia et al., 2012), planning locations of supplementary recycling depots (Lin et al., 2010), and groundwater monitoring and modeling (Babbar and Minsker, 2008; Singh, et al., 2008).

In these applications, a single stakeholder or decision maker was asked to evaluate potential solutions subjectively (e.g., ranking solution preference 1–5). Including expressed preferences in the optimization can reduce the time to convergence and reach solutions that better represent the subjective preferences of decision makers (Roy, 1990; Munda, 1993; Klau et al., 2010; Babbar and Minsker, 2008; Klau et al., 2009). Interaction also allows incorporation of human skills in areas where humans outperform computers, such as visual perception, strategic thinking, and the ability to learn (Klau et al., 2010). These characteristics make interactive optimization a suitable choice for GI design, where many of the benefits to ecosystem and human wellbeing can be difficult to quantify mathematically.

These cited studies have included preferences from only one decision maker. This project would be the first to incorporate the preferences of many stakeholders via ranking aggregation techniques to create a novel framework for collaborative interactive optimization. The ranking aggregation problem has been studied extensively (Wang, et al., 2005; Fields, et al., 2012) and includes approaches such as weighted sums, simple group consensus, distance measurements, and alternative frameworks.

Information Visualization. With a diversity of stakeholders, visualization is essential for communication within the design loop (Ware, 2008). Visualization scenarios can be classified based on several characteristics: size of audience (individual, small group, public forum), level of interactivity (moments to months to incorporate changes and give feedback), and symbolic vs. representational (map vs. image). In the scope of this project, we narrow the focus to two scenarios:

1. *Design planning:* A small, map-savvy team, which may still represent diverse interests (regulators, developers, city managers, activists) can use an interactive-map-based tool to explore what-if scenarios for selecting GI treatments in various combinations of public and private lands. In our experience, providing an effective tool for small group interaction is more important than visual quality. (Zimmons and Panter, 2003; Sonnenwald et al. 2008)

2. *Public presentation:* Whether at a neighborhood association meeting or individual consultation, residents will primarily be concerned with effects of candidate GI treatments on their quality of life, including visual impact and their values as green spaces. A system that allows residents to give feedback on their preferences and on data corrections to models (e.g, where a downspout actually drains), turns their individual experiences and anecdotes into opportunities for engagement rather than distrust. (Moore & Stilgoe, 2009).

The task of creating useful visualization and design tools is complicated by the diversity of disciplines and stakeholders, by the sizes of data sets and models, and by the bewildering array of interaction technologies. Success requires an interdisciplinary team that is able to formulate a common vocabulary of data, design options, and design goals, and make this accessible to a larger group of stakeholders. This is best done through rapid development of prototypes that are anchored in specific scenarios of stakeholder interaction to mediate between conflicting goals, such as accurate modeling and interactive response times (Brooks 2010). In fact, the team interactions are the first prototype of stakeholder interactions, so the team should leverage the tools it builds for itself to support interaction with larger groups. In the tasks below, we specify initial tools, but must retain flexibility because, as Brooks observes, user reactions to prototypes are "almost invariably" surprising. The methods and tools will be tested by our own team and a few collaborating partners to identify the most promising approaches for rigorous testing in parallel neighborhoods of the test watersheds, as described in the evaluation plan.

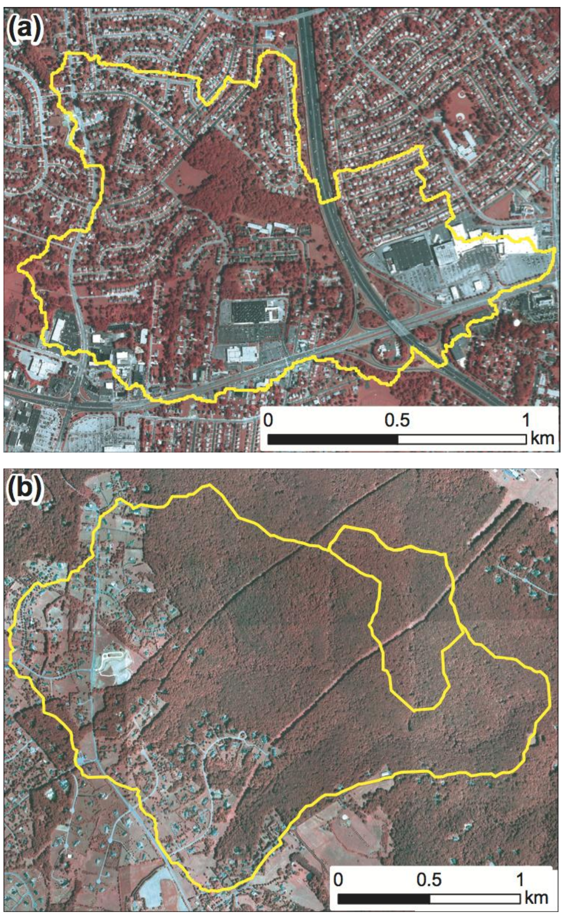
Cyberinfrastructure. A project that involves combining two hydrologic models and two machine learning algorithms, interacting with a variety of stakeholders to incorporate their preferences and values, requires software tools that support capturing *scientific workflows* and merging *data ontologies* in a way that is accessible and reproducible.

Scientific workflows chain computational steps that access, analyze, model, and visualize data in a provenance-preserving manner, allowing users to automatically capture and archive various execution configurations with associated data inputs/outputs (Deelman et al., 2005; Kooper et al., 2007; Marini et al., 2010; Moreau et al., 2012). For example, Pegasus (Deelman et al., 2004 and 2005), a large NSF-funded workflow system primarily for high-performance computing resources, allows workflows to seamlessly execute on desktops, clusters, grids, and Clouds. Cyberintegrator, an open source workflow system developed at the National Center for Supercomputing Applications (NCSA) (Kooper et al., 2007; Marini et al., 2010), supports execution of a heterogeneous set of command-line tools, such as native Java or C code, Matlab code, Excel code, graphical user interface driven code (McHenry et al., 2009; McHenry et al., 2011), and remote services, in heterogeneous computing environments (Kooper et al., 2007; Marini et al., 2010). We will use CyberIntegrator because of its ease of integrating heterogeneous software developed by multiple teams and its exploratory interface to build workflows without programming, enabling easier access by students and other domain model developers.

Data within and from the models must have a shared meaning or ontology. iRODS, the integrated Rule-Oriented Data System, is another community-driven, open source, data grid software solution that provides flexible distributed data curation via customizable collections of rules for actions during data ingestion and replication (Rajasekar 2010). iRODS’ power lies in the policy engine that overlays the file store. When multiple independent systems can rely on a centralized, stateful engine, they can work together without having to interact directly with one another. This project will couple iRODS’ capabilities with the CyberIntegrator workflow system to assist with rapid development and deployment of software prototypes that are flexible and interactive to respond to diverse user needs, reusable, and community curated.

Research Plan.

We will develop the GI design framework shown in Figure 1 and evaluate its performance in catchments that have been the subject of research in the Baltimore Ecosystem Study (BES), where community green infrastructure is being rigorously studied and implemented in diverse neighborhoods. The catchments are in distinctly different residential areas (Figure 2) that have been monitored for more than a decade, including continuous stream outflow, weekly stream chemistry, and periodic "synoptic" sampling (sampling multiple locations along the stream network). Dead Run (Figure 2a) is medium density, with separate sanitary and storm sewer drainage, up to 40% impervious cover, and bisected by Highway I-695. In contrast, residential areas in Baisman Run (Figure 2b) contain low density (2 acre zoning) land cover with ~30% lawn cover and septic sanitary systems, with ~5% impervious surface, and were developed from prior agricultural land use over the last three decades. The third catchment, Watershed 263 (Figure 2c), is in an older underserved neighborhood in southwestern Baltimore with numerous abandoned properties, and is the subject of extensive restoration efforts (Hager et al. 2013). Differences in opportunities for GI and residential preferences are probable across these three catchments.



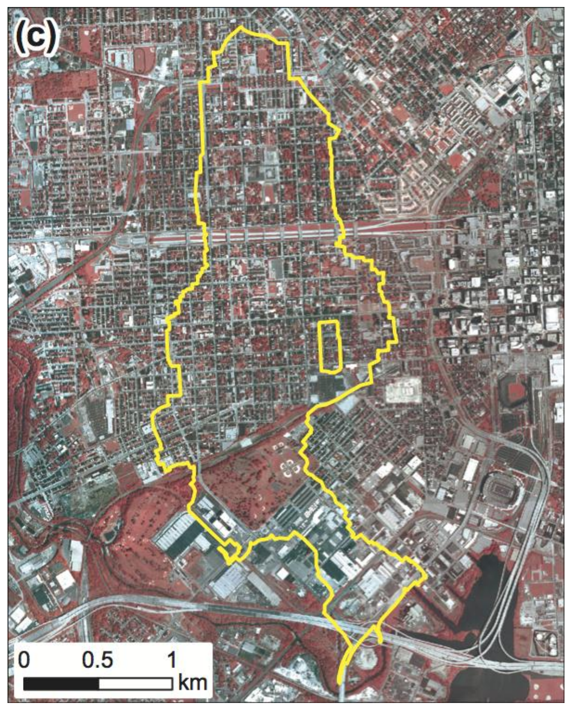


Figure 2. Three catchments instrumented by the Baltimore Ecosystem Study in Baltimore County: (a) Dead Run 5, (b) Baisman Run, and (c) Watershed 263.

The BES project team [note that University of North Carolina (UNC) Principal Investigator (PI) Larry Band serves as co-PI of the BES project] has previously surveyed homeowners in these areas on lawn care practices and sensitivity to environmental issues, and estimates of fertilizer application rates are available. Consultant Neely Law (Center for Watershed Protection) has been actively working on catchment restoration planning and GI assessment in each of these sites. Ongoing and completed work in this area will provide a baseline for comparison with new strategies created in this project using the crowd-sourced design framework. Three specific research tasks are outlined below. The evaluation section describes further tasks that will be undertaken in similar neighborhoods within each of the three catchments to rigorously assess and adapt the framework based on interactions with stakeholders.

Task 1: Create integrated models to predict hydrologic, human, and ecosystem impacts of green infrastructure designs from site to catchment scales (Research Questions 1 and 2).

We will create an integrated modeling system, shown in Figure 3, which evaluates the holistic impacts of green infrastructure from site (patch and land parcel) to neighborhood to catchment scale. We will couple the probabilistic catchment-scale network-based Illinois Urban Hydrologic Model (IUHM) (Cantone and Schmidt, 2011) with the mechanistic Regional Hydro-Ecologic Simulation System (RHESSys - Band et al 1993, 2001; Tague and Band 2004)) to link catchment- and finer-scale data and processes describing the stream and sewer network, catchment properties, green infrastructure practices, and the ecological response of specific green infrastructure practices based on the physical, chemical, and biological properties of the practices.

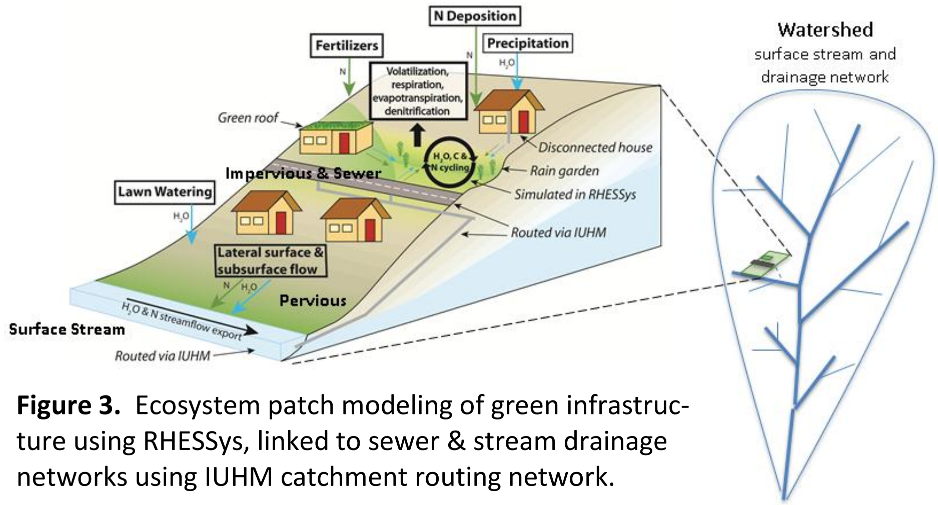


Figure 3. Ecosystem patch modeling of green infrastructure using RHESSys, linked to sewer & stream drainage networks using IUHM catchment routing network.

Human benefits of green-space designs will be estimated using supervised machine learning algorithms trained to predict human impacts of green-space images from previous research in the environmental psychology field (Kaplan & Kaplan 1989; S. Kaplan, 1995), as shown in Figure 4. These preference models will provide initial evaluations of candidate GI designs, thereby reducing the burden on stakeholders in assessing numerous potential designs that may be far from optimal or highly similar to designs that have already been evaluated. Further details on these components are given in the three subtasks below.

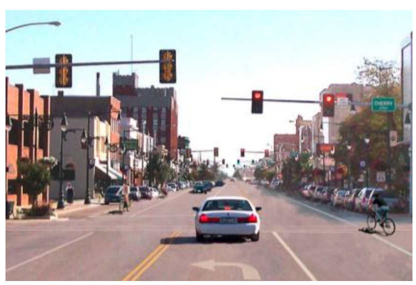


Figure 4(a). Low human preference setting.

Figure 4(b). Low human preference setting.

Sub-Task 1.1. Develop multi-scale model of GI impacts on water, carbon, and nitrogen cycling. To incorporate the transient and dynamic development and feedback of land cover patterns and networks GI components as part of an integrated site-, neighborhood-, and catchment-scale drainage system, we will couple RHESSys' distributed patch-scale ecohydrological modeling with IUHM's catchment-scale advanced hydraulic routing and drainage functions to simulate curb and pipe flow, including both green and more standard grey infrastructure (Figure 3).

A key GI modeling strategy is the decision of where and how to switch between the RHESSys ecosystem model for simulation of site-scale hydrological and biogeochemical processes, and the routing component in the IUHM network. At small, source-area (patch) scales we will simulate run-on infiltration by routing impervious area, or roof and downspout drainage, to pervious areas, such as lawns, swale drains or rain gardens, to increase infiltration, canopy interception and transpiration, and nutrient retention. At the parcel scale with fine-scale variability in land cover, we assume that it is not necessary to simultaneously solve full flow equations as the time scales typically associated with receiving pervious areas may be much greater than the small impervious routing times. This is fortunate, since the ability to represent and scale these connections to larger catchment areas is limited, as it depends on very small-scale architectural and landscape features. Therefore, we will specify the routing of runoff between patches within parcels through a combination of high-resolution lidar analysis and design specifications for high-resolution connectivity between patches (e.g., roof to down spout to rain garden), and derive the storage and release dynamics of each patch in the drainage sequence. We can then concentrate on the biogeochemistry of the patch network elements, conditioned on the flow sequence.

To do this, a mechanistic computation of coupled transport and cycling of water, carbon and nitrogen will be carried out within each patch using RHESSys and linked by the design flow topology to proximal stream and sewer drainage with IUHM during each time step. This has the advantage of explicitly representing short- to long-term ecosystem dynamics (e.g. carbon assimilation, organic matter decomposition, mineralization, etc.) within small source areas and the net retention effects of specific landscape and drainage sequence designs. Effectively, this replaces the simple runoff calculations usually used within stormwater models (such as the runoff block in SWMM), which assume uniformity in runoff producing areas, with an ecosystem patch dynamics model based on process-based ecohydrological cycling modules. RHESSys incorporates full carbon and nitrogen cycling, including photosynthetic assimilation, photosynthate allocation and growth, maintenance and growth respiration, and decomposition and mineralization, as well as nitrification, denitrification, uptake and immobilization. In addition, subtle effects of topographic heterogeneity in rain gardens, swale drains and convergence/divergence of surface and subsurface flow on soil water patterns and resulting ecosystem cycling can be represented. This approach should capture the impacts of small-scale landscape architectural and engineering patterns of runoff source areas, while also coupling carbon and nitrogen cycling impacts of the small scale patch dynamics.

In order to scale from neighborhoods and small catchments to larger urban and urbanizing catchments, we will modify the approach of Cantone and Schmidt (2011) to produce an urban ecohydrological model based on a probabilistic drainage sequence “holding time” and retention rate cascade for water, carbon and nitrogen. While Cantone and Schmidt (2011) extended the Geomorphological Instantaneous Unit Hydrograph (GIUH) of Rodriguez-Iturbe and Valdes (1979) for urban stormwater moving through flowpaths of pervious, impervious surfaces, engineered drainage infrastructure and streams, we will incorporate coupled carbon and nitrogen sources, cycling, and retention processes from RHESSys. The expanded ecohydrological process representation will be coupled with IUHM following the approach of Tang and Schmidt (2013), who have recently extended the model to include impervious to pervious transitions, with explicit representations of green roof water balance and dynamics.

Subtask 1.2. Pattern analysis for model parameterization. In order to apply the integrated model developed in Task 1.1 to any particular catchment, model parameters for existing landscapes and candidate GI treatments must be estimated. For existing landscapes, a high-resolution object classification of the landscape is needed to extract and represent features such as roofs, lawns, canopy, and impervious cover (e.g. drives, sidewalks, roads). This type of classification is increasingly available in urban areas with high-resolution airborne lidar and multispectral data and drainage design specifications that provide outflow and connectivity to engineered and natural drainage systems

The team will use LAStools, which stream large Lidar sets through processing modules and give seamless support for working on small patches within the larger set (Isenburg et al. 2006a,b). Current modules can easily smooth, identify planes, and perform other geometric operations. This project will use these operations to develop new modules capable of detecting relevant GI-related objects from Lidar data (e.g., gardens or paths); the modules will then be used to estimate model parameters (e.g., infiltration parameters). The priority for which tools are developed will be determined by data availability, importance to the model results (determined through model sensitivities at the Baltimore catchments), and importance to stakeholder preferences. We will start with simple approximations and refine the parameters and flow topology as more specific details emerge from the stakeholders using appropriate visualization tools (i.e., crowd-sourcing the model creation – for details see Task 2).

Sub-Task 1.3 Develop GI human preference model. The human benefits of particular GI designs will be evaluated by training supervised machine learning algorithms to predict human preferences and benefits of green-space images, either current or candidate designs. This will enable significant numbers of candidate designs to be evaluated without overly burdening stakeholders. The initial training data set consists of urban green space images with human preference ratings (an example is shown in Figure 3) derived from a research study led by co-PI William Sullivan. To generalize the experimental images into a predictive model of human preferences for any GI design image, the features that lead to high or low preference ratings will be extracted using image segmentation algorithms (Anami et al., 2010) and then used to train a machine learning model that predicts human preference, as shown in Figure 5.

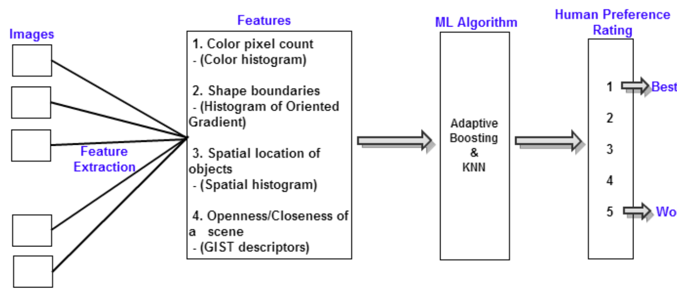


Figure 5. Human preference model.

To identify which image features to include in the model, the table below maps image segmentation algorithms to Kaplan and Kaplan (1989)'s human preference matrix for urban green spaces, which gives GI characteristics that are most linked to human wellbeing. The column on the left are landscape features that attract people and engage them longer by promoting human understanding, while the column on the right are features that encourage human exploration of the landscape.

Understanding	**Exploration**
Coherence Color histogram identifies green shapes and their layouts.	*Complexity* GIST descriptor identifies openness.
Legibility HOG identifies distinctive shapes of trees and pathways.	*Mystery* Spatial histogram identifies spatial locations of features (e.g., paths partially hidden by trees) and GIST descriptor identifies their openness.

The algorithms that will be tested for their performance in predicting human preferences include: (1) color histogram that identifies color features in the GI image (Vailaya et al., 1998); (2) histogram of oriented gradients (HOG) that identifies object shapes and boundaries, such as trees and paths (Anami et al., 2010); (3) spatial histogram that extracts the spatial location of objects in the image (Lazebnik et al., 2006) and (4) GIST descriptor that identifies the low dimensional features of the scene representing the openness, closeness, naturalness, and roughness in GI images (Oliva and Torralba , 2001).

Once relevant features are extracted using these algorithms, supervised machine learning models will be trained to predict human preferences using these features. We will use the adaptive boosting algorithm (AdaBoost, based on Freund and Schapire, 1999) which uses multiple weak classifiers (decision trees and k-Nearest Neighbors) to obtain a combined strong classifier. The predictions will be verified and adaptively improved through stakeholder and designer interaction as defined in the evaluation plan.

Task 2: Develop interactive methods for crowd-sourcing green infrastructure design (Research Question 3).

To implement the models developed in Task 1 for crowd-sourced GI design as shown in Figure 1, new methods will be developed for stakeholders and designers to interact with the models and generate novel designs that meet multiple objectives. Subtask 2.1 will develop new methods for multi-stakeholder interactive optimization of the designs, shown on the left in Figure 1. Subtask 2.2 will develop new methods for visualizing key model parameters and current GI designs and their predicted human and ecosystem performance, shown on the right in Figure 1, to provide crowd-sourced models and user rankings for the interactive optimization. These novel methods will initially be tested with the project team and a few collaborating partners (see support letters) and the most promising methods will then be rigorously tested as described in the evaluation plan.

Subtask 2.1. Develop multi-stakeholder interactive optimization methods. We will couple the models developed in Task 1 with an interactive, multi-objective optimization model to aid in identifying optimal GI designs and deployment locations that minimize costs and maximize human and ecological benefits. The interactive optimization will extend PIs' previous work: Snoeyink with developing interactive catchment analysis tools for forest management in collaboration with a Vancouver company, Facet Decision Systems (www.facet.com), resulting in their Cause and Effect product for stakeholder modeling (McAllister'99), and Minsker in groundwater monitoring design and model calibration, by enabling multiple stakeholders to interact with and vote for potential designs.

The human preference models developed in Task 1.3 will be used to generate plausible initial designs for stakeholder evaluation and learn stakeholder preferences over time, which reduces user fatigue (Singh et al. 2008). As stakeholder evaluations of candidate designs are received, ranking aggregation techniques will be used to combine crowd-sourced rankings of GI designs from multiple stakeholders into a single overall ranking of each design. A multi-objective genetic algorithm (GA) will then be used to generate new designs for further evaluation and evolve candidate designs toward those with high human and ecological benefits. Users will also be able to propose or modify designs that would be added to the population for evaluation and further modification by other users or the automated GA. Several multi-objective GAs will be tested for this purpose, including NSGA-II and SPEA (Deb et al., 2000; Zitzler et al., 1998).

To aggregate the stakeholder rankings, four methods applied to triage prioritization by Fields et al. (2013) will be evaluated and adapted to accommodate specific needs of the GI design problem: (1) Borda-Kendall method (BK), which is the most widely used method for ranking aggregation; (2) estimation of utility intervals (Wang et al., 2005); (3) weighted averaging operator (OWA) weights (Wang, Luo, &

Hua, 2007b); and (4) weight-determining mathematical programming models (Wang et al., 2007a). These methods were selected for their flexibility and adaptability to any type of problem (Fields et al., 2013).

To reduce computational effort during the optimization and enable more rapid evaluation of candidate designs, we will extract both "reduced" models, such as a simplified network to quickly detect poor strategies (e.g., GI that drains poorly and will lead to significant flooding), and "metamodels" that build a library of candidate treatments from previously evaluated designs and their approximate performance as functions of spatial and environmental parameters (e.g., the retention and filtration characteristics of rain gardens of a particular size). By recording these in the project data system (GI Datagrid, defined in Task 3.4 and the data management plan), these findings can become a resource for GI design, and can significantly speed optimization by rapidly rejecting candidate designs that do not meet design criteria, based on previous results (Yan et al. 2006, 2011).

Subtask 2.2. Develop interactive GI visualization and model parameterization.

Ware (2012) points out that problem solving with diagrams or maps is very different than without, that part of the thesis of active perception is that our brains do not try to make a model of the entire world, but is content to use the world, and therefore diagrams or maps, as external storage that can be accessed by our visual processing. When working with multiple stakeholders, a map or virtual image thus becomes a common memory as well as a common basis for communication. Thus, standard infrastructure will be used to create visualizations; the novelty comes in what map- and image-based visualizations will best support collaborative design.

To support interactive crowd sourcing of model parameters (e.g., allowing users to identify down spout locations) and optimization of GI designs optimization, we will leverage commercial and entertainment efforts to present the world and to merge real and virtual images. For example, we will initially build maps using kml for Google Maps and Google Earth, and create plan views from the depth imagery of Google Street View -- these are images that come with distances from the camera for each pixel, which Google uses for transforming their panorama views. We will use various technologies to obtain depth images for areas that have not been imaged, including a DeltaSphere camera owned by UNC (Nyland et al. 2000). By capturing and synthesizing depth images, it becomes possible to replace or merge real imagery with virtual images of possible GI treatments, model parameters (e.g., drainage locations), and estimated impacts of GI treatments (e.g., images of reductions in downstream nutrient concentrations).

The novel visualization contributions will be its use in interactive optimization, crowd-sourcing the models, and closing the loop on elucidating preferences. To address the two GI-design visualization scenarios mentioned in the background section, we will create interactive map-based displays tuned for small groups of map-savvy stakeholders (Google Maps), and image augmentation tuned for the general public (Google Street View). While both audiences will benefit from both visualizations, the purpose of the former is to support collaborative evaluation of possibilities selected by optimization, as well as tuning the optimization for data and preference updates. The purpose of the latter is clear communication of the possible visual and human wellbeing effects of proposed GI, and evaluation of visual preferences with a larger group of stakeholders.

For map-savvy stakeholders, interactivity is a key to exploring the design space and gaining confidence in the model (Sonnenwald et al. 2008). To achieve it, the models will have to produce not just numbers, but also descriptions of the topological networks that they construct. These will also need to be annotated with spatial and visual properties of the regions that appear in the network so that new images can be synthesized. Fortunately, the fidelity of an image need not be high to give an impression of the parameters and impact of a GI treatment. (Zimmons and Panter, 2003). One major research question will be how much state must be exported from the models to the visualization client to support the level of interaction needed to support effective collaborative GI design. Visualization of possible actions to improve environmental and water quality is also appealing to students of all ages. This project will produce GI demonstrations and case studies that can be widely disseminated via the Web.

Task 3: Implement modeling and crowd-sourced design methods in a cyberinfrastructure framework (Research Question 4). The design framework will incorporate four cyberinfrastructure components to enable rapid, flexible, and re-usable implementations of the models/optimization algorithms with interacting stakeholders/decision makers: 1) a scientific workflow system, 2) a distributed data management system, 3) a Web application (Web-based user interface), and 4) underlying compute resources (Figure 6).

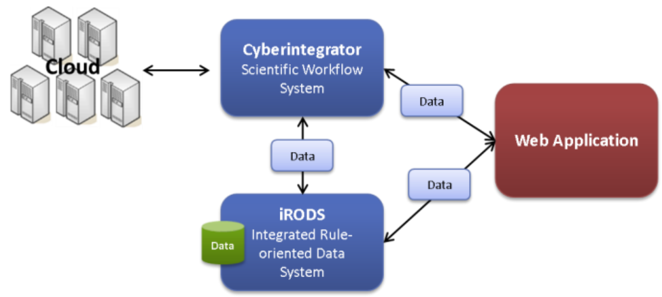


Figure 6. Cyberinfrastructure framework for interactive GI design.

We will use the Cyberintegrator workflow system to provide rapid access and execution of the green infrastructure models and interactive optimization and visualization algorithms and create an integrated crowd-sourced design framework that interacts with users via a Web interface (see Figure 1). Data used and generated from the workflows and users' interactions will be stored, accessed, and managed via iRODS, creating a *Green Infrastructure Datagrid (GI Datagrid)*. A constructed Web application will provide the user interface for visualizing alternative GI designs and interactive user inputs such as ratings and new or modified design ideas. The information that the Web application shows will be generated by the scientific workflows and retrieved from iRODS. Compute resources will be accessed by the workflows to run the models and generate and evaluate the CI designs. More details are given in the four Subtasks below.

Subtask 3.1: Link Integrated Rule-Oriented Data System (iRODS) with CyberIntegrator workflow system. The input/output data accessed within Cyberintegrator will be stored and managed in iRODS, which in turn will be accessed via the Web application to present the information and bring interactive user inputs to the models built as scientific workflows.

In order to integrate iRODS and Cyberintegrator, the data types, formats, usages, and metadata will first be identified from the models and tools developed in Tasks 1 and 2. Second, several possible access mechanisms between Cyberintegrator, iRODS, and the Web application, such as local access via file system and Web services, will be tested and evaluated. The iRODS rule engine will be used to control these access mechanisms and keep the system running smoothly and efficiently.

Subtask 3.2: Implement and test modeling and crowd-sourced design workflows. The scientific models and methods from Tasks 1 and 2 will be implemented as scientific workflows. Developing the workflows requires the following information: (1) identifying model input data and parameters; (2) identifying model output data; (3) identifying dependencies among the models; and (4) identifying the execution environment and its requirements. This information will be used to develop workflows that are responsive to both researchers' and stakeholders' needs for rapid and efficient access and re-use as the models and methods evolve in the future. The workflows will be tested with sample input data, parameters, and output data from Tasks 1 and 2 to ensure that the workflows are producing the same model results. The test results will be used to improve the workflows and the models. The workflows will be deployed on

NCSA's and RENCI's high-performance computing clusters, as well as leverage their participation in Open Science Grid (OSG) to facilitate access to national OSG resources.

*Subtask 3.3: Develop and implement the Web application.*The Web application will connect the interactive stakeholder input discussed in Tasks 2 with the workflows developed in Task 3.2 (Figure 7). The results from the workflow execution will be visualized as described in Subtask 2.2. In order to develop an efficient and user-friendly user interface, we will collect user requirements from the stakeholders in terms of interface design and functionality and adapt the interface based on project team and stakeholder feedback.

Subtask 3.4: Develop and implement Green Infrastructure Datagrid (GI Datagrid). Although our primary aim is the creation of models and interactive tools, this does require curation of a fair amount of data; moreover, the stakeholder interactions will generate valuable research data on human responses to GI designs and the interactive computational methods, visualizations, and cyberinfrastructure. These data management needs, described in more detail in the Data Management Plan, will be met by creating GI Datagrid using iRODS, which will be federated or catalogued as appropriate with other recent DataNet efforts (e.g., DataONE, SEAD, Terra Populus, and DataConservancy).

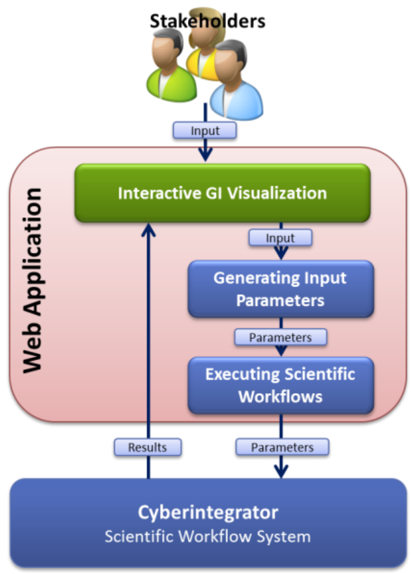


Figure 7. Web application architecture

Evaluation Plan (Objectives 3 and 4).

Konikow and Bredehoeft (1992), while supporting Popper's view that scientific models by definition cannot be validated, only invalidated, answer why models in hydrology are still useful:

> *They are a means to organize our thinking, test ideas for their reasonableness, and indicate which are the sensitive parameters … We are commonly surprised by model outputs; they provide new insights that we would not get otherwise. They serve to sharpen our professional judgment. In the end, action … will be a judgment; a professional judgment by the scientific community and a judgment by society.*

To evaluate these judgments and Research Question 1, we will calibrate flow and nutrient models and validate predictions with the extensive data collected by the Baltimore Ecosystem Study, separating data sets into calibration and testing sets. The models will be implemented in a probabilistic mode, with parameter input distributions also drawn from BES data, resulting in distributions of GI performance. Sensitivity analyses will be used to identify the most critical parameters for probabilistic evaluation. The results will then be compared with ongoing national data synthesis at the Chesapeake Biological Laboratory at the University of Maryland (see letter of support) on the variability in GI nitrogen removal due to factors such as size, age, and position within treatment trains (Koch et al., in preparation).

Through partnerships with the Center for Watershed Protection (CWP), Blue Water Baltimore (BWB), Parks and People Foundation, and Baltimore City's Healthy Harbor Initiative, we will access and interface with community organizations to evaluate Research Questions 2-4, including the machine learning predictions of human GI preferences (Question 2) , the efficacy of the interactive design framework for promoting consensus among diverse stakeholders (Question 3), and the extent to which interactive cyberinfrastructure encourages stakeholder engagement through exploration of the GI design space

(Question 4). The design team and its community partners will initially evaluate the entire suite of methods and tools developed in this project and identify the most promising approaches for more rigorous evaluation with stakeholders.

Novel methods and controls (e.g., collaborative design without the interactive optimization support) will be evaluated in parallel sets of neighborhoods selected across the three watersheds. To minimize confounding factors in answering the research questions, the neighborhood sets will be selected to ensure that the same methods are tested across diverse land use and socio-economic strata and that diverse methods are tested in the same land use and socio-economic strata. In these neighborhoods, design teams and residents will be engaged in evaluating candidate GI designs for their neighborhood and encouraged to explore “what-if” scenarios, provide feedback on model parameters and results for their area, rate others’ designs, and propose their own designs. We will also work at monitored sites where GI has been recently implemented to assess both the simulated and actual performance of the GI and predicted perception by residents. Results will be used to improve the models and framework.

Intellectual Merit.

The models developed in this project will be the first to integrate criteria for human and ecosystem wellbeing with site- and catchment-scale hydrologic processes, a key advance for improving understanding and implementation of green infrastructure design. Stormwater flow and nutrient transport will be modeled from site to catchment scales to support optimizing GI designs at neighborhood scales. The probabilistic nature of the model will allow it to apply in settings with incomplete data and address environmental variability; tools for dynamic update will allow stakeholder refinement through crowd-sourced modeling, a novel approach. The machine learning model of human preferences for green space designs will create the first GI design model that considers both human and ecosystem impacts. Advancing interactive optimization approaches into a crowd sourcing method is novel and will have applications in many other types of engineering design where community input early in the design process is recognized as important for acceptance (e.g., Guest et.al, 2009). Map and image visualization will allow stakeholders from planners to homeowners to give inputs and rank proposed solutions according to their values and preferences, identifying which visualization approaches are most supportive in achieving consensus in collaborative design. The project will also provide the first evaluation of interactive cyberinfrastructure for improving stakeholder engagement in collaborative design.

Broader Impacts.

The project team will work closely with governmental and non-governmental organizations and community members in Baltimore, ensuring that the results will provide significant benefits to community stakeholders. By coordinated, collaborative planning for GI, this project will increase not only water quality but also the greening of urban spaces, with benefits to human and ecosystem health and wellbeing, particularly in underserved and high-density neighborhoods within Watershed 263 where impacts are likely to be greatest. Making the models accessible through simple Web interfaces can help educate and persuade people of all ages of the benefits of GI and build confidence in model predictions. The project will also provide hands-on experiential learning about real-world sustainability problems for the graduate research assistants who will develop and evaluate the framework in Baltimore.

Results of Prior NSF Support:

Lawrence Band (UNC PI): Baltimore Ecosystem Study, Human settlements as ecosystems: Metropolitan Baltimore from 1797 – 2100 (DEB-9714835, \$4.2M, 1997 – 2004; Phase II -- DEB-0423476, \$4.2 M, 2004 – 2010; Phase III – DEB- 1027188; \$5.6M, 2011 – 2016). Research on the Baltimore Urban Long-Term Ecological Research (LTER) site is governed by the overarching question: What are the effects of adaptive processes aimed at sustainability in the Baltimore socio-ecological system? The project spans the Baltimore Metropolitan region and focuses on long-term stream, watershed and social survey monitoring, riparian processing of nutrients and carbon, and stream restoration. BES education programs engage youth, educators, and young scientists in investigations of the urban environment. Band's focus has been on watershed ecosystem and hydrologic dynamics, emphasizing coupled carbon, water and nitrogen cycling and export in forest through urban catchments. In the last 5 years Band has published 12 journal articles, supervised 2 PhD students (1 completed) and two masters students (1 completed).

Kenton McHenry (co-PI): National Archives and Records Administration (NARA) supplement to NSF PACI Cooperative Agreement (SCI-9619019, $2,973,713, 2010-2013). In 2010 the project addressed the problem of Understanding Preservation and Reconstruction of Electronic Records. The work resulted in a suite of tools to address the large number of file formats archives must address, including the Conversion Software Registry, Software Servers, Polyglot, and Versus. In the following year, efforts addressed several problems related to Understanding Data-Intensive and CPU Intensive Services to Support Preservation and Reconstruction of Electronic Records. Additionally, we explored moving From Raw Census Images to Searchable Information at the Hundred Terabyte Scale. In this work we explored practical means of providing searchable access to handwritten data within large archives of images and produced an open source framework by which to do this on digitized data collections. Currently we are working to integrate these developments into a usable and accessible piece of cyberinfrastructure for the automatic and social curation of large collections of data. This work has led to 11 publications, dozens of presentations across a variety of venues, and supported 3 graduate students and 5 undergraduate students.

Principal Investigator (PI) Barbara Minsker: Cost-Effective Risk Based Corrective Action Design for Contaminated Groundwater (Award Number BES 99-03889, $212,977, 8/15/99 – 8/14/02). The objective of this project was to investigate the relationships between human health risk and corrective action design under conditions of uncertainty through coupled optimization and simulation modeling. The models developed in the project combined a noisy genetic algorithm with a numerical contaminant fate and transport model and a human health risk assessment model to identify robust, minimum cost remediation designs. Novel multiscale parallel genetic algorithms were developed, along with guidelines for effective parameter setting of noisy genetic algorithms. The findings are also broadly applicable to the many scientific and engineering disciplines that solve optimization problems with uncertainty and computationally-intensive numerical models. The project resulted in four MS theses and 4 journal papers. These findings contributed towards the PI's receipt of the American Society of Civil Engineers Walter L. Huber Civil Engineering Research Prize in 2003.

Jack Snoeyink (co-PI): Degree-Driven Design of Geometric Algorithms (CISE AF 1018498, $492,552, 8/1/10 – 7/31/13) This project is developing a paradigm for designing geometric algorithms that are guaranteed correct because they take into account not only the running time and memory, but also the precision needed to execute them. In addition to new algorithms for Boolean operations (union, intersection, difference) for polygons, Voronoi diagrams, and distance transforms, this project creates tools and a library of robust geometric predicates and constructions and the beginnings of a geometric workbench/visual debugger, and will conclude with a book on this design paradigm. The project has already produced one PhD thesis, one MSc thesis, and two undergraduate research projects, and is supporting further work at graduate and undergraduate levels. The PhD graduate, Dr. David Millman, is applying the work to atomic reactor design at Bettis Laboratories.