

# **Invasive Species Forecasting System: A Decision Support Tool for the U.S. Geological Survey – FY 2005 Benchmarking Report v. 1.6**

*NASA Applied Sciences Program Benchmarking Team for Invasive Species:  
United States Geological Survey  
Goddard Space Flight Center  
John C. Stennis Space Center  
University of Arizona  
University of Missouri*

National Aeronautics and  
Space Administration

John C. Stennis Space Center  
SSC, Mississippi 39529-6000

---

December 2005

Invasive Species Forecasting System: A Decision Support Tool for the U.S. Geological Survey –  
FY 2005 Benchmarking Report

Acknowledgments

This report was compiled by the NASA Applied Sciences Directorate at the John C. Stennis Space Center, Mississippi. The authors and contributors to this report include the following individuals:

- Tom Stohlgren, USGS Fort Collins Science Center
- John Schnase, NASA Goddard Space Flight Center
- Jeffrey Morisette, NASA Goddard Space Flight Center
- Neal Most, Innovim (GSFC)
- Ed Sheffner, NASA Headquarters
- Charles Hutchinson, University of Arizona
- Sam Drake, University of Arizona
- Willem Van Leeuwen, University of Arizona
- Verne Kaupp, University of Missouri

Participation in this work by Science Systems and Applications, Inc., at SSC was supported under NASA Task Order NNS04AB54T.

The use of trademarks or names of manufacturers is for accurate reporting only and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

## Table of Contents

Executive Summary .....	v
1.0 Introduction .....	1
2.0 Description of the Initial State: .....	2
2.1 Background and History .....	2
2.2 Data Limitations .....	3
2.3 Analysis and Modeling Tools .....	5
2.4 Decision Support Tools .....	8
2.5 Pieces of the Puzzle .....	9
3.0 Partnership Development .....	10
3.1 Basis for Partnership .....	10
3.2 How the Work Was Undertaken .....	10
4.0 ISFS Framework Description .....	10
4.1 Status of the Framework .....	12
4.2 Implementation of ISFS at NIISS .....	12
4.3 Products and Services .....	14
4.4 Target Customers .....	15
4.5 Benchmarking Performance Metrics .....	16
5.0 Impact of ISFS on NIISS .....	17
5.1 First Code Improvement – Parallel Kriging .....	17
5.2 Parallel Kriging Results on Ideally Sized Test Cases .....	19
5.3 Evaluation of Community and Advanced Improvement Goals .....	20
5.4 Progress Toward Milestones .....	21
6.0 Verification and Validation .....	22
7.0 Future ISFS Benchmarking Approach - DDP Process .....	23
7.1 Overview .....	23
7.2 DSS Benchmarking – Risk Management Approach .....	24
7.3 DDP Benchmarking Conclusions .....	26
8.0 Bibliography .....	27
8.1 Selected Presentations .....	28
8.2 Web Sites .....	28
8.3 Software Engineering Documentation .....	29
Appendix A. ISFS Elements and Process Flow .....	31

## Tables

Table 1 - Representative products developed out of the ISFS. The image on the left is a CONUS Habitat Suitability Map for Tamarisk. The image on the right is the corresponding Predictive Confidence Map for the tamarisk prediction.....	17
Table 2 - Test case timing results (elapsed wall clock seconds).....	19
Table 3 - Scaling efficiencies for test cases. ....	20
Table 4 - Baseline scenario timing and scaling results. ....	21
Table 5 - ISFS operational improvements summary.....	22
Table 6 - Candidate Performance and usability metrics under consideration (State 1→State 2.....	25

## Figures

Figure 1. Differences in vegetation community mapping in a typical FORT study area using a 2-ha (bottom) and 0.2-ha (top) minimum mapping unit. ....	4
Figure 2 – Initial multi-phase design using remote sensing and vegetation plots to create GIS-based predictive spatial models.....	5
Figure 3 – Spatial modeling scheme developed by Mohammed Kalkhan (Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO; see Chong et al., 2001). ....	7
Figure 4 – An example of a modeled surface for total plant species richness in the Cerro Grande burn area, New Mexico, and an error surface for the map of non-native species richness for the same area. The “red” standard errors represented about 10% of the mean number of predicted non-native species in a 30 m x 30 m cell. ....	8
Figure 5 – Simple kriging diagrams from Systat statistical software for a 750-ha area in Rocky Mountain National Park, Colorado. ....	8
Figure 6 – An early schematic of a automated system for handling requests for invasive species information linking relational databases and spatial modeling via the World Wide Web.....	9
Figure 7 – The three primary components of the Invasive Species Forecasting System. ....	11
Figure 8 – The TEMPEST Server and the ROCKY Cluster are the hardware backbone of the ISFS.....	14
Figure 9 – Representative products developed out of the ISFS. The image on the left is a CONUS Habitat Suitability Map for Tamarisk. The image on the right is the corresponding Predictive Confidence Map for the tamarisk prediction. ....	15
Figure 10 – Scaling curves for test cases on Medusa.....	20
Figure 11 – Scaling curves for test cases on Medusa.....	21
Figure 12 – DDP tool initial GUI.....	25
Figure 13 – The benchmarking process of the enhancements between a state (n-1) and the following consecutive state (n) follows a cyclic pattern. ....	27
Figure A-1. The elements of, and data flow within, the Invasive Species Forecasting System. ....	31

## Executive Summary

The National Institute of Invasive Species Science (NIISS), through collaboration with NASA's Goddard Space Flight Center (GSFC), recently began incorporating NASA observations and predictive modeling tools to fulfill its mission. These enhancements, labeled collectively as the Invasive Species Forecasting System (ISFS), are now in place in the NIISS in their initial state (V1.0). The ISFS is the primary decision support tool of the NIISS for the management and control of invasive species on Department of Interior and adjacent lands. The ISFS is the backbone for a unique information services line-of-business for the NIISS, and it provides the means for delivering advanced decision support capabilities to a wide range of management applications.

This report describes the operational characteristics of the ISFS, a decision support tool of the United States Geological Survey (USGS). Recent enhancements to the performance of the ISFS, attained through the integration of observations, models and systems engineering from the NASA are benchmarked, i.e., described quantitatively and evaluated in relation to the performance of the USGS system prior to incorporation of the NASA enhancements. This report benchmarks Version 1.0 of the ISFS.

During 10 years of field and laboratory research, the USGA/NIISS team made significant contributions to the advancement of landscape-scale surveys and modeling. New multi-scale methods for measuring plant diversity and several spatial analysis programs had been developed and tested. The research team published papers challenging commonly held paradigms in invasive species science, and on the integration of remotely sensed data, field data, and spatial statistics to accurately quantify and map native and non-native species diversity at landscape scales.

Prior to collaboration with NASA, the tools developed by USGS were limited by the difficulty acquiring environmental data over large areas and the computational limitations of data processing to generate predictive maps that were both sufficiently extensive and timely. The USGS partnership with NASA has improved to a great extent the computational performance of the ISFS and the integration of image data with other large datasets.

Through the implementation of four joint projects beginning in 2001, USGS and NASA worked together to deploy the ISFS as a Web-based software framework that supports the data management and modeling activities needed to produce large-scale, predictive maps. It has been implemented as a workflow management system that addresses the most crucial technological bottlenecks in the production of large-scale predictive maps. The ISFS architecture comprises three primary components: 1) the Invasive Species Data Service (ISDS), 2) the Invasive Species Analysis and Modeling Service (ISAMS), and 3) the Invasive Species Decision Support Service (ISDSS). The focus of the early work, documented in this report, has been implementing the ISAMS library of modeling routines on a high-performance cluster computer.

The benchmarking described in this report focuses on measuring improvements in computational performance of the core USGS modeling codes used in ISAMS. Two "canonical" study sites were evaluated to measure code improvements: the Cerro Grande Fire Site in Los Alamos, NM (CGFS), and the Rocky Mountain National Park, CO (RMNP). Benchmark tests were evaluated against Community Improvement Goals, the quantitative improvements in the underlying model that had been agreed upon as minimal advances needed to improve core capabilities. Parallel kriging model runs for the two study sites show improved run times for the CGFS test of 2 minutes and 7 seconds, and for the RMNP test of 17.6

Invasive Species Forecasting System: A Decision Support Tool for the U.S. Geological Survey –  
FY 2005 Benchmarking Report

seconds, which is 25.2 times faster than the baseline run time of 7 minutes and 23 seconds. Initial benchmark test results indicate attainment of the Milestone 1 Community Improvement goal of a 25x speed up on a 32-processor cluster with greater than 75% efficiency. Overall, USGS reports average model run time improvement from an average of 21 days prior to the NASA collaboration to an average of 2 minutes in 2005.

Current plans call for benchmarking additional ISFS performance characteristics in FY 2006 using systematic benchmarking techniques and tools, including the Defect Detection and Prevention (DDP) software tool developed by the Jet Propulsion Laboratory (JPL). The outcome of the DDP process allows for the quantification of NASA contributions (mitigation factors) in terms of enhanced attainment of Objectives or reduction of Risks between State 1 and State 2, or between State 2 and future “to be” states.

## 1.0 Introduction

This report describes the operational characteristics of the Invasive Species Forecasting System (ISFS), a decision support tool of the United States Geological Survey (USGS). Recent enhancements to the performance of the ISFS, attained through the integration of observations, models and systems engineering from the National Aeronautics and Space Administration (NASA), are benchmarked, i.e., described quantitatively and evaluated in relation to the performance of the USGS system prior to incorporation of the NASA enhancements. This report benchmarks Version 1.0 of the ISFS.

Non-indigenous invasive species pose a formidable threat of natural disaster in the 21st century. The direct cost to the American economy to mitigate invasive species is estimated at \$100-200 billion per year, greater than all other natural disasters combined. The spread of invasive species is growing as globalization increases the mobility of pest and disease organisms. This issue is now one of extraordinary interest to the American public, generating diverse stakeholder support ranging from land management agencies, state, local and tribal governments, the agricultural industry, conservation organizations, and private landowners. The scale, growth rate, and impact of the invasive species problem merit a coordinated national response that capitalizes on the vast scientific and technical capacity of the federal government. An historic confluence of science and technology makes this a particularly opportune time to advance our understanding of, and ability to manage, the invasive species threat.

The National Institute of Invasive Species Science (NIISS) is a USGS-led consortium of governmental and non-governmental partners. The Institute was created in response to high priority needs for invasive species research and management identified by USGS and Department of Interior clients, such as the National Park Service, U.S. Fish and Wildlife Refuge System, Bureau of Land Management, Bureau of Reclamation, tribes, other land managers, and the general public.

The vision for the Institute is to provide national leadership in the area of invasive species science and work with federal agencies and other stakeholders to disseminate and synthesize current, accurate data and research to detect, predict, and reduce the effects of harmful non-native plants, animals, and diseases in ecosystems and natural areas throughout the United States. Its mission is to provide unique services to federal agencies and other stakeholders to improve the nation's response to the invasive species threat. Science and stakeholder support are the primary drivers for the Institute's programs and activities, and it is committed to nurturing the interdisciplinary cooperation and organizational collaborations necessary to meet the invasive species challenge from local to national scales.

NIISS was formed in 2001. It is based administratively in the Department of Interior, USGS Fort Collins Science Center. The Fort Collins Science Center (FORT) is one of many laboratories in USGS's national network of science centers, field stations, and state cooperative units. NIISS has a unique position within the scientific community as an authority in space-based decision-support for invasive species management. Its core competencies include both science research and early detection/rapid response (ED/RR) of invasive species. Research at NIISS focuses on 1) the development and testing of cost-efficient field sampling methods and protocols, and 2) the integration and synthesis of existing information to quantify patterns of invasions, guide management activities, and identify gaps in information and improve the comparability of data across various sources. The objectives of this research are to make better use of existing data, identify habitats vulnerable to invasion, identify the highest priority invasive species, and better coordinate all aspects of invasive species science, such as prevention, early detection and rapid response, monitoring and research, and outreach.

Early detection and rapid response are crucial aspects of the national approach to the invasive species threat. The Institute has a strong ED/RR foundation. Its work, in cooperation with the National Biological Information Infrastructure (NBII), focuses on the technological and social aspects of developing this network by building key partnerships, designing “smart” surveys that can be used by a broad cross-section of the community, and creating web-based data-sharing and analysis technologies and new decision support tools.

Forecasting, the need to locate invasive species and forecast their potential spread, is a crucial capability for both science research and ED/RR. The USGS, through the NIISS and its predecessors at Fort Collins, develop methods for predicting the spread of invasive species using a combination of ground data and environmental parameters as the essential elements. Prior to collaboration with NASA, the tools developed by USGS were not efficient because of the difficulty acquiring environmental data over large areas and the limitations of data processing to generate predictive maps that were both sufficiently extensive and timely.

The USGS is currently collaborating with the National Aeronautics and Space Administration (NASA) to enhance the capabilities of the USGS invasive species forecasting tools. These enhancements, labeled collectively as the Invasive Species Forecasting System (ISFS), are now in place in the NIISS in their initial state (V1.0). The ISFS is the primary decision support tool of the NIISS for the management and control of invasive species on Department of Interior and adjacent lands. Early detection and monitoring protocols and predictive models developed at the NIISS are being used to process NASA and commercial data to create on-demand, regional-scale assessments of invasive species patterns and vulnerable habitats. The ISFS is the backbone for a unique information services line-of-business for the NIISS, and it provides the means for delivering advanced decision support capabilities to a wide range of management applications.

Following this Introduction, Section 2 of this report describes the Initial State of the ISFS, recounting the computational and data analysis structures and characteristics used by the USGS prior to the NASA-derived enhancements. Section 3, Framework Description, details the three elements of the ISFS, how the benchmark operational metrics were derived and the status of version 1.0 of the ISFS in the NIISS. The Partnership Development part in Section 4 describes the partnership between NASA and USGS, how the NASA effort was undertaken and completed, and the population of the ISFS with critical datasets. Section 5 captures the Impact of the ISFS on the NIISS provides the initial benchmark results, including a description of the enhancements to USFSGS capabilities through the ISFS with specific examples. Section 6, Verification and Validation, is a brief description of the approach to the analysis of products from the ISFS to confirm that they meet specifications and user needs. Finally, Section 7 provides an initial description of the Next Benchmarking plan, utilizing a NASA derived tool known as Defect Detection and Prevention (DDP).

## **2.0 Description of the Initial State:**

### **The State of the Invasive Species Research Program Prior to 2000**

#### **2.1 Background and History**

The U.S. Geological Survey has an obligation to meet the science needs for the land management agencies in the Department of the Interior. The highest priority recurring requests for science support from land managers involve several aspects of invasive species science, specifically regarding prevention,



early detection, rapid response, surveys and monitoring, control, and restoration. Explicit in these requests is the need to establish a geographic-based data management system to input, store, evaluate, and retrieve information, maps, and ancillary data important in the active management of harmful invasive species. However, a lack of baseline data on natural resources combined with a long history of modest research and resource management budgets often resulted in the simplification of questions being asked by land managers.

For example, managers in Rocky Mountain National Park, Colorado, asked the FORT research team for systematic surveys of native and non-native plant species diversity in various habitats in the 106,000-ha Park, before they could evaluate and set priorities for invasive species management. The questions were simple. Which vegetation types contained the greatest number and foliar cover of invasive species? Were there general patterns of invasion related to elevation and moisture gradients? Were infestations concentrated or widely dispersed? Based on plot-scale research, which non-native species had the greatest frequency and cover, and in which habitat types?

Such questions can be resolved with a simple stratified random sampling design, using large plots in various common and rare vegetation plots. The survey could be enhanced considerably by using a combination of remote sensing (using Landsat Thematic Mapper imagery to stratify the habitat types), a nested vegetation plot design (to measure foliar cover in small plots and species richness in large plots), and a geographic information system linked to a relational data base to evaluate field data and present results to land managers in a useful format (maps and tables).

Early survey results led to unexpected and important findings that would revolutionize approaches to measuring plant diversity in large, natural areas. Native species richness was found to be very patchily distributed such that there were a few “hotspots” of native plant diversity in a sea of low and moderate diversity habitat types. Contrary to theory and existing paradigms on plant invasion, the hotspots for native plant diversity were also found to be hotspots for non-native plant diversity. Lastly, rare habitat types, and those with high light, nitrogen, water, and warm temperatures were consistently more invaded than stressful habitats. Thus, high resolution mapping of the environment was essential in mapping hotspots of native plant diversity and diversity. The detection of rare habitat features such as wetlands, riparian zones (stream sides), nitrogen rich meadows, etc., became essential to successful invasive species mapping. The most important lessons learned after the first few years of field sampling in the Park were: (1) only a very small percentage, usually less than 1%, of a park or large natural area can be actually surveyed with plot techniques; and (2) knowing that field observations were limited to <1% of the area, estimating what was in the other 99% of the study area became much more important. Spatial interpolation from points (even large plots) to the landscape scale would be essential to provide managers with reliable, quantified patterns of native and non-native plant diversity. However, there were many limitations related to data, analysis and modeling tools, and decision support systems that prevented major advances in landscape ecology.

## **2.2 Data Limitations**

Field studies in Rocky Mountain National Park were limited first by unorganized or inadequate data from past field studies. Field sampling methods in the Park often relied on small quadrat and transect techniques published in the 1960s. The methods generally included a 15-30 meter transect with 0.2 m x 0.5 m quadrats placed every 1.5 m to record foliar cover by species or cover by gross physiological type (e.g., grass, shrub, herb). These designs often captured the most common species in an area, but missed up to 50% of the locally rare or patchily distributed species (Stohlgren et al., 1998). Transect locations often were subjectively selected – many in flat spots, near roads, and close to facilities. Data were often

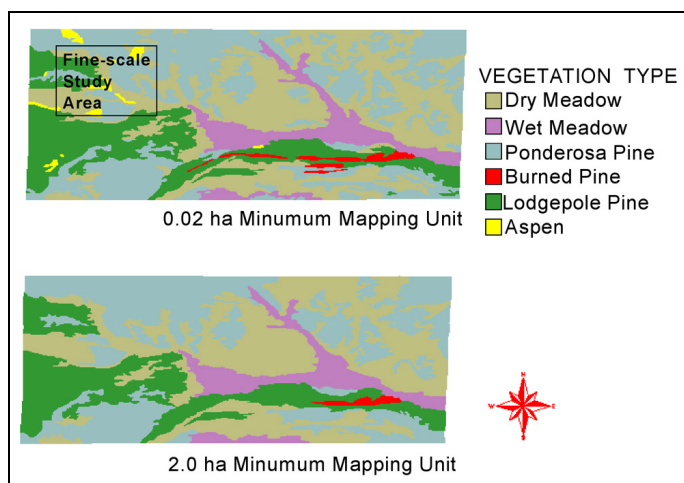
recorded on paper datasheets, and summarized by pocket calculators or simple spread sheets on personal computers. The transect studies were often related to very specific objectives, such as the evaluation of forage availability for deer or elk. Ancillary data on soil characteristics, slope, aspect, canopy cover, were rarely obtained.

Species lists were available from over 80+ years of herbarium collections. However, some areas of the park were surveyed much more completely than other areas. In addition, no accepted way exists to evaluate the completeness of the lists, because taxonomists have different levels of expertise, survey in subjectively selected areas, and many plant species are cryptic (as seed or very small, hidden mature plants), rare, or too patchily distributed to be collected in most surveys (Stohlgren et al., 1997a). While species lists and checklists are important to Park visitors, they have limited application for resource managers. Resource managers need to know the locations and abundance of common and rare species, including threatened and endangered species, as well as the locations and abundance of harmful invasive species – both inside and outside of the Park boundaries.

Remote sensing data was limited to a few select scenes of Landsat Thematic Mapper (TM) imagery over the Park, and some color infrared photography. Digital elevation models with 30 m resolution were available. Combined, these data were used to classify the park's vegetation types into 30 or more common vegetation types, designated by the dominant one or two plant species, usually tree species. The minimum mapping unit (i.e., the smallest area recognized and delineated) was 2-ha (Figure 1). Rare vegetation types such as aspen and wet meadows were greatly under-represented by this approach, yet these types were particularly important habitats for native plant diversity and invasion (Stohlgren et al., 1997b).

Soils information were very coarse, limited to soil type maps with a 5-10 ha minimum mapping unit, which could only crudely be rectified with the vegetation maps. Little information existed on patterns of soil texture, moisture, or nutrients.

In 1995, new multi-scale vegetation plot methods were published by FORT that were more conducive to capturing rare and common species, and were linked to measurements of foliar cover, soils characteristics, and other ancillary data important in modeling patterns in plant species diversity. The nested plot, improved in 1998 (Stohlgren et al., 1998) was 20 m x 50 m containing one 100-m<sup>2</sup> subplot in the center, two 10-m<sup>2</sup> subplots in opposite corners, and ten 1-m<sup>2</sup> subplots systematically arranged around the edge of the 1000-m<sup>2</sup> and 100-m<sup>2</sup> plots.



**Figure 1. Differences in vegetation community mapping in a typical FORT study area using a 2-ha (bottom) and 0.2-ha (top) minimum mapping unit.**

Foliar cover and average height by plant species was recorded in the ten 1-m<sup>2</sup> subplots, along with the nearest percent cover of soil, litter (detached dead plant material), duff (attached or standing dead plant material), woody debris, water, dung, bare soil, and rock. Cumulative plant species presence was recorded in the 10-, 100- and 1000-m<sup>2</sup> plots. Ancillary data typically included the slope and aspect of the plot, and elevation derived from a Digital Elevation Model. Five soil samples were generally taken with a 2.5-cm

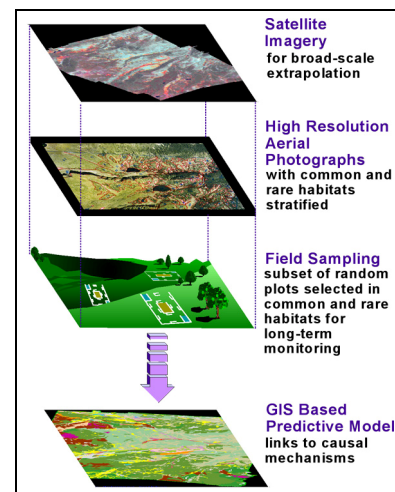
diameter core to a depth of 15 cm in the center and in the corners of the 1000-m<sup>2</sup> plot and combined to represent the plot, and later analyzed for soil texture (percent sand, silt, and clay). The samples were typically air dried, sieved, and analyzed for soil C and N. Thus, we were beginning to capture integrated information on plant diversity and soils in an unbiased, stratified-random fashion for later modeling. The large nested plots were labor intensive, which combined with cost considerations, greatly limited sample size. In Rocky Mountain National Park, Colorado, five years of study resulted in fewer than 200 plots, representing a sampling intensity equal to <0.02% of the Park. A 5-year survey in the Grand Staircase-Escalante National Monument, Utah, resulted in about 350 plots, or <0.004% of the two-million-acre landscape.

Information on the location and distribution of animal species, insects, and pathogens in Rocky Mountain National Park was almost non-existent. Some surveys of elk, deer, birds and butterflies were conducted in subjectively selected, local areas for short time periods (usually less than three years). Several agencies, non-government organizations, universities, and students had collected a variety of data on various aspects of species biology, ecology, and disturbance (e.g., fires, flooding), and aspects of global climate change, but only a few of the hundreds of databases collected have been synthesized (Baron, 2002). Typically, data collected for different reasons by different investigators, result in wildly different formats, data structures, and data quality. This Park is not alone. Most studies in national parks throughout the United States have resulted in little effort in standardizing taxonomy, survey methods, and data formats (Stohlgren et al., 1995). Thus, it was nearly impossible in Rocky National Park, the Grand Staircase-Escalante National Monument, or elsewhere to accurately quantify patterns of native biodiversity and the invasions of non-native plants, animals, and diseases at the appropriate fine-resolution scales due to data limitations.

### 2.3 Analysis and Modeling Tools

Improvements in the quality of data collected in Rocky Mountain National Park (and other parks) did not address the severe limitations in analysis and modeling tools available to most landscape-scale ecologists. Reliance on single-processor, stand-alone personal computers was commonplace. Typical Microsoft-based database software programs included DBASE, FoxPro, Lotus, and Excel. Microsoft Access was added to this arsenal about 5 years ago. Statistical packages included SAS, SYSTAT, and S-Plus, however, S-Plus was best configured for UNIX systems. Geographic information systems (GIS) relied on ARC/INFO products on a single processor Sun workstation poorly linked to relational databases and with limited capability for analyzing spatial data.

Despite these limitations, the FORT research team was able to develop landscape-scale models of the patterns of native and non-native plant species richness relative to several environmental factors (Figure 2). Typically, field data from the 1-m<sup>2</sup> subplots or 1000-m<sup>2</sup> plots on species richness and foliar cover would be entered from data sheets or palm top recording devices into an Excel spreadsheet and combined with soils data from the Modified-Whittaker nested vegetation plots and reflectance values from the Landsat TM imagery (often a few selected bands), while



**Figure 2 – Initial multi-phase design using remote sensing and vegetation plots to create GIS-based predictive spatial models.**

high resolution aerial photography would be used for selecting study plot locations with a stratified random design.

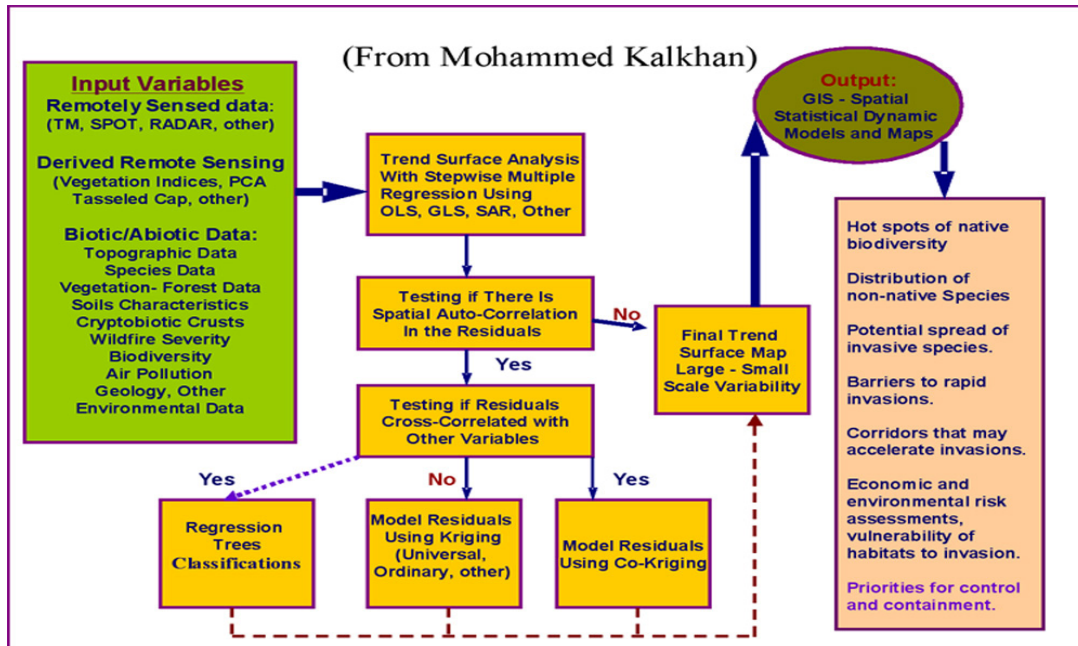
The initial intent was to use both satellite imagery and aerial photography in a multi-phase sampling design in which limited high resolution imagery would be correlated to coarser resolution satellite imagery over a larger study area. All variables would be screened for normality with a scatterplot option in a statistical package, and transformed if needed. One copy of the spreadsheet would be loaded into a statistical package on a Windows-based personal computer and used to develop tables of the mean richness and cover of native and non-native species by vegetation type, or correlations between native and non-native species richness or cover. The second copy of the Excel spreadsheet would be uploaded into S-plus on the Unix-based workstation for spatial analyses as described below.

Early attempts at spatial modeling, as shown in [Figure 3](#), began by manually incorporating various GIS data such as remote sensing, vegetation type, soil characteristics (e.g., texture, nutrients), and topographic variables into Excel spreadsheets. In most cases, multiple regression analysis was first used to explore the variability in species richness and the presence of non-native species as a function of the geographical location, elevation, slope, aspect, and Landsat TM bands 1-7 (Stohlgren, 2005).

Stepwise regression typically was used to identify the best linear combination of independent variables to include in a trend surface model of geographical variables and measures of species richness to describe large-scale spatial variability in the study area (Kallas, 1997; Metzger, 1997; Chong et al., 2001; Stohlgren, 2005).

The algorithms used in most of the studies included ordinary least-squared, Gaussian least-squared, and auto-regressive models described in Kallas (1997), Metzger (1997), and Chong et al. (2001). Because the data may or may not be spatially auto-correlated, it was very important to produce an “error term” for the models, as well as testing the residuals of the regression models with semivariograms (Cressie, 1985). This is easy with small datasets, but computationally demanding for large datasets. The number of computations escalated exponentially when residuals had to be analyzed for every model for both spatial autocorrelation and cross-correlation with the geographical variables – especially when the number of geographic variables was greater than 10 (often the case). Severe computational bottlenecks arose when inverse distance sampling was used to define the spatial weights matrix and to estimate the residuals using ordinary kriging. S-plus sorting routines were very slow when the matrix was even moderately sized. The strategy was elegant ([Figure 3](#)) and the mathematical concepts in the models were innovative and important, but the process of running the models was frustrating and slow, often taking up to three weeks to get the results of a single model of interest. The computational bottleneck was a function of slow, inefficient commercial software routines, serial programming, and single-processor computers.

What did waiting three weeks for results mean for the research program? Since the preliminary vegetation and soils data showed high cross-correlations and co-occurrence in space with native species richness, and since native plant species richness must first be modeled to predict patterns of non-native species richness, predicted surfaces might be needed for nitrogen, elevation, canopy closure, and a few selected bands of the TM imagery before the surface for native species richness could be accurately predicted. In turn, the modeled surface for native species was an input for the model on non-native species richness. Four to six model runs might be needed in succession, and each might take two to three weeks, creating the map of greatest concern up to four months after the initial computations. A simple error found in the original dataset, such as incorrectly transforming one of the variables or the emergence of a new variable of interest (e.g., solar radiation) would require starting the process anew.



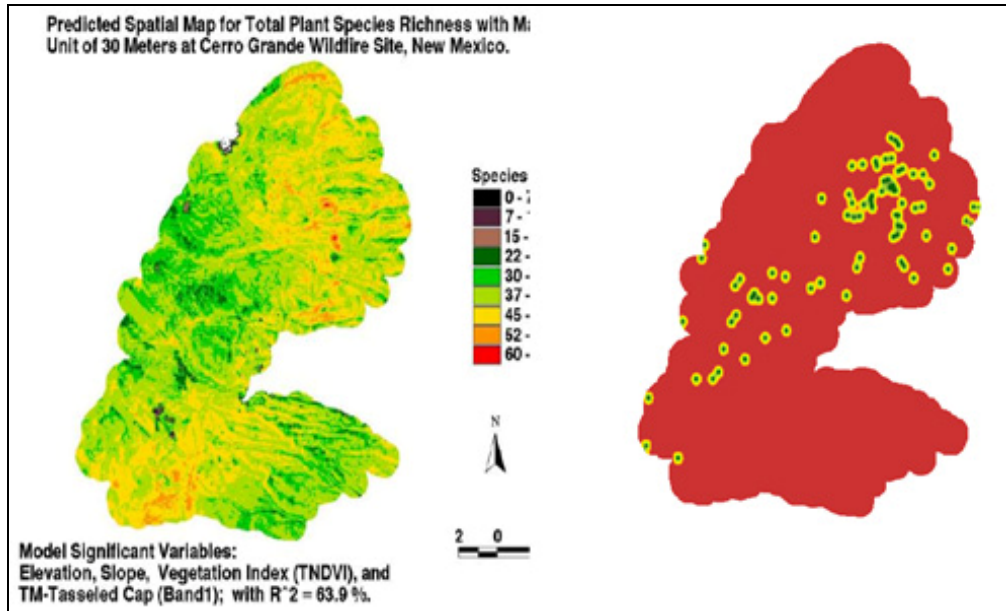
**Figure 3 – Spatial modeling scheme developed by Mohammed Kalkhan (Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO; see Chong et al., 2001).**

Computational bottlenecks created several additional unanticipated constraints on the research. First, sample sizes had to remain “small,” which could also be obtained by reducing the size of the sample area under consideration. This, in turn, reduced the ability to work in large areas or complex terrain. Modeling projects would not be considered for very large landscapes with many vegetation types with complex environmental gradients that would necessitate several hundred vegetation plots and multiple environmental factors or remote sensing variables. The process also limited the number of independent variables that could conveniently be added as predictor variables, so some important factors might be left out of some models. Sometimes, the resolution of the models could be decreased (increased cell-sizes) to speed up the calculations, but the final map might be less useful for on-the-ground managers. Finally, often only one “example” dependent variable would be selected, such as native or non-native species richness rather than attempt to model all 50 non-native species in the dataset.

In addition, the models tested were still simple compared to other more complex spatial analysis routines such as co-kriging (Chong et al., 2001). Little attempt was made to completely explore components of error propagation in the iterative modeling process. Nor was the computing power available to fully evaluate error terms and spatial autocorrelation throughout the process. Instead, simple error surfaces would be created for select mapped products as shown in Figure 4.

Despite the limitations of the modeling efforts as described, the products were exciting from theoretical and ecological standpoints. Modest field sampling efforts produced accurate maps of native and non-native species richness. Often the sampling intensity represents far less than 0.01% of a landscape. Predictions of native and non-native species richness were also possible, even with a modest number of variables (often including high light, high soil moisture, high nitrogen, and warm temperatures). The same general patterns were proven to occur in different biomes (Rocky Mountains, pine forests in the southwest and arid desert areas). Contrary to the existing paradigm, hot spots of native species richness were shown, theoretically, to be heavily invaded by non-native species.

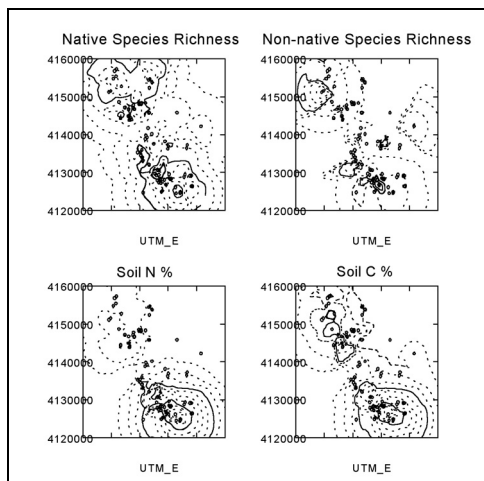




**Figure 4 – An example of a modeled surface for total plant species richness in the Cerro Grande burn area, New Mexico, and an error surface for the map of non-native species richness for the same area. The “red” standard errors represented about 10% of the mean number of predicted non-native species in a 30 m x 30 m cell.**

However, time was of the essence. When there was sufficient time for a detailed analysis, the schematic in Figure 3 was used to guide spatial analyses (Chong et al., 2001). Often, in haste, or to meet an immediate client need, simple kriging was used to rapidly report important patterns to land managers. In Figure 5, data from one 750-ha area of Rocky Mountain National Park were used to demonstrate the

spatial overlap of native and non-native species richness, and with soil characteristics. In this case, kriging from a commercial software package was used – simple interpolation between points, reducing a complex, multivariate problem into a single variable map.



**Figure 5 – Simple kriging diagrams from Systat statistical software for a 750-ha area in Rocky Mountain National Park, Colorado.**

Constraints in computational time and funding, and unreasonable deadlines for results and products often forced the substitution of simple models for complex, more realistic and accurate models. The approach was inadequate for larger areas.

## 2.4 Decision Support Tools

In 2000, an informal and rudimentary decision support system existed, based primarily on personally responding to individual telephone or email requests for specific information on invasive species. Meanwhile, FORT clients and customers wanted more models over larger areas. They wanted individual species maps and constantly updated maps in the few areas where they had successfully

controlled harmful invasive species. They wanted FORT to expand modeling efforts to include non-native animals, insects, and wildlife and plant diseases. They wanted in-house capabilities and training to run the spatial models as new field information became available. They wanted immediate access to the tools that were beginning to develop. In response to this myriad of requests, the research team began to “dream design” a better system to be more responsive to clients and customers (Figure 6).

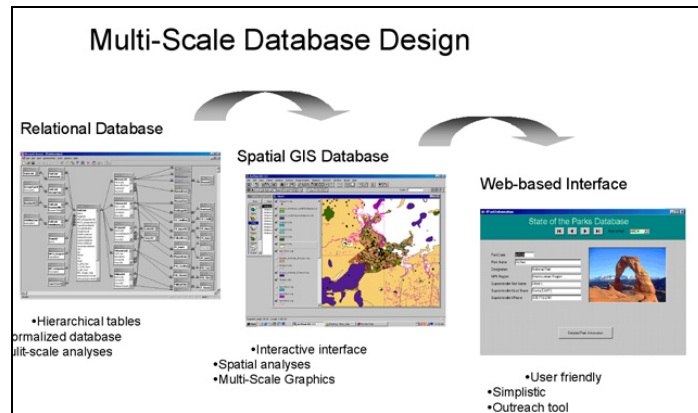
This early design effort envisioned a time in not too distant future where data on the invasive species abundances and distributions would be shared openly by dozens of agencies, non-government organizations, tribes, universities, and the museum communities in every county in the 50 states. Expansion to international datasets was also envisioned to carefully evaluate potential harmful invasive species. The idealized design included specifically identified predictive modeling capabilities to:

1. Document current distributions of species despite modest sampling/survey efforts;
2. Predict “potential species distributions and abundances based on current distributions and abundances;
3. Develop “probability maps” to guide early detection survey methods;
4. Measure current and predict future rates of spread of invasive organisms;
5. Link current and future distributions maps to “costs” (environmental, economic, and human-health; or impacts) for species-specific or area specific risk analyses,
6. Monitor the cost-effectiveness and benefits of control and restoration activities to evaluate “what works where”; and,
7. Provide rapid access to data and information on invasive to all interested parties and increase public awareness of invasive species issues.

A dedicated funding stream for this endeavor had yet to be identified – and few of the “details” in designing, testing, and institutionalizing such a system were known or anticipated.

## 2.5 Pieces of the Puzzle

During more than 10 years of field and laboratory research, the FORT/NIISS team made significant contributions to the advancement of landscape-scale surveys and modeling. New multi-scale methods for measuring plant diversity had been developed. The team now includes taxonomists and landscape ecologists on staff or at nearby Colorado State University. New software for palmtop computers had been developed and tested to improve the accuracy of field records while speeding up data transfer and analysis. Several spatial analysis programs were designed and tested in different biomes. The research



**Figure 6 – An early schematic of a automated system for handling requests for invasive species information linking relational databases and spatial modeling via the World Wide Web.**

team published papers challenging commonly held paradigms in invasive species science, and on the early attempts to integrate remote sensing data, field data, and spatial statistics to accurately quantify and map native and non-native diversity at landscape scales. In short, the team had built a welcoming view of how technology might improve the research and ultimately its value to society.

## **3.0 Partnership Development**

### **3.1 Basis for Partnership**

In the United States, the U.S. Geological Survey has the lead role for delivering invasive species scientific information on Federal lands. USGS technical and scientific capabilities directly support management of Department of Interior lands and waters by documenting, monitoring, and predicting the establishment and spread of invasive species. USGS studies the ecology of invading species and vulnerable habitats to support prevention, early detection, assessment, containment, and where possible eradication of new invaders and investigates the physical properties, composition, and hydrology of geologic substrates to identify lands vulnerable to invasion of exotic plants. The USGS NBII has several regional programs delivering invasive species information and is establishing a national node for invasive species.

NASA has a potentially complementary role to play in helping USGS understand and manage invasive species. NASA's Science Mission Directorate currently provides observations and measurements from Terra, QuikSCAT, Landsat 7, Jason and other missions of key ecosystem attributes needed to predict invasive species distributions. A number of planned missions in the near- to mid-term will expand these observations and measurements to include critical three-dimensional structure derived from SAR and LIDAR technologies. In addition, NASA provides the expertise in high-performance computing, systems engineering, and Earth system modeling needed to assure the successful transfer of the system into operational use.

### **3.2 How the Work Was Undertaken**

NASA collaboration in the ISFS resulted from four projects funded through competitive solicitations:

- “Biotic Prediction: Building the Computational Technology Infrastructure for Public Health and Environmental Forecasting”—John L. Schnase, Goddard Space Flight Center, funded through NASA Contract No. GSFC-CT-1
- “Predicting Regional-Scale Exotic Plant Invasions in Grand Staircase-Escalante National Monument”—John L. Schnase, Goddard Space Flight Center, funded through NASA Research Announcement NRA-00-OES-08
- “The Invasive Species Data Service: Towards Operational Use of ESE Data in the USGS Invasive Species Decision Support System”—John L. Schnase, Goddard Space Flight Center, funded under NASA Cooperative Agreement Notice CAN-02-OES-01
- “Value Added Products from Vegetation and Precipitation Time-Series Data Sets in Support of Invasive Species Prediction”—Jeffrey Morisette, Goddard Space Flight Center, funded under NASA Research Announcement NRA-03-OES-02

## **4.0 ISFS Framework Description**

The Invasive Species Forecasting System (ISFS) is a Web-based software framework that supports the data management and modeling activities needed to produce large-scale, predictive maps. It has been



implemented as a workflow management system that addresses two of the most crucial technological bottlenecks in the production of large-scale predictive maps: (1) the difficulty of finding, downloading, and integrating observational data about invasive species with the other types of satellite and environmental data used in predictive modeling, and (2) the computational demands of producing landscape- and regional-scale predictions that require large amounts of heterogeneous data as input. Since large-scale predictive maps are essential to almost all aspects of invasive species science, management, and policy decision-making, ISFS represents a significant advance in national capabilities, a state-of-the-art decision support tool of value to USGS and a wide range of public- and private-sector stakeholders.

ISFS has been built using a Web services approach. Web services allow capabilities to be “published” to an Internet community. Standard programming interfaces allow these published capabilities to be invoked through various client applications. Components of the service can be hosted anywhere on the Internet, and the functions offered by a service can be combined in different ways to build new applications. Web services have become the industry-standard approach to distributed computing and convey several advantages to the ISFS. Web services align ISFS with prevailing COTS technologies for Web-based computing, can be scaled to accommodate future growth, and allow ISFS capabilities to be tailored to new uses. The ISFS architecture comprises three primary components (Figure 7):

- Invasive Species Data Service (ISDS) — ISDS provides methods to ingest, catalog, visualize, and merge data from different sources. In particular, it allows USGS biological field data to be integrated with NASA satellite data of particular value to invasive species modeling. ISDS produces ISAMS-compatible datasets (see below) that can be used for ISFS-modeling or other purposes, and new datasets can be registered with the system at any time.
- Invasive Species Analysis and Modeling Service (ISAMS) — ISAMS allows users to conduct an interactive session in which data and models are brought together to produce predictive maps. ISAMS consists of a library of modeling routines that know how to ingest ISDS’s merged datasets and run them on high-performance cluster computers. Like ISDS, ISAMS can be extended by registering new participating modeling routines.
- Invasive Species Decision Support Service (ISDSS) — ISDSS consists of a library of decision support tools that use ISDS and ISAMS output. The primary tool developed so far is T-Map, a specialized modeling and mapping tool for tamarisk.

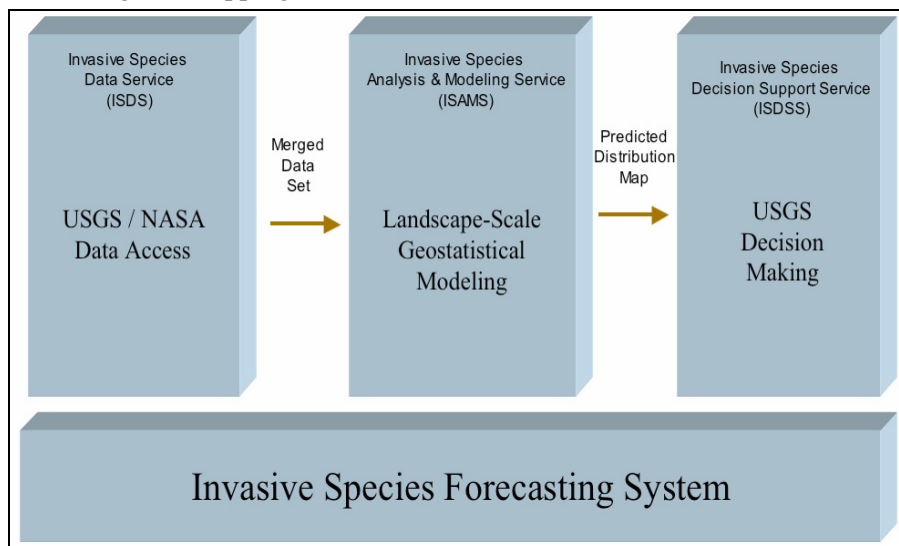


Figure 7 – The three primary components of the Invasive Species Forecasting System.

## 4.1 Status of the Framework

Over the past three years, the USGS/NASA ISFS team has developed an enterprise-level architecture and methodology for an operational system that can be maintained and extended without destabilizing core components. The ISFS currently is deployed at the Institute in an experimental mode. It has been used to produce models and maps for three of the Institute's major long-term study areas: the Cerro Grande Fire Site (CGFS) near Los Alamos, Rocky Mountain National Park (RMNP), and Grand Staircase-Escalante National Monument (GSENM). It also has been used to produce a National Tamarisk Habitat Suitability Map for the Continental United States. Over the next 24 months, the team will finish ISFS server and client development, harden the system for operational use, finish populating its underlying data service, complete verification and validation, and further benchmark the improvements gained by using ISFS in USGS and partner agencies.

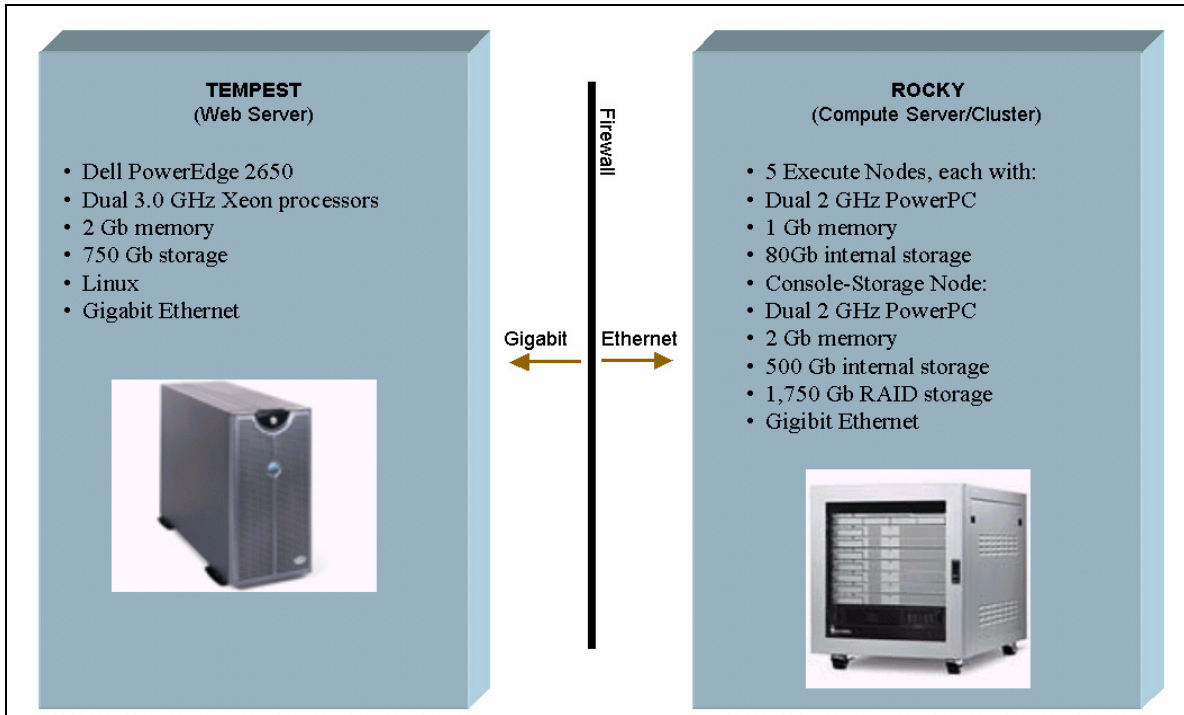
## 4.2 Implementation of ISFS at NIISS

The delivered system includes the following:

- ISFS software – ISFS V1.0 is a production-grade server application supporting a basic suite of internal operations. It exposes a subset of those operations as Web services. ISFS has a comprehensive Web interface for use by administrators and a simpler Web interface for use by the public. The major functions supported by the core system include:
  - User authentication and authorization
  - Persistent storage of system metadata and user-generated output
  - ISDS/ISAMS/ISDSS library maintenance and registration operations
  - ISFS kernel components, such as job controller, scheduler, and data manager
  - SOAP-compliant Web service functions
- ISDS library – An initial collection of biological, satellite, and environmental data is provided with ISFS V1.0. In addition, ISDS contains a library of merged datasets from all model runs produced during the development process. Metadata documentation accompanies each dataset. The core ISDS dataset suite consists of the following:
  - Ecological Field Data
    - Tamarisk Presence/Absence point data for the Continental United States
    - Species Richness profile by county for the Continental United States
    - GODM core ISDS subset, including field data from NIISS study sites and research datasets contributed by NIISS partners
  - Satellite Imagery
    - Continental United States:
      - 3-year time-series summary statistics of MODIS MOD13Qv4 Vegetation Index 16-day composites at 250 m resolution
      - 96-day IGBP Land Cover mosaic of MODIS MOD12Q1
      - 1 km incremental Distance to Streams from 2002 ESRI hydrology product

T. Stohlgren, J. Schnase, J. Morisette, N. Most, E. Sheffner, C. Hutchinson, S. Drake, W. Van Leeuwen, and V. Kaupp.

- Soil Characteristics from Natural Resources Conservation Service (NRCS) STATSGO database
- Regional Datasets:
  - SSURGO Soils for Colorado and other states (may become a national dataset)
  - SRTM 30 m elevation for Colorado (may become a national dataset)
  - 1 km Landsat NDVI and Tassel Cap for selected states
  - 5 m EOS-Hyperion hyperspectral reflectance for selected study sites
  - 15 m Aster multispectral reflectance for selected study sites
- Ancillary GIS layers
  - Manmade boundaries - cities, counties, roads, national parks, BLM land
  - Natural boundaries - lakes and streams
  - Topographic – 24 k elevation, aspect, slope
  - Demographic - human population
  - Environmental - precipitation and temperature
- Computed Products - The ISDS library also includes a collection of “starter kit” merged datasets and model arrays, maps, JPGs, GeoTiffs, model reports, and metadata for all work associated with RMNP, CGFS, GSENM, the National Tamarisk Map, and the new study sites at LaJunta, Debeque, and Gunnison
- ISAMS library – An initial collection of modeling algorithms is provided with ISFS V1.0. Comprehensive metadata documentation accompanies each routine. The core ISAMS modeling suite consists of the following:
  - Ordinary Least-Squares Regression (+/- kriging)
  - Logistic Regression
- ISDSS library – An initial Web service decision support tool called *T-Map* is provided with ISFS V1.0. *T-Map* is an interactive modeling and mapping tool specialized for use on Tamarisk. *T-Map* is the primary interface for delivering Tamarisk data, the National Tamarisk Habitat Suitability Map, and other Tamarisk-related products.
- Hardware – An initial hardware configuration (Figure 8) consisting of the *Tempest* Web server and *Rocky*, an Apple Xserve G5 cluster computer, is provided with ISFS V1.0.
- Documentation – The core ISFS documentation suite consists of the following delivered in paper and electronic form:
  - Standard software engineering documents, including the *ISFS Requirements Document*, *ISFS Software Design Document*, and *ISFS Test Plan*
  - Standard system administration documents, including the *Installation Manual* and the *System Administration and Configuration Manual*
  - The *ISFS Users Guide* and the *ISFS Programmers Guide*



**Figure 8 – The TEMPEST Server and the ROCKY Cluster are the hardware backbone of the ISFS.**

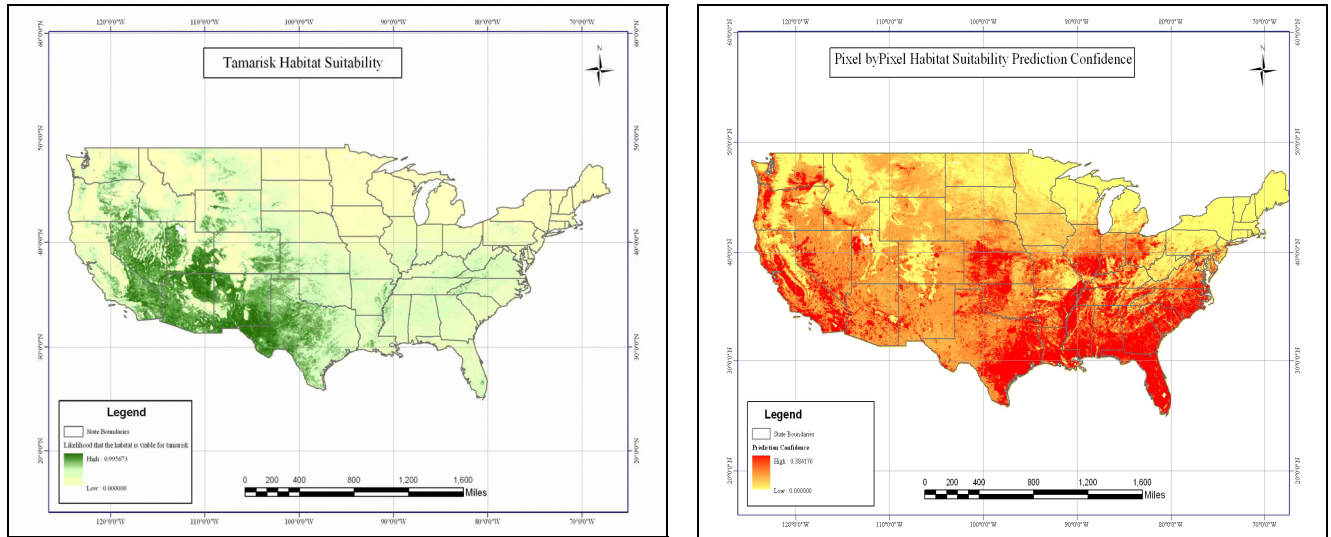
### 4.3 Products and Services

The ISFS can support a variety of products and services, including:

- Models and maps – Dynamic printed and electronic predictive maps are the primary products generated by the ISFS (Figure 9). These maps can show areas at risk of invasion, habitats suitable for invasive, and the expected distribution of a species or other biological resource. Maps can be combined with other types of data, such as stream data, road locations, and demographic data, to further aid in analyses. An important feature of the geo-statistical methods used in ISFS is their ability to produce “confidence” maps showing the statistical error present in a prediction.
- Merged datasets – ISFS’s Invasive Species Data Service (ISDS) can be used to easily combine field data with other types of satellite data to produce geospatially-registered merged datasets for use in other applications.
- Research and consulting – Collaborative research with other government and non-governmental agencies is an important aspect of the Institute’s mission and a significant source of funding. The ISFS is designed to support the work processes commonly used by the Institute. As a result, the Institute’s graduates students and scientists will be able to perform their core modeling activity more effectively and at reduced cost. The Institute also will be more competitive in obtaining research and consulting work because of this unique advantage.
- National data and model repository – Field data drives predictive modeling. The ISFS’s Invasive Species Data Service is designed to store, deliver, and act as a library for raw data as well as the computed predictive maps generated by model runs. It thus creates an opportunity for the Institute to become a national repository for predictive models and associated data. This complements and

extends the Institute’s work to develop a global database of field data on non-native plants, animals, and pathogens.

- Application development and private partnerships – With access to software engineering expertise and high-performance computing, the availability of ISFS’s Web services protocol, and its rich connections to the invasive species community, the Institute could pursue new types of work based on application development. Licensing or other public-private partnership arrangements are likely to be the best way to realize this opportunity. While this is perhaps the most adventurous use of ISFS capabilities, the funding potential for invasive species products and services in the coming years makes application development worth considering.



**Figure 9 – Representative products developed out of the ISFS. The image on the left is a CONUS Habitat Suitability Map for Tamarisk. The image on the right is the corresponding Predictive Confidence Map for the tamarisk prediction.**

#### 4.4 Target Customers

ISFS development has been driven by customer needs. Technical requirements for ISFS were extracted through conversations with scientists at the USGS Fort Collins Science Center. Community needs assessments occur on an ongoing basis as a series of one-on-one and group meetings in which potential federal and non-federal ISFS users responded to standard questionnaires in an interview setting. These studies consistently reveal strong, across-the-board interest in ISFS as a national capability. The following demonstrate the types of customers and work that the ISFS is designed to support:

- Federal agencies, responsible for managing public land and preserving natural habitats, that perform research and report on ecological and biological health or publish scientific results and data to the public. Examples include
  - U.S. Fish and Wildlife Service
  - National Park Service
  - Bureau of Land Management
  - U.S. Geological Survey

- U.S. Forest Service
- Department of Energy.
- State and local governmental institutions that work with natural resource planning and preservation, such as state departments of natural resources and county weed management agencies.
- Local, national, and international natural resource conservation groups, such as Nature Conservancy and the Natural Heritage Programs.
- Private companies performing services such as landscaping, retail, commercial and municipal construction, aerial photography, pesticide and herbicide production and sales, weed and exotic species mapping, logging and agriculture.

#### 4.5 Benchmarking Performance Metrics

The strategy for benchmarking NASA enhancements to the ISFS will focus on the following:

- Measuring improvements in computational performance of the core, USGS modeling codes used in ISFS's Analysis and Modeling Service (OLS Regression with kriging)
- Measuring improvements in data access and cost-reduction gained by the use of ISFS's Data Service
- Measuring the impact of ISFS products and services on target customers, starting with researchers at the Institute.

During this initial phase, benchmarking has focused on improvements in computational performance.

Two “canonical” study sites are being evaluated to measure code improvements: the Cerro Grande Fire Site in Los Alamos, NM, and the Rocky Mountain National Park, CO. These two sites provide contrasting ecological settings and analysis challenges and vary in the types and scales of data used, areas covered, and maturity of the investigation. As described in detail in the *Baseline Software Design Document* (NASA/USGS, 2002b), three factors influence the performance of ISFS model code: 1) the size of the output surface area over which kriging occurs (area), 2) the total number of sample points in the data set (pts), and 3) the number of “nearest neighbor” (nn) sample points from the total data set actually used to compute a kriged value for any given point in the output area.

When the NASA team first began working with colleagues at USGS, a scalar, single-processor run of this model using S-plus took approximately two weeks. The major computational bottleneck in the model is the kriging routine. Solving for the weights in the equations that form the ordinary kriging system uses LU decomposition with backsubstitution to do matrix inversions. The overall computational complexity of ordinary kriging is thus  $O(n^3)$  and the time required to compute a result is strongly influenced by the number of sampled data points used to estimate the residual surface across the entire study area. The overall goal for code improvement was to reduce processing times and increase the amount of data handled by the model. As described in BP-BSD-1.3, increasing the amount of data handled by the model translates into either increasing spatiotemporal resolution or increasing coverage.

The team first wanted to accomplish quantitative improvements in the underlying model that had been agreed upon by the user community as minimal advances needed to improve core capabilities. These goals, referred to as “Community Improvement Goals,” were driven by the fact that a 16-node cluster was built in the USGS facility. NASA, however, has access to greater computational capabilities that can apply this modeling approach to some important and challenging problems that have been previously



unapproachable. Thus, the NASA’s clusters were targeted for use to attain more challenging performance improvement goals concurrent with accommodating basic needs. These complementary challenges are considered “Advanced Improvement Goals.” Table 1 provides a summary of the baseline performance characteristics of the core USGS model code as well as the various performance goals anticipated over the course of this effort.

**Table 1 - Representative products developed out of the ISFS. The image on the left is a CONUS Habitat Suitability Map for Tamarisk. The image on the right is the corresponding Predictive Confidence Map for the tamarisk prediction.**

BASELINE SCENARIO		Sec	Min	Hrs	Days	Notes
CGFS base 079 pts 79 nn 01x area (S-Plus) (Version 0.0)		-	-	-	-	
CGFS base 079 pts 18 nn 01x area (S-Plus) (Version 0.0) (USGS Actual)		1209600.0	20160.0	336.0	14.0	
CGFS base 079 pts 18 nn 01x area (S-Plus) (Version 0.0) (NASA Estimate)		1608426.0	26807.1	446.8	18.6	
CGFS base 079 pts 18 nn 01x area (FORTRAN) (Version 0.1)		114.5	1.9	0.0	0.0	
CGFS base 079 pts 79 nn 01x area (FORTRAN) (Version 0.1)		3702.6	78.4	1.3	0.1	A
RMNP base 1800 pts 18 nn 01x area (FORTRAN) (Version 0.1)		443.0	7.4	0.1	0.0	B
RMNP base 1800 pts 1180 nn 01x area (FORTRAN) (Version 0.1) (est.)		6812384.0	113539.7	1892.3	78.8	
COMMUNITY IMPROVEMENT (CI) GOALS	x baseline	Sec	Min	Hrs	Days	
CGFS base 079 pts 79 nn 01x area (Version 1.0 -F)	25.0	188.1	3.1	0.1	0.0	C
CGFS base 790 pts 79 nn 01x area (Version 2.0 -G)	25.0	188.1	0.3	0.1	0.0	D
CGFS base 790 pts 79 nn 10x area (Version 2.0 -G)	2.5	1881.0	31.4	0.5	0.0	E
RMNP base 1800 pts 18 nn 01x area (Version 2.0 -G)	25.0	17.7	3.1	0.0	0.0	C
ADVANCED IMPROVEMENT (AI) GOALS	x baseline	Sec	Min	Hrs	Days	
CGFS base 079 pts 79 nn 01x area (Version 1.0 -F)	200	23.5	0.4	0.0	0.0	F
RMNP base 1180 pts 1180 nn 01x area (Version 2.0 -G)	1000.0	6812.4	113.5	1.9	0.1	G
RMNP base 11800 pts 1180 nn 01x area (Version 2.0 -G)	1000.0	6812.4	113.5	1.9	0.1	H
RMNP base 11800 pts 1180 nn 100x area (Version 2.0 -G)	10.0	681238.4	11354.0	189.2	7.9	I

A: Proposed CGFS canonical baseline using FORTRAN kriging routine.

B: Proposed RMNP canonical baseline using FORTRAN kriging routine.

C: Milestone 1 CI Goal -speed up- 75% efficiency, 32 node cluster = 25x speed up

D: Milestone 2 CI Goal -increased resolution- “sliding window” adaptive selection of 10% of 10x nn from 1x area

E: Milestone 2 CI Goal -increased coverage- “sliding window” adaptive selection of 10% of 10x nn from 10x area

F: Milestone 1 AI Goal -speed up- 75% efficiency, 256+ node cluster = 200x speed up

G: Milestone 2 AI Goal -speed up- 75% efficiency, 1024+ node cluster = 1000x speed up

H: Milestone 2 AI Goal -increased resolution- “sliding window” adaptive selection of 10% of 100x nn from 1x area

I: Milestone 2 AI Goal -increased coverage- “sliding window” adaptive selection of 10% of 100x nn from 10x area

## 5.0 Impact of ISFS on NIIS

### 5.1 First Code Improvement – Parallel Kriging

Kriging is a spatial interpolator that determines the best linear unbiased estimate of the value at any given pixel in an output surface or image using a weighted sum of the values measured at arbitrary sample locations. It determines the weights and the spatial continuity of the data as measured by the variogram. The scalar kriging algorithm is a double loop over all rows and for each pixel within the row. At each

pixel the algorithm determines the  $n$  nearest neighbor sample points and computes the  $(n \times n)$  distance matrix containing the Euclidean distance between each sample points, and also computes the  $(n \times 1)$  distance vector from the pixel to each of the sample points. The Euclidean distances are converted to statistical distances by applying the variogram model to create a covariance matrix and vector. The kriging weights are obtained by multiplying the inverse of the covariance matrix by the covariance vector. The computationally expensive part of kriging is the inversion of the covariance matrix, which is done at each pixel since the nearest neighbor sample points can vary across the kriged surface.

The steps to estimate the value at each pixel are independent of all other pixels. The algorithm is therefore “elegantly parallel” and highly amenable to parallel implementation via domain decomposition – a processor is simply assigned to each a section of the output kriged surface or image. The domain is decomposed along the rows only, i.e. each processor works with full rows of the output surface. This means the inner loop over columns can be left unaltered. The parallelization could decompose into contiguous rows, effectively giving each processor a strip of the output image. Instead consecutive rows are assigned to separate processors. Thus, for a kriging  $512 \times 512$  image using 32 processors, the first processor would be assigned rows 1, 33, 65, . . . 449 and 481, while the last processor would calculate rows 32, 64, 96, . . . , 480 and 512.

Both domain decompositions are equally load balanced if the number of sample points used in the covariance matrix is always the same at each pixel. This is the case now, but plans are to implement an adaptive scheme that will use more points in densely sampled regions and fewer points in sparsely sampled areas. Significant load imbalance would result if sparsely sampled rows are assigned to one processor while assigning densely sampled rows to another processor.

Parallel kriging is implemented in FORTRAN using MPI, the Message Passing Interface. The code employs a ‘node 0’ controller process and a collection of worker nodes. Prior to execution an input data file is copied to each node containing the dimensions and cell spacing of the output kriged surface, the variogram parameters that describe the spatial structure, and the series of plant diversity measurements (UTM X and Y coordinates and the number of plant species at each location). Each node reads this input data file, computes the kriged estimates for its assigned rows, and then sends each row to node 0. Node 0 only receives the data from the worker nodes, assembles the kriged surface in memory, and writes the final kriged estimates to its local disk.

Computation and communication overlap to increase parallel efficiency. When the first row has been calculated, an asynchronous send (MPI ISEND) of this row is issued to node 0. Since this is a non-blocking send, the processor proceeds to calculate the second row. At the end of this row, a wait is issued (MPI WAIT) to insure that the first row has been received by node 0 before proceeding. For the smallest kriged surface tested ( $512 \times 512$ ), the compute time for each row is over 4 seconds, thus the first row has more than sufficient time to be received and the wait call should also return “immediately” (in reality, the latency time MPI’s implementation of the MPI ISEND and MPI WAIT calls). Meanwhile, node 0 posts a serial set of asynchronous receive calls (MPI IRECV) for each row sent by the worker nodes, followed by a series of waits (MPI WAITS). When the waits are finished, each row of data is copied into the appropriate location within the output kriged array on node 0.

This parallel implementation has been evaluated on two clusters at the NASA Goddard Space Flight Center. The code was designed for use on the Medusa cluster, on which was met the Community Improvement Goals. Medusa is a 64-node, 128-processor, 1.2 GHz AMD Athlon cluster with 1 GB of memory per node and 2.3 TB of total disk storage. Each node is connected to the others with dual-port



Myrinet. Node 0 is `frio.gsfc.nasa.gov`, a Linux PC with a single 1.2 GHz AMD Athlon processor and 1.5 GB memory, which resides on one of team member’s desks and is connected to the Medusa cluster via fiber Gigabit Ethernet. Typically logon is only into Node 0, and to the user it appears that all calculations are done on Node 0. Thunderhead has also been used to evaluate the Advanced Improvement Goals. Thunderhead is a 512-processor, 2.4 GHz Pentium 4 Xeon cluster with 256 GB of memory and 20 TB of disk storage.

## 5.2 Parallel Kriging Results on Ideally Sized Test Cases

Performance results are presented for four test problems to evaluate the efficiency of the parallel implementation. These tests held the number of input data points constant at 79 (the size of the field sample data set for the Cerro Grande Fire Site), while the output kriged image size varied from 5122, 7682, 10242, to 20482. These problem sizes are ideally sized in the sense that an equal number of rows are assigned to each processor in all cases. In each case the area kriged was held constant and the pixel size was decreased as the problem size increased.

Table 2 shows the results of this timing study. The processing times shown are elapsed wall-clock time in seconds. As expected, the kriging time increases in direct proportion to the area of the output kriged surface (e.g. the 20482 problem ran 16x longer than the 5122 case). The processing times decreased nearly linearly as the number of processors was increased, as shown in Figure 10. The scaling efficiency for N processors is defined as the ratio of the 1-processor to N-processor wall-clock times divided by N. The efficiencies obtained were excellent, shown in Table 3, ranging from 96–98% when using 32 processors and over 99% when using 16 or fewer processors. The scaling efficiencies dropped slightly for the 64-processor tests, but were still greater than 97% for the 20482 problem.

**Table 2 - Test case timing results (elapsed wall clock seconds).**

Number of Medusa Processors	Size of Kriged Image			
	2048 <sup>2</sup>	1024 <sup>2</sup>	768 <sup>2</sup>	512 <sup>2</sup>
65	583.9	147.5	84.0	38.4
33	1150.4	289.8	163.8	73.8
17	2285.1	573.89	324.0	144.7
9	4558.4	1142.0	642.8	287.2
5	9083.9	2277.4	1281.3	571.5
3	18190.4	4556.3	2562.0	1140.9
2	36252.7	9079.6	5107.0	2269.1

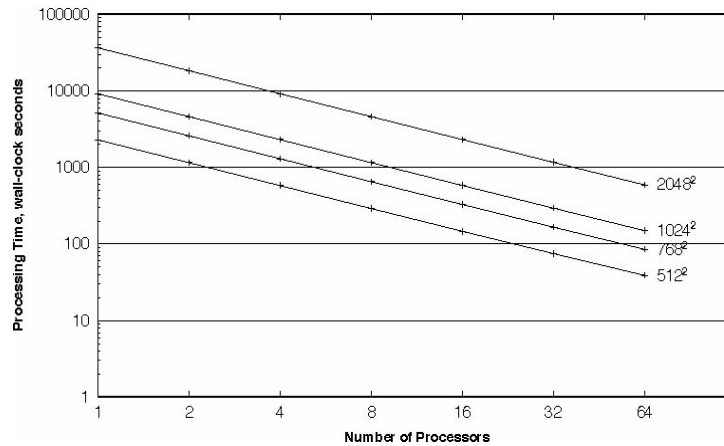


Figure 10 – Scaling curves for test cases on Medusa.

Table 3 - Scaling efficiencies for test cases.

Number of Medusa Processors	Size of Kriged Image			
	2048 <sup>2</sup>	1024 <sup>2</sup>	768 <sup>2</sup>	512 <sup>2</sup>
65	97.0%	96.2%	95.0%	92.3%
33	98.5%	97.9%	97.4%	96.0%
17	99.2%	98.9%	98.5%	98.0%
9	99.4%	99.4%	99.3%	98.8%
5	99.8%	99.7%	99.6%	99.3%
3	99.6%	99.6%	99.7%	99.4%
2	100.0%	100.0%	100.0%	100.0%

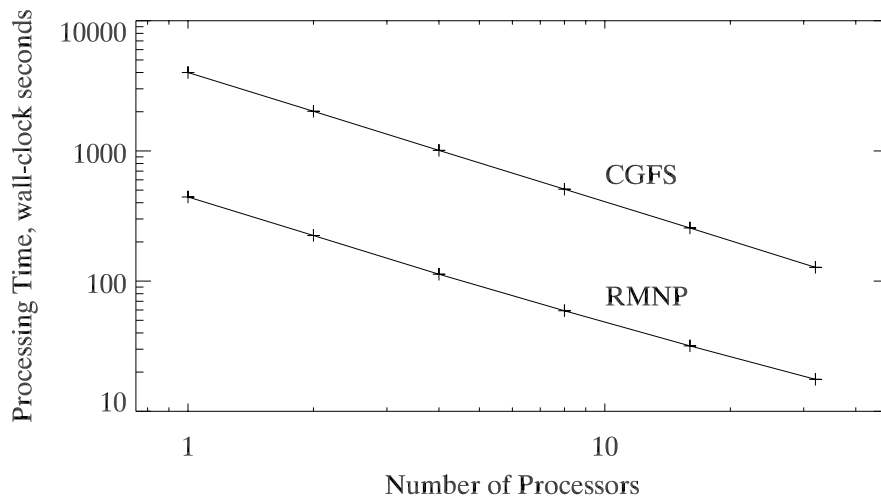
### 5.3 Evaluation of Community and Advanced Improvement Goals

The evaluation of the performance of the parallel kriging on the baseline scenarios defined in [Table 1](#) follows. These differ from the ideally sized test cases evaluated above because a different number of rows are assigned to each processor. As such they represent real operational scenarios that estimate an arbitrary area at a given resolution and use all available processors. Seldom in such scenarios will the number of rows or processors be a power of two, and there will be “left over” rows that lead to load imbalance and reduced scaling efficiencies.

[Table 4](#) shows the performance results for both the CGFS and RMNP baseline scenarios, which are plotted in [Figure 11](#). The run times improve for the CGFS test to 2 minutes and 7 seconds, which is more than a minute better than our Milestone 1 Community Improvement goal of 3 minutes and 8 seconds. Similarly, the RMNP test improves to 17.6 seconds, which is 25.2 times faster than the baseline run time of 7 minutes and 23 seconds. This case can now be run interactively. As expected, however, the scaling efficiency drops as the run time is reduced due to the parallel overhead and load imbalance.

**Table 4 - Baseline scenario timing and scaling results.**

Number of Medusa Processors	Baseline Scenario			
	CGFS (715 rows, 652 cols)		RMNP (1041 rows, 1186 cols)	
33	127.5	90.1%	17.6	76.3%
17	255.9	87.1%	31.8	81.9%
9	508.2	82.9%	59.2	83.1%
5	1009.1	75.1%	113.0	66.0%
3	2016.0	62.7%	223.9	98.9%
2	4000.4	52.8%	442.8	50.0%



**Figure 11 – Scaling curves for test cases on Medusa.**

## 5.4 Progress Toward Milestones

Test results indicate attainment of the Milestone 1 Community Improvement goal of a 25x speed up on a 32-processor cluster with greater than 75% efficiency. The results also document the general scaling behavior of the kriging algorithm and point to the limits in scalability expected as more nodes are allocated to the canonical data sets. The project has successfully completed Milestone 1 (First Code Improvement) according to plan.

The USGS ISFS team provided the following summary of noted improvements in their operational capability that have been realized as a direct result of the collaboration with NASA ([Table 5](#)).

**Table 5 - ISFS operational improvements summary.**

<b>Benchmark</b>	<b>Pre-NASA Partnership (Time Zero: 1995–2000)</b>	<b>Post-NASA Partnership (Current – 2005)</b>
Average Model Run Time	21 days	2 minutes: improvements in code, parallel processing, and software engineering
Typical Cumulative Model Run Time	4 months	8 hours: improvements in error checking, quick outputs and visualizations, and software engineering
Maximum modeled area	750 ha to 65,000 ha	State of Colorado: 104,000 square kilometers
Maximum number of remote sensing types	One (Landsat TM)	5 (Landsat TM, MODIS, ASTER, Hyperion, SRTM)
Number of Platforms Supported	2 (IMB PC; Sun Workstation)	3 (IMB-PC; Sun Workstation; Apple G5)
Long-term Strategic Planning	Absent	National Institute of Invasive Species Science
Database Interoperability	Low (local use)	High (integrated system design can link to global scale datasets and remote sensing types).
Number of Entities Using or Relying on the ISFS	0 (did not exist)	4 (CSU, USGS, State of Colorado, BLM )

## 6.0 Verification and Validation

Verification is defined as the process of evaluating a system or component to determine whether the products of a given development phase satisfy the conditions imposed at the start of that phase. Validation is defined as the process of evaluating a system or component during or at the end of the development process to determine whether it satisfies specified requirements. Verification asks "Did we build the system right?" and validation asks "Did we build the right system?"

Since NASA's role is to enhance system capabilities, ISFS V&V will focus on engineering aspects of the project and the direct impact that engineering enhancements are having on the end user. Broader cultural impacts of the ISFS within USGS and the extended invasive species research community will be evaluated over time as the system becomes more widely used. Outcomes relating to ecological and remote sensing science will be validated and reviewed according to accepted practice for each community and will not be included in ISFS V&V.

This first phase of work has concentrated on parallel algorithm improvements and the construction of an integrated web-based environment for invoking ISFS capabilities. The primary system goal has been to generate statistical modeling and mapping output from the ISFS that matches the results obtained by USGS prior to re-coding. The primary user-need relating to this engineering enhancement was for typical USGS analyses to be completed in a more automated way and in less time.

From a V&V perspective this system goal is being addressed in two ways. First, the ISFS statistical modeling output is compared to the statistical output from the original USGS analysis. This output

different evolutionary states of the DSS and the display and comparison of their respective residual risk profiles and objectives attainment.

## 7.2 DSS Benchmarking – Risk Management Approach

Benchmarking is a formal process intended to document how the integration of NASA data and technology improve the capabilities and efficiency of the ISFS. This process involves a direct comparison of the limited State 1, or “as is” condition with the enhanced State 2, or “target” condition,

At State 1, the DSS will be benchmarked via appropriate metrics to indicate its “as is” performance characteristics as a baseline for the system. Similarly, at State 2, its “target” performance will be measured against the same metrics but it will also be considered with respect to an enhanced suite of performance measures. The performance of the ISFS at State 2 will be compared to its capability at State 1 via the common set of indicators, and the improvements and changes in the DSS will be measured and documented.

Usability indicators and defined performance metrics, critical for assessing changes to the DSS, will be obtained through questionnaires and interviews, and the DDP risk assessment software will be employed as described above. The current ISFS description documents provide critical information about the State 1 characteristics of the ISFS and will aid in the set-up of the benchmarking process and selection of performance indicators. The benchmarking process does not involve developing the State 2 enhanced DSS from the State 1 system, but is generally conducted in parallel with the developmental process and is primarily associated with measuring the State 1-2 performance characteristics and documenting the results achieved. It usually involves the following tasks to assess performance-enhancing contributions to a DSS:

1. Involve users and experts in all stages of the enhancement process (“As is”→ “To be” OR State 1→State 2).
2. Characterize State 1 or baseline the DSS including mandate, requirements/objectives and DSS functioning.
3. Select performance and usability indicators to analyze and compare the “As is” and “To be” systems using an approach that involves user questionnaires and a software tool (DDP) adapted for benchmarking DSS’s.
4. Evaluate “As is” DSS performance
5. Plan assimilation process (transition), and formulate enhancements to the DSS (“To be”).
6. Develop enhanced prototype with benchmarking partners and document information about the state-of-the-art DSS and its science, data and technology inputs.
7. Optimize prototype based on user and expert guidance and existing constraints.
8. Evaluate performance of the prototype enhanced DSS (“To be”) based on user feedback.
9. Compare performance of “As is” and “To be.”

Using a systems engineering approach, the DDP process is intended to facilitate benchmarking over the entire DSS project life cycle beginning with baseline system evaluation, continuing to data, and science and technology assimilation decisions all the way through operations. The DDP process aims to determine (1) the DSS requirements/objectives and their importance, (2) the risk factors and their impact

includes estimated model parameters. Each parameter from the ISFS output is compared to the corresponding parameter from the USGS analysis. Values are considered acceptably similar if they match to within four significant digits. Second, the mapping output products from ISFS are compared to the maps resulting from the USGS single-processor analyses. In both cases the maps are raster data products; so that pixel-by-pixel subtraction can be utilized to create a “difference image” that can be used to check the consistency of the two output products. The output is considered acceptable if the “difference image” has no pixels with an absolute value greater than 0.1% of the mean of the USGS output image.

We address the second goal of automation and speed by noting the computational time involved with the USGS original analyses and the processing time required by ISFS according to standard means of quantifying parallel algorithm improvements, as shown in Section 5. At this stage, V&V confirms that ISFS is meeting engineering specifications and engineering-related user needs as planned.

## **7.0 Future ISFS Benchmarking Approach - DDP Process**

### **7.1 Overview**

This section is intended to show the benchmarking approach and the scope of the benchmarking effort that is scheduled to be performed in beginning in FY 2006. It is anticipated that the changes in and enhancements to the ISFS, such as the application and utilization of NASA EOS data, will be measured and evaluated using systematic benchmarking techniques. A team of interested government personnel and science- and applications-oriented academicians has been established to guide the process. Broadly, the benchmarking process will begin by a) defining DSS requirements or objectives, b) identifying risks or obstacles to the achievement of those objectives, and c) analyzing sets of mitigations that can be used to minimize or eliminate the risks and achieve a demonstrable level of objectives attainment. All phases of the benchmarking process will be documented and the benchmarking activities will conclude by applying performance metrics to measure the difference between the “as is” and the enhanced “to be” states and documenting the results.

The main objectives of this section are to describe a benchmarking strategy and develop a protocol that can:

1. Help define significant, mission-oriented ISFS requirements;
2. Provide insight into the risks that stand in the way of achieving them; and,
3. Identify mitigations to offset the risks as critical factors for the adaptation and incorporation of enhancements to the DSS.

Information will be gathered to benchmark the ISFS in its present-developmental and future-enhanced states. Both qualitative, using data from questionnaires and interviews, and quantitative techniques will be used. A primary tool to be used for analysis is an interactive risk management software suite, the Defect Detection and Prevention (DDP) application developed by NASA's Jet Propulsion Laboratory. The DDP will be employed to quantify the effectiveness of the enhancements provided by incorporating new generation NASA information products into the ISFS. This will be done using risk balance and attainment of objectives as performance indicators. The main steps in using the DDP as a primary analysis tool in the benchmarking process are: 1) to identify and formulate ISFS DSS requirements/objectives, 2) estimate the impact of risks/obstacles on the requirements/objectives, and 3) evaluate the effectiveness of mitigation factors or solutions that alleviate risks and enhance attainment of objectives. The DDP supports sensitivity analysis, and this will allow for the evaluation of mitigation scenarios corresponding to

on attaining requirements/objectives, and (3) the effectiveness of mitigating inputs that help retire risks and attain the objectives.

Based on work to date with NASA and ISFS, a partial list of candidate performance and usability metrics is given in Table 6.

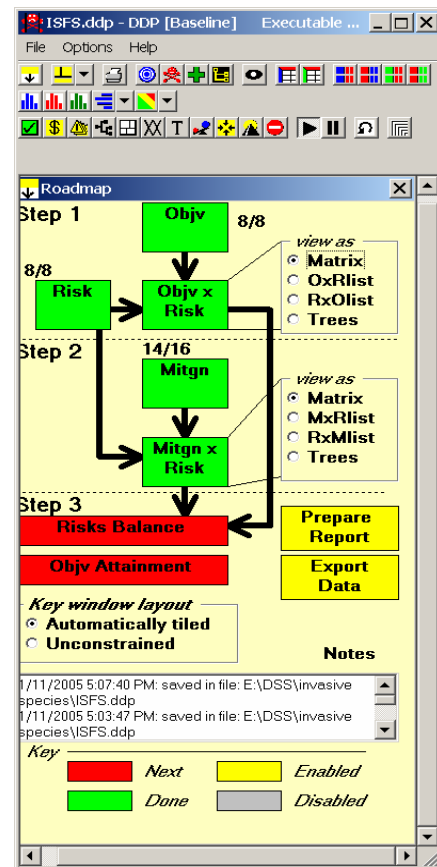
**Table 6 - Candidate Performance and usability metrics under consideration (State 1→State 2).**

DSS Usability Metrics (State 1→State 2)	DSS Performance Metrics (State 1→State 2)
Ease of use (GUI, tools)	Attainment of mandate, objectives and requirements
Frequency of use (e.g. Web statistics)	Cost effectiveness (data sources, DSS tools)
System statistics	User needs (data quality, consistency, timeliness, data mining)
Learning curve (training)	Organizational needs (information technology, expertise)
Workload (more data, better models)	System performance (accuracy, timeliness, system statistics, processing speed and costs)
User needs (consistent data, value-added products)	Bottlenecks (network)
User tasks (interactive, automated)	
Documentation	

Figure 12 provides a look at the DDP tool interface as it initializes. The user is guided to enter the DSS objectives, risks and mitigations as row and column headings in the Impact matrix (Objectives x Risks) and the Effectiveness matrix (Mitigations x Risks), populate the matrices with numerical cell values, and perform analyses to obtain residual risk profiles and objectives attainment levels.

The DDP process is accomplished in six processing steps:

1. Identify all goals, requirements or Objectives of the ISFS for state 1 (+ state 2) and rate their relative importance.
2. List the Risk factors/obstacles (their a priori likelihood of occurring is initially assumed to be equal) which affect attainment of Objectives and hinder ISFS functionality (e.g. lack of data, usefulness of tools).
3. Establish Mitigations (e.g., assimilated NASA products and technology) that reduce the Risks and improve attainment of Objectives.
4. Appraise the impacts of the Risks on the ISFS Objectives (i.e., if a Risk occurs, how much loss in attainment of Objectives would it cause).
5. Assess the effectiveness of the Mitigations at preventing or reducing the likelihood and impact of the Risks/obstacles.



**Figure 12 – DDP tool initial GUI.**

6. Balance the residual Risk profile with a selection of Mitigations and run Objectives-attainment scenarios.

Interviews, discussions and surveys will allow for the discovery of ISFS goals and objectives, risk factors that could get in the way of attaining the goals and mitigation factors that are in place, or are planned, to reduce the risks.

The impact of risks on meeting requirements and the effectiveness of mitigations in reducing risks will be obtained by designing a survey for the ISFS user community. The answers to this survey will then be translated into relative values of 0, 0.1, 0.5, 0.9, and \* for None, Low, Medium, High and Unknown respectively. Consequently, the geometric means of these values will be entered into the DDP tool, resulting in an Objective x Risk impact matrix, and a Mitigation x Risk effectiveness matrix. The matrices will be used to evaluate risk balance scenarios and attainment of Objectives. In conjunction with the survey, the relative importance of the ISFS's requirements and objectives will be determined through discussion with ISFS users and collaborators.

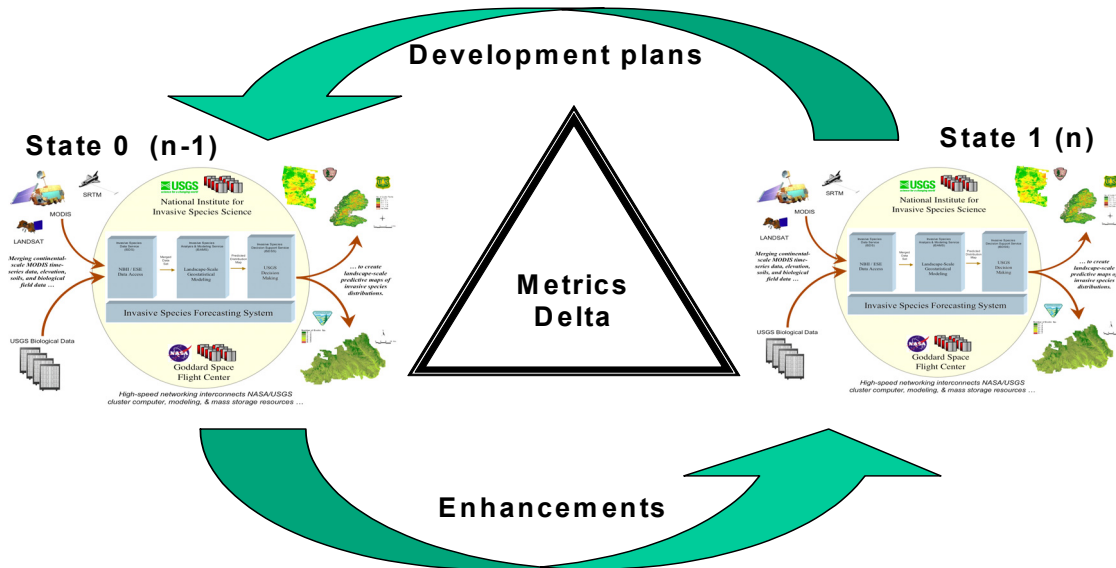
The purpose of the DDP tool is to systematically quantify the performance differences between State 1 and State 2 of the DSS. The DDP tool and an associated elicitation process is employed to prioritize the Objectives, to weight the impact of the Risks, and to evaluate the effect of various Mitigation combinations at reducing these Risks while attaining the Objectives. By simple mathematical operations, the potential impact of a given Risk and the effectiveness of a collection of Mitigations results in a residual Risk profile and visualization of attainment of specific Objectives. The software suite also supports a sensitivity analysis permitting identification of the requirements having the greatest degree of impact on attaining a desired outcome or result. The outcome of the DDP process allows for the quantification of NASA contributions (mitigation factors) in terms of enhanced attainment of Objectives or reduction of Risks between State 1 and State 2, or between State 2 and future "to be" states.

Results from the currently ongoing benchmarking activities will be used to determine the contribution of the proposed enhancements to the system and evaluate it with respect to the current state of the DSS. It should be noted that the benchmarking process is best done by including all aspects of the DSS (management, users, tools, Information Technology, and data sources) as the assimilation of a product will affect overall system performance and that of individual components of the DSS (e.g. user interface, knowledge and problem processing system).

### **7.3 DDP Benchmarking Conclusions**

The DDP software is a powerful tool to consistently manage and analyze risk and benchmark the ISFS. The DDP tool is also useful in eliciting constructive discussions and helping direct the benchmarking thought process in a risk management context. Since NASA enhancements to the ISFS are currently at various stages of implementation (e.g. between State 0 and State 1), the process/approach presented here can be complemented with annual repeat surveys or when most State 2 enhancements are implemented. [Figure 13](#) shows the cyclic nature of the benchmarking process. The approach to measuring the difference in performance of the DSS between State 1 and State 2 will be presented in a subsequent benchmarking report and will be based on a comparison of performance indicators like Risk balance and Objective attainment using the current ISFS-DDP application as a baseline.





**Figure 13 – The benchmarking process of the enhancements between a state (n-1) and the following consecutive state (n) follows a cyclic pattern.**

## 8.0 Bibliography

- Alley, N., T. J. Stohlgren, P. Evangelista, and D. Guenther. 2004. Iterative model development for natural resource managers: a case example in Utah's Grand Staircase-Escalante National Monument. *Geographic Information Sciences*, In Press.
- Baron, J. S., ed. 2002. *Rocky Mountain futures: An ecological perspective*. Washington, DC: Island Press. 325 p.
- Bashkin, M., T. J. Stohlgren, Y. Otsuki, M. Lee, P. Evangelista, and J. Belnap. 2002. Soil characteristics and exotic species invasions in the Grand Staircase-Escalante National Monument, Utah. *Applied Soil Ecology*, 22:67-77.
- Chong, G. W., R. M. Reich, M. A. Kalkhan, and T. J. Stohlgren. 2001. New approaches for sampling and modeling native and exotic plant species richness. *Western North American Naturalist*, 61:328-335.
- Cressie, N. 1995. Bayesian smoothing of rates in small geographic areas. *Journal of Regional Science*, 35:659-673.
- Kallas, M. 1997. Armillaria root disease on the Black Hills National Forest. MS thesis, Dept. of Forest Sciences, Colorado State University, Fort Collins, CO.
- Metzger, K. 1997. Modeling small-scale spatial variability in stand structure using remote sensing and field data. MS thesis, Dept. of Forest Sciences, Colorado State University, Fort Collins, CO.
- Schnase, J., T. J. Stohlgren, and J. A. Smith. 2002a. The national invasive species forecasting system: A strategic NASA/USGS partnership to manage biological invasions. *Earth Observation Magazine*, 11:46-49.

- Schnase, J. L., J. A. Smith, T. J. Stohlgren, S. Graves, and C. Trees. 2002b. Biological invasions: A challenge in ecological forecasting. In *Proceedings of the International Geoscience and Remote Sensing Symposium, 2002* (IGARSS '02), June 24-28, Toronto, Canada, IEEE 2-2 International Vol. 1, pp. 122-124.
- Stohlgren, T. J., Falkner, M. B., and L. D. Schell. 1995. A modified-Whittaker nested vegetation sampling method. *Vegetation*, 117:113-121.
- Stohlgren, T. J., G. W. Chong, M. A. Kalkhan, and L. D. Schell. 1997a. Rapid assessment of plant diversity patterns: A methodology for landscapes. *Environmental Monitoring and Assessment*, 48:25-43.
- Stohlgren, T. J., M. B. Coughenour, G. W. Chong, D. Binkley, M. A. Kalkhan, L. D. Schell, D. J. Buckley, and J. K. Berry. 1997b. Landscape analysis of plant diversity. *Landscape Ecology*, 12:155-170.
- Stohlgren, T. J., K. A. Bull, and Y. Otsuki. 1998. Comparison of rangeland vegetation sampling techniques in the Central Grasslands. *Journal of Range Management*, 51:164-172.
- Stohlgren, T. J., D. Binkley, G. W. Chong, M. A. Kalkhan, L. D. Schell, K. A. Bull, Y. Otsuki, G. Newman, M. Bashkin, and Y. Son. 1999. Exotic plant species invade hot spots of native plant diversity. *Ecological Monographs*, 69:25-46.
- Stohlgren, T. J., Y. Otsuki, C. A. Villa, M. Lee, and J. Belnap. 2001. Patterns of plant invasions: A case example in native species hotspots and rare habitats. *Biological Invasions*, 3:37-50.
- Stohlgren, T. J. 2005. *Measuring Plant Diversity: Lessons from the Field*. Oxford University Press (in production).

## 8.1 Selected Presentations

- Stohlgren, T. 2002. Ecological Forecasting of Invasive Species: High Performance Computing Needs. In the Earth Science and Technology Session of the American Astronautical Society 40th Goddard Memorial Symposium, March 20-21, NASA Goddard Space Flight Center, Greenbelt, MD. (INVITED).

## 8.2 Web Sites

For information about the National Institute for Invasive Species Science: <http://www.niiss.org>

For information about the USGS Fort Collins Science Center: <http://fort.usgs.gov>

For information about the USGS Biological Resources Division: <http://biology.usgs.gov/>

For information about the National Biological Information Infrastructure: <http://www.nbi.gov>

For information about the Invasive Species Forecasting System: <http://InvasiveSpecis.gsfc.nasa.gov>

T. Stohlgren, J. Schnase, J. Morisette, N. Most, E. Sheffner, C. Hutchinson, S. Drake, W. Van Leeuwen, and V. Kaupp.

For information about NASA Goddard Space Flight Center: <http://www.gsfc.nasa.gov>

For information about NASA's Applied Sciences Program: <http://www.earth.nasa.gov/eseapps/>

### 8.3 Software Engineering Documentation

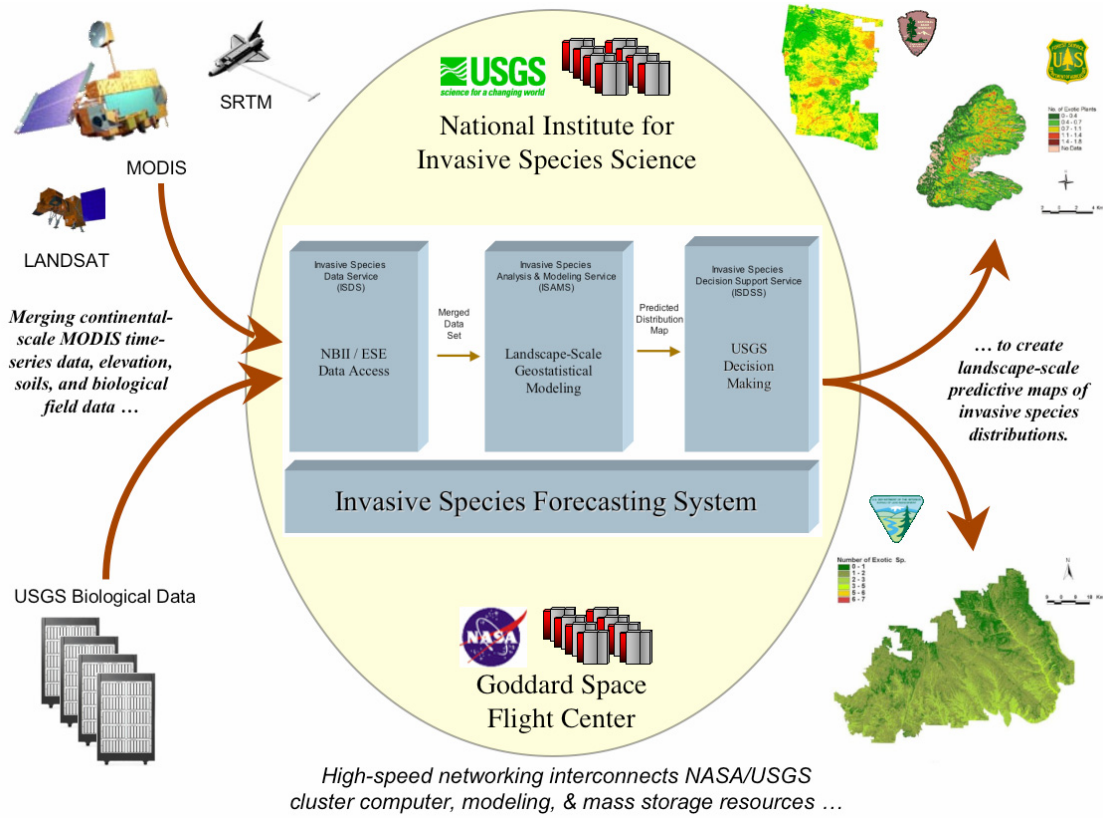
NASA/USGS. 2002a. Biotic Prediction: Baseline Software Design Document. BP-BSD-1.3, GSFC CT-1, December 4. 59 p. <http://bp.gsfc.nasa.gov/pub/BP-BSD-1.3.pdf> (accessed February 27, 2006).

NASA/USGS. 2002b. Biotic Prediction: Concept of Operations. BP-CONOP-1.9, Task Agreement GSFC CT-1, December 4. 30 p. <http://bp.gsfc.nasa.gov/pub/BP-CONOP-1.9.pdf> (accessed February 27, 2006).

Document Title	Version	Date
Software Requirements Trace Matrix ISFS_SRTM_2.0.doc	2.0	2004-11-30
Software Requirements Document ISFS_SRD_2.0.doc	2.0	2004-11-30
Test Plan (ISFS-TP)	2.0	2005-01-04
Configuration and Operations Guide ISFS_Config_Ops_Guide_v1.3.doc	1.3	2004-09-26

*This page intentionally blank*

## Appendix A. ISFS Elements and Process Flow



**Figure A-1. The elements of, and data flow within, the Invasive Species Forecasting System.**

*This page intentionally blank*

**REPORT DOCUMENTATION PAGE**

*Form Approved  
OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> 29-12-2005		<b>2. REPORT TYPE</b> Benchmarking Report v. 1.6		<b>3. DATES COVERED (From - To)</b> 2000 - December 2005	
<b>4. TITLE AND SUBTITLE</b> Invasive Species Forecasting System: A Decision Support Tool for the U.S. Geological Survey - FY 2005 Benchmarking Report				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Tom Stohlgren (1) John Schnase, Jeffrey Morisette (2) Neal Most (3) Ed Sheffner (4) Charles Hutchinson, Sam Drake, Willem Van Leeuwen (5) Verne Kaupp (6)				<b>5d. PROJECT NUMBER</b> SWR BA55-TMRK-00	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> (1) USGS, Fort Collins Science Center, Bldg C, 2150 Centre Avenue, Fort Collins, CO 80526-8118; (2) NASA, Mailstops 606.0/614.5, Goddard Space Flight Center, Greenbelt, MD 20771; (3) Innovim, Mailstop 606.3, Goddard Space Flight Center, Greenbelt, MD 20771; (4) NASA Headquarters, Mail Suite 5L79, 300 E Street SW, Washington, DC 20546-0000; (5) Arizona Remote Sensing Center, The University of Arizona, Suites 208B/205, 1955 E. 6th Street, Tucson, AZ 85721; (6) Interdisciplinary Center for Research in Earth Science Technologies, University of Missouri-Columbia, 349 Engineering Building West, Columbia, MO 65211				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Applied Sciences Directorate, National Aeronautics and Space Administration, Code MA00, Bldg. 1100, John C. Stennis Space Center, MS 39529				<b>10. SPONSORING/MONITOR'S ACRONYM(S)</b> NASA ASD	
				<b>11. SPONSORING/MONITORING REPORT NUMBER</b> SSTI-2220-0060 (New)	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> Publicly available STI per NASA Form 1676					
<b>13. SUPPLEMENTARY NOTES</b> NASA Applied Sciences Program Benchmarking Report; see <a href="http://aiwg.gsfc.nasa.gov/">http://aiwg.gsfc.nasa.gov/</a>					
<b>14. ABSTRACT</b> The National Institute of Invasive Species Science (NIISS), through collaboration with NASA's Goddard Space Flight Center (GSFC), recently began incorporating NASA observations and predictive modeling tools to fulfill its mission. These enhancements, labeled collectively as the Invasive Species Forecasting System (ISFS), are now in place in the NIISS in their initial state (V1.0). The ISFS is the primary decision support tool of the NIISS for the management and control of invasive species on Department of Interior and adjacent lands. The ISFS is the backbone for a unique information services line-of-business for the NIISS, and it provides the means for delivering advanced decision support capabilities to a wide range of management applications. This report describes the operational characteristics of the ISFS, a decision support tool of the United States Geological Survey (USGS). Recent enhancements to the performance of the ISFS, attained through the integration of observations, models, and systems engineering from the NASA are benchmarked; i.e., described quantitatively and evaluated in relation to the performance of the USGS system before incorporation of the NASA enhancements. This report benchmarks Version 1.0 of the ISFS.					
<b>15. SUBJECT TERMS</b> National Institute of Invasive Species Science (NIISS); Invasive Species Forecasting System (ISFS); United States Geological Survey (USGS); NASA Applied Sciences Program; benchmarking; decision support tool					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			Rodney McKellip
U	U	U	UU	32	<b>19b. TELEPHONE NUMBER (Include area code)</b> (228) 688-2984