BIOTIC PREDICTION

Building the Computational Technology Infrastructure for Public Health and Environmental Forecasting

Second Annual Report

(January 1 – December 31, 2003)

BP-AR2-1.3

Task Agreement: GSFC-CT-1

February 20, 2004

1 Introduction

Summary

This project is the development of a high-performance, computational technology infrastructure needed to analyze the past, present, and future geospatial distributions of living components of Earth environments. This involves moving a suite of key predictive, geostatistical biological models into a scalable, cost-effective cluster computing framework; collecting and integrating diverse Earth observational datasets for input into these models; and deploying this functionality as a Web-based service. The resulting infrastructure will be used in the ecological analysis and prediction of exotic species invasions. This new capability, known as the Invasive Species Forecasting System, will be deployed at the USGS Fort Collins Science Center and extended to other scientific communities through the USGS National Biological Information Infrastructure program.

Document Overview

This Annual Report describes research and project accomplishments during the reporting period of 2003. The Report is divided into six sections. Section 1 briefly describes the project and its referenced documentation. Section 2 describes second year progress, including scientific and technical accomplishments, progress toward milestones, current status, and plans for the coming year. Section 3 describes this year's publications and presentations related to the project. Section 4 provides contact information and information about the project's website. Section 5 provides a list of references. Finally, Appendix A summarizes the project's overall milestone schedule.

Referenced Documents

Document Title	Version	Date
Software Design (BP-SDD)	1.2	2003-12-02
Concept of Operations (BP-CONOP)	1.9	2002-12-04
Software Requirements (BP-SRD)	1.6	2003-11-30
Software Requirements Trace Matrix (BP-SRTM)	1.0	2003-11-30
Baseline Software Design (BP-BSD)	1.3	2002-11-25

2 Second Year Progress

Scientific and Technical Accomplishments

This year we have produced updated versions of the Software Requirements Document (BP-SRD), Software Requirements Trace Matrix (BP-SRTM), and the Software Design Document (BP-SDD). We also have produced a preliminary version of the Software Test Plan and Procedures (BP-TP). Our most important technical accomplishment this year is development of parallel kriging and the achievement of our first phase of model code performance improvement.

Model Code Performance Improvement

We are working with two “canonical” study sites: the Cerro Grande Fire Site in Los Alamos, NM (CGFS), and the Rocky Mountain National Park, CO (RMNP). 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 (BP-BSD-1.3), three factors influence the performance of ISFS model code: the size of the output surface area over which kriging occurs (area), the total number of sample points in the data set (pts), and 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 we first began work 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 is 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. We first wish to accomplish quantitative improvements in the underlying model that have been agreed upon by the user community as minimal advances needed to improve core capabilities. These goals are driven by the fact that we are building a 32-node cluster in the USGS facility. We refer to these as “Community Improvement Goals.” The ESTO/CT program, however, provides access to greater computational capabilities that can be used to apply this modeling approach to some important and challenging problems that heretofore have been unapproachable. We would therefore like to use CT’s clusters to attain more challenging performance improvement goals at the same time we are accommodating basic needs. We refer to these complementary challenges as “Advanced Improvement Goals.” Table 1 provides a summary of the baseline performance characteristics of the model code as well as the various performance goals anticipated over the course of the project.

Table 1. Current Performance Characteristics and Improvement Goals.

BASELINE SCENARIO		**Sec**	**Min**	**Hrs**	**Days**	
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	61.71	1.03	0.04	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	148.1	2.47	0.04	0.0	C
CGFS base 790 pts 79 nn 01x area (Version 2.0 -G)	25.0	148.1	2.47	0.04	0.0	D
CGFS base 790 pts 79 nn 10x area (Version 2.0 -G)	2.5	1481.0	24.68	0.41	0.0	E
RMNP base 1800 pts 18 nn 01x area (Version 2.0 -F)	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	18.5	0.31	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 F CI Goal -speed up- 75% efficiency, 32 node cluster = 25x speed up

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

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

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

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

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

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

Milestone F — 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 we determine the n nearest neighbor sample points and compute the (n x n) distance matrix containing the Euclidean distance between each sample points, and also compute the (n x 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. We obtain the kriging weights 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; we simply assign to each processor a section of the output kriged surface or image. We chose to decompose the domain along the rows only, i.e. each processor works with full rows of the output surface. This means we can leave unaltered the inner loop over columns. We could decompose into contiguous rows, effectively giving each processor a strip of the output image. Instead, we chose to assign consecutive rows to separate processors. Thus, for a kriging 512 x 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 soon we plan 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 we assigned sparsely sampled rows to one processor while assigning densely sampled rows to another processor.

We have implemented parallel kriging in FORTRAN using MPI, the Message Passing Interface. Our code employs a 'node 0' controller process and a collection of worker nodes. Prior to execution we copy to each node an input data file 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.

We overlap the computation with the communication to increase parallel efficiency. When the first row has been calculated, we issue an asynchronous send (MPI ISEND) of this row to node 0. Since this is a non-blocking send, the processor proceeds to calculate the second row. At the end of this row, we issue a wait (MPI WAIT) to insure that the first row has been received by node 0 before proceeding. For the smallest kriged surface we tested (512 x 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.

We have evaluated our parallel implementation on two clusters at the NASA Goddard Space Flight Center. Our code was designed for use on the Medusa cluster, on which we met our 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 1.5GB memory, which resides on one of our desks and is connected to the Medusa cluster via fiber Gigabit Ethernet. We typically only log into Node 0, and to the user it appears that all calculations are done on Node 0. We have also used Thunderhead to evaluate our 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.

Parallel Kriging Results on Ideally Sized Test Cases

We begin by presenting performance results for four test problems to evaluate the efficiency of the parallel implementation. We 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 512^2, 768^2, 1024^2, to 2048^2. 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 2048^2 problem ran 16x longer than the 512^2 case). The processing times decreased nearly linearly as the number of processors was increased, as shown in Figure 5. We define the scaling efficiency for N processors as the ratio of the 1-processor to N-processor wall-clock times divided by N. The efficiencies we 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 2048^2 problem.

Table 2. Test Case Timing Results (Elapsed Wall Clock Seconds)

Number of Medusa Processors	Size of Kriged Image			
	2048^2	1024^2	768^2	512^2
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 5. Scaling Curves for Test Cases on Medusa

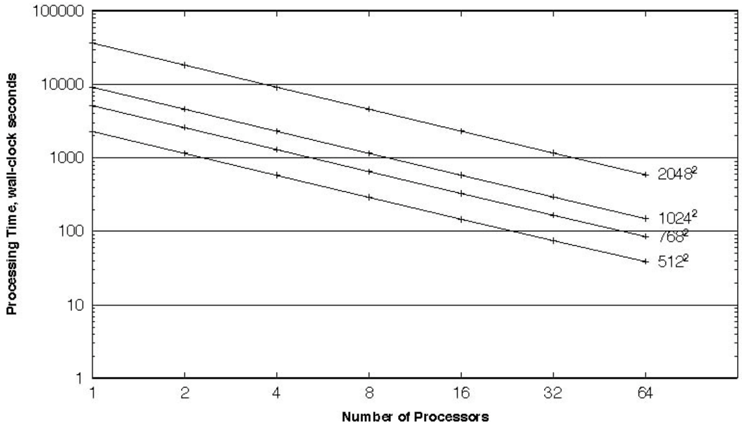


Table 3. Scaling Efficiencies for Test Cases

Number of Medusa Processors	Size of Kriged Image			
	2048^2	1024^2	768^2	512^2
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%

Evaluation of Community and Advanced Improvement Goals

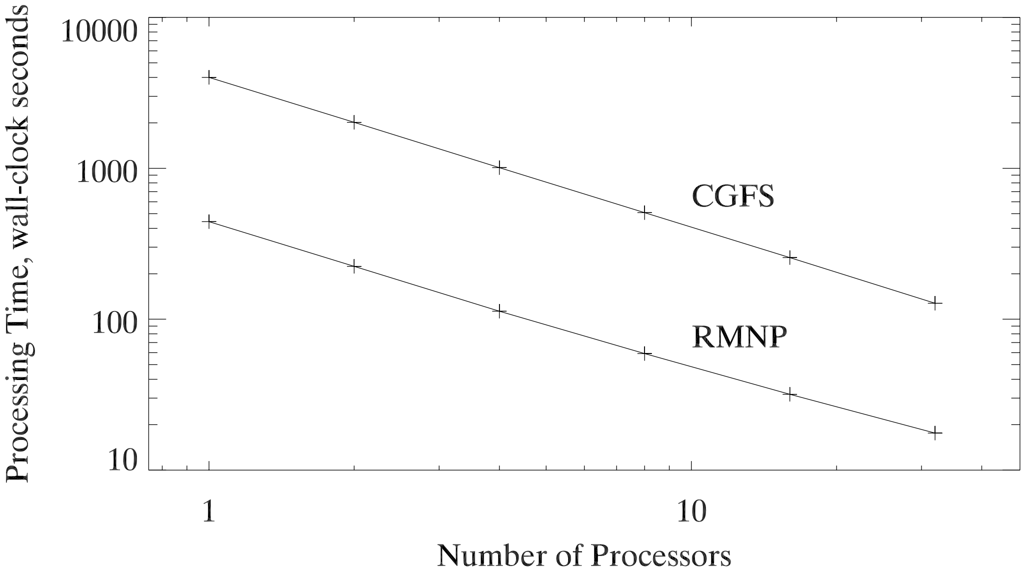
We now evaluate the performance of the parallel kriging on the baseline scenarios defined in Table 1. These differ from the ideally sized test cases evaluated above because a different number of rows are assigned to each processor. As such they represent real operational scenarios that estimate an arbitrary 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. 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%

We show in Table 4 the performance results for both the CGFS and RMNP baseline scenarios, which we plot in Figure 5. We improved the run time for the CGFS test to 2 minutes and 7 seconds, which compares favorably to our Milestone F Community Improvement goal of 2 minutes and 28 seconds. We improved the RMNP test 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. We can clearly see, however, that the scaling efficiency drops as the run time is reduced. This is expected and due to the parallel overhead and the load imbalance.

Figure 5. Scaling Curves for Test Cases on Medusa



Progress toward Milestones

These results indicate that we have achieved our Milestone F 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 one might expect as more nodes are allocated to the canonical data sets. The project thus completed Milestone F (First Code Improvement) according to plan.

Current Status

Milestone F deliverables are currently under review by the CT Program office. Preliminary feedback has been used to make significant modifications to the software engineering documentation, and approval of these modifications is forthcoming.

GSFC's transition to IFMP and the associated procurement embargo has significantly delayed progress toward our Opt Milestone (Installation of Linux Cluster). Parts were ordered in September 2003 for two 16-node/32-processor clusters (see Appendix B for specifications). Both clusters will be built at Goddard; one will be deployed here, the other at USGS's Fort Collins Science Center. We now anticipate that March 2004 is the earliest we will be able to complete this Milestone, and we are in the process of re-negotiating the delivery date of the Opt Milestone.

This year, we opened the project's website to the public. The public response has been very favorable. (See: http://InvasiveSpecies.gsfc.nasa.gov).

The project has completed staffing for next year's work. The core NASA/USGS team now includes Neal Most (Project Manager, IntelView), David Kendig (Sr. Programmer/Analyst, SSAI), Nathan Pollock (Jr. Programmer/Analyst and Webmaster, SSAI), Jim Closs (Sr. Technical Writer, SSAI), and Donal Hogan (Sr. Technical Writer, IntelView).

Plans for Coming Year

Work this year will focus on four major areas:

- Customer-Driven Collaborative Design and Outreach – We will continue to engage collaborators in our partner agency in design and review meetings. The input received during these sessions will continue to guide the development of the next version of the system and associated software engineering documentation. We have begun to organize these meetings as Science Team Meetings, the next one of which is planned for Spring 2004 in Fort Collins, Colorado.

- Milestone G – Second Code Improvement – Through a series of intermediate software releases, the project will improve the parallel performance of the baseline software system and create an "adaptive" implementation of the core kriging algorithms used in our models. Jeff Pedelty, Jeff Morisette, David Kendig, and John Dorband will lead this effort with assistance from the CT program office

- Construction and Deployment of the "FireAnt" and "Rocky" Clusters – Per modified schedule and plans, we hope to build and deploy our project clusters by the spring 2004 science team meeting.

- New Data/Algorithm/Science Development – Recently funded activities relating to this invasive species project will substantially broaden the impact of our modeling activities. PI Schnase has been funded under NASA's REASoN CAN to build an invasive species data service which will become the ingest subsystem of our overall architecture. In addition, CoI Stohlgren has been funded under NASA's IDS NRA to apply our new modeling capabilities to a large-scale problem: mapping the distribution of tamarisk and Canada thistle across the Western US.

3 Publications and Presentations

This section lists this year's major publications and presentations relating to the invasive species project.

Scientific Publications

1. Stohlgren, T.J., G. W. Chong, L.D. Schell, K.A. Rimar, Y. Otsuki, M. Lee, M.A. Kalkhan, and C.A. Villa. 2002. Assessing vulnerability to invasion by non-native plant species at multiple scales. *Environmental Management* 29:566-577.
2. Schnase, J., T.J. Stohlgren, and J. A. Smith. 2002. The national invasive species forecasting system: A strategic NASA/USGS Partnership to manage Biological Invasions. *Earth Observation Magazine* 11:46-49.
3. Stohlgren, T.J. 2002. Beyond Theory of Plant Invasions: lessons from the field. *Comments on Theoretical Biology* 7: 355-379.
4. Bashkin, M., T.J. Stohlgren, Y. Otsuki, M. Lee, P. Evangelista, and J. Belnap. 2002. Soil characteristics and plant exotic species invasions in the Grand Staircase-Escalante National Monument, Utah, USA. *Applied Soil Ecology* 22: 67-77.
5. Stohlgren, T. J., T. T. Veblen, K. Kendall, W. L. Baker, C. Allen, A. Logan, and M. Ryan. 2002. Pages 203-218. Montane and subalpine ecosystems. In: Rocky Mountain Futures: an Ecological Perspective. J. Baron (eds). Island Press, Washington DC.
6. Stohlgren, T.J., D. Barnett, and J. Kartesz. 2003. The rich get richer: Patterns of plant invasions in the United States. *Frontiers in Ecology and the Environment* 1:11-14.
7. Barnett, D., and T.J. Stohlgren. 2003. A nested intensity sampling design for plant diversity. *Biodiversity and Conservation* 2(2): 255-278.
8. Kalkhan, M.A., E.J. Martinson, P.N. Omi, G.W. Chong, M.A. Hunter, and T.J. Stohlgren. 2003. Fuels, fire severity, and invasive plants within the Cerra Grande Fire, Los Alamos, NM. Proceedings of the Tall Timbers Fire Ecology Conference, Tallahasse, FL. (In Press).
9. Omi, P.N., E.J. Martinson, M. Kalkhan, T.J. Stohlgren, G.W. Chong, and M.A. Hunter. 2003. Integration of spatial information and spatial statistics: a case study of invasive plants and wildfire on the Cerra Grande Fire, Los Alamos, New Mexico, USA. Proceedings of the Tall Timbers Fire Ecology Conference, Tallahasse, FL. (In Press).
10. Guenther, D.A., T.J. Stohlgren and P. Evangelista. 2003. Relict sites compared to grazed landscapes in the Grand Staircase-Escalante National Monument, Utah. Proceedings of the Fifth Biennial Conference of Research on the Colorado Plateau Conference, Flagstaff, AZ. (In Press).
11. Fornwalt, P.J., M. Kaufmann, L.S. Huckaby, J.M. Stoker, and T.J. Stohlgren. 2003. Non-native plant invasions in managed and protected ponderosa pine/Douglas-fir forests of the Colorado Front Range. *Forest Ecology and Management* 177: 515-527.
12. Pedelty, J.A., J.T. Morisette, J.L. Schnase, J.A. Smith, T.J. Stohlgren, and M.A. Kalkhan. 2003. High performance geostatistical modeling of biospheric resources in the Cerro Grande wildfire site, Los Alamos, New Mexico and Rocky Mountain National Park, Colorado. xxx
13. Evangelista, P., D. Guenther, T J. Stohlgren, and S. Stewart. 2003. Fire effects on cryptobiotic soil crusts in the Grand Staircase-Escalante National Monument, Utah. Proceedings of the Fifth Biennial Conference of Research on the Colorado Plateau Conference, Flagstaff, AZ. (In Press).
14. Stohlgren, T.J., C. Crosier, G. Chong, D. Guenther, and P. Evangelista. 2004. Habitat matching by non-native plant species. Biodiversity and Distributions (to be submitted).
15. Chong, G.W., Y.Otsuki, T.J. Stohlgren, D. Guenther, and C. Villa. 2003. Evaluating plant invasions from both habitat and species perspectives. *To be submitted.*
16. Stohlgren, T.J., T. Chase, R.A. Pielke Sr,. and J. Graham. 2004. Mapping spatial anomalies in Ecology. *Frontiers in Ecology and the Environment* (In Review).
17. Stohlgren, T.J., D. Guenther, P. Evangelista, and N. Alley. 2004. Patterns of plant rarity, endemism, and uniqueness in an arid landscape. *Ecological Applications* (In Review).
18. Simonson, Sara, David Barnett, Thomas Stohlgren, Michael Ielmini, and staff of the USFWS National Wildlife Refuge System. 2004. Invasive Species Survey: Invasion of the National Wildlife Refuge System. (Draft Report).
19. Waters, M.A. 2003. Species richness, vegetation cover, and disturbance relationships in an arid ecosystem. Masters Thesis, Colorado State University, Fort Collins, CO.

20. Alley, N. 2003. Iterative model development for natural resource managers: a case example from the Grand Staircase-Escalante National Monument, Utah. Masters Thesis, Colorado State University, Fort Collins, CO.
21. Crosier, C. 2004. Data synergies and invasive plant species distributions in Colorado. Draft title for Ph.D. Dissertation, Colorado State University, Fort Collins, CO. (To be completed January 2004).
22. Chong, G.W., T.J. Stohlgren, C. Crosier, S. Simonson, G. Newman, and E. Petterson. 2003. Ecological Effects of the Hayman Fire: Part 7: Key Invasive Nonnative Plants. Pages 244-249. In Graham, R.T. (Technical Editor). Hayman Fire Case Study. Gen. Tech. Rep. RMRS-GTR-114. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 396 p.
23. Kotliar, N.B., S. Simonson, G.W. Chong, and D. Theobold. 2003. Ecological Effects of the Hayman Fire: Part 8: Effects on Species of Concern. Pages 250-262. In Graham, R.T. (Technical Editor). Hayman Fire Case Study. Gen. Tech. Rep. RMRS-GTR-114. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 396 p.

Web Sites

National Institute of Invasive Species Science:
http://nrel.colostate.edu/Stohlgren/projects/niiss/niiss.html

Invasive Species Information Node:
http://invasivespecies.nbii.gov

NASA/USGS Invasive Species Forecasting System:
http://InvasiveSpecies.gsfc.nasa.gov

Invited Science Talks

“The Rich Get Richer: Don’t Let Small-scale Experiments Fool You” Conference on Invasive Plants in Natural and Managed Systems: Linking Science and Management and 7th Annual International Conference on the Ecology and Management of Alien Plant Invasions, November 3-8, 2003, Fort Lauderdale, Florida (Stohlgren).

“A NASA-USGS Invasive Species Forecasting System” Conference on Invasive Plants in Natural and Managed Systems: Linking Science and Management and 7th Annual International Conference on the Ecology and Management of Alien Plant Invasions, November 3-8, 2003, Fort Lauderdale, Florida (J. Schnase, NASA).

“The National Institute of Invasive Species Science” The Nature Conservancy Annual Meeting in Fort Lauderdale, FL, November 2, 2003 (C. Crosier)

“Mapping and Modeling Non-native Species Invasions: Ecological Forecasting in the 21st Century,” Women in Science and Engineering Conference, September 2003, Alabama (Stohlgren).

“Information Management for the National Institute of Invasive Species Science” Long-Term Ecological Research Program Annual Meeting in Seattle, WA, September 2003, (G. Newman)

“Plant Invasions: a National and International Perspective,” New England Weed Society, October 2003, Framingham, MA (Stohlgren).

“Patterns of Non-native Species Invasions in the United States” June 24-26, 3003, University of Wyoming 2003 Ecological/Economic Conference of Bioinvasions (Stohlgren)

“Predicting Plant Invasions in the United States and the World” Canadian Botanical Society Keynote Speaker, June, 2003, Nova Scotia, Canada (Stohlgren, co-author)

“The National Institute of Invasive Species Science,” one of three talks selected for the Department of Interior’s Senior Executive Service Retreat, Virginia, May 7-8, 2003.

4 Contact Information

Team Members

Robert Baker	SSAI	10210 Greenbelt Road Suite 500 Lanham, MD 20706	301-867-2073 (w) 301-867-2191 (fax)	Robert Baker@sesda.com
Lori Bruce	MSU		662-325-8430	bruce@ece.msstate.edu
Jim Closs	SSAI	10210 Greenbelt Road Suite 500 Lanham, MD 20706	301-867-6252	james_closs@ssaihq.com
Catherine Crosier	CSU/NREL		970-491-5630	crosier@nrel.colostate.edu
John E. Dorband, PhD	NASA	NASA Goddard Space Flight Center Building 28, Room S206 Code 935 Greenbelt, MD 20771	301-286-9419 (w) 301-286-1634 (fax)	John.E.Dorband.1@gsfc.nasa.gov
Kristin Eickhorst	UMAINE		207-581-5711	snoox@umit.maine.edu
Michael T. Frame	USGS	Biological Resources Division US Geological Survey Reston, VA 20192	703-648-4164 (w) 703-648-4224 (fax)	mike frame@usgs.gov
Jim Graham	CSU/NREL		970-491-5835	jim@nrel.colostate.edu
David Herring	NASA	NASA Goddard Space Flight Center Building 33, Room A325 Code 913.0 Greenbelt, MD 20771	301-614-6219 (w) 301-614-6307 (fax)	David.D.Herring@gsfc.nasa.gov
Mohammed A. Kalkhan,	CSU	Colorado State University Natural Resource Ecology Laboratory NREL-A244 Fort Collins, CO 80523-1499	970-491-5262 (w) 970-491-1965 (fax)	mohammed@nrel.colostate.edu
David Kendig	SSAI	Goddard Space Flight Center Global Change Master Directory Code 902 Greenbelt, MD 20771	301-867-2084 (w) 301-614-6695 (fax)	David.J.Kendig@gsfc.nasa.gov
Jacqueline J. Le Moigne,	NASA	Goddard Space Flight Center Building 28, Room W186 Code 935 Greenbelt, MD 20771	301-286-8723 (w) 301-286-1777 (fax)	lemoigne@backserv.gsfc.nasa.gov
Jeffrey T. Morisette, PhD	NASA	Goddard Space Flight Center Building 33, Room G325 Code 923 Greenbelt, MD 20771	301-614-6676 (w) 301-614-6695 (fax)	Jeffrey.T.Morisette.1@gsfc.nasa.gov
Neal Most	Intelview	NASA Goddard Space Flight Center Building 28, Room W254, Code 930 Greenbelt, MD 20771	301-286-6747 (w)	nmost@intelview.com
David Obler	SSAI	10210 Greenbelt Road Suite 500 Lanham, MD 20706	301-867-2151 (w) 301-867-2191 (fax)	David.Obler@gsfc.nasa.gov
Jeffrey A. Pedelty, PhD	NASA	Goddard Space Flight Center Building 33, Room G325 Code 923 Greenbelt, MD 20771	301-614-6609 (w) 301-614-6695 (fax)	Jeffrey.A.Pedelty.1@gsfc.nasa.gov

Robin M. Reich, PhD	CSU	Department of Forrest Sciences Colorado State University Ft. Collins, CO 80523	970-491-6980 (w) 970-491-6754 (fax)	robin@cnr.colostate.edu
John L. Schnase, PhD	NASA	Goddard Space Flight Center Building 28, Room W230D Code 930 Greenbelt, MD 20771	301-286-4351 (w) 301-286-1777 (fax)	schnase@gsfc.nasa.gov
Robert Simmon	SSAI	10210 Greenbelt Road Suite 500 Lanham, MD 20706	301-614-6201	simmon@climate.gsfc.nasa.gov
James A. Smith, PhD	NASA	Goddard Space Flight Center Building 33, Room G125D Code 920 Greenbelt, MD 20771	301-614-6020 (w) 301-614-6015 (fax)	James.A.Smith.1@gsfc.nasa.gov
Thomas J. Stohlgren, PHI	USGS	Midcontinent Ecological Science Cen US Geological Survey Ft. Collins, CO 80523	970-491-1980 (w) 970-491-1965 (fax)	toms@nrel.colostate.edu
Curt A. Tilmes	NASA	Goddard Space Flight Center Building 32, Room S36C Code 922 Greenbelt, MD 20771	301-614-5534 (w) 301-614-5269 (fax)	Curt.Tilmes@gsfc.nasa.gov
Asad Ullah	SSAI	10210 Greenbelt Road Suite 500 Lanham, MD 20706	301-867-2019	asad_ullah@ssaihq.com

5 Appendix A

Document/System Access

ESTO/CT milestone schedule deliverables for this project are available at http://ltpwww.gsfc.nasa.gov/BP/deliverables.html. The baseline system, along with complete documentation, are available in the project's sourcecode store, located at: http://tamarisk.sesda.com/cgi-bin/bp/viewcvs.cgi/BP/?sortby=date. Users may log on to the system to run the baseline program (please contact Neal Most at 6-6747 for userid and password). In addition, a tarfile is available from both the website and the ISFS home directory that can be used to build the baseline environment on a different machine.

6 Appendix B – Cluster spec – from website

ISFS Production Cluster Computing Environment

Twin clusters: FIREANT & ROCKY, comprised of:

- 2 development nodes
- 16 compute nodes
- 1 storage node

System Specs

Processor:	16-node, 32 1.4Ghz AMD Opteron 240DP
Speeds:	2.8 Gflop double, 5.6 Gflop single peak
Memory:	16 GB SDRAM
Diskspace:	120 GB local/execute node ,
Network:	1.44 TBytes raided storage node

Hardware

Motherboard:	Tyan Thunder K8S (S2880GNR)
Processor:	AMD Opteron 240 DP; 1.4Ghz; 1MB L2 cache
Memory:	Kingston 1GB 333MHz DDR PC2700 Reg ECC DIMM
Harddrive:	Seagate 120GB Speed 7200rpm 8.5 ms seek time
Monitor	SAMSUNG Syncmaster 240T, 24IN LCD, 1920X1200 60HZ
Graphics Card	MSI FX5600-VTDR128 (MS-8912) 128MB, GeForceFX nVIDIA
Raid Card Controller	3WARE Escalade 7506-12, 12 channel Ultra ATA, DiskSwitch
DVD+RW Drive	TDK IndiDVD 440N black 4x2x12 Int IDE DVD+R/- R/+RW/-RW
Floppy drive	Samsung 1.44mb 3.5" floppy Black
Keyboard+Mouse	Logitech Wheel Mouse Optical PS/2 and USB. Black Dell keyboard PS/2 and USB 104keys Black
PC Case Thermaltake	XASER III V2420A (SILVER) 12-Bay All Aluminum ATX Super Tower Chassis w/ Large Side Window & Locks, 7 fans