Prospectus 9

MULTISCALE ADVANCED RASTER MAP ANALYSIS SYSTEM
Geographical Surveillance for Hotspot Detection, Delineation, and Prioritization: Spatial Scan Statistics for Irregularly Shaped Clusters and Early Warning System.
DEVELOPMENT OF REMOTE SENSING METHODS FOR CROP BIOTERRORISM

G. P. Patil
Center for Statistical Ecology and Environmental Statistics
Department of Statistics
The Pennsylvania State University
University Park, PA 16802
http://www.stat.psu.edu/~gpp

Research and Outreach Team

Ganapati P. Patil, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, The Pennsylvania State University
Richard L. Tutwiler, Image Processing and Computer Vision, Applied Research Laboratory, The Pennsylvania State University
Paul Backman, Department of Plant Pathology, The Pennsylvania State University
Barbara Christ, Department of Plant Pathology, The Pennsylvania State University
A. Scott Lewis, Control Theory, Vibration, and Image Processing, Applied Research Laboratory, The Pennsylvania State University
Joseph P. Stitt, Pattern Recognition, Signal and Image Processing, Applied Research Laboratory, The Pennsylvania State University
C. Taillie, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, The Pennsylvania State University
Eric D. Warner, Remote Sensing, Environmental Resources Research Institute**,** The Pennsylvania State University
David P. Weingartner, Department of Plant Pathology, University of Florida

July 23, 2002

DEVELOPMENT OF REMOTE SENSING METHODS FOR CROP BIOTERRORISM
Spatial Scan Statistics for Irregularly Shaped Clusters and Early Warning System.
Geographical Surveillance for Hotspot Detection, Delineation, and Prioritization:
MULTISCALE ADVANCED RASTER MAP ANALYSIS SYSTEM

Development of Remote Sensing Methods for Crop Bioterrorism

Table of Contents

ABSTRACT

This prospectus has the specific aim of developing novel remote sensing methods for the detection of crop bioterroism. A cross disciplinary team has been assembled from Penn State University to address this problem. The team is formed of members with extensive experience in plant pathology, entomology, biology, computer science, electrical engineering, environmental and ecological statistics, and inferential geoinformatics. The objective is to utilize hyperspectral sensing technology as a coarse detection system to optimize time and resources of ground-based teams that monitor the US agriculture and forest lands with real-time PCR, and sentinel plant technologies. The research thrust is a three-pronged approach consisting of:

i. Development of signatures for key pathogens using ground-based, portable hyperspectral cameras.
ii. Development of a Photosynthesis mapping operator that will create a deformable spectral signature mapping based on the biophysical and biochemical interactions of the plant with the pathogen.
iii. Confirm signatures developed in i. under field conditions utilizing an airborne hyperspectral imaging platform.

Development of Remote Sensing Methods for Crop Bioterrorism

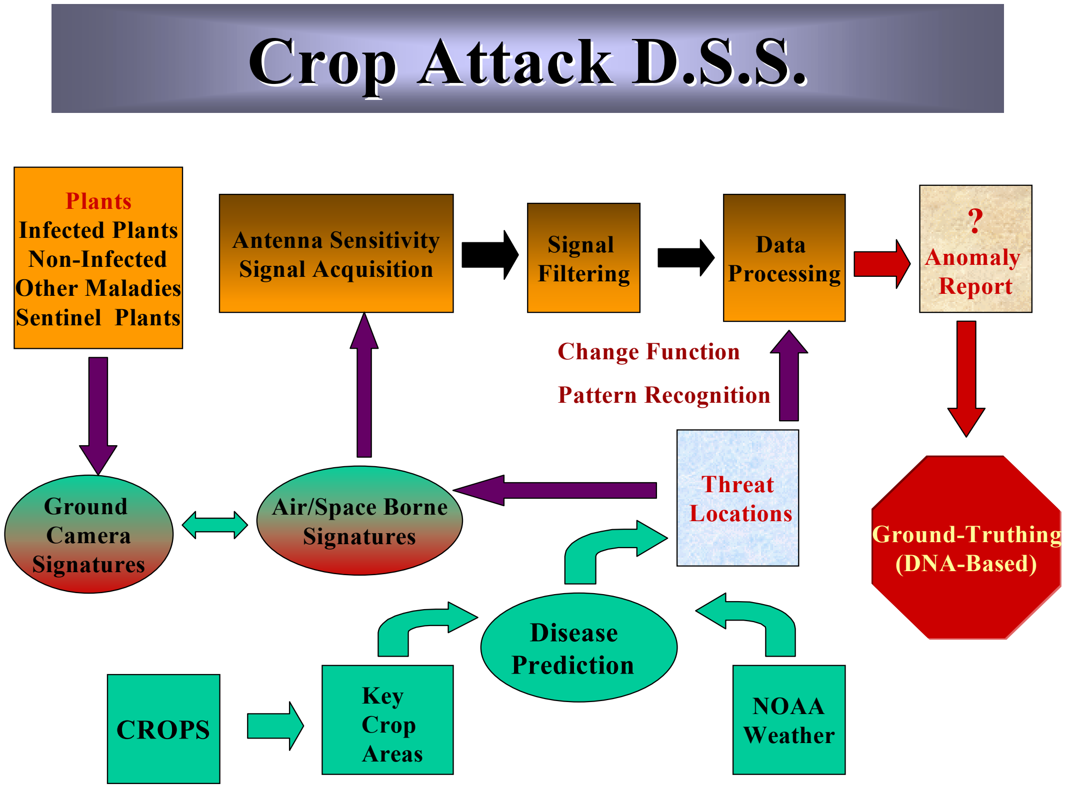
RESEARCH PLAN

American Agriculture is a widely dispersed enterprise, covering many different environments crossing the entire continent and extending into the Caribbean and Pacific Ocean. Its impact on our country ranges to supplying the food for all our tables, to employing 16% of our workforce in Ag-related enterprises, to producing more than 25% of our export value, as well as generating 16% of the US GDP. Disruption of agriculture and the food system could be catastrophic to the nation's stability.

A shrewd enemy, familiar with the epidemiology of plant pests, could easily and quickly cause huge disruptions with virtually undetectable quantities of microbial pathogens. Testimony to the House Government Reform, National Security, Veterans Affair, and International Relations Subcommittee this past fall pointed to the high probability that crop terrorism could and would occur. Reference was made to the fact that Cuba and the former Soviet Union satellite states already had capabilities in this arena, including genetically engineered bioweapons. This threat has been recognized by our Homeland Defense Office, but is little understood by the vast majority of the American population. To this end, USDA and DoD have begun to invest in DNA-based technologies for the rapid detection of high threat pathogens. Also, DARPA has invested in sentinel plant technology, that causes genetically altered plants to produce reporter proteins that are produced in response to infestations by particular high-threat races of pathogenic species. Unfortunately, it is not obvious as to how ground-based teams could monitor the whole of US agriculture and forest lands with real-time PCR, or, in the case of sentinel technologies, how would it be possible to survey all of the dispersed sentinel plants that would be planted all across the United States? The logistics of ground-based monitoring are daunting, and point to a strong need to monitor US cropland by remote sensing. Remote sensing, to be effective, should resolve zones of infection that are as small as 5m in diameter. Otherwise, infection zones extending much beyond that size become increasingly difficult to contain. This is particularly true of wind-borne pathogens that can quickly spread infective propagules miles beyond their site of production. Unfortunately, resolution at this scale would require responding to only one pixel of information generated by present day satellites. Improvements in the next generation of sensors will have to be made, as will the techniques for handling the huge volumes of high-resolution data.

The other difficulty has been the development of high-quality signatures that can be detected from airborne or space-borne platforms. To-date, reliable detectable signatures have been difficult to obtain. However, most of these efforts have occurred primarily with engineer-dominated research teams. We propose to defeat this impasse by integrating biologists with engineers to generate high quality signatures, determine signal strength on unique portions of these hyperspectral signatures, and to develop goals for instrument resolution for next generation platforms.

The data developed would be integrated into a decision support system (below) that would link the sensor platform with a list of key target areas. These key areas would be further limited by using NOAA weather information to indicate which of the key areas would be likely to develop a sustainable epidemic if the were to receive a bioterrorist inoculation. Only these threat areas would be scanned to reduce the scope of the monitoring problem. If signatures and patterns of disease development were consistent with an incipient epidemic, ground-truthing teams would visit the site for confirmation of high-threat pathogen release. The following figure depicts the process from signature development through identification of threat areas, signal acquisition and processing to the development of an anomaly report.



System Block Diagram

TECHNICAL/ MANAGEMENT

<u>TECHNICAL: The objectives, tasks and subtasks of the proposed 9 month project would be the following:</u>

1. To develop signatures for key pathogens using ground-based, portable hyperspectral cameras. These signatures would be developed from healthy plants, and diseased plants infected with single and multiple pathogens (both visible and latent infections) and pests. Similar processes would be developed for sentinel plants, particularly for those using green fluorescent protein as a reporter.
    a. Task 1. Develop signatures of leaf blight (necrosis causing) and a blight plus yellowing pathogen, and finally a yellows inducing insect in greenhouse grown plants.
        i. Determine effects of plant varieties on signature quality
        ii. Determine quenching effects of key pesticides

        iii. Determine if multiple pathogens with similar visible signatures can be differentiated by a hyperspectral camera

2. Evaluate and enhance hardware and processing algorithms to filter, enhance, and resolve crop pathogen signatures developed from greenhouse and field research
    a. Task 1.: Photosynthesis Mapping Operator Development
        i. Biophysical/Biochemical Analysis
        ii. Photon transport simulation utilizing existing Bi-directional Canopy Reflectance Models
        iii. Signature verification with ground based hyperspectral data sets
    b. Task 2: Spectral signature integration with un-mixing algorithm and detection/classification algorithms
        i. Integrate Photosynthesis Mapping Operator with deconvolution un-mixing algorithm.
        ii. Verification with Task 1 iii. data.
        iii. Integrate results of Task 1 iii. with non-linear correlator algorithm for spatial and spectral detection/classification
    c. Task 3: Decision Support System Development
        i. Integration of Task 2 iii. output with Multi-Scale Advanced Raster Map Analysis System
        ii. Selection of Optimal Scan Statistic for hotspot detection, delineation, and prioritization
3. Field plantings of a test crop will be evaluated by hyperspectral sensors both from the ground and from two altitudes.
    a. Task1: Confirm signatures developed in 1 under field conditions
        i. Establish field plots in Florida during January
        ii. Infect or infest plots with multiple pests
        iii. Evaluate with ground and air-borne hyperspectral cameras
    b. Task 2. Determine if differential signatures can be found among when plants are affected by multiple pests
4. Provide goals for the next generation of sensors relating to the early and reliable resolution of high-threat pathogens from air and space-borne platforms.

Scientific and Engineering Practices

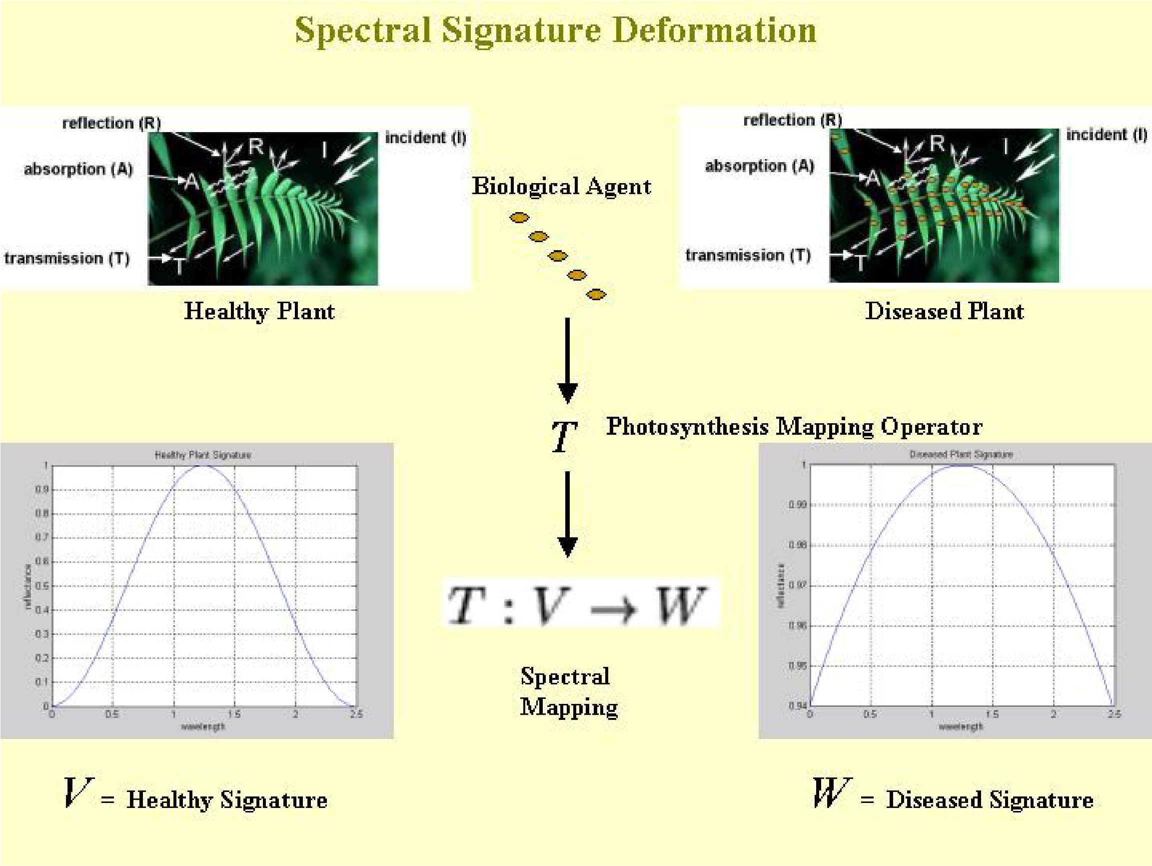
For the purposes of this short-term research project, we felt that much of the biological work would have to be conducted in greenhouses. Since plant pathologists and entomologists can artificially infect or infest test crops with pathogens or insects in greenhouses, it was considered that this would be the best place to start to secure signatures of key crop pests. In a similar manner, later trials would evaluate the ability to identify the presence of one key disease when the plant was infected with multiple pests. Similar experiments would be conducted with various cultivars (plant varieties) to see if horticultural differences might obscure the signatures detected.

We are proposing to use the potato crop and the late blight disease as key crop and pathogen, and early blight disease, insect damage, and varieties as probable confusing issues. Our reasons are that we have experience and have developed previous hyperspectral data for this crop and this pest complex. There are several diseases and pests that could be used to test the ability to sort out different types of stresses, i.e. disease, pests and environmental factors. From our previous experience, we know that the signatures are likely to change depending on crop stage and the potato variety. We also have developed signatures for potatoes diseased with early blight. We have the ability to grow potatoes and the ability to initiate the diseases early and late blight. We have all necessary cultures of the pathogens needed to initiate disease in our greenhouse-grown plants.

Objective. 1, Task 1. Potato seed of three different potato varieties will be planted in pots and allowed to emerge. These varieties will differ in maturity classification and differ in reaction to pathogens as well as drought stress. Treatments will consist of: no fertilizer, fertilizer applied at planting and after emergence liquid fertilizer applied weekly (high fertility), plants provided with minimum water, plants provided with high levels of water. Plants will be photographed with a hand held Hyperspectral camera to record hyperspectral data at several growth stages of the plants. On each of the three varieties there will be healthy plants with no fungicide applied and healthy plants with weekly application of fungicide and insecticides that are typically applied to the crop. Plants will be photographed with a hand held camera to record images at several growth stages of the plants.

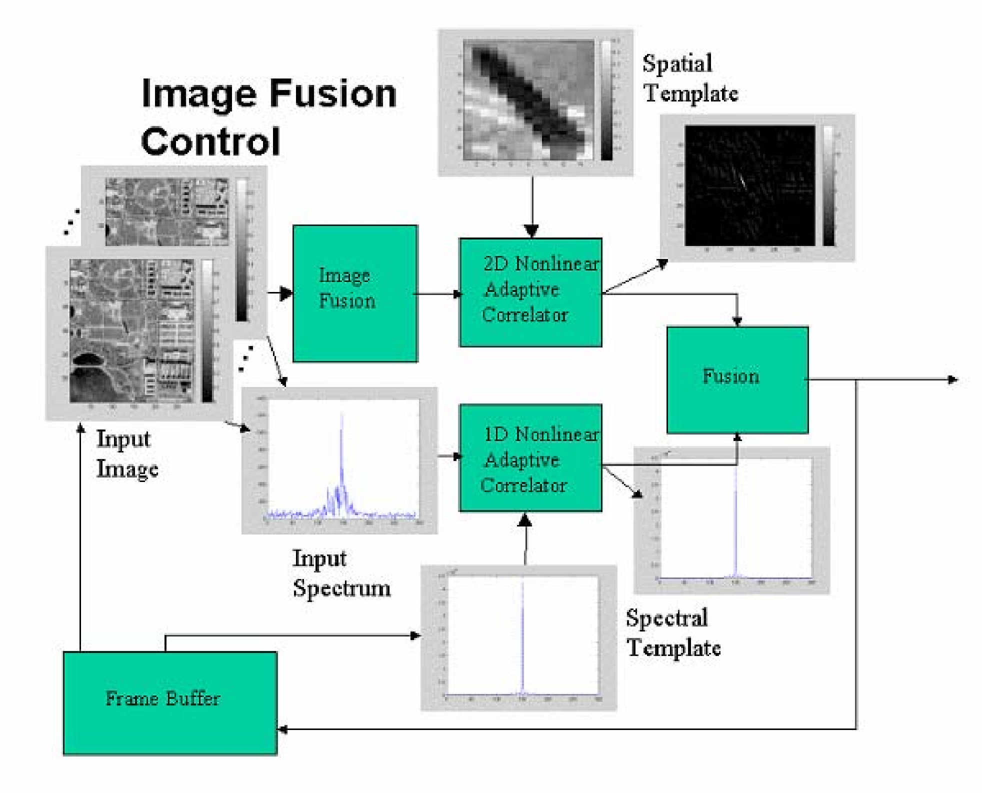
Plants of each of the three varieties will be inoculated with early blight pathogen (*Alternaria solani*, which causes yellowing and "dry" circular leaf lesions) and some with the late blight pathogen (*Phytophthora infestans*, which causes a "wet irregular lesions and a general blighting. We will take pictures within 48 hrs of inoculation of the two pathogens to determine if latent infection can be detected. Latent infections are those where the pathogen has infected but there are no visible symptoms appearing. We will subsequently take additional hyperspectral images as symptoms appear and disease progresses.

Objective 2, Task 1. The key to correct classification and detection of plant infection is to understand the spectral signatures of a healthy and diseased plant. The basic research thrust is to develop an understanding of the biological interactions that occur in the plant and how the radiation scattering process is affected. A conceptual diagram of the process in illustrated in the figure below. The operation is analogous to proper template selection for the matched filtering process. Variation in tissue optical properties is wavelength dependent. For green leaves the smallest variation is in the visible region (VIS = 400nm-700nm), and the near infrared (NIR = 700nm – 1300nm). It is envisioned that if the biochemical phenomena can be analytically mapped to the photon transport mechanism that defines the absorptive and reflective properties of the plant then a better model for spectral signatures could be utilized in the detection/classification process. The reflectance is divided into three major parts: (1) radiation that is not scattered by vegetative tissues, (2) singly scattered radiation, and (3) radiation that undergoes multiple scattering. Existing radiative photon transport algorithms (SAIL, MODTRAN4) and code that can simulate the interaction of photons with multiple plant components will be utilized and modified if necessary to accommodate the new features. The following input parameters will be used in the reflectance computations. Leaf and NPV area index, leaf and NPV angle distributions, leaf and NPV hemispherical reflectance and transmittance and soil reflectance. In addition, according to the specific areas of coverage for the experiments, solar and viewing geometry for the sensing system will be incorporated into the modeling. The outputs of the modeling development will be correlated with the data collections of both the ground and air based systems used in the experiments to verify the validity. These specular features will be the inputs for the un-mixing process as well as the non-linear correlator algorithm described next for both spatial and spectral localization.



Photosynthesis Mapping Operator

Objective 2, Task 2. Standard algorithms suites from software packages TNTmips and ENVI 3.5 will be utilized in the un-mixing process and for automatic endmember selection. In addition to the standard techniques mentioned above several in-house deconvolution algorithms developed on the theory of Blind Source Separation and Independent Component Analysis will also be tested for unsupervised spectral un-mixing. The deconvolved spectral signatures of the spectral un-mixing stage will be processed by a non-linear correlator algorithm for detection, classification, and localization of hypothesized targets. Specifically, the nonlinear correlation algorithm is simply a conventional correlation between an image space and object template with an additive nonlinear corrective filter term. The nonlinear filter is an error measure between the actual and ideal correlation coefficients. Currently, a set of two parameters is utilized in the nonlinear filter to control the percent of tolerable object-object template mismatch. The parameter update and optimization procedure is based on the solution of two simultaneous equations that are derived from the statistical distribution of the non-linear correlator error matrix histogram. Constructing simple foundations for the nonlinear correlator by minimizing internal calculations and parameter-input eases operating maintenance and algorithm enhancements. The non-linear correlator procedure is depicted in the figure below.



Non-Linear Correlator Block Diagram

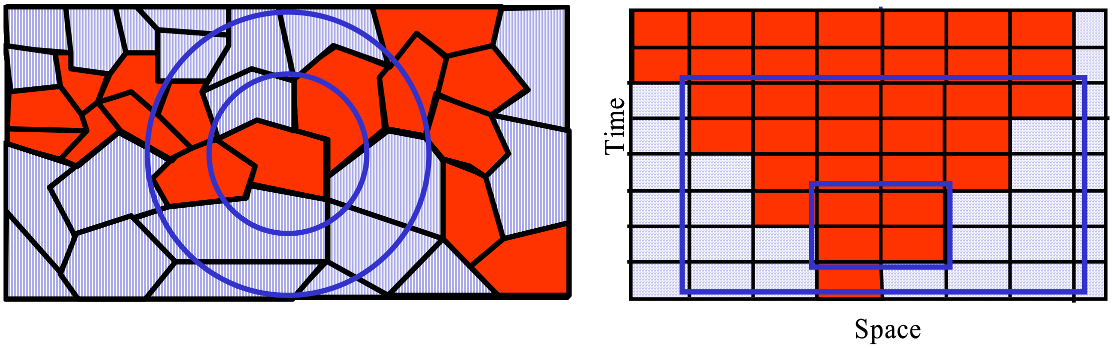
The inputs to the algorithm are a spatial image and a spectral template. The spatial input could be any of the two dimensional images that constitute the makeup of the spectral bands from the hyperspectral imaging platform, or a fusion of a select group of bands. The fusion processes is accomplished by a decomposition of the image via Discrete Wavelet Transform techniques. The structure of the non-linear correlator algorithm can be utilized to work with both one and two dimensional data sets. The spectral signature template would be input via the Photosynthesis Mapping Operator. The spatial signature template could be downloaded from a library of reference signatures or actually selected and cropped from a set of existing images or specific spectral bands. The nonlinear correlator is simply a conventional correlation $h(x, y)$ summed with an adaptive filter term $b(x, y)$:

$$Y(x,y) \equiv h(x,y) - b(x,y). \tag{1}$$

The nonlinear filter term $b(x, y)$ should suppress false hits raised by conventional correlation and accentuate any true hits. The process is intended to localize object matches so that other more traditional image processing techniques may then be used to help the overall object recognition scheme. Consequently, $b(x, y)$ weighs the difference between the ideal and actual correlation contributions. The outputs of the nonlinear correlator will be processed by the Multi-Scale Advanced Raster Map Analysis System.

<u>Objective 2, Task 3.</u> In geographical disease surveillance, two important problems are (*i*) the identification of areas (clusters) with exceptionally high response and (*ii*) the determination of whether the high response can be attributed to chance variation (false alarm) or is statistically real. The spatial scan statistic (Kulldorff and Nagarwalla 1995; Kulldorff 1997) has quickly become a popular method for detection and evaluation of disease clusters, and is now widely used by many health departments and academic scientists both nationally and internationally. When applied in space-time, the scan statistic can provide early warning of outbreaks and can monitor the spatial spread of an outbreak.

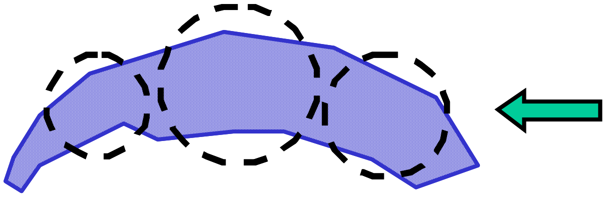
Three basic properties of the scan statistic are the geometry of the area being scanned, the probability distribution generating responses under the null-hypothesis of chance variation, and the shapes and sizes of the scanning window. Depending on the application, different response distributions are chosen and the test statistic is evaluated through Monte Carlo sampling. For purposes of the present proposal, available scan statistic software suffers two limitations. First, circles have been used for the scanning window, resulting in low power for detection of irregularly shaped clusters. Alternatively, an irregularly shaped cluster may be reported as a series of circular clusters. Second, reflecting the epidemiological origins of the scan statistic, response distributions have been taken as discrete (specifically, binomial or Poisson).



Cholera outbreak along a river flood-plain
- Small circles miss much of the outbreak
- Large circles include many unwanted cells

Outbreak expanding in time
- Small cylinders miss much of the outbreak
- Large cylinders include many unwanted cells

Circular spatial scan statistic zonation (left) and cylindrical space-time scan statistic zonation (right)



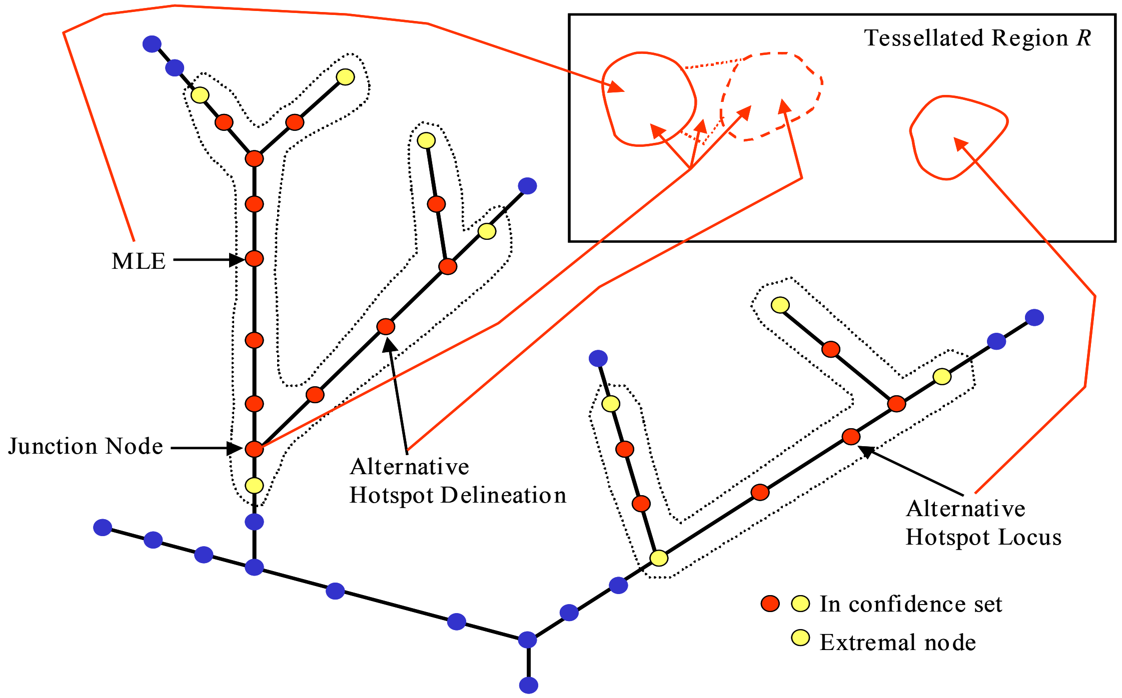
Circular zones may represent a single cluster as multiple clusters

Circular scan statistic represents a single actual cluster as a series of small clusters.

The first limitation will be addressed by abandoning fixed-shape scanning windows. Instead, candidate hotspot zones will be of arbitrary shape and will be derived as connected components of upper level sets of the response surface. These candidate zones form a (combinatorial) tree under set inclusion, the scan statistic and its likelihood ratio value is computed for each node of the tree, and statistical significance of each node as a hotspot is assessed through Monte Carlo simulation (Patil et al, 2002abc).

The second limitation will be overcome by extending the scan statistic methodology to continuous response distributions. We will focus on three parametric families of distributions: gamma distribution, lognormal distribution, and scaled beta distribution. The first two families apply to responses that can range from zero to infinity, while the third is for bounded responses. The overall approach is to model the mean and relative variance in terms of the area of the candidate hotspot zone. In general, one expects the relative variability to decrease as this area increases (i.e., as the effective sample size increases) and we will use an inverse power function to represent this behavior. In turn, the mean and variance are functions of the parameters of the response distribution, so that the likelihood function can be written down and parameters estimated by maximum likelihood.

Likelihood methods are employed throughout, with the important benefits that confidence values can be associated with identified hotspots and plausibility of alternative hotspot delineations can be assessed.

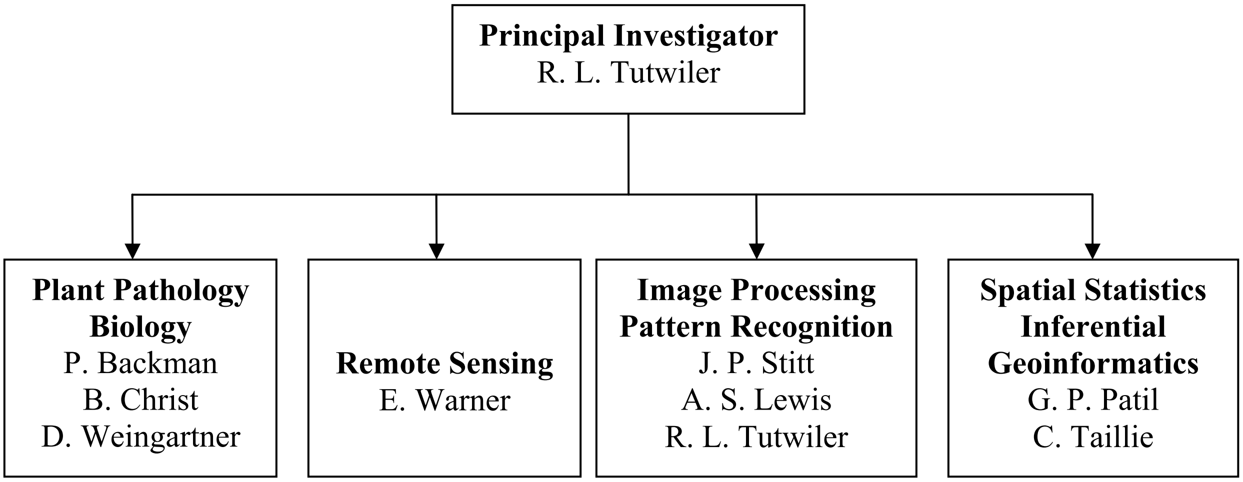


A hotspot confidence set with two connected components is shown on the ULS tree. The connected components correspond to different hotspot loci while the nodes within a connected component correspond to different delineations of that hotspot – all at the appropriate confidence level.

<u>**Objective 3, Task 1 and 2.**</u> Field trials will be planted during January in the St. John's region of Florida. Typically a one square acre field will plant 68-72 rows 200 ft long of potatoes. We will plant 3 varieties based on maturity classification. Seed of Kennebec (late), Atlantic (medium-late) and Norland (early) will be planted in blocks approximately 22 rows per variety.
Within each variety there will be sub-blocks that are used to maintain healthy plants, establish early blight, establish late blight and establish a leafhopper infestation. The cultures of the early and late blight pathogens will be cultured and induced to produce spores. The spores will be collected from artificial media and placed in water for inoculation.

The hand-held 128 channel hyperspectral camera will be used to record images of the different plots several times throughout the growing season. We will engage a private company to fly over the fields and record images of the field. This will allow us to compare signatures gathered at ground level verses those at several altitudes by airplane.

<u>**Objective 4, Task 1.**</u> Based on results of greenhouse and field trials, we will be able to determine if there are unique signatures for key potato pests. Further, since we are evaluating multiple pests we should be able to determine if pests with similar symptom types can still produce unique signatures that will allow the detection of a "key pathogen". Field plots will be infested with the pathogens and pest combinations discussed in Objective 1. Futher, we will determine if there is significant signal strength for this to be detected by airborne platforms. The data from different altitudes should allow us to determine the rate of signal attenuation that should provide information on necessary resolution.



PROJECT MANAGEMENT STRUCTURE

Dr. Richard L. Tutwiler is the principal investigator at ARL in image processing/computer vision research. Several projects have included; An image analysis Bubble Distribution Detection System for the characterization of hydrodynamic propulsion experiments, IR focal plane array testbed for the diagnostic monitoring of rotating machinery, A Range-Doppler map image analysis system, Image analysis/pattern recognition system for the analysis of melt pool characteristics and parametric control of multiple electron beam guns for Electron Beam-Physical Vapor Deposition processing, An advanced image segmentation techniques utilizing fuzzy clustering and PIC (Prototype Intelligent Controller) architecture are currently being designed for detection/classification and localization of objects in side scan SONAR images. His current research efforts involve using artificial life multi-agent software systems based on complexity theory and reinforcement learning to develop an artificial retina function.

Dr. Paul Backman has published more than 200 papers, book chapters, proceedings and research abstracts in the broad area of plant pathology. These include numerous papers on disease epidemiology, disease assessment, chemical and biological controls for plant diseases, and decision support systems for farmers. He has served as a Director of a research institute at Auburn University, and as an Associate Dean of Research at Penn State University.www.personal.psu.edu/pab24.

Dr. Barbara Christ has published more than 200 papers, book chapters, proceedings, research abstracts and trade journal articles in plant pathology and plant breeding. Her expertise is specifically on the potato crop developing potato varieties and researching several pathogens including early and late blight. She served as the project manger of a grant entitled "Early Detection of Potato Late Blight Using Hyperspectral Remote Sensing".

Dr. David P. Weingartner's research interests and expertise is in the area disease and nematode management strategies for vegetable crops grown in north Florida. He has developed a management matrix for nematodes and soil borne diseases affecting potatoes grown in north Florida. The plan includes cover crops, resistant cultivars, and use of various chemicals. This systems has been adapted and implemented a late blight forecast and spray advisory system.

Dr. Joseph P. Stitt interests include pattern recognition, signal and image processing, and research in the areas of biomedical ultrasound and neuroengineering.

Dr. A. Scott Lewis interests include sliding mode control, vibration and control theory, image processing and pattern recognition.

Dr. David Warner is actively engaged in GIS and remote sensing research. Most recently he has worked on issues related to land use mapping, assessing the agricultural impacts of urban land expansion, and the detection of potato blight with hyperspectral data collected from an airborne platform.

Dr. Ganapati Patil has published more than 300 papers, 30 books, and several white papers in mathematical, statistical, computational, data analytical and interpretational areas in theory and practice. He has authored 35 papers on inferential geoinformatics and multiscale advanced raster map analysis system with emphasis on remote sensing, landscape patterns analysis, change detection and accuracy assessment, geospatial modeling and simulation devices. He is involved with spatial scan statistics for irregularly shaped clusters and early warning systems for geographical surveillance for hotspot detection, delineation, and prioritization for monitoring, assessment, and management. He is Distinguished Professor of Mathematical and Environmental Statistics, Director of the Penn State Center for Statistical Ecology and Environmental Statistics, Editor-in-Chief of the international journal, *Environmental and Ecological Statistics*. For additional information, http://www.stat.psu.edu/~gpp .

Dr. Charles Taillie has published more than 150 papers, 5 books, and several lead papers in mathematical, statistical, computational, data analytical and interpretational areas in theory and practice. He has authored 25 papers on inferential geoinformatics and multiscale advanced raster map analysis system. He has been senior research associate collaborating with Professor G. P. Patil at the Penn State Center for Statistical Ecology and Environmental Statistics for 25 years, participating effectively in its pioneering initiatives, including biosurveillance using tree-structured spatio-temporal scan statistics. For additional information, http://www.stat.psu.edu/~gpp .

Project Timeline

Workplan for Spectral Imaging Study

Month Task	1	2	3	4	5	6	7	8	9
Hire People	x								
Buy Equipment	x	x							
Objective 1									
Task 1									
i. Plant variety effects		x	x	x					
ii. Quenching effects		x	x	x					
iii. Multiple Pathogen effects			x	x	x				
Objective 2									
Task 1									
i. Biophysical/Biochemical Analysis	x	x	x	x	x				
ii. Photon transport simulation			x	x	x	x	x	x	
iii. Signature verification				x	x	x	x	x	x
Task 2									
i. Un-mixing alog. integration	x	x	x	x					
ii. Verification with Task 1 iii. data			x	x	x	x			
iii. Non-linear correlator algorithm integration									
with Task 1 iii. data		x	x	x	x	x			
Task 3									
i. Integration of Task 2 iii. output				x	x	x	x	x	x
with multi-scale advanced raster map analysis system									
ii. Selection of Optimal Scan Statistics for hotspot detection				x	x	x	x	x	x
delineation, and prioritization									
Objective 3									
Task 1									
i. Establish field plots				x					
ii. Infect field plots					x	x	x		
iii. Evaluate field plots				x	x	x	x	x	
Task 2									
i. Determine signatures						x	x	x	x
Project Meetings	x	x	x	x	x	x	x	x	x
Monthly Reports	x	x	x	x	x	x	x	x	x

REFERENCES

Ingram, R. N., and Tutwiler, R. L. (2002). An Adaptive Nonlinear Correlation Algorithm for Object Recognition: Applications in Hyperspectral Image Processing. Technical Report, Applied Research Laboratory, Autonomous Control and Intelligent Systems Division, The Pennsylvania State University. pp. 1-30.

Patil, G. P. (2002a). Multiscale Advanced Raster Map Analysis System: Geographical Surveillance for Hotspot Detection, Delineation, and Prioritization: Spatial Scan Statistics for Irregularly Shaped Clusters and Early Warning System. A Research and Outreach Prospectus. http://www.stat.psu.edu/~gpp/PDFfiles/Prospectus-8.pdf.

Patil, G. P, Bishop, J., Myers, W. L., Taillie, C., Vraney, R., and Wardrop, D. H. (2002b). Detection and Delineation of Critical Areas Using Echelons and Spatial Scan Statistics with Synoptic Cellular Data. http://www.stat.psu.edu/~gpp/PDFfiles/TR2002-0501.pdf.

Patil, G. P., Balbus, John, Biging, Gregory, JaJa, Joseph, Myers, W. L., and Taillie, C. (2002c). Multiscale Advanced Raster Map Analysis System: Definition, Design and Development. http://www.stat.psu.edu/~gpp/PDFfiles/TR2002-0203.pdf.