Biotic Prediction

Building the Computational Technology Infrastructure
for Public Health and Environmental Forecasting

Test Plan

BP-TP-1.1

Task Agreement: GSFC-CT-1

August 22, 2003

Contents

List of Tables

List of Figures

1 Introduction

1.1 Invasive Species Forecasting System (ISFS)

This project is developing the 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 will be deployed at the USGS Mid-continent Ecological Science Center and extended to other scientific communities through the USGS National Biological Information Infrastructure program.

1.2 Test Plan Overview

This Test Plan has been prepared in accordance with NASA/GSFC's "Recommended Approach to Software Development Revision 3". The sections included are as follows:

Section 1. Introduction
1.1 Invasive Species Forecasting System (ISFS) — Abstract describing the project.
1.2 Test Plan Overview — List describing sections of this Test Plan.

Section 2. Test Procedures
2.1 Test Objectives — Purpose, type and level of testing.
2.2 Testing Guidelines — Testing procedures.
2.3 Evaluation Criteria — Guidelines to determining the success of the tests.
2.4 Error Correction — Errors and retesting procedures.

Section 3. Test Summary
3.1 Environmental Prerequisites — Datasets and computer results required.
3.2 System Summary — Summary of expected tests to be validated.
3.3 Requirements Trace — Matrix mapping requirements and functional specifications to one or more test items.

Section 4. Test Descriptions
4.1 Test Name — Use cases associated with the tests.
4.2 Purpose of this Test — Summary of the capabilities to be tested.
4.3 Method — Step-by-step procedures for conducting the test.
4.4 Test Input — Test input to be used.
4.5 Expected Results — Description of the expected outcome.
4.6 Actual Results — Description of the observed results in comparison to the expected results.

Section 5. Regression Tests

2 Test Procedures

2.1 Test Objectives

The objective of these tests is to evaluate the integrity of the ISFS and validity of the results. By following the steps in Section 4.3 for Use Case 1.0, the test will result in a metadata file (Figure 3. Metadata) and an image (Figure 4. Output Image). The test plan will note the expected results and how they are achieved.

2.2 Testing Guidelines

All tests should be conducted independently of each other to verify the results of each test. Testing includes selecting a study site and model array and applying nearest neighbor criteria and a kriging routine to be calculated with each test.

2.3 Evaluation Criteria

The expected results of testing Use Case 1.0 should be compatible with the metadata and image files demonstrated in Section 4.5.

2.4 Error Correction

If login to the system is not successful, the user is not authenticated and the application displays an error message Your Account name or password is invalid.

3 Test Summary

3.1 Environmental Prerequisites

To test the system, the following conditions must exist:

1. Logon/password verification is established for the tester.
2. Project servers must be accessible.
3. Connection to the Goddard cluster must be available.
4. Tester must have access to the Internet using a web browser.

3.2 System Summary

The current system handles communication through the firewall between the development server "webserv" and the project's data server "frio" which is the interface with the Goddard cluster.

The subsystem can be broken into the three areas of Client Browser, Web Applications and the Business Tier.

3.3 Requirements Trace

The following matrix refers to the system requirements as they are satisfied by the respective software builds. Milestone F requirements are noted in bold the matrix:

Table 1. Verifi cation Matrix

ID	**Requirement**	**Version**
3.	**Functional Requirements**	
3.1	**User Interface**	
3.1.1	**Profile Database**	
3.1.1.1	The system shall maintain a profile database of users.	Milestone G
3.1.2	**Roles**	
3.1.2.1	The system shall support various roles that control access to various capabilities of the system.	Milestone G
3.1.3	**Graphical User Interface**	
3.1.3.1	The system shall include a Graphical User Interface (GUI) to support user interaction with the system.	Milestone G
3.1.3.2	The GUI shall dynamically construct personalized web pages based on the Profile Database and the User s Role.	Milestone G
3.1.3.3	The GUI shall display predictive map and uncertainty map output.	Milestone G
3.1.3.4	The GUI shall invite all system users to register in order to use the system based on their roles.	Milestone G
3.1.3.5	The system shall allow the model builder to create new models.	Milestone G
3.1.3.6	The system shall allow the model user to select from an assortment of modeling techniques and to modify model parameters.	Milestone G

ID	Requirement	Version
3.1.4		
3.1.4.1	The user shall have the option of saving run results with annotations in personal repository.	Design Goal, future builds
3.2	**Ingest**	
3.2.1	**Validation**	
3.2.1.1	The system shall verify integrity, but is not required to validate the quality of the data before ingest.	**Milestone F**
3.2.1.2	The system will accept data from an authoritative source.	Design Goal, future builds
3.2.1.3	The system shall log all data sources.	Milestone G
3.2.2	**Field Data**	
3.2.2.1	The system shall provide standard templates for ingesting field data in a tabular form.	Milestone G
3.2.2.2	The templates shall include all required fields to be captured.	Milestone G
3.2.2.3	The templates shall be in an accessible format.	Milestone G
3.2.3	**Remote Sensing Data**	
3.2.3.1	The system shall support ingest from external satellite data archives.	
3.2.3.2	The system shall support ingest of user-supplied satellite data or airborne imagery from digital files.	Milestone G
3.2.3.3	The system shall support ingest of user-supplied data.	**Milestone F**
3.2.3.4	The system shall support ingest of user-supplied data files for ancillary layers.	Milestone G
3.2.4	**Data Acquisition**	
3.2.4.1	The system shall support secured ftp-push for ingest of user supplied data.	Milestone G
3.2.4.2	The system shall support automated secured ftp pull from external archives.	Design Goal, future builds
3.2.4.3	The system shall provide accounting and logging by requiring users name and password.	Milestone G
3.2.4.4	The system shall maintain a list of external archives and required data sets.	Design Goal, future builds
3.2.4.5	The interface between the system and each external archive will be thoroughly documented.	Design Goal, future builds
3.2.5	**Monitoring and Reporting**	
3.2.5.1	The ingest subsystem shall monitor the number and volume of data brought into the system.	Milestone G
3.2.5.2	The ingest subsystem shall produce data reports broken down by location, user, and external archive.	Milestone K
3.2.5.3	The system shall generate and display associated metadata describing output files, runtime parameters, and performance statistics.	Milestone K
3.3	**Pre-processing**	
3.3.1	**Merge Data**	

ID	Requirement	Version
3.3.1.1	The system shall merge ingested datasets.	Design Goal, future builds
3.3.1.2	The system shall perform resampling if the input data are not at the same resolution.	Design Goal, future builds
3.3.1.3	The data shall be converted to a common analysis format.	Milestone G
3.4	**Modeling**	
3.4.1	The system shall allow the ability to specify response and explanatory variables from the available databases.	Milestone G
3.4.2	The system shall provide graphical techniques to explore the relationships between these variables.	Milestone G
3.4.3	The system shall have the ability to fit models through least squares (or other) optimization routines.	**Milestone F**
3.4.4	The system shall provide screening techniques to quantitatively assess which explanatory variables are related to the response variable.	Milestone G
3.4.5	The system shall calculate geospatial statistics.	**Milestone F**
3.4.6	The system shall incorporate spatial structure into the modeling.	Design Goal, future builds
3.4.7	The system shall be able to output model results and relevant model diagnostics.	**Milestone F**
3.4.8	The system shall allow the model builder to create new models.	Milestone G
3.4.9	The system shall allow the model user to select from an assortment of modeling techniques and to modify model parameters.	Milestone G
3.5	**Post-processing**	
3.5.1	**Reprojecting data**	
3.5.1.1	The system shall display an image of the output data.	**Milestone F**
3.5.1.2	The system shall produce a data file suitable for reprojection by external COTS utilities.	**Milestone F**
3.5.2	**Data Overlay**	
3.5.2.1	The system shall overlay output data with other layers as requested.	Milestone K
3.5.3	**Metadata**	
3.5.3.1	Output data shall be packaged with appropriate metadata.	**Milestone F**
3.5.3.2	Each output data set shall be assigned a unique identifier.	**Milestone F**
3.6	**Archive**	
3.6.1	**Database**	
3.6.1.1	The Archive shall have a database that will store pointers to the archived files.	Milestone G
3.6.1.2	The database shall be able to refer to internal (stored in the local File Store) and external files.	Milestone G
3.6.1.3	All files (internal and external) shall be indexed with a unique file ID.	Milestone G
3.6.2	**Internal File Store**	
3.6.2.1	Files shall be stored in a logically arranged directory structure.	Milestone G
3.6.3	**External Files**	

ID	Requirement	Version
3.6.3.1	For external files, the archive system shall store a pointer that can be used to retrieve and stage the files for subsequent processing.	Milestone G
4.	**Performance Requirements**	
4.1	**CT Project Scaling Milestones**	
4.1.1	Deliver canonical products 25X faster than the baseline implementation & Design Goal, future builds	**Milestone F**
4.1.2	Accommodate 10X more sample data at 25X the time required in the baseline implementation and 10X larger area at 2.5X the time required in the baseline implementation.	Milestone G
4.1.3	Deliver canonical products 200X faster than the baseline implementation on a 256-node cluster accomodating 100X more sample data 1000X faster than the baseline and 100X larger area 10X faster than the baseline on a 1024-node cluster.	Milestone G

4 Test Descriptions

4.1 Test Name

The primary test of the system is to follow the steps listed in Use Case 1.0 in Section 4.4.

4.2 Purpose of this Test

Testing Use Case 1.0 will verify the integrity of the system and validate the results of the metadata and output image files.

4.3 Method

Run through Use Case 1.0 using different criteria for the nearest neighbor and kriging options each time the test is run. Use Case 1.0 is demonstrated in the following table:

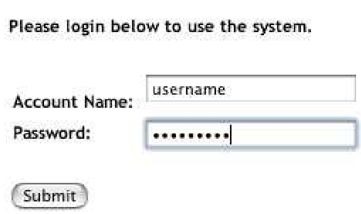
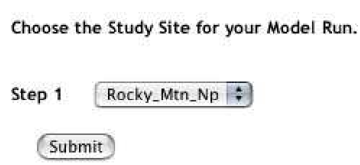
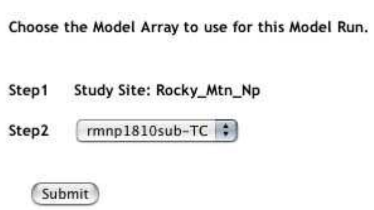
<table>
<tr><th>Graphical User Interface</th><th>Back-end processes</th></tr>
<tr><td>
</td><td>1. User enters account name and password to enter the system.
2. User is authenticated in lookup table and enters the new system.
OR
User is not authenticated and the jsp displays an error message "Your Account name or password is invalid."</td></tr>
<tr><td>
</td><td>1. A list of Study Sites from the DB (or from file system) that are ready to be submitted as a model run are displayed.
2. User selects study site and submits selection.</td></tr>
<tr><td>
</td><td>1. A list of Model Arrays (MA) from DB (or from file system) is displayed.
2. User selects model array and submits selection.</td></tr>
<tr><td colspan="2">Table 2 Continued on next page</td></tr>
</table>

Table 2. Graphical User Interface

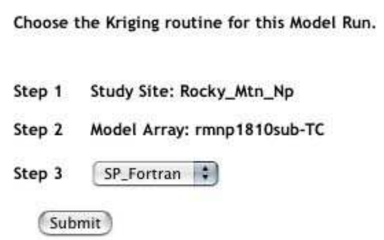
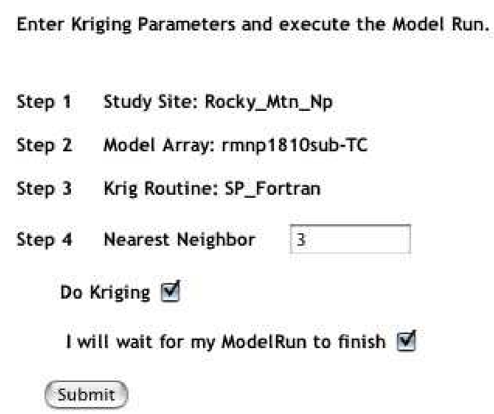
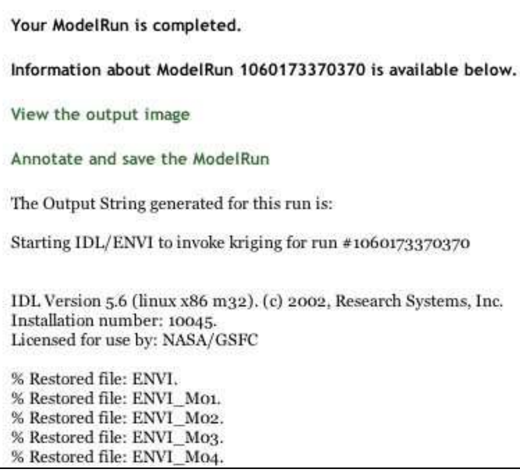
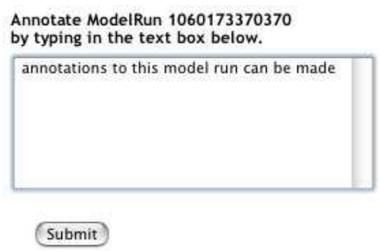
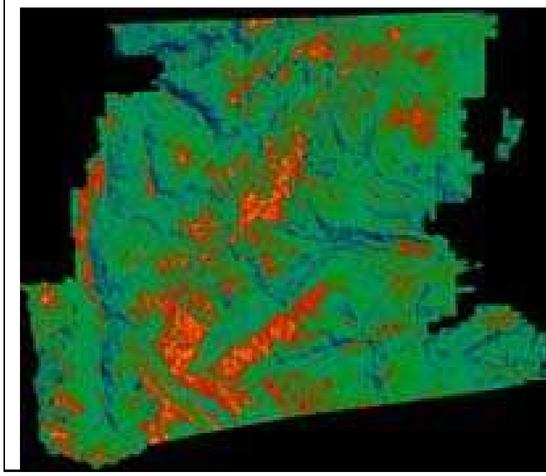
<table>
<tr><td colspan="2">Table 2 Continued from previous page</td></tr>
<tr><th>Graphical User Interface</th><th>Back-end processes</th></tr>
<tr><td>
</td><td>1. A list of Kriging routines from DB (or from file system)is displayed.
2. User selects Kriging routine and submits selection.</td></tr>
<tr><td>
</td><td>1. Use enters number of nearest neighbors to be calculated.
2. User selects whether to incorporate Kriging.
3. User selects whether to wait for the ModelRun to be completed.
4. User selects Kriging routine and submits selection and submits selections.</td></tr>
<tr><td>
</td><td>1. The metadata for the run is displayed (possibly requiring a call to the DB to get it). User can select to “View the Output Image” or “Annotate and save the ModelRun.”</td></tr>
<tr><td>
</td><td>1. When user selects to Annotate and save the ModelRun, a text box is displayed for annotations.
2. User annotates and submits text.</td></tr>
<tr><td>
</td><td>The user comments are saved to a field within the metadata record.</td></tr>
<tr><td colspan="2">Table 2 Continued on next page</td></tr>
</table>

Table 2 Continued from previous page	
Graphical User Interface	**Back-end processes**
	When user selects to View the output image, the image file from the ModelRun is drawn on the screen.

4.4 Test Input

The steps in Section 4.3, Use Case 1.0 detail the steps for the input and output (I/O) of data for a model run. The following I/O steps are computed on the project s server "frio" from the reading in of data to the output of results:

1. Read in tabular data
2. Compute the distance matrix
3. Map tabular data to remote sensing data
4. Perform stepwise regression
5. Fit squares
6. Engage kriging routine
7. Output results variables
8. Rasterize and output graphical presentation of data as .jpeg (for higher quality images the project is interested in saving these images as GeoTiff).

4.5 Expected Results

SSelections from Use Case 1.0 should output a metadata file and image. The results of running Use Case 1.0 include a metadata file and image similar to the following:

Your ModelRun is completed.

Information about ModelRun 1060715013738 is available below.

View the output image

Annotate and save the ModelRun

The Output String generated for this run is:

```
Starting IDL/ENVI to invoke kriging for run \#1060715013738

IDL Version 5.6 (linux x86 m32). (c) 2002, Research Systems, Inc.

Installation number: 10045.
Licensed for use by: NASA/GSFC

% Restored file: ENVI.
% Restored file: ENVI_M01.
% Restored file: ENVI_M02.
% Restored file: ENVI_M03.
% Restored file: ENVI_M04.
% Restored file: ENVI_M05.
% Restored file: ENVI_M06.
% Restored file: ENVI_M07.
% Restored file: ENVI_M08.
% Restored file: ENVI_D01.
% Restored file: ENVI_D02.
% Restored file: ENVI_D03.
% Restored file: ENVI_CW.
% Restored file: ENVI_IDL.
% Restored file: ENVI_IOU.
% Compiled module: GETINPUT.
% Compiled module: ISFS.
% Compiled module: DOVARIO.
% Compiled module: KRIG_GAUSS.
% Compiled module: KRIG_EXPON.
% Compiled module: KRIG_SPHERE.
% Compiled module: JPKRIG.
% Compiled module: BINARY.
% Compiled module: MAKE_KB2D_INPUT.
% Compiled module: MAKEKRIG.
% Compiled module: STAT_REGR_OUT.
% Compiled module: MY_STEPWISE.
% Compiled module: EXPF.
% Compiled module: GAUF.
% Compiled module: SPHERF.
% Compiled module: XML2STRUCT::INIT.
% Compiled module: XML2STRUCT::CHARACTERS.
% Compiled module: XML2STRUCT::STARTELEMENT.
% Compiled module: XML2STRUCT::ENDELEMENT.
% Compiled module: XML2STRUCT::GETSTRUCT.
% Compiled module: XML2STRUCT__DEFINE.
% Error opening file. File: getRunID
```

```
% Compiled module: GETRUNID.
% Loaded DLM: XMLSAX. Opening input file
% ../data/input/1060715013738_params.xml
Unknown tag name: ISFS_RUN
Unknown tag name: krigRoutine
Unknown tag name: modelArray
Unknown tag name: modelArray
Unknown tag name: servletWrapper
Unknown tag name: servletWrapper
Unknown tag name: willWait
Unknown tag name: willWait
The input values for the Model run are:
** Structure ISFS, 6 tags, length=56, data length=52:
RUNID STRING ’1060715013738’
DOKRIG INT 1
KRIGROUTINE STRING ’1
SP_Fortran’ NNEIGHBORS INT 1
STUDYSITE STRING ’Rocky_Mtn_Np’
FNAME STRING ’../data/input/1060715013738.ma’
INPUTDIR STRING = ’../data/input/Rocky_Mtn_Np/’
BASENAME STRING = ’’
***************************************************************
Read the input file containing the merged field and RS data
***************************************************************

number of variables = 16
number of lines of data = 1800
% Compiled module: STRPARSE.
variables:
column 0: variable = tplant
column 1: variable = band1
column 2: variable = band2
column 3: variable = band3
column 4: variable = band4
column 5: variable = band5
column 6: variable = band6
column 7: variable = band7
column 8: variable = elev
column 9: variable = slp
column 10: variable = absasp
column 11: variable = xutm
column 12: variable = yutm
column 13: variable = tc1
column 14: variable = tc2
column 15: variable = tc3
location of xutm = 11
```

```
location of yutm = 12

**************************************************************
Computing distance matrix
**************************************************************

% Compiled module: W.
max before rescaling = 36405.2
min before rescaling = 2.25375
max after rescaling = 16153.2
min after rescaling = 1.00000

**************************************************************
Determine boundaries of study area
**************************************************************

map: xl, xu, yl, yu = 422594.75 458132.74 4450614.29 4481848.75
field: xl, xu, yl, yu = 426481.00 456026.35 4451880.00 4476650.00
krigsize = 1186 1041

**************************************************************
Prepare for modeling steps
**************************************************************

elem = 1 2 3 4 5
6 7 8 9 10 13
14 15

**************************************************************
Perform stepwise regression
**************************************************************

% Compiled module: STDDEV.
% Compiled module: MOMENT.
% Compiled module: REGRESS.
% Compiled module: T_CVF.
% Compiled module: T_PDF.
% Compiled module: IBETA.
Final Statistics
Variablen i.d.Gl.: band5 elev band7 band2 band1 slp band6
Var m R % S(m) Beta Temp Tprob
band5 2.41e-01 12.2 1.73e-02 1.36e+00 13.913 100.0
elev 4.43e-03 8.4 4.45e-04 2.96e-01 9.962 100.0
band7 -3.83e-01 7.8 3.96e-02 -1.21e+00 9.671 100.0
band2 -2.52e-01 7.0 9.11e-02 -2.88e-01 2.765 99.7
band1 1.56e-01 5.9 6.04e-02 2.65e-01 2.585 99.5
```

```
slp -5.52e-02 3.5 1.54e-02 -7.90e-02 3.590 100.0
band6 5.46e-02 2.2 1.93e-02 9.57e-02 2.832 99.8
Const -18.937 R 24.2% F_emp 81.665
% Compiled module: F_PDF.
F_prob 100.0 J 7 DF 1792 n 1800

***************************************************************
Calculate and plot the variogram of the residuals to the fit
***************************************************************

min, max of residuals = -10.625395 17.661105
GAMV Version: 2.000
data file = ./residuals.dat
columns for X,Y,Z = 1 2 0
number of variables = 1
columns = 3
trimming limits = -1.00000002E+21 1.00000002E+21
output file = ./gamv.out
number of lags = 20
lag distance = 1000.
lag tolerance = 0.
number of directions = 1
azm, atol, bandwh = 0. 180. 9999.
dip, dtol, bandwd = 0. 90. 90.
flag to standardize sills = 0
number of variograms = 1
tail,head,type = 1 1 1
xltol is too small: resetting to xlag/2
Variable number 1 Number = 1800
Average = 5.59815394E-09
Variance = 25.2956276
Variogram 1 Semivariogram : tail=Residual head=Residual
GAMV Version: 2.000 Finished
initial fit guess = 32.7029 9982.05
Gaussian fit coefficients = 26.9169 542.588
dokrg = 1

***************************************************************
Perform the kriging of the residuals
***************************************************************

NUGGET FLOAT = 10.0000
RANGE FLOAT = 542.588
SILL FLOAT = 26.9169
% Compiled module: READCOL.
% Compiled module: NUMLINES.
```

```
% READCOL: Format keyword not supplied - All columns assumed floating
% point
% Compiled module: GETTOK.
% Compiled module: REPCHR.
% READCOL: Skipping Line 1
% Compiled module: STRNUMBER.
% READCOL: 22 valid lines read
% Compiled module: MEAN.
% Compiled module: CURVEFIT.
percent done = 0
percent done = 5
percent done = 10
percent done = 15
percent done = 20
percent done = 25
percent done = 30
percent done = 35
percent done = 40
percent done = 45
percent done = 50
percent done = 55
percent done = 60
percent done = 65
percent done = 70
percent done = 75
percent done = 80
percent done = 85
percent done = 90
percent done = 95
percent done = 100
Done with kriging
ncols: 1186
nrows: 1041
xllcorner: 422594.750000
yllcorner: 4450614.500000
cellsize: 29.965000
max is 17.661104 and min is -10.625395
ncols: 1186
nrows: 1041
xllcorner: 422594.750000
yllcorner: 4450614.500000
cellsize: 29.965000
max is 7.337152 and min is 4.472137
% Restored file: ENVI_UTL.
```

```
**************************************************************
Apply the model to estimate total plants over the study area
**************************************************************

InEQ = 4 7 6 1 0 8 5
number of significant variables = 7
coefficients = 0.24134670
0.0044325691
-0.38294379
-0.25177076
0.15618168
-0.055248675
0.054570118
opening and reading image file band5
opening and reading image file elev
opening and reading image file band7
opening and reading image file band2
opening and reading image file band1
opening and reading image file slp
opening and reading image file band6
Min and max of the derived total plant image = 0.00000 33.4452
% LOADCT: Loading table BLUE/GREEN/RED/YELLOW
% Loaded DLM: JPEG.
0: fid_name = ../data/input/1060715013738.ma
1: fid_name = ../data/input/1060715013738.ma
2: fid_name = ../data/input/1060715013738.ma
3: fid_name = ../data/input/1060715013738.ma
4: fid_name = ../data/input/1060715013738.ma
5: fid_name = ../data/input/1060715013738.ma
6: fid_name = ../data/input/1060715013738.ma
7: fid_name = ../data/input/1060715013738.ma
8: fid_name = ../data/input/1060715013738.ma
9: fid_name = ../data/input/1060715013738.ma
10: fid_name = ../data/input/1060715013738.ma
FINISHED
% Program caused arithmetic error: Floating underflow
```

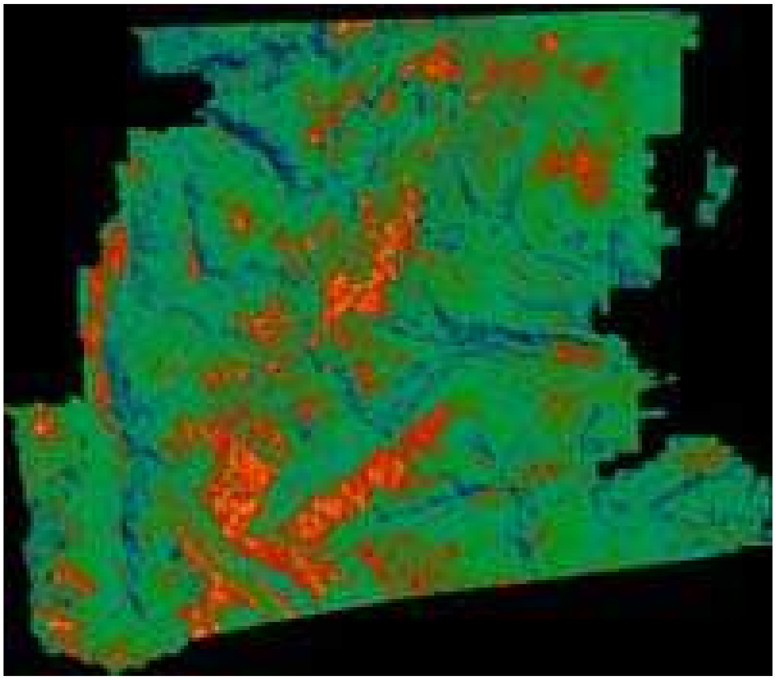
	This sample image file resulted from the test run for this document.

Figure 1. Output Image

4.6 Actual Results

The actual results of the model run will vary depending on the choice of the nearest neighbor and kriging routine options selected for each combination of study site and model array.

5 Regression Test Descriptions

The regression test section will be required as the project grows to include regression testing.