

A cyber-enabled spatial decision support system to inventory Mangroves in Mozambique: coupling scientific workflows and cloud computing

Wenwu Tang, Wenpeng Feng, Meijuan Jia, Jiyang Shi, Huifang Zuo, Christina E. Stringer & Carl C. Trettin

To cite this article: Wenwu Tang, Wenpeng Feng, Meijuan Jia, Jiyang Shi, Huifang Zuo, Christina E. Stringer & Carl C. Trettin (2017) A cyber-enabled spatial decision support system to inventory Mangroves in Mozambique: coupling scientific workflows and cloud computing, International Journal of Geographical Information Science, 31:5, 907-938, DOI: [10.1080/13658816.2016.1245419](https://doi.org/10.1080/13658816.2016.1245419)

To link to this article: <http://dx.doi.org/10.1080/13658816.2016.1245419>



Published online: 19 Oct 2016.



Submit your article to this journal [↗](#)



Article views: 131



View related articles [↗](#)



View Crossmark data [↗](#)



A cyber-enabled spatial decision support system to inventory Mangroves in Mozambique: coupling scientific workflows and cloud computing

Wenwu Tang^{a,b}, Wenpeng Feng^{a,b}, Meijuan Jia^{a,b}, Jiyang Shi^{a,b}, Huifang Zuo^{a,b,c}, Christina E. Stringer^d and Carl C. Trettin^d

^aCenter for Applied GIScience, University of North Carolina at Charlotte, Charlotte, NC, USA; ^bDepartment of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, NC, USA;

^cDepartment of Education Leadership, University of North Carolina at Charlotte, Charlotte, NC, USA;

^dCenter for Forested Wetlands, USDA Forest Service Southern Research Station, Cordesville, SC, USA

ABSTRACT

Mangroves are an important terrestrial carbon reservoir with numerous ecosystem services. Yet, it is difficult to inventory mangroves because of their low accessibility. A sampling approach that produces accurate assessment while maximizing logistical integrity of inventory operation is often required. Spatial decision support systems (SDSSs) provide support for integrating such a sampling design of fieldwork with operational considerations and evaluation of alternative scenarios. However, this fieldwork design driven by SDSS is often computationally intensive and repetitive. In this study, we develop a cyber-enabled SDSS framework to facilitate the computationally challenging fieldwork design that requires the efficacious selection of base camps and plots for the inventory of mangroves. Our study area is the Zambezi River Delta, Mozambique. Cyber-enabled capabilities, including scientific workflows and cloud computing, are integrated with the SDSS. Scientific workflows enable the automation of data and modeling tasks in the SDSS. Cloud computing offers on-demand computational support for interoperation among stakeholders for collaborative scenario evaluation for the fieldwork design of mangrove inventory. Further, this framework allows for harnessing high-performance computing capabilities for accelerating the fieldwork design. The cyber-enabled framework provides significant merits in terms of effective coordination among science and logistical teams, assurance of meeting inventory objectives, and an objective basis to collectively and efficaciously evaluate alternative scenarios.

ARTICLE HISTORY

Received 10 July 2015

Accepted 24 September 2016

KEYWORDS

Mangrove inventory; Zambezi Delta; spatial decision support system; scientific workflow; cloud computing

1. Introduction

Spatial decision support systems (SDSSs) (Densham 1991, Sugumaran and Degroote 2010) provide an integrative computer-based framework that assist spatially explicit decision-making. SDSS have a variety of applications in natural resource assessments (Shim *et al.* 2002, Sugumaran *et al.* 2004), land management and planning (Verstegen *et al.* 2012), and

transportation management (Jankowski and Richard 1994, Ray 2007). SDSS involves the integration of diverse data and modeling components, yet a significant challenge facing conventional SDSS is that they are often ill-suited to automating the integration of these components and providing efficacious computational support. In this article, we present a SDSS framework that couples cyber-enabled capabilities, including scientific workflows, cloud computing, and high-performance computing, to gain insight into this challenge. The development of this SDSS framework is placed within the context of inventorying biomass stocks in an extensive mangrove forest in Mozambique.

Mangroves are wetland forests growing in (sub)tropical coastal saltwater environments (Valiela *et al.* 2001, Kathiresan and Rajendran 2005). Mangroves, as one of the most productive terrestrial ecosystems (Kathiresan and Rajendran 2005), are characterized by a strong linkage among coastal ecosystem health, community livelihoods, and global climate change through their effect on the carbon cycle (Fatoyinbo *et al.* 2008). However, anthropogenic activities, including deforestation and conversion to agricultural use, decrease values of ecosystem services from mangroves (Alongi 2002). Although mangroves occupy a low proportion of land surface, 1% reported from Fatoyinbo and Simard (2013), these wetlands have very high carbon density among terrestrial ecosystems (Donato *et al.* 2011). Carbon storage in these forests is significant at a global scale (Fatoyinbo and Simard 2013) and ensuring the integrity of those stocks can mitigate, for example, greenhouse gas emissions (see Kathiresan and Rajendran 2005). Accurate estimates of mangrove biomass and associated carbon stock are thus critical for providing a verifiable basis for the monitoring of these natural resources. While remote sensing data are increasingly available for the estimation of mangrove biomass and carbon at regional or higher levels, mangrove inventory as a field-based approach is of necessity to ensure the accuracy of the estimation for validation purpose. Conducting fieldwork of mangrove inventory has decades of history (see Alongi 2002, Fatoyinbo and Simard 2013).

While the importance of mangrove inventory has been well recognized, a rational fieldwork design for mangrove inventory often requires the integration of data, geoprocessing, and spatial analysis capabilities for alternative scenario evaluation. Although SDSS enables this integration, a number of these capabilities may be involved for the fieldwork design of the inventory. Conventional SDSS are often ill-suited to handling the coupling and computation of these diverse data and analytical capabilities, possibly implemented in alternative platforms by different developers. Therefore, conventional SDSS are facing, as Sugumaran and Degroote (2010) highlighted, challenges from technical, technological, social-organizational, and educational dimensions. First, the diverse data and spatial analytical capabilities required by SDSS often complicate the integration of the processing, analysis, and visualization of data involved (Sugumaran and Degroote 2010). SDSS users often spend considerable efforts to manually, instead of automatically, integrate these capabilities together (technical challenges). Second, users oftentimes need to reuse and share the geoprocessing and analytical capabilities in SDSS. However, the reuse and sharing of these capabilities are often inadequately supported in conventional SDSS (Medeiros *et al.* 2005). Third, the utilization of SDSS for scenario building and analysis to inform decision-making often requires considerable computing resources. Conventional SDSS are often built within standalone computing environments with limited computing support (Sugumaran and Degroote 2010, Keenan 2013). All of these above pose significant challenges for SDSS.

The advancement of cyberinfrastructure (or e-Science; Atkins *et al.* 2003, Wang 2010, Yang *et al.* 2010) has made cyber-enabled computational capabilities (e.g., high-performance computing, scientific workflows, grid computing, and cloud computing) available. Cyber-enabled capabilities create an opportunity for addressing the challenges associated with the coupling, reusing, and computation of data and analytical capabilities in SDSS. This is particularly the case for the SDSS developed for the fieldwork design of mangrove inventory. Cyber-enabled capabilities, represented by scientific workflows and cloud computing, hold great promise for the integration and sharing of data and models of mangrove inventory in a seamless, efficient, and automatic manner.

The major goals of this study were thus to (1) develop a cyber-enabled SDSS framework to support and facilitate the design of field-based inventory of mangrove forests and (2) investigate and demonstrate the capability of cyber-enabled computational technologies, represented by scientific workflows, cloud computing, and high-performance computing, for resolving computational challenges facing the inventory design. Scientific workflows offer a means of enabling and promoting automated, reusable, and sharable fusion of analytical and modeling capabilities for mangrove inventory. Cloud computing provides a platform for on-demand computational resources for group-based spatial decision-making often required by the fieldwork design. The integration of these cyber-enabled approaches in the SDSS provides unique and efficacious support for scenario building and implementing a robust and objective inventory design. Further, our SDSS provides an easy-to-use and cost-efficient platform for stakeholders to leverage cloud-based and high-performance computing for efficient spatial decision-making.

Our study area is the Zambezi River Delta, Mozambique, a large deltaic landform encompassing diverse assemblage of mangrove communities. The mangrove research project in this study area involves collective efforts among USDA Forest Service, US Agency for International Development (USAID), NASA, government agencies and Universities where the field survey was conducted in Mozambique, and the University implementing this SDSS. The fieldwork design of this mangrove inventory in our study area is associated with communication among 30–50 professionals from different domains. That is, a solution that supports group-based spatial decision-making is needed, which drives the development of the cyber-enabled SDSS framework that couples scientific workflows and cloud computing in this study.

The rest of this article is organized as follows. In [Section 2](#), we conduct reviews on mangrove inventory and spatial sampling, SDSS with scientific workflows and cloud computing. [Section 3](#) introduces the study area and data. [Section 4](#) presents the cyber-enabled SDSS framework for the fieldwork design of mangrove inventory. In [Section 5](#), we discuss the implementation of the framework. We present results of scenario analysis and give discussion in [Section 6](#). This article ends with conclusions in [Section 7](#).

2. Background

2.1. Mangrove inventory and spatial sampling design

Field assessments of mangrove forests are of importance for large-extent monitoring and assessment of mangrove biomass and carbon using, for example, remote sensing

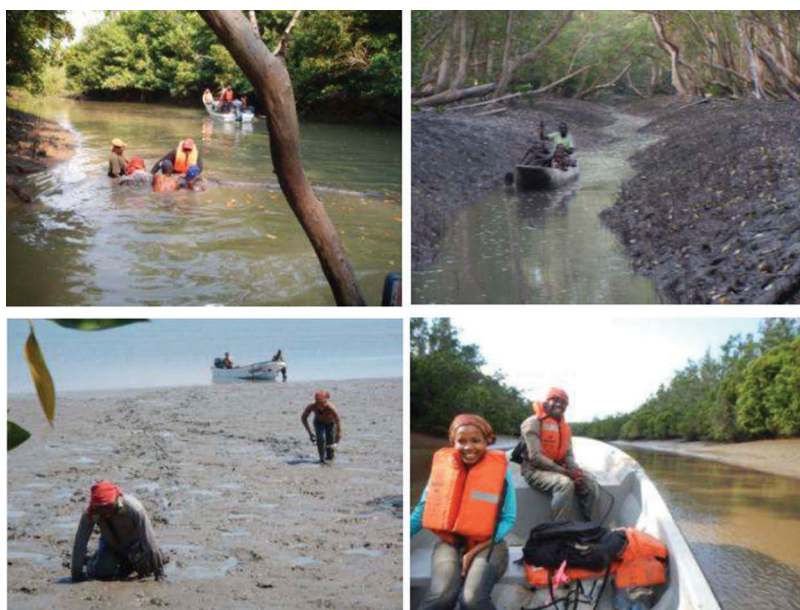


Figure 1. Physically challenging inventory in mangrove forest (photos were taken in the study area: Zambezi Delta in Mozambique; Photo credit: Dr. Carl Trettin).

approaches. Mangrove inventory is, however, extremely difficult due to the challenging physical environment (Fatoyinbo and Simard 2013; also see Figure 1 for photos associated with the mangrove inventory reported in this article). Suggested protocols for inventorying mangroves, based on traditional forest inventory procedures, have recently been published (Kauffman and Donato 2012). Due to their remote locations and accessibility issues, frequent field assessments were infeasible and most assessments of carbon pools in mangroves were based on short synoptic campaigns in selected forest areas (Krause *et al.* 2004). The synoptic approach includes a combined use of, for example, field survey, remote sensing, and geographic information systems (GIS). Consequently, the assessment of mangrove forests represents a challenge because of lack of (sufficient) considerations for spatial variability within the entire forest or assessment area. Such considerations are, however, important for designing an effective and meaningful forest inventory, vital to the estimation of carbon stocks in mangrove forests (Fatoyinbo *et al.* 2008). Ideally, the design for sample plot location incorporates characteristics of mangrove forest, while also considering costs, logistics, and operational factors (see Jena *et al.* 2012).

The design of sample locations in mangrove inventory falls within the category of spatial sampling, which allows for taking into account spatial characteristics of environmental systems of interest. The spatial sampling design (sampling plans) can be random, stratified random, or systematic (see Haining 2003). For detailed introduction to spatial sampling, the readers are directed to Haining (2003) and Delmelle (2009). The importance of arranging sample plots in a spatially optimal way has been acknowledged, and this optimal arrangement can be transformed into a facility location problem (see Church and Velle 1974, Daskin 1995, Owen and Daskin 1998). Albareda-Sambola *et al.*

(2009) designed an algorithm based on mixed integer programming to address dynamic Capacitated Facility Location Problems (CFLP). Jena *et al.* (2012) proposed a mixed integer programming approach to optimize the spatial locations of logging camps including their number and capacity. Jena *et al.* (2012) expanded the CFLP approach to a Camp Size and Location Problem by integrating such considerations as multiple periods, commodities, and capacities as well as optimal utilization of existing facilities.

2.2. Spatial decision support systems with scientific workflows and cloud computing

SDSS provide a digital computing platform that integrates the functionality of decision support systems and GIS for tackling ill- or semi-structured spatial decision problems. SDSS typically consist of four components (Armstrong *et al.* 1986): database management, modelbase management, visualization and report generation, and user interface. These components are integrated through loose, tight, or full coupling strategies (see Sengupta and Bennett 2003). Interactions among these components allow for organizing and processing geographically referenced data as well as developing spatial models based on these data to better inform decision-making. Consequently, alternative spatial decision solutions are derived and presented to decision-makers (Jankowski and Nyerges 2001). The advancement of computational science (e.g., Internet and web technologies, artificial intelligence; see Sugumaran and Sugumaran 2005) remarkably pushes the development of SDSS. The coupling of geoprocessing and modeling capabilities in SDSS allows for promoting the understanding of the nature of complex spatial decision problems. Alternative types of spatial models, including statistics, optimization, and simulation, have been employed to provide enhanced support in these SDSS.

One of the challenges facing SDSS is how to explicitly represent, manage, and share the set of data processing and models involved. The utility of scientific workflows (Taylor *et al.* 2007) in resolving this challenge has been recognized (Seffino *et al.* 1999, Tuot *et al.* 2008). Conceptually, a scientific workflow can be regarded as a directed acyclic graph (DAG), in which nodes correspond to tasks and edges represent the interrelationship among these tasks that are sequentially organized (Taylor *et al.* 2007). Scientific workflows, characterized by their sharable, extensible, and reusable features, allow for the seamless integration and automation of data and computational tasks. Scientific workflows support the representation, organization, and execution of a sequence of tasks or processing steps for the resolution of a specific computational problem, which can automate and improve reuse efficiency, and reduce the complexity of operations (Taylor *et al.* 2007). Therefore, scientific workflows hold great potential in assisting the resolution of ill- or semi-structured spatial decision problems. For example, Seffino *et al.* (1999) presented a workflow-based SDSS (WOODSS) to capture user interactions and record decision procedures in agri-environmental planning activities. Seffino *et al.* expanded the architecture of SDSS by adding WorkflowBase to enhance the participation of decision-makers. Tuot *et al.* (2008) developed a web-enabled SDSS based on a workflow management platform, Kepler (Ludäscher *et al.* 2006), to support location-aware decision-making associated with biomass yield modeling.

The computational challenge of SDSS (see Sugumaran and Degroote 2010) calls for support from cloud computing. Extended from distributed computing, cloud computing provides on-demand virtualized computing capabilities (e.g., compute, network, and storage) (Mell and Grance 2011, Yang and Huang 2013). With support from enabling technologies (e.g., virtualization and distributed file systems), cloud computing is scalable, elastic, and cost efficient. The extensive availability of cloud computing platforms (e.g., Amazon EC2, Google Cloud, and Microsoft Azure for public clouds) has encouraged the development and use of cloud computing solutions for scientific discovery in different domains (Huang *et al.* 2013, Yang and Huang 2013), particularly the domain of GIScience (e.g., see Kim and Tsou 2013, Yue *et al.* 2013). Scholars (though a few) have begun to explore the application of cloud computing in DSS in general and SDSS, in particular. For example, Jun and Jun (2011) proposed a cloud-based DSS for business study. Sun (2013) developed a web-based environmental DSS within cloud computing environments for watershed management. While the capabilities of cloud computing in assisting spatial decision have been advocated (see Yang *et al.* 2011, Keenan 2013), this form of study for SDSS remains at an early stage.

3. Study area and data

Our study area is the Zambezi River Delta (Figure 2), located in the central Mozambique (approximate longitude and latitude: 36° 28' E, 18° 34' S). The Zambezi River, the fourth

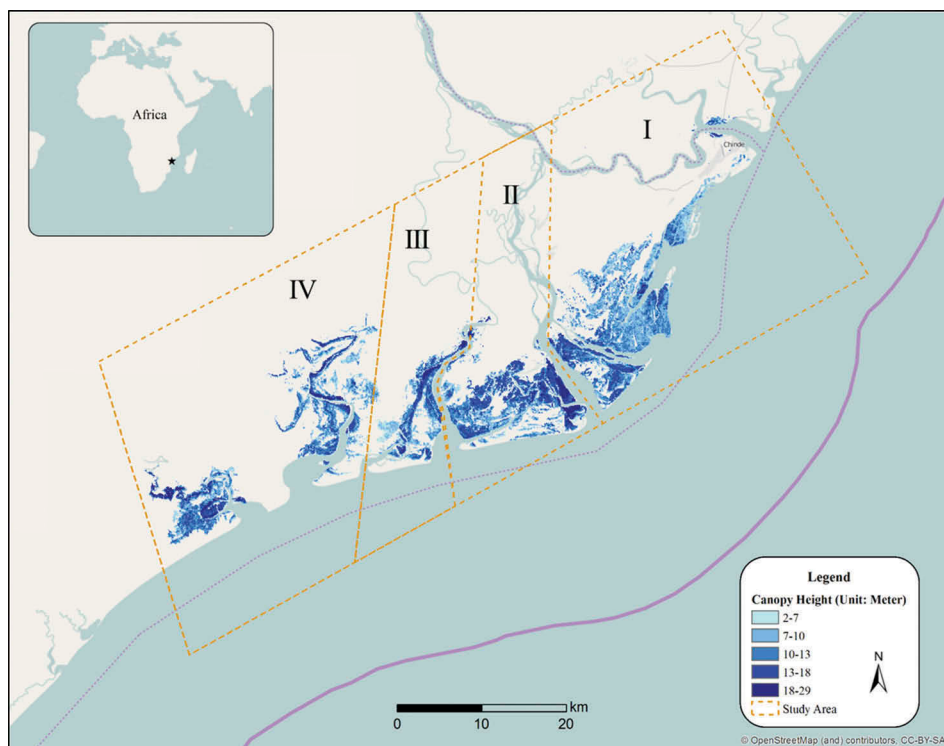


Figure 2. Map of study area of the Zambezi River Delta, Mozambique (base map: OpenStreetMap).

Table 1. List of data collected for the study area.

Data name	Data types	Data source
Canopy height of mangroves	Raster	NASA
Stream network	Vector in polygon	Digitized from remote-sensing images from Google Earth
Human settlement	Vector in point	Digitized from remote-sensing images from Google Earth
Cities	Vector in point	World Wildlife Fund-Mozambique
Roads	Vector in polyline	World Wildlife Fund-Mozambique
Coast lines	Vector in polyline	World Wildlife Fund-Mozambique
Mangrove types	Vector in polygon	World Wildlife Fund-Mozambique

Table 2. Canopy height classes of mangroves.

Class	Height (m)
1	2–7
2	7–10
3	10–13
4	13–18
5	18–29

longest river system in Africa, originates in Zambia and flows 2650 km, discharging into the Indian Ocean through the Zambezi River Delta. The region is typically characterized by a tropical climate exhibiting distinct wet and dry seasons, with a mean annual rainfall of 1130 mm and average air temperatures ranging from 16°C to 23°C (Bento and Beilfuss 2000). The delta functions as critical habitat for diverse plant and animal species. The biodiversity of the Zambezi delta makes it invaluable for wetland conservation (Bento and Beilfuss 2000).

The spatial extent of the Zambezi Delta mangroves is 302.7 km², derived from the data of canopy heights of mangroves collectively estimated from NASA Shuttle Radar Topography Mission (SRTM) data, Ice, Cloud, and land Elevation Satellite/Geoscience Laser Altimeter Systems (ICESat/GLAS) LiDAR data, and Landsat Enhanced Thematic Mapper Plus (ETM+) data (see Fatoyinbo and Simard 2013). There are nine mangrove species inhabiting the Zambezi Delta; the dominant species are *Avicennia marina* (Forssk.) Vierh., *Bruguiera gymnorhiza* (L.) Lam., *Ceriops tagal* (Perr.) C.B. Robinson, *Rhizophora mucronata* Lam., and *Sonneratia alba* Sm. The other species occurring within the Delta include *Heritiera littoralis* Ait., *Lumnitzera acemose* Willd., *Xylocarpus granatum* Koenig, and *Xylocarpus moluccensis* Koenig.

A suite of datasets has been collected for this study (see Table 1). Canopy height data for the Delta were obtained from NASA (raster format; spatial resolution: 89 m by 89 m; see Fatoyinbo and Simard 2013). We classified canopy heights of mangroves into five classes among which the mangrove coverage is balanced (see Table 2). We digitized human settlement locations and stream channels in our study area (from Google Earth). We also collected other types of spatial data, including roads, cities, and coastal boundaries.

4. Methods

4.1. Framework design

In this study, we developed a cyber-enabled SDSS framework (see Figure 3) by coupling scientific workflows and cloud computing to facilitate the design of physically

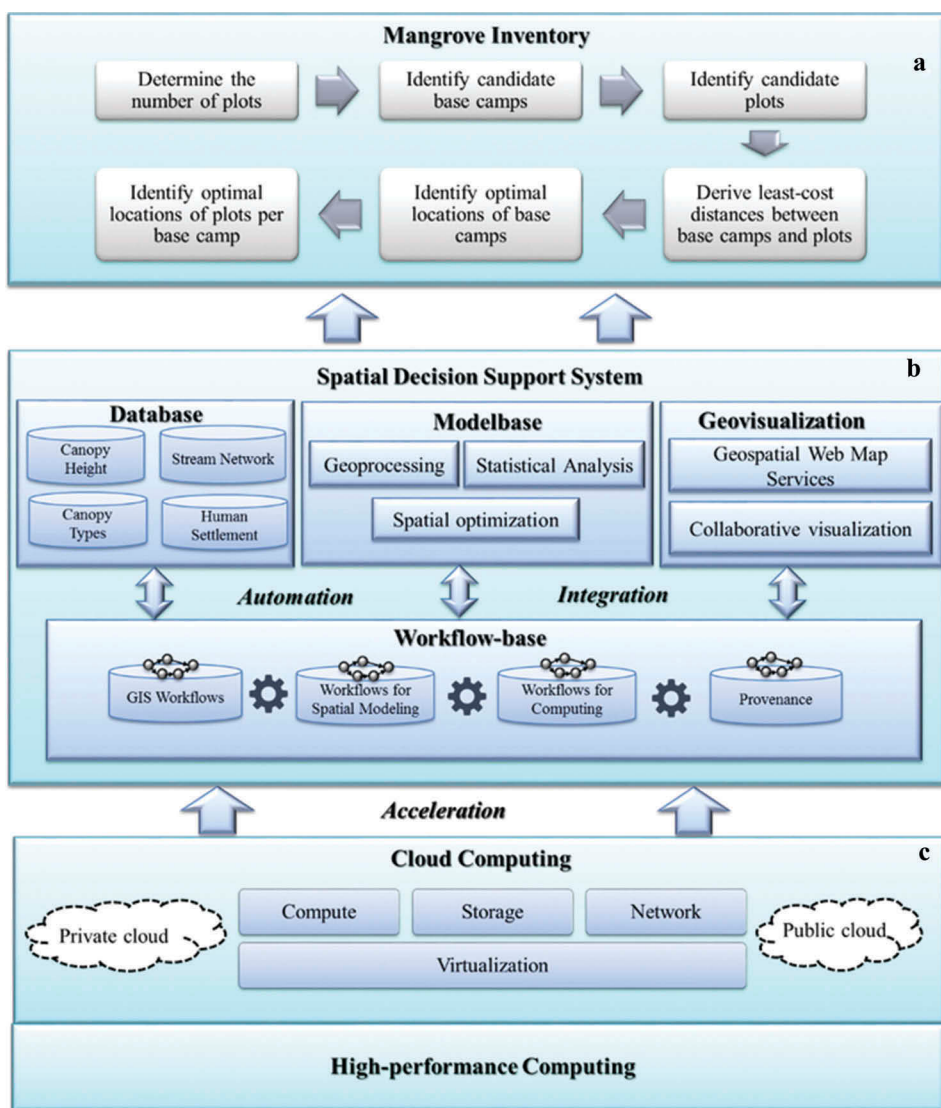


Figure 3. Conceptual architecture of the SDSS framework for mangrove inventory (a: process for identifying optimal locations of base camps and plots; b: modules required for the SDSS; c: cloud computing capabilities).

challenging fieldwork of mangrove inventory. The framework allows for the identification of optimal base camps and plots from candidate locations informed by empirical data.

The objectives of developing this SDSS framework are to (1) identify the locations of base camps that maximize the coverage of potential places for mangrove survey (ensure the success of mangrove inventory) while keeping the number of base camps as minimal as possible (lower the logistics cost of the trip; construction of base camps consumes considerable costs and efforts in mangrove forests), (2) determine the locations of plots that best represent alternative canopy height types of mangroves (ensure

the coverage of mangrove diversity, e.g., mangroves in different growing stages) and are evenly distributed among the base camps (warrant the balance of logistical needs). We used spatial optimization to solve this problem (see Deb 2001). Evaluation metrics used to optimize the locations of base camps and plots are: the total number of candidate plots covered by base camps and the total number of candidate plots covered by plots per base camp. Because warranting the safety and logistics is the top priority for conducting fieldwork in physically challenging mangrove forests, our fieldwork design places a high preference on the selection of optimal locations of base camps. Thus, our spatial optimization problem can be solved in two steps: search for optimal locations of base camps and then optimize the locations of plots to maximize the coverage of these plots.

The determination of the locations of base camps and plots for the study region requires the following steps (see Figure 3(a)): determination of the number of plots, identification of candidate base camps, identification of candidate plots, derivation of least-cost distances between base camps and plots, and identification of optimal locations of base camps and plots. For the implementation of these steps, we designed four modules in this framework (see Figure 3(b)): database management, modelbase management (see Bennett 1997), workflow management, and geovisualization. The spatial data of our study region (e.g., distributions of canopy height and type, stream network, and human settlement) are maintained in the data management module. The modelbase management module provides such functionality as statistical analyses, spatial analyses, and spatial optimization. Spatial analyses here provide analytical capabilities for geoprocessing and analysis, for example, the generation of least-cost distances serving as an important input for spatial optimization.

The geovisualization module relies on web-based mapping for the visualization of model input and outputs (e.g., locations of base camps and plots). Web-based GIS and geovisualization (see Peng and Tsou 2003, Fu and Sun 2011) facilitate interactions among stakeholders and domain experts, and support the generation of alternative scenarios to cope with diverse needs of fieldwork and stakeholder preferences. Stakeholders and domain experts could understand and discuss collaboratively for the solutions and provide their feedback.

The workflow management module is developed for coupling and automating the functionality of the SDSS for mangrove inventory. The workflow module organizes scientific workflows for different types of functionality required by the SDSS, including geoprocessing, spatial analysis, spatial optimization, geovisualization, and computation. Functionality provided by or implemented in different computing platforms or software can be encapsulated into individual scientific workflows that automate the associated computation. These individual workflows can be further integrated into composite workflows to automate the coupling of diverse functionality required by the entire SDSS for the fieldwork design of mangrove inventory (see Figure 4). While these workflows may be executed by different users, associated provenance is recorded and managed in database. Therefore, the use of scientific workflows enables the representation and management of interrelationships and dependencies among diverse data and models for the SDSS of mangrove inventory. Further, workflows of computing support the deployment and management of computing tasks on high-performance computing platforms besides local computing environments.

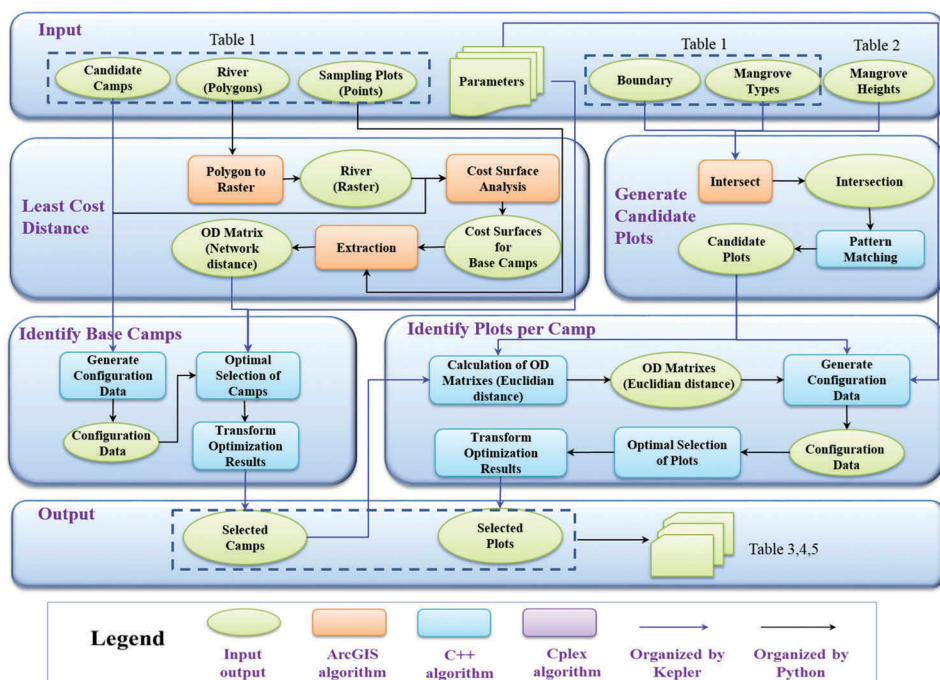


Figure 4. Scientific workflows for the selection of base camps and associated plots (OD: origin-destination).

The cloud computing layer (Figure 3(c)) provides foundation for implementing and automating computational capabilities required by the framework of mangrove inventory. Two types of service models, Infrastructure as Service (IaaS) and Software as a Service (SaaS), support the framework in need of on-demand computation for collaborative spatial decision. The data, models, scientific workflows, and provenance were maintained in virtual machines that can be instantiated from private or public clouds (i.e., IaaS). In this study, we implemented a private cloud infrastructure (discussed later) to support the dynamic organization of a pool of virtual machines. We used a load-balancing engine to diverge user requests into different virtual machines. The size of the virtual machine pool is associated with the number of user requests. Further, users can request an instance of virtual machine to improve or extend the functionality of geoprocessing, spatial optimization, and scientific workflows in the SDSS. It is convenient for users to adapt the entire SDSS into specific use cases via virtual machines.

Next, we discuss in detail the steps for the identification of base camps and plots, and the encapsulation of these steps into scientific workflows (see Figure 4).

4.2. Scientific workflows for geoprocessing

Base camps need to be allocated to locations close to human settlements (for food and fresh water supply) and stream channels (for transportation) as well as within appropriate ranges of potential sampling plots (lowering travel cost). Thus, we determine the set of candidate base camps using buffer analysis. Buffer width is a parameter that users

can specify. For example, a 1-km buffer width corresponds to 54 locations for candidate base camps (see [Figure 5\(a\)](#)).

Candidate plots are those sample plots supporting the estimation of the above-ground biomass. Candidate plots are associated with base camps identified in the previous step. To identify candidate plots, we relied on the intersection of two spatial datasets: mangrove types and canopy heights (in 2012). Considering travel cost, a criterion for identifying candidate plots is that these plots are located within a buffer (e.g., 1 km width) of the stream network. These candidate plots were then categorized into five classes regarding canopy height (see [Table 2](#) and [Figure 5\(b\)](#)). The selection criterion is: each qualified candidate plot should comprise at least four grid cells (89 m by 89 m in spatial resolution) containing the same attribute of height class and spatially forming a square pattern. A pattern matching algorithm was applied to search qualified candidate plots. The process of identifying candidate plots is encapsulated using an individual scientific workflow. 1630 candidate plots were then identified.

Given the number of base camps, and number of plots for each base camp, our objective is to maximize the coverage of plots and minimize travel cost. Thus, we need to generate origin–destination matrix to evaluate travel cost between base camps and candidate plots. In our study region, the only long-distance transportation tool available is boat. Hence, the computation of travel cost is based on the stream network. We derived travel distance based on stream network (widths were considered) and origin–destination matrices using least-cost surface analysis. Stream channels in our study area as a delta vary in width (from 5 to 4000 m; four major river systems connected with many branches). Consideration of stream width information is very important because (1) it determines whether a boat can pass through the waterway and (2) a large stream width brings in extra travel cost for side-to-side travel. If we use original vector data, calculation of interpoint distance needs to be conducted on a continuous space constrained by varied-width stream channels (i.e., these channels need to be in polygonal features instead of networks). Network analysis for least-cost routing (without consideration of stream width) is thus inappropriate for this case. Further, routing analysis on continuous space (extensively studied in robot motion; see Canny 1988, Latombe 2012) often runs into a computational intractability issue when the configuration pattern of the space is complex (polygons of stream channels are concaved). Considering these factors, we chose to use the raster-based cost surface analysis in our study area (i.e., an approximate approach instead of exact). Corresponding to candidate base camps, a collection of raster-based least-cost surfaces was generated by using GIS-based scientific workflows. Each cost surface represents the shortest distances from each single location to the corresponding base camp. Thus, we could extract the least-cost travel distance to generate distance matrices.

4.3. Scientific workflow for identifying optimal locations of base camps

Our assumption was that the investigation of one plot is completed within 1 day, and that crews will return to the corresponding base camp after completing fieldwork in each plot. A user-specified parameter of daily travel distance (e.g., 8 km) can be used: only those candidate plots within this threshold to a base camp are eligible for plots of the base camp.

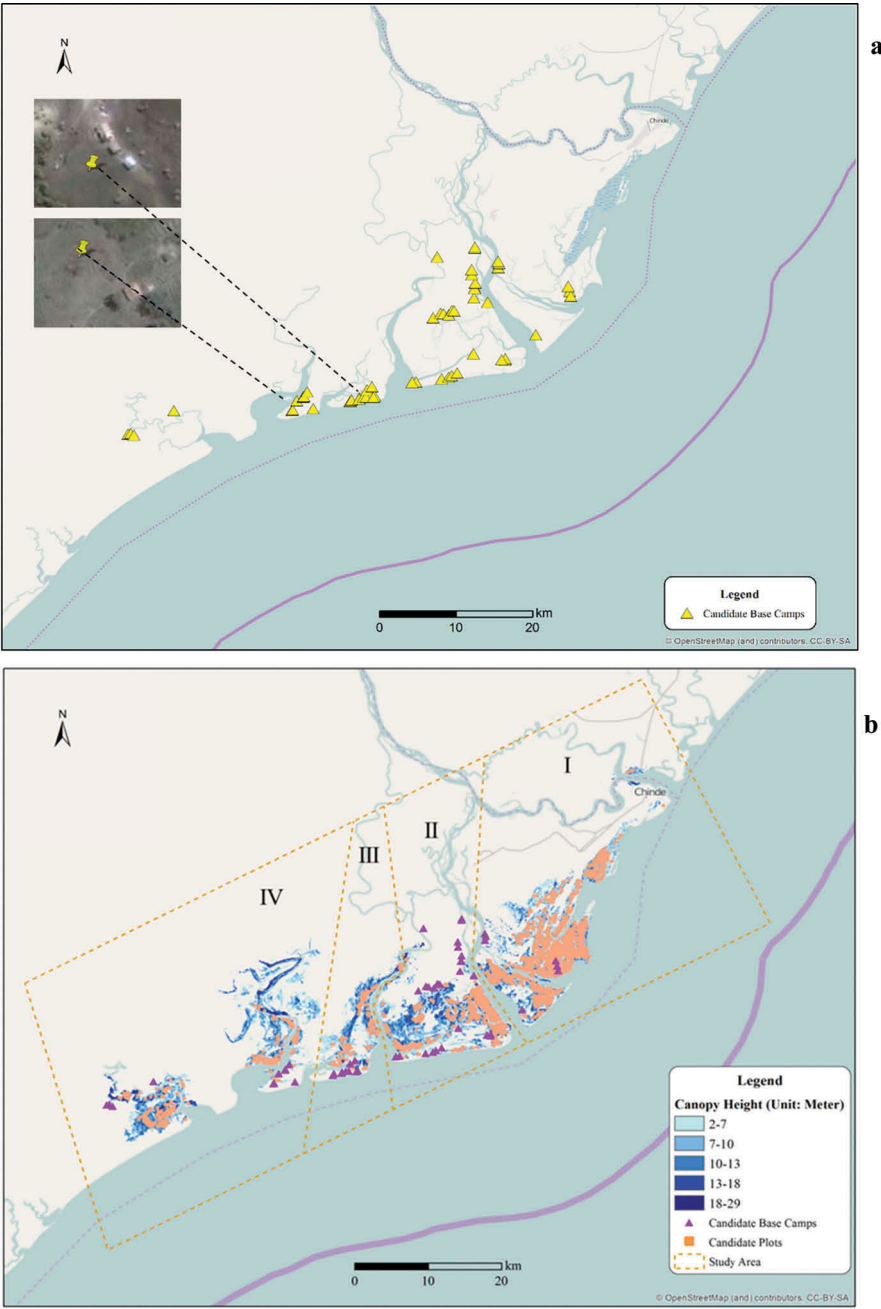


Figure 5. Spatial distribution of candidates of base camps and plots (a: candidate base camps; b: candidate plots).

The selection of optimal locations of base camps is essentially a location-allocation problem, which can be solved using spatial optimization algorithms (see Densham and Rushton 1992, Daskin 1995, Ligmann-Zielinska *et al.* 2008). The objective is to identify a number (noted as p_c) of base camps (facilities) that cover a maximum number of

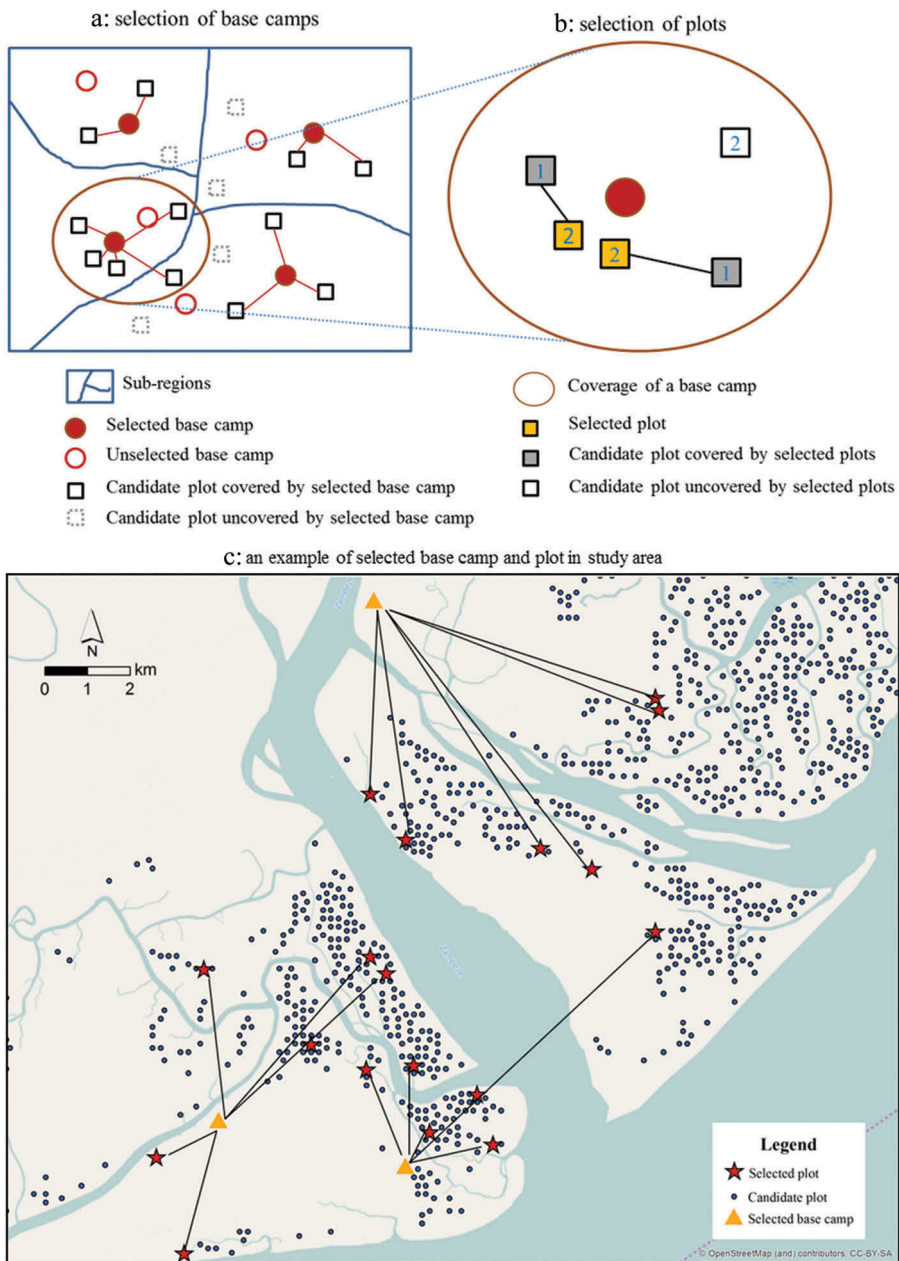


Figure 6. Optimization of the selection of base camps and plots (a: selection of base camps; b: selection of plots for each base camp; c: an example of selected base camps and plots in study area; candidate plots selected by the optimization algorithm as survey plots are referred to as selected plots; candidate plots within the coverage area of a selected plot are referred to as covered candidate plots).

candidate plots (demands) in the entire study area (see Figure 6(a)). That is, the number of candidate plots to be visited is maximized by selecting p_c base camps. This optimization process for identifying the optimal locations of base camps is formulated as follows:

$$\text{Max} : \sum_{i=1}^n \sum_{j=1}^m X_{ij}. \quad (1)$$

Subject to

$$X_{ij} \leq a_{ij} Y_j \quad \forall i = 1, 2, \dots, n \quad \forall j = 1, 2, \dots, m$$

$$\sum_{j=1}^m X_{ij} \leq 1 \quad \forall i = 1, 2, \dots, n$$

$$\sum_{j=1}^m Y_j = p_c$$

$$\sum_{N_{js}} Y_j \geq 1 \quad \forall js = 1, 2, \dots, 4$$

$$X_{ij} = 0, 1 \quad \forall i = 1, 2, \dots, n \quad \forall j = 1, 2, \dots, m$$

$$Y_j = 0, 1 \quad \forall j = 1, 2, \dots, m,$$

where:

- i : index of candidate plots ($i = 1, 2, \dots, n$)
- j : index of candidate base camps ($j = 1, 2, \dots, m$)
- N_{js} : set of subregions, $js = 1, 2, 3, 4$ in this study
- p_c : number of base camps to be opened
- a_{ij} : binary variable to tell whether candidate plot i is covered (1) or not (0) by candidate base camp j ; this information comes from the origin–destination matrix based on stream network distance.
- X_{ij} : binary decision variable that shows whether a candidate plot is covered by a base camp or not. X_{ij} is 1 if candidate base camp j is chosen to be a base camp and candidate plot i be covered by j ; otherwise, $X_{ij} = 0$.
- Y_j : binary decision variable that shows whether a base camp is chosen or not. Y_j is 1 if candidate base camp j will be chosen to be a base camp; otherwise, Y_j is 0.

While identifying the optimal locations of base camps, we need to ensure that candidate plot i cannot be covered by candidate base camp j if j is not open ($Y_j = 0$) or i is not within the daily travel distance of site j ($a_{ij} = 0$). Each candidate plot can be covered by at most one base camp (the second constraint). Further, we aim to have at least one base camp chosen in each subregion (survey team's preference).

We encapsulate the process of identifying optimal locations of base camps into an individual scientific workflow. This workflow needs the input of origin–destination matrix that maintains least-cost distance between candidate plots and candidate camps, which is the output from the workflow of least-cost distance. The workflow mainly includes three steps: (1) reading data and parameters (e.g., number of base camps) and generating a configuration file based on the model formulation above, (2) invoking the optimization model to solve the spatial optimization defined in the configuration file, and (3)

transforming optimization results into the locations of base camps. The information of identified optimal base camps will be fed into the workflows for identifying optimal location of plots per base camp.

4.4. Scientific workflow for identifying optimal locations of plots per base camp

Once locations of base camps are determined, we need to identify the optimal locations of plots for each base camp (see Figure 6(b)). In this study, we assume that one plot represents its surrounding area with a certain radius (e.g., 2 km). Our goal is to locate a number of plots to maximize the coverage of candidate plots within the daily travel distance of a base camp (see Equation (2); also see Murray *et al.* 2007). Please see Appendix 1 for detailed formulation of this optimization process for each base camp.

$$\text{Max} : \sum_{i=1}^n \sum_{k=1}^n Z_{ik}, \quad (2)$$

where Z_{ik} is a binary decision variable showing whether a location k is covered by candidate plot i or not. Location k cannot be covered by candidate plot i if i is not open ($X_i = 0$) or k is not within the daily travel distance of plot i ($b_{ik} = 0$). Each location can only be covered by at most one plot. The number of plots in each height class should not exceed the averaged number of plots per class – the number of plots among five canopy height classes needs to be balanced.

The implementation of scientific workflows to identify optimal locations of plots per base camp is similar to the process of identifying optimal base camps. Because plots for each camp need to be determined, the scientific workflow is run for each camp. The optimized locations of plots for each camp are transformed into GIS data and geospatial web services.

5. Implementation

The SDSS framework for the selection of optimal base camps and associated plots for the design of mangrove inventory is computationally repetitive and intensive. First, the fieldwork design of mangrove inventory requires collaboration from multiple groups of decision-makers, for example, fieldwork team, development team, and local communities (i.e., multiusers). A potentially large number of repetitions of model executions in response to different parameter sets are needed while these collaborators are working together for fieldwork design (model automation is needed; i.e., implication of scientific workflows). Second, to ensure the models used for the selection of base camps and plots are robust, model evaluation using, for example, uncertainty analysis based on Monte Carlo perturbation is often needed. Third, geoprocessing and spatial optimization for the selection of base camp and associated plots may consume considerable computing resources for alternative parameter combinations. For example, a single run of our SDSS in the study area (including cost surface analysis, selection of base camps, and selection of plots per camp) may need up to 42 h of sequential computing time (see Experiment section for detail). Fourth, our SDSS requires alternative data processing and analytical functionality from multiple software packages running on different computing platforms or operating systems. We thus implemented a cyber-enabled platform for mangrove

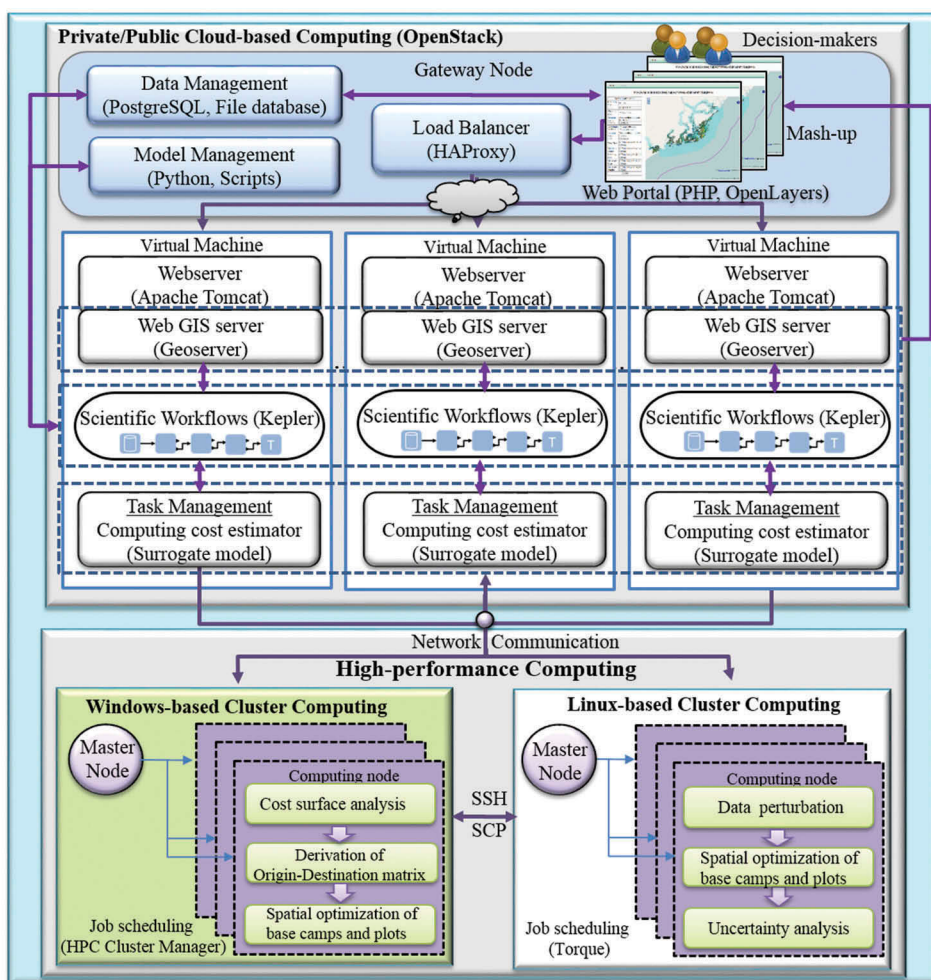


Figure 7. Cyber-enabled implementation of the spatial decision support systems framework.

inventory, Mangrovelnv (see Figure 7), which integrates scientific workflows, cloud computing, and high-performance computing to resolve the computationally repetitive and intensive challenges.

5.1. Web portal and cloud-based implementation

To facilitate the concurrent access of our SDSS for different stakeholders involved in fieldwork design, a Web GIS-based portal was implemented for Mangrovelnv (see Figure 8). Users can specify parameters and data required by the SDSS through this web-based frontend. Once computation of spatial optimization for locations of base camps and plots are completed, results can be accessed and visualized using Web GIS interface (see Figure 8). We chose Apache Tomcat (<http://tomcat.apache.org/>) within Linux as the web server for our portal. The Web GIS server that publishes GIS data into geospatial web services is GeoServer (<http://geoserver.org>). GeoServer REST API was

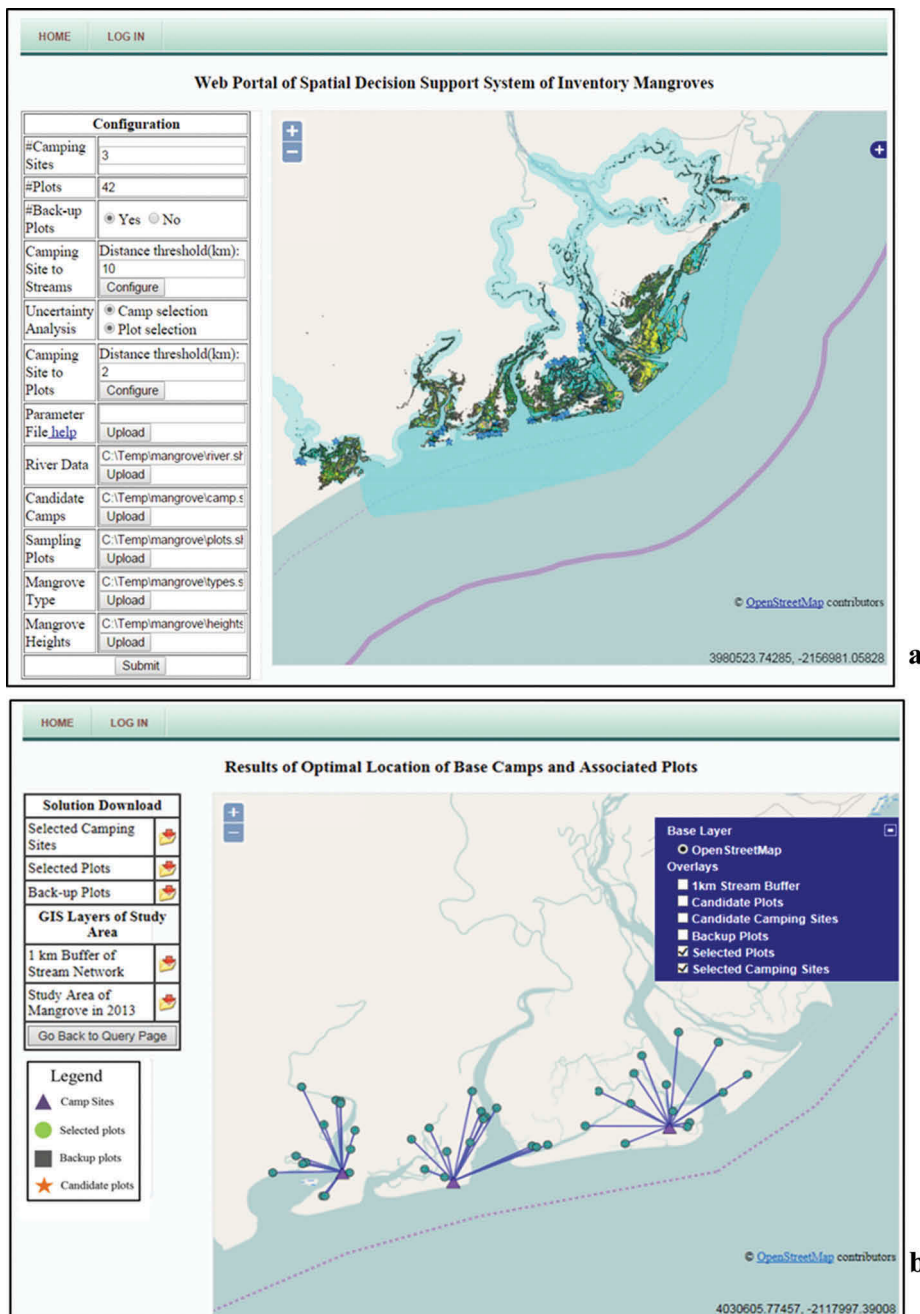


Figure 8. Web GIS portal of the spatial decision support system of mangrove inventory (a: configuration interface; b: results of identified base camps and plots; base map: OpenStreetMap).

used to automate the configuration and deployment of GIS data into web services. Server-side scripting language PHP was chosen to develop the web interface that supports alternative scenario analysis. Furthermore, an open-source JavaScript library, OpenLayers (<http://openlayers.org/>), was employed for mashing up web map services.

Mangrovelnv integrates a set of software packages and independently developed functionalities for identifying base camps and plots for mangrove inventory. We thus used cloud-based virtual machines (IaaS) for the integration of these different model and analytical functionalities. An open-source cloud platform, OpenStack (<https://www.openstack.org/>), was chosen to develop the private cloud for on-demand computation required by the SDSS. Our private cloud comprises a cluster controller and multiple computing nodes (see Tang and Feng *in press*). A dashboard of OpenStack was installed on the cluster controller to manage virtual machines and computing resource pool. Because multiple stakeholders may use Mangrovelnv concurrently, we use a load-balancer, HAProxy (High Availability Proxy; see <http://haproxy.1wt.eu/>), to balance the workload among virtual machines. HAProxy is an open source load balancer that supports a series of load-balancing algorithms including round robin. All of the key models and tools were encapsulated into virtual machines, which can be easily deployed, reused, and extended.

5.2. Workflow implementation

The workflow management system used in Mangrovelnv to couple the entire analytical and modeling process is Kepler (<https://kepler-project.org/>). Kepler is an open-source platform for designing and executing scientific workflows that connects alternative functionality and data flows. Kepler enables users to build workflows by reusing functional and data components. Further, Kepler provides an easy-to-use interface for the encapsulation of other software packages, such as R, Matlab, and ArcGIS. Kepler has been applied to alternative geospatial applications (Tuot *et al.* 2008, Pratt *et al.* 2010, Zhang 2012).

In this study, we used scientific workflows capabilities in Kepler to couple the following sub-modules implemented in different software platforms (see Figure 4): (1) GIS data processing for least-cost surface analysis, (2) generation of candidate plots, (3) selection of base camps, (4) selection of plots per camp, and (5) conversion of model outcomes for access and visualization. We realized operations between Kepler and the web portal using SSH (Secure Shell; an encrypted network protocol) and a PHP library phpseclib (<http://phpseclib.sourceforge.net/>).

The scientific workflow for geoprocessing is associated with GIS algorithms, including cost surface analysis. This is supported by ESRI ArcGIS 10 and its Python API: ArcPy. ArcPy was in charge of organizing, executing, and reusing ArcGIS functionality. We implemented other GIS analyses, including pattern matching and derivation of origin-destination matrices, using C/C++ programming language. While these scientific workflows are linked to the web portal, they can be deployed and run within standalone computing environments (e.g., laptops) when Internet access is limited.

5.3. High-performance computing support

Cost surface analysis in ArcGIS is computationally demanding: about 30 min are needed for generating cost surfaces of 54 candidate camps; multiple runs are further needed for uncertainty analysis. Thus, Mangrovelnv supports the deployment and execution of scientific workflows on Windows-based HPC clusters (the cluster used here has 40

CPUs; Operating system: Windows Server 2012 R2). We partitioned the entire computation of cost surface analysis into a set of smaller computing jobs (e.g., based on number of runs). A single computing job is executed on a CPU. We used Microsoft HPC 2012 R2 as the job manager to submit and monitor computing jobs on the Windows-based HPC cluster. This job manager allows for remotely managing computing jobs through scientific workflows.

Considering the computationally demanding nature of scientific workflows for the selection of optimal base camps and plots, Mangrovelnv supports the deployment of these workflows on Linux HPC clusters (#CPUs for the one used here: 720). Therefore, these workflows can be executed concurrently to accelerate the selection. Linear programming for spatial optimization was provided by CPLEX (CPLEX 2016). We generated the inputs of CPLEX with the linear programming format. These input files were produced with Python script for each parameter combination. We parallelized the selection of optimal camps and plots: multiple CPUs concurrently execute the same procedure of spatial optimization model based on different parameter combinations. Each CPLEX run was encapsulated into one computing job and submitted to the Linux HPC cluster through open-source job scheduler, Torque (<http://www.adaptivecomputing.com/products/open-source/torque/>).

To better guide the use of computing resources for the selection of base camps and plots, we constructed empirical models of computing time (Kriging-based surrogate models; see Forrester *et al.* 2008) in response to number of plots and daily travel distance. Figure 9 shows examples of response surfaces of computing time for a single model run of selection of plots per camp for the two variables (Kriging approach with spherical-based semivariograms was used; 100 randomly sampled parameter sets; see Experiment section for detail). A single run for the selection of plots for the third base camp may need up to 4.2 h (see Figure 9(c)). This is because the number of candidate plots associated with the base camp is large (leading to a large search space for the optimization algorithm). Based on these surrogate models, computing time for submitted jobs can be estimated in advance to leverage efficiently cyber-enabled computing power. As more jobs are completed, these models of computing time in Mangrovelnv can be further improved for refined estimation on computing cost.

6. Experiments and discussion

The cyber-enabled SDSS framework and the Mangrovelnv implementation allow for conducting experiments for the fieldwork design of mangrove inventory. In this study, we design experiments to examine the capability of the SDSS framework with respect to scenario analysis and computing performance.

6.1. Scenario analysis and model evaluation

For the fieldwork of mangrove inventory, the cost for establishing and maintaining base camps is typically high. Accordingly, the number of base camps to be established should be limited. Our experiment consists of five scenarios, S_1 , S_2 , S_3 , S_4 , S_5 , by using 3, 4, 5, 6, and 7 base camps, respectively. The buffer width for identifying candidate base camps and candidate plots close to the stream network is 1 km (maximum hiking distance from

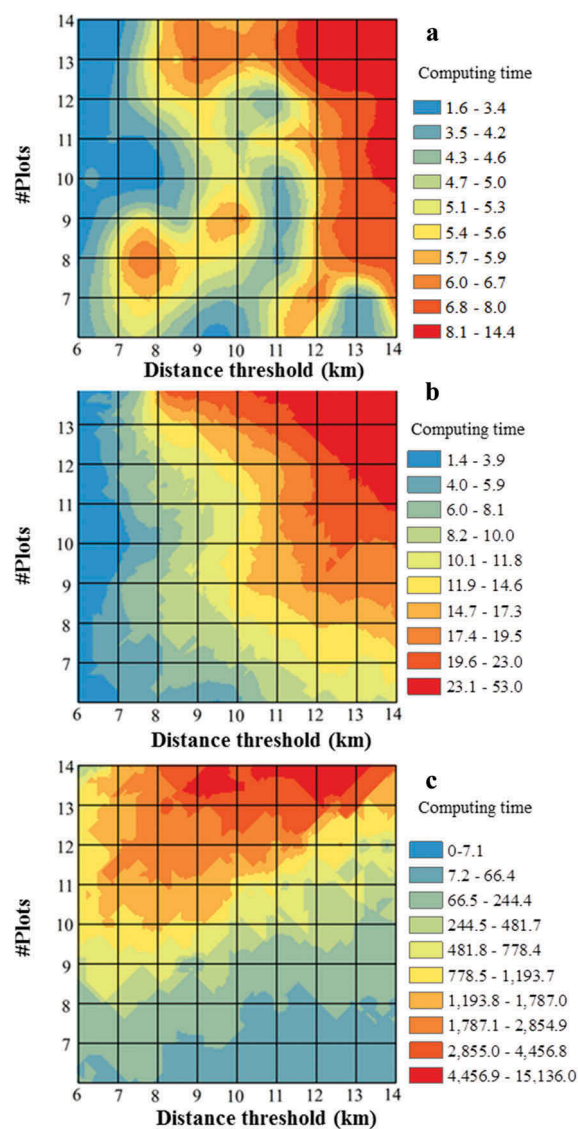


Figure 9. Response surfaces of computing time for the selection of optimal plots per base camp in a scenario (the scenario of using 3 base camps was used; a: for base camp 1, b: for base camp 2; c: for base camp 3; computing time unit: seconds).

river to base camps or plots). In total, 54 candidate base camps are thus identified. The radius of coverage area of a plot is 2 km. Based on the information from the 2012 survey, a range of 33–47 was suggested for the total number of plots. The total number of plots was chosen as 42 (recommended by the survey team). For each scenario, we ran the scientific workflows in the SDSS framework to determine the optimal locations of base camps and associated plots.

To provide better spatial decision support for the selection of base camps and plots, we used a surrogate modeling approach (Forrester *et al.* 2008). Surrogate modeling is a

simulation-based inversion approach for optimal design by using response surfaces (i.e., model of outcome; see Forrester *et al.* 2008). Surrogate modeling typically includes three steps: selection of parameter samples, construction of surrogate models (response surfaces), and model appraisal (evaluation of model quality). In this study, surrogate models of optimized objective values were constructed for the selection of plots for each candidate base camp in response to two variables: daily travel distance and number of plots. For example, three surrogate models were built for the scenario of three base camps. 100 samples of parameters were randomly drawn from the parameter space of the two variables (uniform distribution; 5–15 km for travel distance; 5–15 for number of plots per camp). Response surfaces of optimized objective values (see Figure 10) were constructed using Kriging approach (spherical semivariogram used). We applied a jackknife-based cross-validation approach (i.e., leave-one-out validation; see Efron and Efron 1982) to evaluate the performance of the surrogate models (model appraisal). Root-mean-square errors for the cross validation of the three surrogate models of objective values are 3.43, 6.33, and 7.34. This indicates these surrogate models are reliable for the selection of plots associated with their base camps. As these response surfaces show, objective values of spatial optimization tend to be high when the daily travel distance or number of plots per camp is high. This provides scientifically informed decision support for the inventory team to set the daily travel distance as long as possible. The travel distance was set to 8 km, which is the maximum distance that ensures the timely return of the inventory team within 1 day (estimated from boat's average speed).

Table 3 reports the optimization results for the five scenarios. Figure 11 depicts the spatial distributions of base camps and their associated plots for these scenarios. Scenarios with more camps tend to have higher objective values in terms of the coverage of candidate plots. Averaged coverage per camp for scenario S_1 and S_2 are close and higher than those for other scenarios. This suggests scenario S_1 and S_2 (3 or 4 base camps) are suitable for fieldwork (mangrove coverage per camp is maximized).

With respect to determining the locations of plots for base camps, the number of plots per camp for scenario S_1 is 14 (given by the spatial optimization model). One of our operational criteria is to balance the distribution and number of plots in each camp so as to cover alternative height classes and thus obtain more adequate information of mangroves. The results in Table 3 suggest that scenario S_1 tends to be superior to others (S_2 – S_5). Scenario S_1 considers the spatial distribution of plots for different height classes within and across the three camps more reasonably – the arrangement of plots for each class tends to be balanced. This facet is beneficial for capturing information within and across camps as even as possible while assuring the coverage of each height class for each camp. For example, in scenario S_2 , one camp does not cover plots with height class 5 (see Figure 11). However, in scenario S_1 , both camps capture plots with height class 5. It is thus warranted to assure the presence of each height class for plots associated with each camp for scenario S_1 . Therefore, plots for each mangrove height class are accessible from each camp.

The overall objective values of scenarios S_2 – S_5 are larger than that of scenario S_1 . However, scenarios S_2 – S_5 are incapable of covering every type of height class for each camp (see Figure 11). Meanwhile, the number of plots for each height class is uneven across different camps in other scenarios. Therefore, although 14 plots per camp in

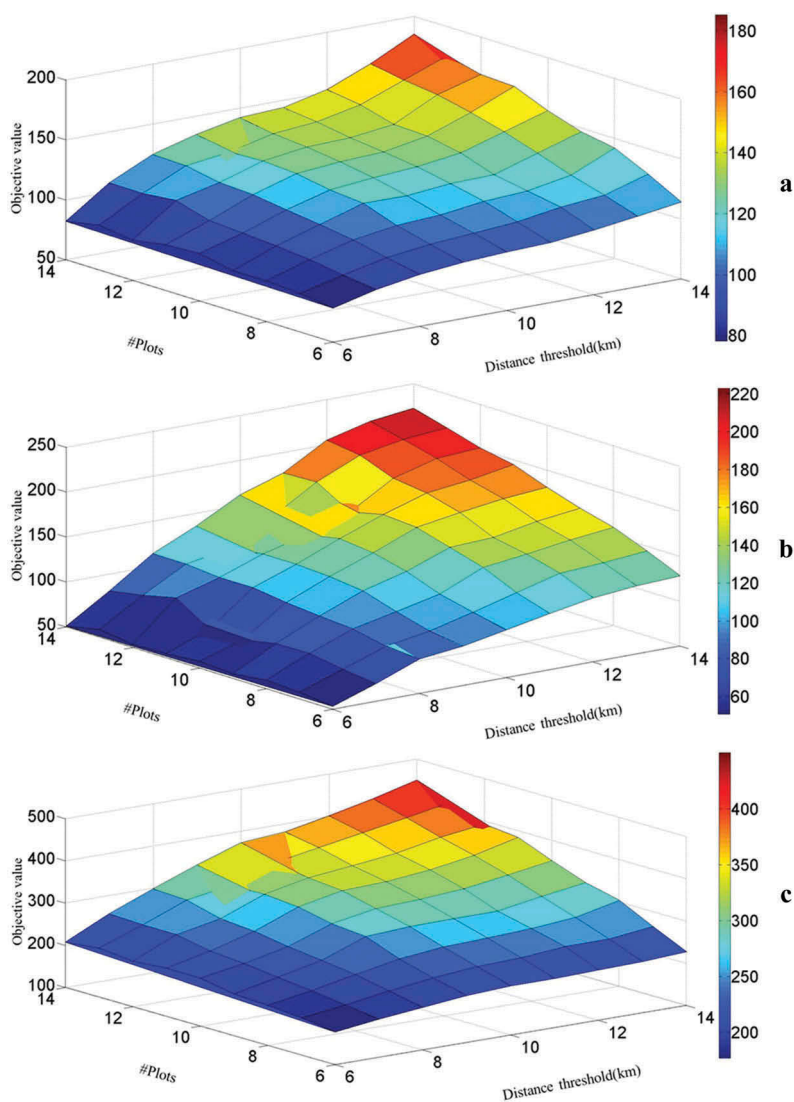


Figure 10. Response surfaces of objective values for the selection of optimal plots per base camp for a scenario (the scenario of using 3 base camps was used; a: for base camp 1, b: for base camp 2; c: for base camp 3).

scenario S_1 reach the upper limit of the number of plots for a camp (crews need to be switched every other week), the choice of only three camps in this scenario induces less construction or operating cost than other scenarios. Further, the spatial distributions of plots are similar across three camps in scenario S_1 , compared to other scenarios. Most plots in camps in other scenarios tend to cluster to each other, limited in providing sufficient support for mangrove inventory. Overall, it is suggested that scenario S_1 is preferred over others in terms of increasing the likelihood of obtaining more complete and balanced field-based information of mangroves and reducing potential cost of the

Table 3. Summary of spatial optimization results of base camps and plots.

#Base camps	Selection of base camps			Objective value (average)	Selection of plots per base camp		
	Base camp	Latitude	Longitude		#Plots	Objective value per plot	Objective value (average)
3	1	-18.905	36.0207		14	102	
	2	-18.9118	36.1043	330	14	126	609
	3	-18.8711	36.2668	(110)	14	381	(203)
4	1	-18.7582	36.2601		11	278	
	2	-18.8711	36.2668	442	11	324	830
	3	-18.9118	36.1043	(110.5)	10	126	(207.5)
	4	-18.905	36.0207		10	102	
5	1	-18.9050	36.0207	531	8	102	870
	2	-18.9128	36.0897	(106.2)	9	45	(174)
	3	-18.8711	36.2668		9	318	
	4	-18.7582	36.2601		8	277	
	5	-18.8955	36.1509		8	128	
6	1	-18.7582	36.2601	574	7	272	874
	2	-18.8711	36.2668	(95.7)	7	184	(145.7)
	3	-18.862	36.2274		7	170	
	4	-18.8955	36.1509		7	101	
	5	-18.9136	36.0851		7	45	
	6	-18.905	36.0207		7	102	
7	1	-18.9249	35.8547	598	6	113	1200
	2	-18.9050	36.0207	(85.4)	6	102	(171.4)
	3	-18.9128	36.0897		6	26	
	4	-18.8711	36.2668		6	184	
	5	-18.7582	36.2601		6	272	
	6	-18.8620	36.2274		6	342	
	7	-18.8955	36.1509		6	161	

fieldwork. Table 4 summarizes results of spatial optimization for determining optimal locations of plots per camp for scenario S_1 .

The algorithm-level selection of base camps and plots as model outputs may be affected by uncertainties in input data and models. Uncertainty analysis could be used to evaluate model robustness in response to, for example, locational accuracy of candidate camps and plots. We applied Monte Carlo perturbation on the location of candidate camps and plots. The perturbation radius is 60 m for base camps (maximal size of digitized residential settlements) and 267 m for plots (up to 3 raster cells), and the direction for perturbation is randomly drawn from a uniform distribution. The number of Monte Carlo perturbations is 100. We then reran the scientific workflows using these 100 perturbed datasets (cost surfaces per camp were regenerated). We used coefficient of variation (the ratio of standard deviation over mean; see Mann 2007) as an index to evaluate model robustness. All of the coefficient of variation values for our scenarios are very low (see Figure 12 for an example of scenario S_1), suggesting our models are robust in response to uncertainties in the locations of candidate camps and plots.

6.2. Computing performance evaluation

The overall scientific workflow in MangroveInv that implements the SDSS framework is computationally demanding. Model evaluation for the robust selection of camps and plots (e.g., using uncertainty analysis) requires even more computational support. We thus used the metric of speedup (see Equation (3); also see Wilkinson and Allen 2004),

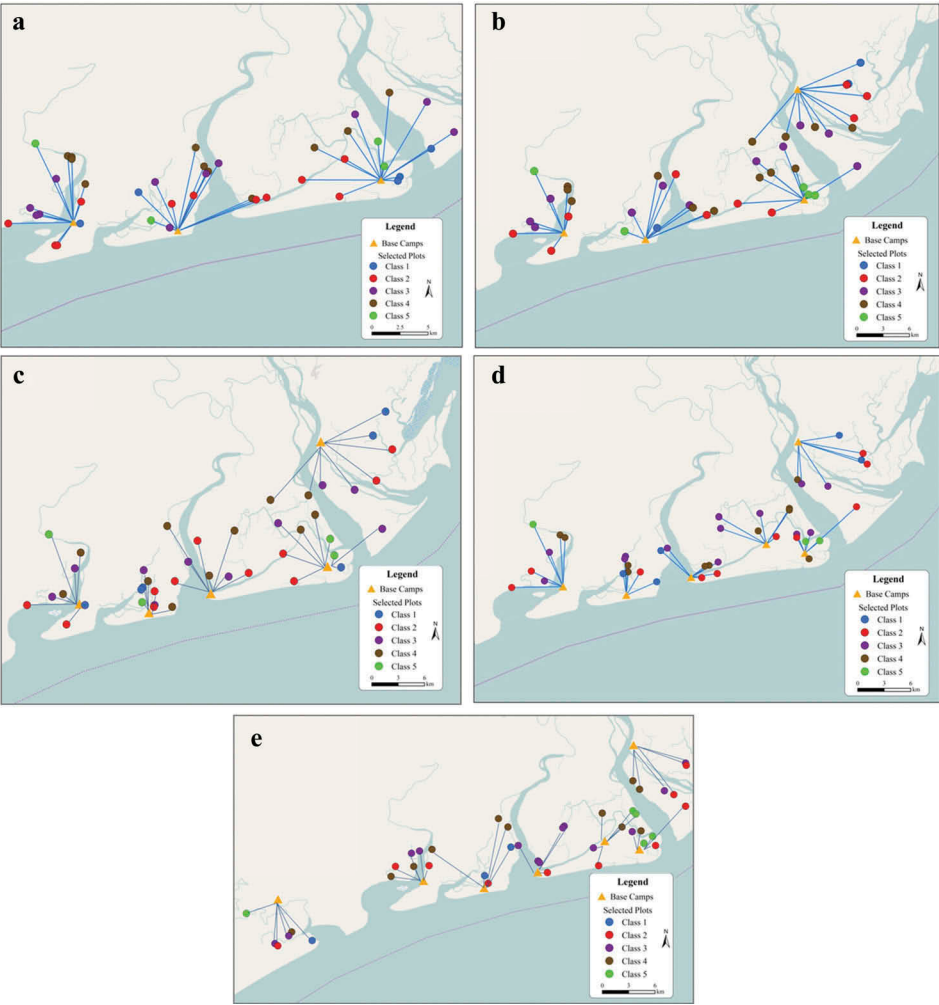


Figure 11. Map of the spatial distributions of base camps and associated plots for different scenarios (a, b, c, d, and e: number of base camps is 3, 4, 5, 6, and 7).

Table 4. Summary of spatial optimization results of plots per base camps for scenario 1.

Base camp	#Plots	Objective value	#Plots in class 1	#Plots in class 2	#Plots in class 3	#Plots in class 4	#Plots in class 5
1	14	102	1	4	4	4	1
2	14	126	1	4	4	4	1
3	14	381	3	3	3	3	2

together with computing time, to evaluate the computing performance of the SDSS framework.

$$S = T_1/T_n, \tag{3}$$

where T_1 is sequential computing time (using 1 CPU). T_n is parallel computing time using n CPUs. The larger the speedup value, the higher the computing performance is.

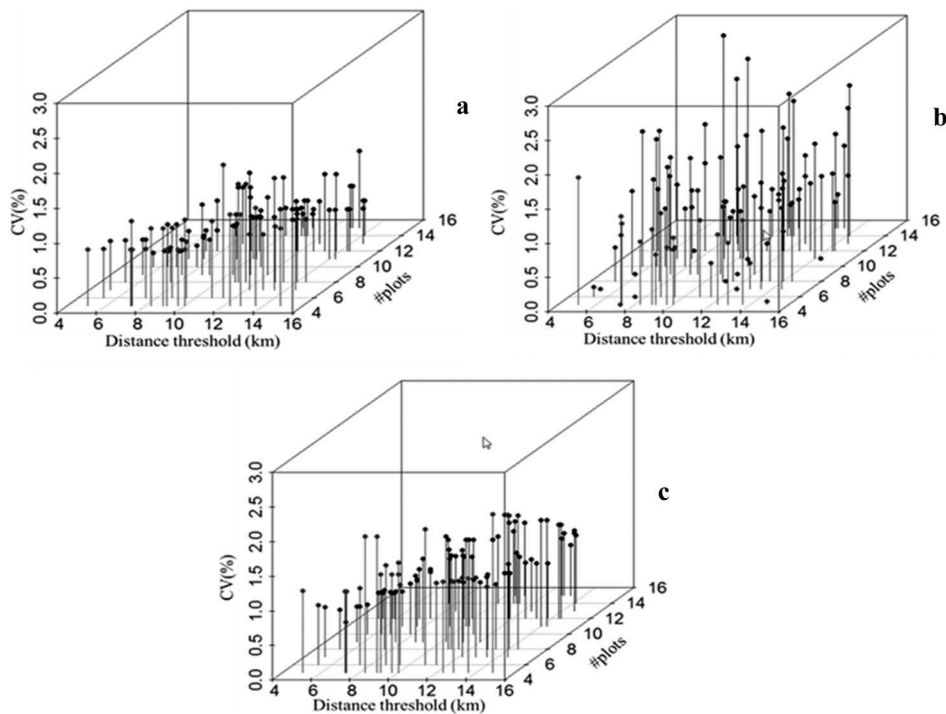


Figure 12. Coefficient of variation for uncertainty analysis of selection of plots for a base camp (the third base camp in the scenario of 3 base camp's was illustrated; 100 Monte Carlo perturbations were used; a: for base camp 1; b: for base camp 2; c: for base camp 3).

Table 5. Summary of computing performance of the cyber-enabled SDSS framework (number of Monte Carlo perturbations for reliability analysis: 100; time units: seconds; the scenario of three base camps was used for the selection of plots per base camp).

Scientific workflows	Computing time (seconds)		Speedup	Computing resources
	Sequential	Parallel		
Cost surface analysis	186,000	5701	32.62	Windows cluster (40 CPUs)
Selection of base camps				
#Camps = 3	102	2	51	Linux cluster (100 CPUs used)
#Camps = 4	100	1	100	
#Camps = 5	844	75	11.25	
#Camps = 6	1397	97	14.40	
#Camps = 7	610	64	9.53	
Selection of plots per camp				
Camp 1	54,841	1614	33.97	
Camp 2	146,419	9816	14.92	
Camp 3	14,991,312	155,030	96.70	
Overall time	15,381,634	172,400	89.22	

15,381,634 s = 4273 h for sequential computing; 172,400 s = 47.89 h for parallel computing.

We focus on scenario S_1 (three camps) for computing performance evaluation. Table 5 reports the computing performance of our experiment for scenario analysis (including uncertainty analysis with 100 Monte Carlo perturbations). Cost surfaces analysis (in ArcGIS) for each candidate base camp was conducted on the Windows cluster (40 CPUs used). The selection of

optimal camps and plots was deployed on the Linux cluster (100 CPUs used). Overall sequential computing time for the experiment of scenario analysis needs about 178 days. The selection of optimal plots per camp is most computationally expensive. Table 5 shows that the sequential computing time for selection of plots for camps (176 days) is about 98.77% of overall computing time. The analysis of cost surfaces is also computationally demanding – about 2 days of sequential runs. The cyber-enabled Mangrovelnv implementation in this study, while automating the entire process, reduces substantially the overall computing time down to 2 days – a speedup of 89.22. The computation of cost surface analysis only needs about 1.58 h on the 40-CPU Windows cluster (otherwise 2 days of sequential run). The most computationally demanding part, the selection of optimal plots for the third base camp only needs 1.79 days of computing time when 100 CPUs were used.

6.3. Discussion

The SDSS framework driven by scientific workflows and cloud computing provides unique and solid support for conducting experiments to analyze alternative scenarios facing the fieldwork design of mangrove inventory. First, the SDSS allows for the robust selection of the optimal locations of base camps and plots. Second, the SDSS offers a means of warranting the safety and logistics of the mangrove inventory, which is very important for the physically challenging mangrove inventory. For our fieldwork, all of the supplies, transportation tools (boats), and local guides have to be prepared and team members trained before the field trip. Members of survey teams have to be switched every other week due to the psychological pressure exposed from extremely challenging physical environments of mangrove forests. Thus, having a well-crafted fieldwork design, particularly, the information on the locations of base camps and plots, assures the success of mangrove inventory. It is thus of critical importance to have such a SDSS framework and cyber-enabled implementation to efficaciously guide the fieldwork design.

Further, the SDSS framework is targeted to support the repeated use in different study regions for mangrove inventory and resolve computational challenges. Users could use the SDSS by providing input data and parameters. The SDSS helps determine the optimal locations of base camps and survey plots subject to their criteria. Users could adapt the spatial models and scientific workflows in the SDSS to what they need by requesting and working on cloud-based virtual machines (the core scientific workflows of the SDSS can be deployed on standalone computers). The integration of scientific workflows and cloud computing in the SDSS framework provides solid support for automating and accelerating the selection of base camp and plots, which is computationally repetitive and intensive (as shown in this study). With support from the cyber-enabled SDSS implementation, significant computing performance gain could be achieved (e.g., monthly level sequential computing time was reduced to a daily level here).

The application of the exact approach for optimization (linear programming here) led to long computation time for plot selection. Though the exact approach guarantees the optimal solution, it brings in computational intensity. The computing time could grow significantly in response to change in inputs. For example, the number of plots per camp and the travel distance threshold as two parameters directly affect the size of search

space for the optimization algorithm. Increase in the number of plots per camp, as experimental results show, leads to combinatorial growth in search space and thus long computing time. Besides cyber-enabled capabilities (focus of this study), another solution to cope with this combinatorial explosion issue is to use heuristics (e.g., machine learning represented by evolutionary algorithms; see Xiao 2008, Tong and Murray 2012). While optimality may not be guaranteed by heuristic approaches, it is efficient for these approaches to search for near-optimal solutions in face of large search space (e.g., when the number of plots per camp is high).

The inventory plan developed using this cyber-enabled SDSS was implemented in October 2013 for the estimation of mangrove carbon stocks in our study area. Plot and base camp locations were computed and evaluated in advance and then loaded into a hand-held GPS unit for navigating to the desired point during the fieldwork. Plots were located by boating to the nearest point accessible by boat, and then the field crews hiked to the designated plot. An accuracy of 6% (international standard: under 10%) was achieved regarding the estimation of mangrove carbon stocks (see Stringer *et al.* 2015). The cyber-enabled SDSS framework and MangroveInv implementation thus proved functionally effective in the fieldwork design of mangrove inventory. Further, this cyber-enabled SDSS framework supports the sharing and update of sampling information of surveyed plots for communication among different users for a purpose of long-term monitoring and assessment of mangrove forests in our study region (through Web GIS and geodatabase; ongoing work).

7. Conclusion

In this study, we presented a cyber-enabled SDSS framework that facilitates computationally intensive fieldwork design of mangrove inventory. The framework functions as an integrative modeling platform for efficacious spatial decision-making by coupling scientific workflows and cloud computing. The SDSS framework plays a critically important role in group-based decision making that assures the logistical integrity and safety of the operation while realizing the assessment objectives of mangrove inventory. Consequently, this framework provides significant merits for effective coordination among experts and stakeholders.

This cyber-enabled SDSS framework allows for integrating diverse spatial data and modeling functionality for robust fieldwork design. The novelty of this framework lies in the coupling of scientific workflows and cloud computing for automating and accelerating computationally intensive fieldwork design of mangrove inventory. Scientific workflows enable the automation and reuse of interconnected but repetitive analytical components, greatly enhancing the efficiency of processing and analyzing GIS data for the fieldwork design. Spatial optimization modeling and the surrogate-based approach are essential in the optimal selection of base camps and plots as well as robustness evaluation. With scientific workflows, these spatial optimization models support the automated generation of alternative optimal solutions with operational constraints in response to parameter combinations. Further, cloud computing, together with high-performance computing, provides scalable computational support for resolving the computational challenges of these scientific workflows in the SDSS. It is flexible to share, reuse, and extend these workflows for organizing a (usually large) pool of data and modeling functionality (e.g., via encapsulation into virtual machines). The cyber-enabled SDSS

can leverage high-performance computing for accelerating the selection of optimal base camps and plots. This substantially lowers the bar of using high-performance computing particularly for experts of mangrove studies often without sufficient computing background.

While the cyber-enabled SDSS framework in this study is pivotal to the efficacious fieldwork design of mangrove inventory, a suite of future directions exist. First, in the current version of the framework, linear programming as an exact approach is used for spatial optimization. Linear programming is often ill-suited to cases in which the number of base camps or plots per base camp is large (combinatorial optimization). In future work, we will introduce heuristic approaches (e.g., evolutionary algorithms) into the spatial optimization process. Second, in this study we have successfully applied the SDSS into the Zambezi delta, Mozambique. We will use more study cases (e.g., ongoing work in Tanzania and West Africa) to further examine the capability of the cyber-enabled SDSS framework in supporting computationally demanding fieldwork design of mangrove inventory.

Acknowledgments

The authors would like to thank Temilola E. Fatoyinbo and Marc Simard from NASA for providing LiDAR-driven canopy height data in this study. This work was supported with funding from USAID – Mozambique through the US Forest Service International Programs Office. Special thanks is given to Dr. Yonggang Wang for his assistance on statistical analysis in this study, Jing Deng, and Dr. Eric Delmelle.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by United States Agency for International Development – Mozambique through the International Programs, US Forest Service.

References

- Albareda-Sambola, M., et al., 2009. The multi-period incremental service facility location problem. *Computers & Operations Research*, 36 (5), 1356–1375. doi:[10.1016/j.cor.2008.02.010](https://doi.org/10.1016/j.cor.2008.02.010)
- Alongi, D.M., 2002. Present state and future of the world's mangrove forests. *Environmental Conservation*, 29 (3), 331–349. doi:[10.1017/S0376892902000231](https://doi.org/10.1017/S0376892902000231)
- Armstrong, M.P., Densham, P.J., and Rushton, G., 1986. Architecture for a microcomputer based decision support system. In: *Proceedings of the second international symposium on spatial data handling*, 5–10 July 1986, Seattle, WA. New York: International Geographical Union.
- Atkins, D.E., et al., 2003. *Revolutionizing science and engineering through cyberinfrastructure: report of the national science foundation blue-ribbon advisory panel on cyberinfrastructure*. Arlington, VA: National Science Foundation.
- Bennett, D.A., 1997. A framework for the integration of geographical information systems and modelbase management. *International Journal of Geographical Information Science*, 11 (4), 337–357. doi:[10.1080/136588197242301](https://doi.org/10.1080/136588197242301)
- Bento, C.M. and Beilfuss, R.D., 2000. *Wattled cranes, waterbirds, and wetland conservation in the Zambezi Delta, Mozambique*. Available from: http://www.savingcranes.org/images/stories/pdf/conservation/Wattled_Cranes_and_Waterbirds_Working_Paper_1.pdf [Accessed 10 March 2015].

- Canny, J., 1988. *The complexity of robot motion planning*. Cambridge, MA: MIT Press.
- Church, R. and Velle, C.R., 1974. The maximal covering location problem. *Papers of the Regional Science Association*, 32 (1), 101–118. doi:10.1007/BF01942293
- CPLEX, 2016. CPLEX. Available from: <http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/> [Accessed 12 October 2016].
- Daskin, M.S., 1995. *Network and discrete location: models, algorithms, and applications*. New York, NY: John Wiley & Sons.
- Deb, K., 2001. *Multi-objective optimization using evolutionary algorithms*. Vol. 16. New York, NY: John Wiley & Sons.
- Delmelle, E., 2009. Spatial Sampling. In: Rogerson, P. and Fotheringham, S. eds. *The SAGE handbook of spatial analysis*. London: SAGE, 183–206.
- Densham, P.J., 1991. Spatial decision support systems. *Geographical Information Systems: Principles and Applications*, 1, 403–412.
- Densham, P.J. and Rushton, G., 1992. Strategies for solving large location-allocation problems by heuristic methods. *Environment and Planning A*, 24 (2), 289–304. doi:10.1068/a240289
- Donato, D.C., et al., 2011. Mangroves among the most carbon-rich forests in the tropics. *Nature Geoscience*, 4 (5), 293–297. doi:10.1038/ngeo1123
- Efron, B. and Efron, B., 1982. *The jackknife, the bootstrap and other resampling plans*. Philadelphia, PA: SIAM: Society for Industrial and Applied Mathematics.
- Fatoyinbo, T.E. and Simard, M., 2013. Height and biomass of mangroves in Africa from ICESat/GLAS and SRTM. *International Journal of Remote Sensing*, 34 (2), 668–681. doi:10.1080/01431161.2012.712224
- Fatoyinbo, T.E., et al., 2008. Landscape-scale extent, height, biomass, and carbon estimation of Mozambique's mangrove forests with Landsat ETM+ and Shuttle Radar Topography Mission elevation data. *Journal of Geophysical Research: Biogeosciences*, 113 (G2). doi:10.1029/2007JG000551
- Forrester, A., Sobester, A., and Keane, A., 2008. *Engineering design via surrogate modelling: a practical guide*. Chichester: John Wiley & Sons.
- Fu, P. and Sun, J., 2011. *Web GIS: principles and applications*. Redlands, CA: ESRI press.
- Haining, R.P., 2003. *Spatial data analysis: theory and practice*. Cambridge, UK: Cambridge University Press.
- Huang, Q., et al., 2013. Using adaptively coupled models and high-performance computing for enabling the computability of dust storm forecasting. *International Journal of Geographical Information Science*, 27 (4), 765–784. doi:10.1080/13658816.2012.715650
- Jankowski, P. and Nyerges, T., 2001. GIS supported collaborative decision making: results of an experiment. *Annals of the Association of American Geographers*, 91 (1), 48–70. doi:10.1111/0004-5608.00233
- Jankowski, P. and Richard, L., 1994. Integration of GIS-based suitability analysis and multicriteria evaluation in a spatial decision support system for route selection. *Environment and Planning B: Planning and Design*, 21 (3), 323–340. doi:10.1068/b210323
- Jena, S.D., Cordeau, J.-F., and Gendron, B., 2012. Modeling and solving a logging camp location problem. *Annals of Operations Research*, 232 (1), 151–177.
- Jun, L. and Jun, W., 2011. Cloud computing based solution to decision making. *Procedia Engineering*, 15, 1822–1826.
- Kathiresan, K. and Rajendran, N., 2005. Mangrove ecosystems of the Indian Ocean region. *Indian Journal of Marine Sciences*, 34 (1), 104–113.
- Kauffman, J.B. and Donato, D., 2012. *Protocols for the measurement, monitoring and reporting of structure, biomass and carbon stocks in mangrove forests*. Bogor, Indonesia: Center for International Forestry Research (CIFOR).
- Keenan, P.B., 2013. Cloud computing and DSS: the case of spatial DSS. *International Journal of Information and Decision Sciences*, 5 (3), 283–294. doi:10.1504/IJIDS.2013.055926
- Kim, I.-H. and Tsou, M.-H., 2013. Enabling Digital Earth simulation models using cloud computing or grid computing—two approaches supporting high-performance GIS simulation frameworks. *International Journal of Digital Earth*, 6 (4), 383–403. doi:10.1080/17538947.2013.783125

- Krause, G., et al., 2004. Mapping land-cover and mangrove structures with remote sensing techniques: A contribution to a synoptic GIS in support of coastal management in North Brazil. *Environmental Management*, 34 (3), 429–440. doi:[10.1007/s00267-004-0003-3](https://doi.org/10.1007/s00267-004-0003-3)
- Latombe, J.C., 2012. *Robot motion planning*. Vol. 124. New York, NY: Springer Science & Business Media.
- Ligmann-Zielinska, A., Church, R.L., and Jankowski, P., 2008. Spatial optimization as a generative technique for sustainable multiobjective land-use allocation. *International Journal of Geographical Information Science*, 22 (6), 601–622. doi:[10.1080/13658810701587495](https://doi.org/10.1080/13658810701587495)
- Ludäscher, B., et al., 2006. Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience*, 18 (10), 1039–1065. doi:[10.1002/\(ISSN\)1532-0634](https://doi.org/10.1002/(ISSN)1532-0634)
- Mann, P.S., 2007. *Introductory Statistics*. 6th ed. Hoboken, NJ: John Wiley & Sons.
- Medeiros, C.B., et al., 2005. WOODSS and the Web: annotating and reusing scientific workflows. *ACM SIGMOD Record*, 34 (3), 18–23.
- Mell, P. and Grance, T., 2011. The NIST definition of cloud computing (draft). *NIST Special Publication*, 800 (145), 7.
- Murray, A.T., et al., 2007. Coverage optimization to support security monitoring. *Computers, Environment and Urban Systems*, 31 (2), 133–147. doi:[10.1016/j.compenvurbsys.2006.06.002](https://doi.org/10.1016/j.compenvurbsys.2006.06.002)
- Owen, S.H. and Daskin, M.S., 1998. Strategic facility location: A review. *European Journal of Operational Research*, 111 (3), 423–447. doi:[10.1016/S0377-2217\(98\)00186-6](https://doi.org/10.1016/S0377-2217(98)00186-6)
- Peng, Z. and Tsou, M., 2003. *Internet GIS: distributed geographic information services for the Internet and Wireless Networks*. Hoboken, NJ: John Wiley & Sons.
- Pratt, A., et al., 2010. Exposing the Kepler scientific workflow system as an OGC web processing service. In: *Proceedings of iEMSs (International Environmental Modelling and Software Society)*, 5–10 July 2010, Ottawa, Canada.
- Ray, J.J., 2007. A web-based spatial decision support system optimizes routes for oversize/overweight vehicles in Delaware. *Decision Support Systems*, 43 (4), 1171–1185. doi:[10.1016/j.dss.2005.07.007](https://doi.org/10.1016/j.dss.2005.07.007)
- Seffino, L.A., et al., 1999. WOODSS—a spatial decision support system based on workflows. *Decision Support Systems*, 27 (1–2), 105–123. doi:[10.1016/S0167-9236\(99\)00039-1](https://doi.org/10.1016/S0167-9236(99)00039-1)
- Sengupta, R.R. and Bennett, D.A., 2003. Agent-based modelling environment for spatial decision support. *International Journal of Geographical Information Science*, 17 (2), 157–180. doi:[10.1080/713811747](https://doi.org/10.1080/713811747)
- Shim, K.-C., Fontane, D.G., and Labadie, J.W., 2002. Spatial decision support system for integrated river basin flood control. *Journal of Water Resources Planning and Management*, 128 (3), 190–201. doi:[10.1061/\(ASCE\)0733-9496\(2002\)128:3\(190\)](https://doi.org/10.1061/(ASCE)0733-9496(2002)128:3(190))
- Stringer, C.E., et al., 2015. Carbon stocks of mangroves within the Zambezi River Delta, Mozambique. *Forest Ecology and Management*, 354, 139–148. doi:[10.1016/j.foreco.2015.06.027](https://doi.org/10.1016/j.foreco.2015.06.027)
- Sugumaran, R. and Degroote, J., 2010. *Spatial decision support systems: principles and Practices*. Boca Raton, FL: CRC Press.
- Sugumaran, R., Meyer, J.C., and Davis, J., 2004. A Web-based environmental decision support system (WEDSS) for environmental planning and watershed management. *Journal of Geographical Systems*, 6 (3), 307–322. doi:[10.1007/s10109-004-0137-0](https://doi.org/10.1007/s10109-004-0137-0)
- Sugumaran, V. and Sugumaran, R., 2005. Web-based spatial decision support systems (WebSDSS): evolution, architecture, and challenges. In: *Third annual SIGDSS workshop: designing complex decision support: discovery and presentation of information and knowledge*, 11 December 2005, Las Vegas, NV.
- Sun, A., 2013. Enabling collaborative decision-making in watershed management using cloud-computing services. *Environmental Modelling & Software*, 41, 93–97.
- Tang, W. and Feng, W., in press. Parallel map projection of vector-based big spatial data: coupling cloud computing with graphics processing units. *Computers, Environment and Urban Systems*. doi:[10.1016/j.compenvurbsys.2014.01.001](https://doi.org/10.1016/j.compenvurbsys.2014.01.001).
- Taylor, I.J., et al., 2007. *Workflows for e-Science: scientific workflows for grids*. Secaucus: Springer-Verlag.
- Tong, D. and Murray, A.T., 2012. Spatial optimization in geography. *Annals of the Association of American Geographers*, 102 (6), 1290–1309. doi:[10.1080/00045608.2012.685044](https://doi.org/10.1080/00045608.2012.685044)

- Tuot, C.J., Sintek, M. and Dengel, A.R., 2008. IVIP - A scientific workflow system to support experts in spatial planning of crop production. In: B. Ludäscher and N. Mamoulis, eds. *Scientific and statistical database management: 20th international conference, SSDBM 2008, Proceedings*, 9–11 July 2008, Hong Kong, China. Heidelberg: Springer Berlin Heidelberg, 586–591.
- Valiela, I., Bowen, J.L., and York, J.K., 2001. Mangrove forests: one of the world's threatened major tropical environments: at least 35% of the area of mangrove forests has been lost in the past two decades, losses that exceed those for tropical rain forests and coral reefs, two other well-known threatened environments. *Bioscience*, 51 (10), 807–815.
- Verstegen, J.A., et al., 2012. Spatio-temporal uncertainty in spatial decision support systems: A case study of changing land availability for bioenergy crops in Mozambique. *Computers, Environment and Urban Systems*, 36 (1), 30–42. doi:10.1016/j.compenvurbsys.2011.08.003
- Wang, S., 2010. A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers*, 100 (3), 535–557. doi:10.1080/00045601003791243
- Wilkinson, B. and Allen, M., 2004. *Parallel programming: techniques and applications using networked workstations and parallel computers*. 2nd ed. Upper Saddle River, NJ: Pearson Prentice Hall.
- Xiao, N., 2008. A unified conceptual framework for geographical optimization using evolutionary algorithms. *Annals of the Association of American Geographers*, 98 (4), 795–817. doi:10.1080/00045600802232458
- Yang, C., et al., 2011. Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing?. *International Journal of Digital Earth*, 4 (4), 305–329. doi:10.1080/17538947.2011.587547
- Yang, C. and Huang, Q., 2013. *Spatial cloud computing: a practical approach*. Boca Raton, FL: CRC Press.
- Yang, C., et al., 2010. Geospatial Cyberinfrastructure: past, present and future. *Computers, Environment and Urban Systems*, 34 (4), 264–277. doi:10.1016/j.compenvurbsys.2010.04.001
- Yue, P., et al., 2013. Geoprocessing in Cloud Computing platforms—a comparative analysis. *International Journal of Digital Earth*, 6 (4), 404–425. doi:10.1080/17538947.2012.748847
- Zhang, J., 2012. A practical approach to developing a web-based geospatial workflow composition and execution system. In: *Proceedings of the 3rd international conference on computing for geospatial research and applications*, 1-3 July 2012, Reston, VA, USA. New York: ACM Press, 21.

Appendix 1

Formulation of the spatial optimization model for the selection of optimal plots for a base camp.

$$\text{Max : } \sum_{i=1}^n \sum_{k=1}^n Z_{ik}$$

Subject to

$$Z_{ik} \leq b_{ik} X_i \quad \forall i = 1, 2, \dots, n \quad \forall k = 1, 2, \dots, n$$

$$\sum_{k=1}^n Z_{ik} \leq 1 \quad \forall i = 1, 2, \dots, n$$

$$\sum_{i=1}^n X_i = p_p$$

$$\sum_{M_{is}} X_i \leq p_b \quad \forall is = 1, 2, \dots, 5$$

$$Z_{ik} = 0, 1 \quad \forall i = 1, 2, \dots, n \quad \forall k = 1, 2, \dots, n$$

$$X_i = 0, 1 \quad \forall i = 1, 2, \dots, n,$$

where

i : index of candidate plots ($i = 1, 2, \dots, n$)

k : index of covered locations; here we use candidate plots within a radius of 2 km; $k = 1, 2, \dots, n$

M_{is} : set of classes, $is = 1, 2, 3, 4, 5$

p_p : number of plots that will be visited

p_b : ceiling of p_p/n_c (n_c : number of canopy height classes; the smallest integer that satisfies $p_b \geq p_p/n_c$)

b_{ik} : whether candidate location k is covered (1) or not (0) by candidate plot i ; this information comes from the origin-destination matrix (based on Euclidean distance).

Z_{ik} : binary decision variable to tell whether candidate covered location k is covered by candidate plot i or not. Z_{ik} is set to 1 if candidate plot i is chosen to be a plot and candidate covered location k is covered by i ; otherwise, Z_{ik} is 0.

X_i : binary decision variable to tell whether candidate plot i is chosen to be a plot or not. X_i is 1 if candidate plot i is chosen to be a plot; otherwise, $X_i = 0$.