HERBARIUM INFRASTRUCTURE DEVELOPMENT AND ECOLOGICAL APPLICATIONS OF SPECIMENS USING GEOGRAPHIC INFORMATION SYSTEMS

A Thesis Presented for the Master of Environmental Science

The University of Tennessee at Chattanooga

Ryan Joseph Miller

May 2008

To the Graduate Council:

I am submitting herewith a thesis written by Ryan Miller entitled "Herbarium Infrastructure Development and Ecological Applications of Specimens Using Geographic Information Systems." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Environmental Science.

[redacted]

Joey Shaw, Ph.D., Major Professor

We have read this thesis and recommend its acceptance:

[redacted]

Thomas P. Wilson, Ph.D.

[redacted]

Andrew D. Carroll

Accepted for the Graduate Council:

[redacted]

Dean of the Graduate School

DEDICATION

This thesis work is dedicated to my parents, John and Jane Miller, for their continual support and encouragement as well as their immense sacrifices towards my education.

ACKNOWLEDGEMENTS

I wish to express sincere appreciation to the many people in the Department of Biological and Environmental Sciences who have assisted me in obtaining my Master's degree. Special thanks is due to my major advisor and thesis committee members (Dr. Joey T. Shaw, Mr. Andrew D. Carroll and Dr. Thomas P. Wilson) who have displayed their commitment to my education through their generous guidance and patience.

Financial support for this project was provided through a USGS National Biological Information Infrastructure (NBII) research grant and scholarships awarded by the southeast chapter of the Geospatial Information and Technology Association (GITA).

Original idea and subsequent advice towards examining the spread of *Baccharis halimifolia* into new physiographic provinces was provided by Dr. L. Dwayne Estes at Austin Peay State University.

Research support was provided by Dr. Mark Schorr (UTC) in regards to guidance on statistical analyses and SAS PROCs. Assistance in specimen data collection was provided by SERNEC collaborators, especially Drs. Zack Murrell (APSU), B. Eugene Wofford (UTK), Alan Weakley and Carol Ann McCormick (UNC), and Wendy Zomlefer (UGA). For assistance in issues related to plant taxonomy, appreciation is extended to Mr. Stacy Huskins. Support during conferences and GIS-related career advice was graciously given by Mr. Charles Marlin (GITA) and Dr. Gary Litchford (UTC).

It is difficult to overstate my gratitude to Ms. Mary Kathryn Petrovic for her unwavering encouragement and cooperation during my graduate student career.

In addition to my brother Andy and sister Sarah, I am indebted to the many colleagues and friends who have given guidance and friendship in the past several years: Justin Conley, Dr. Carl Friese, Dr. Donald Geiger, David Hall, David Herr, Chris Lanka, Andrea Litt, Amy Mahon, Ben Miles, Johnny O'Donnell, Patrick and Leah Wensink, and Shaun Westfall.

ABSTRACT

Geographic Information Systems (GIS) allow for herbarium data to be used for new scientific research and also for the greater efficiency of the traditional uses of herbaria. My research uses the resources of both the herbarium at the University of Tennessee at Chattanooga (UCHT) and partnering herbaria of the Southeastern Regional Network of Expertise and Collections (SERNEC) to examine methods of incorporating new GIS technology into a functional infrastructure in order to enhance botanical research. Evaluations were performed on database and georeferencing software for use in the development of a pilot, GIS-enabled website used to query herbaria specimen information. Using this newly created database infrastructure, two ecological studies were performed using improved SERNEC datasets. In the first study, three common wetland invasive plant species were chosen and analyzed via two different methods to determine the historic rate and pattern of spread as well as to identify periods of invasiveness. Performance of these methods varied in scale with a general, area-corrected approach covering the entire southeastern United States and an associate species-corrected method covering smaller, regional areas of dense historical specimen collection. Results showed positive spread over time for these three species across both methods; however, periods of invasiveness did not coincide between the two methods. The use of two different methods and the subsequent comparison of results show the importance of sampling bias correction, scale selection, and adequate sample sizes for spatiotemporal analyses of plant distributions using herbarium records. The second study describes the spread of *Baccharis halimifolia* L. into new physiographic areas of the

southeast using SERNEC records in GIS. Results were compared against distribution descriptions in the taxonomic literature. The dataset strongly complimented the existing distributions described in the taxonomic literature and, therefore, supports the concept that recent range expansion into previously unoccupied physiographic areas has truly occurred for this species, rather than being an artifact of collection bias. The two studies mentioned above have been submitted for publication in peer-reviewed, scientific journals and are arguments for the use of herbarium data within GIS software. It is of note that similar techniques will be fundamental to future botanical research in the southeast.

PREFACE

As subdisciplines in the biological sciences have developed, herbarium uses have diversified to include applied research for issues pertinent to molecular systematics, ecology, conservation, and biodiversity. The full potential of the latter three disciplines has not been realized because of the inaccessibility and unassimilated nature of most of the data inherent in collections. These data may be intrinsic to the specimen itself (e.g., reproductive state) or recorded on the specimen label (e.g., geographic location of collected specimen or associated taxa). Due to the technological advances made in Geographic Information Systems (GIS) the amount of information contained on a specimen label is magnified as additional information can be associated with the specimen (e.g., soil type, elevation, annual rainfall, minimum winter temperature). This work demonstrates, through the production of a pilot website and two ecological studies, the potential of enhanced botanical research when large herbarium datasets are used within GIS.

The research contained within this thesis is a product of our collaboration with the USGS National Biological Information Infrastructure (NBII), whose mandate is to increase the access of biological data and information, and the Southeastern Regional Network of Expertise and Collections (SERNEC) which seeks to make the data of all southeastern herbaria available online as a single dataset. The first half of my thesis demonstrates the efforts of this collaboration as well as steps taken to modernize the University of Tennessee at Chattanooga herbarium (UCHT). In order to assist SERNEC, the first goal was to combine the various databases from participating university herbaria. These herbaria databases varied in terms of completeness, information entered off specimen labels, general format, and software used. Combining these databases required both the evaluation of software to suite the needs of large scale datasets and the need for a standardized method of databasing so future herbaria data could be combined to the SERNEC database. By using UCHT data, evaluations of both database software and existing protocols for standardizing biodiversity data were performed. After the determination of efficient software packages and standardization techniques that should be incorporated into SERNEC practices was complete, methods of georeferencing were compared in order to assess the feasibility of using new software to efficiently and correctly generate geographic coordinates for specimens with only locality descriptions. Using the information gathered during evaluation of database issues, UCHT practices were modernized to include the use of a new standardized database. The data from UCHT was incorporated into a new a pilot SERNEC database. Using the standardized pilot SERNEC database, the infrastructure was created to host the SERNEC information on a GIS-enabled website. Website development allowed for SERNEC data to be queried and the subsequent results to be mapped out for the user displaying county-level distribution ranges of species.

The second half of my thesis involves two ecological studies that show a "proof of concept" for the potential of enhanced research when large herbaria collections are combined and used within a GIS environment. My first study examines the spatiotemporal spread of three exotic, invasive wetland plant species. By researching questions on an important topic such as invasive plants, our work attempts to show that coupling herbaria datasets with GIS can assist in the analyses of modern ecological issues. Secondly, due to modern changes in climate and land development, some native plant species have begun to exhibit increased geographic distributions. To address this issue, the second study analyzes the range expansion of a native coastal species into new physiographic provinces. These two studies have both been submitted for publication to peer-reviewed regional botany journals.

This thesis is divided into six parts, each designed to be a separate, publishable unit except Part 6. From initial herbaria organization and database development to the final products of a pilot, GIS-enabled website and the submission of two important ecological studies, this work is intended to increase the awareness and use of GIS and herbaria data in biological research.

TABLE OF CONTENTS

List of Tables

Part 2.

Part 4.

List of Figures

Part 1.

Part 3.

Part 4.

Part 5.

GENERAL INTRODUCTION

A herbarium is a collection of dried plant specimens that are classified according to an accepted scheme and available for reference or other scientific study. Herbaria have historically housed physical specimens for taxonomic and systematic research in plants. Moreover, collected specimens have served as reference points for the development of county distribution maps, which are maps illustrating the counties within a given geopolitical area from which a particular plant species has been documented. For example, Figure 1 shows all the counties for which a specimen of red maple (*Acer rubrum* L.) was collected.

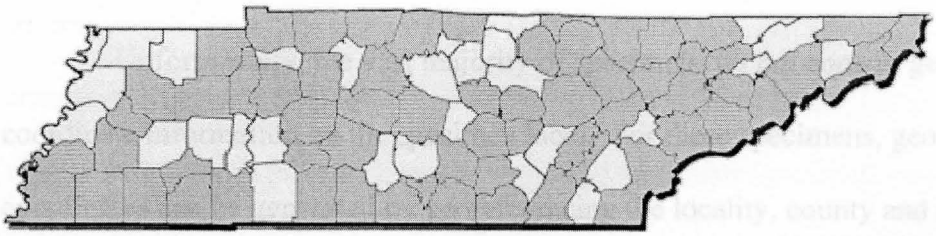


Figure 1. County distribution map of *Acer rubrum* (red maple) in Tennessee. (http://tenn.bio.utk.edu/vascular/). Retrieved 11/08/2006.

The need for herbaria to database specimens, in particular their associated biological characteristics and geographic locations, has been realized for several years; however, most herbaria do not have comprehensive databases because of time and financial constraints. Adding to the challenge is that each unique herbarium collection contains varying degrees of region-specific collection efforts and furthermore, recorded data associated with specimens can vary, especially in regards to the quality of geographic locality descriptions. For example, some of the earliest specimens in the UTC Herbarium (UCHT) contain only the most general description of where they were collected (Fig. 2a), but recent specimens often contain more detailed collection information and even GPS coordinates (Fig. 2b).

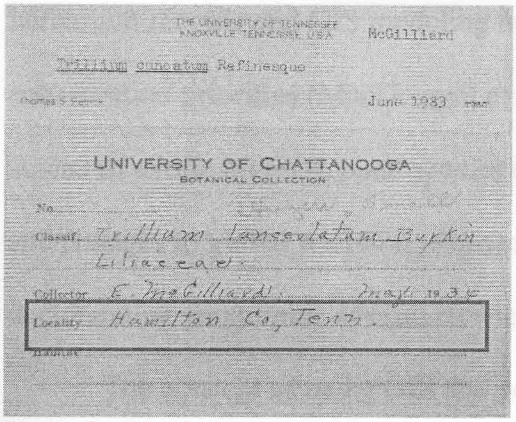


Figure 2a. Specimen with vague description.

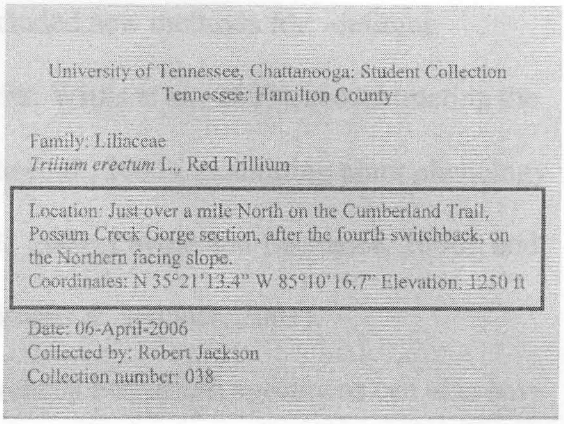


Figure 2b. Specimen with GPS coordinates and locality description.

Unfortunately, the vast majority of specimens do not contain geographic coordinate information on the specimen label. For these specimens, geographic coordinates can be generated by georeferencing the locality, county and state information given. Recent computer programs have been developed to utilize locality descriptions and calculate coordinates in batch format from a database file. These programs (e.g., GEOLocate and BioGeomancer) are relatively new and cannot guarantee a level of reasonable accuracy (See Part 2). Despite these challenges, databasing and georeferencing herbarium specimens is a task worthy of pursuit in order to allow for greater academic, professional, and public access to scientific botanical data. Access to the botanical data is very important to scientists who "…should consider the role of the herbarium not only [locally] at their institution but also on the international scene, and guidelines should be set up to insure maximum availability of information in herbaria both at present and in the future" (Radford et al., 1974, p.31).

GIS allow for herbarium data to be used for new scientific research and also for the greater efficiency of the traditional uses of herbaria. Recent studies using detailed herbarium records and GIS modeling have included new methods for: defining conservation priorities (MacDougall et al., 1998; Willis et al., 2003), reconstructing the spread of exotic invasive plant species (Delisle et al., 2003), examining plant phenology and changes to global warming (Primack et al., 2004; Lavoie and Lachance, 2006), and documenting geographic species richness (Hijmans & Spooner, 2001).

The concept of using GIS and georeferenced herbarium specimens can also have implications on molecular plant taxonomy in regards to conservation efforts of plant genetic resources. By using both GIS and results from molecular analyses of specimens, conservation efforts of plant genetic diversity can include identifying geographic areas of populations of high genetic diversity regarding particular species of concern. Subsequently, population areas at risk of declining genetic diversity, in need of more collecting for genetic analysis, and identifying the seeds and germplasm of species to conserve for biodiversity can also be determined (Guarino et al., 2002).

SERNEC has recognized the growing volume of scientific research being performed using improved herbarium datasets and was founded based upon the concept of increasing the use and access of southeastern herbaria. Goals of SERNEC include digitizing all the herbarium specimens in the southeast using standardized methods and combining these records into a single online, researchable collection. By helping to fund collaborative efforts such as SERNEC, the USGS National Biological Information Infrastructure (NBII) has continued to fulfill its mandate to increase access to the nation's biological information. Currently, SERNEC is working towards the integration of more partnering institutions, managing initial efforts for the creation of a single searchable database, and developing plans for large-scale databasing projects.

Part 1. Database design and standardization for UCHT specimens using Darwin Core elements

Abstract.

The University of Tennessee at Chattanooga herbarium (UCHT) contains approximately 20,000 specimens, primarily from southeastern Tennessee. In order to expand access to this valuable collection, databasing efforts were overhauled in late summer 2006. Serving as a pilot study for the larger goals of the Southeastern Regional Network of Expertise and Collections (SERNEC), the UCHT databasing project has allowed for a better understanding of the needs and best practices associated with the standardized databasing of natural history collections. A common and internationally proposed standard, Darwin Core, was utilized during this process and evaluated for its compatibility with the needs of herbaria. New fields for Darwin Core were determined and proposed to the larger herbarium community. The UCHT database includes over 50 searchable characteristic fields based on the Darwin Core format (e.g., taxonomic information, date collected, locality description, reproductive state, and geographic coordinates). The implementation of this standard was on a Microsoft Access platform using a customized entry form. During databasing, records of issues that arose and rates of entry were kept in order to inform other SERNEC partners about plausible timeframes and efforts needed to digitize herbaria collections. The potential of the digitized UCHT herbarium includes not only greater exposure and use for our own collection but also aides in the progress of SERNEC's effort to link and share data among all southeastern herbaria.

Introduction.

Before advanced scientific analysis of herbarium information, such as using GIS techniques, can be employed using a herbarium's specimens, the data inherent in the collection must first be captured in a digital format using standardization techniques. Darwin Core is a proposed standard currently being used by collection museums participating in the Global Biodiversity Information Facility (GBIF) and Biodiversity Information Standards (previously known as the Taxonomic Database Working Group (TDWG)) forums (GBIF website, 2007; Biodiversity Information Standards (a) website, 2007). The purpose of the Darwin Core standard is to create a universal information format for specimens in natural history collections that would include major biological, morphological, and geographical information for each record (Biodiversity Information Standards (b) website, 2007).

The implementation of Darwin Core, or any other standard, for use in natural history collections, relies heavily on the software used. Software needs include the following criteria: ease of use, customizable to the evolving standard, and the allowance for the query and export of data for use in research. There is a wide variety of software options; common software used for databasing collections has included Microsoft Excel and Access, Specify, and Index Kentuckiensis. The Specify software project is run by the Biodiversity Research Center at the University of Kansas; the software manages specimen data for all types of natural history collections and includes taxonomic fields for most disciplines (Specify website, 2007). Index Kentuckiensis is a database application based on a Microsoft Access platform. It contains a data entry interface as well as the capabilities to print labels for new accessions and create county-level distribution maps (Index Kentuckiensis website, 2007; Jones and Thompson, 2006). Both Specify and Index Kentuckiensis were evaluated against a customized Microsoft Access database. Evaluation of the software choices made by myself, Dr. Joey Shaw, and UCHT herbarium workers determined Microsoft Access as fitting our needs the best.

A database using Microsoft Access was used to encompass all existing and proposed fields of the Draft Botanical Extension to Darwin Core (See Appendix 2). The manual entry of the data from the herbarium specimens was performed by undergraduate workers with supervision by myself and Dr. Joey Shaw with continued review for purposes of data quality assurance.

The larger objective of our databasing effort was to contribute and serve as a pilot project for our collaboration with the Southeastern Regional Network of Expertise and Collections (SERNEC). The unique and primary purpose of SERNEC is to faciliatate and encourage the databasing activities of both large and small herbaria in the southeast. The future vision, and eventual goal, of SERNEC partners is to merge all collaborating herbaria datasets into a single, searchable online database for the southeast. Smaller herbaria such as UCHT are especially important in this collaborative digitization endeavor by SERNEC due to their strong regional focus and expertise which in turn will help fill in county-level distribution gaps for widely distributed species. In addition, UCHT will also be able to provide additional information on endemic plants of southeastern Tennessee previously overlooked or under-represented by larger herbaria outside of our region. With its smaller size, yet important regional accessions, UCHT makes an excellent herbarium to serve as a pilot project for the databasing goals of SERNEC via Darwin Core standardization.

Goals.

The overall objective of this part was to organize and digitize UCHT records into a database using Darwin Core standards and report back to SERNEC any issues that arose during this pilot test of herbarium databasing. To meet this objective, several goals were set including: 1) evaluation of software to use in databasing, 2) use the chosen software to digitize the UCHT collection via Darwin Core standards, 3) determine and propose new fields relevant to herbaria to be added into the Darwin Core standard.

Materials and Methods.

Previous efforts to digitize the UCHT collection were performed using Microsoft Excel with approximately 5,700 records entered in a non-sequential and non-standardized manner with limited information fields established. Previous information fields were restricted to: family, genus, species, infraspecific taxa (i.e. subspecies, variety), state, county, date, and collector. The non-standardized digitization of UCHT inhibited the distribution of the data associated with the collection's approximately 20,000 specimens to other scientists and SERNEC partners.

As of June 2006, efforts of databasing were overhauled to the Darwin Core standardization system. Multiple software programs were evaluated to determine the easiest and most efficient method of databasing the UCHT collection. Databasing software designed expressly for natural history collections, such as "Specify" and "Index Kentuckiensis", were tested and rejected due to issues of user-friendliness, time to train herbarium workers, and GIS compatibility. The use of Microsoft Access software was determined to be the best software for databasing due to user familiarity, easy customization of entry forms, and its compatibility to export database information into GIS software. Using Microsoft Access 2003, a database form was created to include the entire currently proposed Darwin Core Standards version 1.21 with Draft Botanical Extension (See Appendix 2). Additionally, as part of a pilot study we have added and proposed additional fields not yet included in official Darwin Core versions. These fields included: a "GPS" check box to distinguish between verbatim collecting coordinates and those georeferenced at a later date, "Native/Non-native" field to distinguish those records of non-native plants found in Tennessee, and "Species of Special Concern" field to identify those species that are state or federally listed as threatened or endangered.

During implementation of the new standardized digitization efforts, careful records of entry rates and potential problems and issues were kept in order to serve as information regarding this pilot project. The digitization of specimens via Darwin Core standards was performed over four semesters with nearly the complete UCHT collection databased.

Results.

During the first semester, between late August 2006 and early December 2006, five undergraduate students performed the standardized databasing, beginning with an initial file of 6,639 records from the non-standardized excel file that were added into the new Access form. Combined, the five students contributed to seven credit hours of work, correlating to 21 hours of databasing and herbarium work a week. Over approximately 3.5 months, a total of 2,859 accessions were digitized via Darwin Core into the database (Figure 1). During the course of the second semester, early January 2007 through late April 2007, three of the students returned to contribute six credit hours combined, equaling 18 hours of databasing and herbarium work a week. During this second semester, 3,897 accessions were digitized to the Darwin Core Standard (Figure 1).

The Access software chosen was deemed successful through easy integration of the database information into exportable files for use in our GIS analyses (See Parts 4 and 5). Responses to our proposed herbaria-specific additions to Darwin Core from the larger community will be ongoing and a consensus is not expected before the completion of this project.

Discussion.

Herbarium databasing efforts at UCHT involved evaluating several software packages with the final decision to use Microsoft Access. Although the employment of this software worked well for this pilot study, other institutions databasing their collections may find one of the other choices more suitable for their needs. A major determining factor for our decision to use Access was its easy ability to export data into a GIS system; herbaria not interested in future use of the data for GIS analyses may opt to use a different database platform.

Comparison of the rate of entry between the first two initial semesters showed an increased efficiency during the second semester databasing period. This increased efficiency can be attributed to both the experience of the returning herbarium workers as well as having identified and solved several formatting issues within the database.

This work, in serving as a pilot for the larger SERNEC goal of databasing the entire southeastern herbaria collections, has helped to give proof-of-concept to this organization's objectives and has shown the implications for future broader impacts for the scientific community. The eventual goal of a single, complete database of all herbaria specimens in the southeast, using standardized methods of databasing, will allow for a powerful source of information for future biological research. This future database can allow for questions on plant taxonomy, plant distributions and ecology, conservation, and natural resources to be answered more efficiently given the wealth of information detailed and recorded for each individual herbarium specimen. In addition, a benefit of the SERNEC organization's databasing initiative will be in the training of future botanists. With easily retrievable and searchable data on specimens, students can more easily access the important taxonomic and biologic information of plant species.

With the completion of databasing the current UCHT collection using Darwin Core standards, the exposure and use of UCHT can now be expanded. Additionally, our work will serve as an example to collaborating SERNEC institutions by aiding in the understanding of best practices, proposed standards, and software options associated with databasing a herbarium.

Literature Cited.

Biodiversity Information Standards (a). Retrieved 7/23/2007. (http://www.tdwg.org/)

Biodiversity Information Standards (b). Retrieved 7/23/2007. (http://www.tdwg.org/activities/Darwin Core/)

Global Biodiversity Information Facility (GBIF). Retrieved 7/23/2007. "General Fact Sheet". (http://www.gbif.org/press/factsheet).

Index Kentuckiensis Software (2007). Retrieved 1/2/2008. (http://people.eku.edu/jonesron/herbarium/default.php)

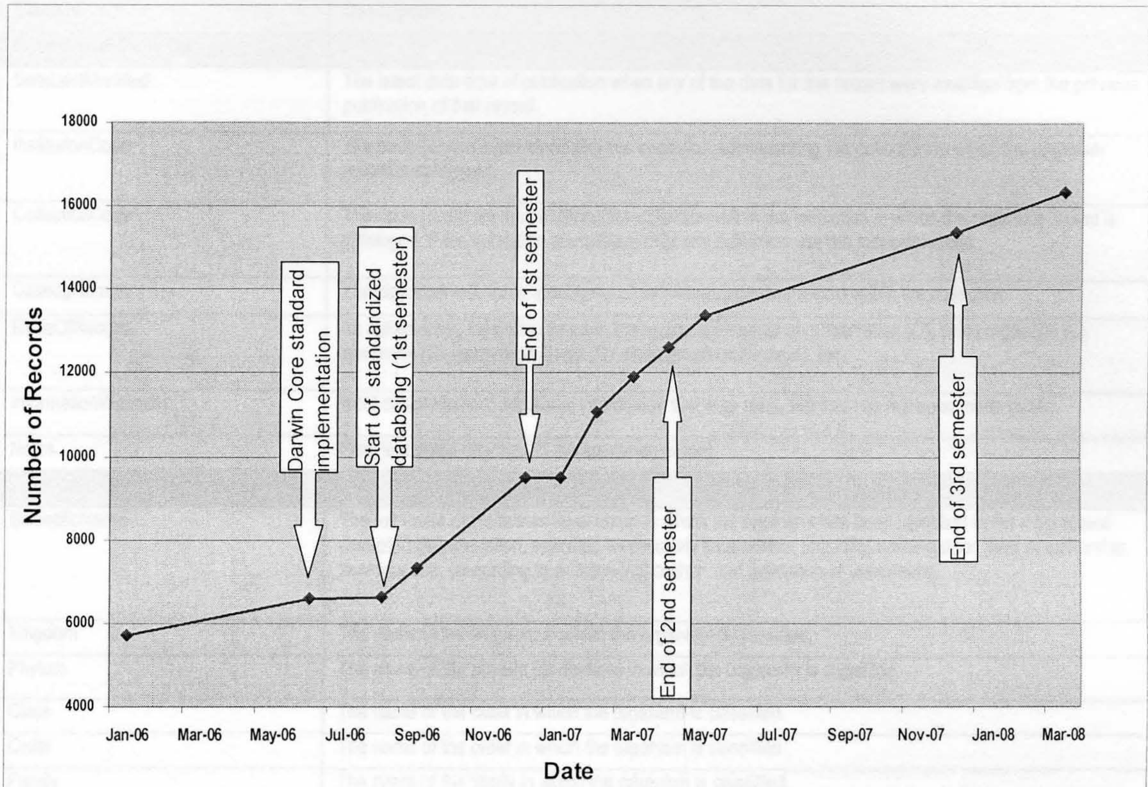
Jones, Ronald L. and Ralph L. Thompson. (2006) Index herbariorum kentuckiensis IV. Journal of the Kentucky Academy of Science 67: 87-92.

Southeastern Regional Network of Expertise and Collections. Retrieved 7/23/2007. (http://glaucon.sunsite.utk.edu/SERNEC/)

Specify Software Project. Retrieved 7/23/2007. (http://www.specifysoftware.org/Specify/)

Figures.

Figure 1. UCHT Digitization Efforts using Darwin Core Standardization Techniques.



Appendix.

Appendix 1. Darwin Core v. 1.21 with Draft Botanical Extension.

Element	Description
Record-level Elements	
DateLastModified	The latest date-time of publication when any of the data for the record were modified from the previous publication of that record.
InstitutionCode	The code (or acronym) identifying the institution administering the collection in which the organism record is cataloged.
CollectionCode	The code (or acronym) identifying the collection within the institution in which the organism record is cataloged. If the institution administers only one collection use the InstitutionCode.
CatalogNumber	The alphanumeric value identifying an individual organism record within the collection.
BasisOfRecord	An abbreviation indicating whether the record represents an observation (O), living organism (L), specimen (S), germplasm/seed (G), photograph (still image), etc.
InformationWithheld	Brief descriptions of additional information that may exist, but that has not been made public.
Notes	Free text notes attached to the specimen record.
Taxonomic Elements	
ScientificName	The full name of the lowest level taxon to which the organism has been identified in the most recent accepted determination, specified as precisely as possible, including name-author, year or authorship, sensu or sec. (according to or following) author, and indication of uncertainty.
Kingdom	The name of the kingdom in which the organism is classified.
Phylum	The name of the phylum (or division) in which the organism is classified.
Class	The name of the class in which the organism is classified.
Order	The name of the order in which the organism is classified.
Family	The name of the family in which the organism is classified.
Genus	The name of the genus in which the organism is classified.
Species	The specific epithet of the scientific name applied to the organism.
InfraspecificRank	The infraspecific rank (subspecies, variety, forma) of the Subspecies (InfraspecificEpithet).
Subspecies	The final infraspecific epithet of the scientific name applied to the object or observation.
ScientificNameAuthor	The author of the ScientificName and the year of publication, if known.
Hybrid	This is a named or unnamed hybrid (hybrid formula).
GraftChimaera	This is a named or unnamed graft chimaera.
NomenclaturalCode	The nomenclatural code under which the ScientificName is constructed.
IdentifiedBy	The name(s) of the person(s) who applied the ScientificName to the object or observation.
YearIdentified	The four digit year in the Common Era calendar in which the object or observation was identified as having the ScientificName.
MonthIdentified	The month in the Common Era calendar during which the object or observation was identified.
DayIdentified	The day of the month in the Common Era calendar during which the object or observation was identified.
TypeStatus	A list of one or more nomenclatural types (including type status and typified taxonomic name) represented by the object.
CollectorNumber	An identifying string applied to the object or observation at the time of collection. Serves as a link between field notes and the object or observation.
FieldNumber	An identifying string applied to a set of objects or observations resulting from a single collecting event.

Collector	The name(s) of the collector(s) of the original data for the object or observation.
YearCollected	The four digit year in the Common Era calendar in which the organism was collected from the field.
MonthCollected	The month of year in the Common Era calendar during which the organism was collected from the field.
DayCollected	The day of the month in the Common Era calendar during which the organism was collected from the field.
VerbatimCollectingDate	The verbatim original representation of the date (and time) information for the collecting event.
JulianDay	The ordinal day of the year (the number of days since December 31 of the previous year) on which the object or observation was collected.
TimeOfDay	The time of day the object or observation was collected from the field, expressed as decimal hours from midnight, local time (e.g., 12.0 = noon, 13.5 = 1:30pm).
Locality Elements	
ValidDistributionFlag	A flag to state that the locality information represents a valid distribution occurrence for the specimen.
ContinentOcean	The continent or ocean from which a specimen was collected.
Country	The full, unabbreviated name of the country or major political unit from which the organism was collected.
StateProvince	The full, unabbreviated name of the state, province, or region (i.e., the next smaller political region than Country) from which the organism was collected.
County	The full, unabbreviated name of the county, shire, or municipality (i.e., the next smaller political region than StateProvince) from which the organism was collected.
Locality	The description of the locality from which the organism was collected.
Longitude	The longitude of the location from which the organism or observation was collected, expressed in decimal degrees
Latitude	The latitude of the location from which the organism or observation was collected, expressed in decimal degrees.
GeodeticDatum	The geodetic datum to which the latitude and longitude refer. If not known, use "not recorded".
CoordinatePrecision	The upper limit of the distance (in meters) from the given latitude and longitude describing a circle within which the whole of the described locality lies. Use NULL where the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates).
Collecting Event Elements	
MinimumElevation	The minimum distance in meters above (positive) or below sea level of the collecting locality.
MaximumElevation	The maximum distance in meters above (positive) or below sea level of the collecting locality.
MinimumDepth	The minimum distance in meters below the surface of the water at which the collection was made; all material collected was at least this deep.
MaximumDepth	The maximum distance in meters below the surface of the water at which the collection was made; all material collected was at most this deep.
CollectingMethod	The name of, reference to, or brief description of the method or protocol under which the collecting event occurred.
Biological Elements	
Sex	The sex of a specimen.
LifeStage	The age class, reproductive stage, or life stage of the biological individual referred to by the record.
PreparationType	A concatenated list of preparations and preservation methods for the object.
IndividualCount	The number of individuals present in the lot or container. Not to be used for observations.
PreviousCatalogNumber	A list of previous or alternative fully qualified catalog numbers for the same object or observation, whether in the current collection or in any other.
References Elements	

DuplicateFrom	For duplicate specimens received from other institutions, the acronym of the sending herbarium.
RelatedCatalogItem	The fully qualified identifier of a related Catalog Item (a reference to another specimen); Institution Code, Collection Code, and Catalog Number of the related Cataloged Item, where a space separates the three subelements.
RelationshipType	A named or coded valued that identifies the kind relationship between this Collection Item and the referenced Collection Item.
ImageURL	A reference to digital images associated with the specimen or observation.

Part 2. EVALUATION OF METHODS FOR GEOREFERENCING DIGITIZED HERBARIUM RECORDS

Abstract.

A major barrier to herbarium specimens being used in more powerful scientific analyses, especially in GIS, is the lack of geographic coordinate information in specimen labels. Locality descriptions of varying quality and scale with the absence of geographic coordinates are the historical norm for a vast majority of existing collections – even with the recent expanded availability of personal GPS units. The process of assigning geographic coordinates to records based upon available locality descriptions is known as georeferencing. Georeferencing herbarium specimens will allow for ecological analyses to be performed on a more powerful scale and thus increasing a collection's value. Evaluation of current georeferencing methods, both automated software options (GEOLocate) and the traditional gazetteer method, were performed using recent field collected specimens that had GPS coordinates recorded during collection. The generated coordinates of each method were compared, via distance measurements in ArcGIS, to those of the actual GPS coordinates recorded during collection. Results of this testing showed GEOLocate had the benefits of low cost and time commitment but a large inaccuracy (>100 km). Traditional gazetteer georeferencing methods allowed for acceptable accuracy (<5 km) with considerable time investment and associated costs. As automated software continues to be developed and improved, it is imperative that a community consensus among herbaria curators is developed so that the development of georeferencing standards and methods meet the needs of high accuracy and low cost.

Introduction.

A vast majority of the UCHT specimens, approximately 95%, do not list geographic coordinates on label information because GPS technology is relatively new and its use is still not ubiquitous during botanical field collection (J. Shaw, University of Tennessee at Chattanooga, personal communication). Despite this, geographic coordinates can still be assigned to specimens through the process of georeferencing. Georeferencing is the process of taking textual descriptions of a location and assigning geographic coordinates to it. Currently, multiple academic and scientific organizations have unique georeferencing standards that their associated members are encouraged to follow (Chapman and Wieczorek, 2006; MaPSTeDI, 2007). In order for herbarium data to be used for higher resolution analyses using GIS technology, the geographic coordinates for a specimen's location is of great importance (Rhoads & Thompson, 1992; Murphey et al., 2004; Beaman and Conn, 2003). The county, state, and locality description of an accession can be used to find and estimate coordinates for an individual specimen using a gazetteer, standard USGS quad maps, digital maps on GIS software like ESRI ArcMap or Google Earth. Several online georeferencing sources also exist such as TopoZone (TopoZone website, 2007) and the USGS Geographic Names Information System (GNIS) (USGS website, 2007); these sources allow for place names to be queried and produce results of geographic coordinates assigned to them. The purpose of georeferencing is to allow for a more definitive, numerical location description (i.e. geographic coordinates and error estimate) of where the specimen was found so that scientists can revisit a chosen accession site as well as perform geospatial analyses using the results from georeferencing.

"The georeferencing process often increases locality precision. During georeferencing, locality descriptions are examined, evaluated and geospatial coordinates are assigned to them. When this process results in a reduction of the size of the geographic area encompassing all possible locations for the point (a decrease in the size of the potential area of geographic error), it increases the precision associated with the locality" (Murphey et al., 2004). Often, however, georeferencing individual specimens can be time consuming and therefore costly for large numbers of historical specimens.

Due to the time requirement to georeference herbarium specimens one at a time, recent software packages such as GEOLocate and online formats such as BioGeomancer Workbench have been developed to batch process databased collections. These two georeferencing programs use the locality text, county, and state data of a specimen to calculate geographic coordinates when they were not supplied by the specimen collector. BioGeomancer Workbench is a collaborative effort, organized by the University of California at Berkeley, consisting of collection museums and geospatial professionals. The program, scheduled for future release, uses the place names within the locality description as well as distance and direction from the place name to calculate the coordinates of a specimen (BioGeomancer website, 2008). The GEOLocate v. 2.03 software was developed by researchers at the Tulane Museum of Natural History. It is similar to BioGeomancer (converting locality, county and state data into latitude and longitude coordinates); however, it also offers a digital map display for verification, error estimate, and potential correction (Rios and Bart, 2003). This software is distributed at no cost through the Tulane Museum of Natural History (Tulane Museum of Natural History, 2007).

Goals.

The georeferencing software package GEOLocate was tested using a sample set of Dr. J. Shaw's personal North American plum collection records (genus *Prunus;* Rosaceae). These 130 records, collected from throughout the eastern U.S. between 1999-2004, contain relevant information useful to this type of evaluation including specific locality descriptions, county and state information, as well as geographic coordinates obtained with a GPS unit in the field during collection. The records were used without the coordinates and entered into GEOLocate; distance measurements of the calculated coordinates to the actual field collected coordinates were done in ArcMap. As a comparison, the manual method of using a gazetteer to estimate coordinates based on the same data was also performed to test the accuracy of this traditional, non-automated method compared to the GEOLocate software package. The gazetteer method and GEOLocate were evaluated with the goal to see which method(s) can be realistically used to give accurate coordinates for historical herbarium records.

Materials and Methods.

<u>*GEOLocate v. 2.03 Evaluation*</u>.

A spreadsheet file was created containing all the original information on the Shaw *Prunus* records. A copy of this file, minus the given geographic coordinates, was used in the evaluation of GEOLocate software. The file was formatted according to GEOLocate instructions for batch processing. Columns and subsequent information in this file were: locality string, county, state, country as well as columns for the generated results such as latitude, longitude, correction status, precision score, error polygon array and multiple results array. This file was imported into the GEOLocate software for the generation of geographic coordinates via batch georeferencing and results were recorded. Of the 130 records, 87 were able to be processed successfully by GEOLocate. Results were then imported into the GIS software ESRI ArcMap along with the actual field coordinates to measure the distance between the "true" field coordinates and those generated by GEOLocate. The average distance between actual and generated coordinates for these 87 records allow for an interpretation and evaluation of the accuracy of GEOLocate.

Gazetteer Method Evaluation.

The same *Prunus* file of 87 records with field coordinates removed used in the previous assessment was employed in the manual georeference method (often referred to as the "gazetteer method") for comparison purposes. The information utilized during this evaluation included locality string, county, state, and country. Several map sources, both electronic and print, were used to match the given location information; sources used for each record varied depending on the ability to match locations with confidence. These sources included printed state atlases and gazetteers, online maps via MapQuest (cite website) and Google Earth (cite software), and ESRI ArcMap document "StreetMap USA" (cite ESRI's Streetmap). The manually generated coordinates via this method were recorded and loaded into ESRI ArcMap along with actual field coordinates. Distance between the method generated and field coordinates was again measured as mentioned above.

Results.

Of the 130 records initially entered, GEOLocate was able to successfully return results for 87. Results of the 87 records gave an average error distance for GEOLocate generated coordinates of 101.85 km. The minimum error was 0.065 km and the maximum was 944 km (See Table 1). An important observation of GEOLocate results was the occurrence of generated points for records outside of the county stipulated in the data file. Testing results of the manual gazetteer method showed it to be much more accurate. The results showed an average error of 4.82 km, with a minimum error of .007 km and a maximum error of 56 km (See Table 2).

Discussion.

The decision to georeference herbarium specimens that lack geographic coordinates depends on several factors including: time and associated costs to georeference as well as the accuracy calculated coordinates. The evaluation of georeferencing methods determined that GEOLocate software produces an average error well over an acceptable level for most ecological analyses (average distance over 100 km of actual location). In regards to assigning geographic coordinates for each record in a natural history collection, such as the UCHT herbarium, georeferencing via GEOLocate batch processing would allow for rapid coordinate generation at virtually no cost but produce a low accuracy for most specimens. Comparatively, the manual method produced more accurate results (average within 5 km of actual location); however, the manually georeferencing specimens can be time consuming. It was determined that approximately 15 records could be georeferenced by the gazetteer method on average. A collection the size of UCHT (approximately 20,000 accessions) could take up to 1,000 hours to fully georeference manually. This commitment of time and associated costs is a major deterrent to most herbaria with limited resources. The rate observed in this evaluation of the gazetteer method could be faster, however, if records are pre-sorted by location to allow for more efficient use of map sources; in addition, the gain of experience over time with this method will also lead to faster processing.

During the evaluation period common problems related to the information on locality labels and georeferencing were observed. These problems included 1) ambiguous locality descriptions (e.g. 'near Smith farm'), 2) linear features and place names (e.g. rivers or roads), 3) changing political boundaries and duplicate names for a single place, 4) historical place names no longer used, 5) broad descriptions (e.g. 'Hamilton county'), 6) misspellings, and 7) contradictory descriptions (e.g. 'Chattanooga, TN; Marion county'). The wide variety of quality and information found on herbarium labels and the problems associated with this variety can lead to the calculation of inaccurate coordinates during georeferencing.

Results of this analysis have shown the high inaccuracy of automated georeferencing software currently available and the significant commitment of time and money to manually georeference herbaria. Future decisions for UCHT and other herbaria to georeference their collections will depend on advancements in automated georeferencing software and/or funding resources.

Literature Cited.

Beaman, Reed S. and Barry J. Conn. (2003) Automated geoparsing and georeferencing of Malesian collection locality data. Telopea 10: 43-52.

BioGeomancer Project. (Retrieved 7/1/2006) BioGeomancer Software. (http://www.biogeomancer.org/software.html).

Chapman A.D., and J. Wieczorek (eds.). 2006. Guide to Best Practices for Georeferencing. Global Biodiversity Information Facility: Copenhagen.

ESRI. (2006) StreetMap USA. (http://www.esri.com/data/streetmap/about/overview.html). Redlands, California

Mountains and Plains Spatio-Temporal Database Informatics Initiative (MaPSTeDI). (Retrieved 7/25/2007) "Georeferencing protocols". (http://mapstedi.colorado.edu/georeferencing-protocols.html).

Murphey, Paul C., Robert P. Guralnick, Robert Glaubitz, David Neufeld, and J. Allen Ryan. (2004) Georeferencing of museum collections: a review of problems and automated tools, and the methodology developed by the Mountain and Plains Spatio-Temporal Database-Informatics Initiative (Mapstedi). PhyloInformatics 3: 1-29.

Rhoads, A.F. and L. Thompson. (1992) Integrating herbarium data into geographic information system: requirements for spatial analysis. Taxon 41: 43-49.

Rios, Nelson E. and Henry L. Bart. (2003) "GEOLocate User's Manual". Tulane Museum of Natural History: Belle Chase, LA.

TopoZone. (Retrieved 7/24/2007). (http//:www.topozone.com).

Tulane Museum of Natural History. (Retrieved 7/20/2007) "Overview". (http://www.museum.tulane.edu/geolocate/default.aspx)

United States Geological Survey. (Retrieved 7/24/2007) "Geographic Names Information System". (http://geonames.usgs.gov/pls/gnispublic).

Tables.

Table 1. Comparison of generated GEOLocate coordinates to field GPS coordinates.

Record	JSH#	Field_Lat	Field_Lon	GeoLocate_Lat	GeoLocate_Lon	Measured_Distance
1	773	30.36289900	-84.36968923	30.36167000	-84.36167000	1.84
2	776	31.55285960	-83.52585034	31.53736700	-83.50861000	1.92
3	779	31.96938160	-83.29321138	34.92278000	-85.41528000	239.83
4	784	34.15395082	-85.21640947	34.05402200	-85.25500000	4.68
5	786	34.40019085	-85.21008625	34.25694000	-85.16472000	5.42
6	809	35.99507458	-84.26965674	35.98944000	-84.27389000	0.48
7	810	35.99421141	-84.27132198	35.98944000	-84.27389000	0.30
8	811	36.08076889	-83.89971726	35.95041400	-84.19953400	33.85
9	815	35.89726992	-84.72232821	35.86306000	-84.75416900	3.64
10	823	36.48360324	-85.50974021	36.48417000	-85.50417000	0.63
11	824	36.48288659	-85.50842584	36.48417000	-85.50417000	0.53
12	829	36.10753935	-84.81126197	36.10250000	-84.80720000	0.48
13	831	35.75358247	-83.46154086	36.38889000	-88.35917000	546.69
14	833	35.75155665	-83.43565545	35.88417000	-85.00278000	174.09
15	834	37.91311730	-80.20356886	37.91750000	-80.21306000	1.08
16	835	37.87357399	-80.23048826	39.25167000	-79.61722000	73.47
17	836	39.68206767	-78.45085281	38.22028000	-75.58944000	323.38
18	837	39.73314799	-78.34821548	39.72367400	-78.33020900	2.03
19	846	40.77727832	-77.95599271	40.79972000	-77.94639000	1.21
20	847	40.77983304	-77.95286031	40.79972000	-77.94639000	0.90
21	848	40.77986271	-77.95286341	40.79972000	-77.94639000	0.91
22	853	42.87211130	-73.02005706	43.41667000	-72.91667000	21.55
23	861	42.85327834	-72.83245808	42.85044800	-72.85473400	2.54
24	862	42.86379235	-72.75141246	43.87639000	-72.17694000	72.74
25	868	42.94680496	-72.64105137	42.94667000	-72.67944000	4.28
26	869	42.90591778	-72.41696748	43.52417000	-71.70139000	82.54
27	874	42.86689910	-70.81785793	42.84306000	-70.81806000	0.88
28	875	42.79617604	-70.83998867	42.81250000	-70.87778000	4.23
29	876	41.73779162	-70.38184390	41.73944000	-70.37917000	0.31
30	878	42.06493453	-70.16294179	42.05833000	-70.17917000	1.86
31	882	41.14182464	-72.31767143	41.16028000	-72.23417000	9.33
32	884	35.72895175	-85.38957305	35.66583000	-85.35583000	3.80
33	904	28.75413182	-81.87875482	28.81056000	-81.87806000	0.89
34	905	28.78000927	-81.88790912	28.81056000	-81.87806000	1.21
35	909	34.30832854	-85.85606695	34.26389000	-85.86056000	0.70
36	910	32.80048954	-88.02665172	31.44417000	-86.61306000	158.82
37	911	32.72476114	-88.11519412	31.44417000	-86.61306000	164.37
38	913	32.12513488	-88.87583985	32.01071900	-88.48056000	43.96
39	916	30.11348030	-95.94931872	31.29667000	-104.35583000	941.96
40	918	30.18328588	-95.92658825	31.29667000	-104.35583000	944.79
41	919	29.87434560	-96.19824163	29.88218800	-96.15694000	4.65
42	920	29.86955401	-96.19500001	29.88218800	-96.15694000	4.26
43	922	28.70880333	-97.17322653	31.20333000	-97.02960200	37.66

Record	JSH#	Field_Lat	Field_Lon	GeoLocate_Lat	GeoLocate_Lon	Measured_Distance
44	923	28.70799146	-97.17693309	31.20333000	-97.03469300	38.31
45	927	29.27492084	-98.31989846	31.64361000	-96.85861000	167.07
46	928	29.87053880	-98.40918895	31.80222000	-97.09139000	150.16
47	929	30.00907464	-99.11680455	30.04722000	-99.14000000	2.80
48	932	30.74789825	-99.22717419	30.26694000	-97.74278000	166.84
49	933	30.74635338	-98.41188591	30.73932000	-98.43083000	2.14
50	934	30.75018198	-98.17427102	30.53610800	-94.84608000	371.57
51	935	30.52802660	-97.62081027	31.16694000	-95.61817500	225.81
52	936	30.71951394	-96.98201453	30.65528000	-97.00111000	2.33
53	937	30.71570420	-96.98241301	30.65528000	-97.00111000	2.22
54	938	30.69950759	-96.98281752	30.65528000	-97.00111000	2.08
55	939	30.68202211	-96.94127815	31.80222000	-97.09139000	22.50
56	940	30.72141152	-96.83227041	31.80222000	-97.09139000	32.54
57	941	30.99348793	-96.52475003	34.14417000	-102.62442900	688.04
58	943	32.59272790	-93.80437327	32.43067200	-93.75000000	6.16
59	944	32.77319253	-93.84529363	32.81823300	-93.84500000	0.34
60	945	32.86174449	-93.87356043	32.88604900	-93.87833000	0.56
61	948	34.81959497	-92.33886237	34.74639000	-92.28944000	5.24
62	952	35.78592706	-88.40193601	36.27361000	-86.89556000	168.51
63	953	36.06853123	-83.73866628	36.06111000	-83.74583000	0.81
64	954	36.06250087	-83.68230770	35.64639000	-84.03778000	40.19
65	955	36.02268322	-83.76261505	35.30167000	-88.33233400	513.30
66	892	36.08077291	-83.89953721	36.03194000	-83.93750000	4.19
67	959	34.90872670	-85.10968362	34.30694000	-83.41167000	190.09
68	960	35.72278225	-84.82709111	36.56056000	-82.29000000	284.29
69	962	36.26051151	-82.18652909	36.28399400	-82.23189100	5.03
70	964	36.26042409	-82.21241467	35.80917000	-88.54508600	709.80
71	965	36.27798703	-82.31842194	36.27546100	-82.31889000	0.07
72	967	36.41583455	-82.49037081	36.51666400	-82.49543500	1.58
73	971	36.56387807	-85.52223050	36.56288700	-85.51428200	0.89
74	979	44.56300487	-73.00074868	44.61194000	-73.00917000	1.85
75	993	36.63919936	-84.56031734	36.47201700	-84.66890100	12.46
76	994	36.63351978	-84.55029921	36.45000000	-84.77917000	25.80
77	995	36.63358969	-84.55239334	36.45000000	-84.77917000	25.64
78	996	36.65314752	-84.55757847	36.45000000	-84.77917000	25.82
79	997	36.65165906	-84.55519893	36.45000000	-84.77917000	25.21
80	1004	43.43431262	-89.73230422	43.41694000	-89.70361000	3.21
81	1005	43.43094628	-89.70168823	43.41694000	-89.70361000	0.25
82	1016	33.89400839	-81.04232982	33.61556000	-81.10222000	9.59
83	1018	33.46360395	-80.55038335	33.61556000	-81.10222000	60.80
84	1022	34.22995498	-81.43633107	33.61556000	-81.10222000	38.87
85	966	36.50666000	-82.41638800	36.31333000	-82.35361000	38.81
86	885	36.02198333	-83.77105000	36.27361000	-86.89556000	350.00
87	886	36.02375000	-83.75515000	35.30167000	-88.32522500	512.37
					AVERAGE	**101.85 Km**

Table 2. Comparison of generated gazetteer coordinates to field GPS coordinates.

Record	JSH#	Field_Lat	Field_Lon	Gazetteer_Lat	Gazetteer_Lon	Measured_Distance
1	773	30.36289900	-84.36968923	30.35704100	-84.36592900	0.43
2	776	31.55285960	-83.52585034	31.54971700	-83.54618700	2.25
3	779	31.96938160	-83.29321138	31.96780200	-83.29165100	0.20
4	784	34.15395082	-85.21640947	34.15477500	-85.21622300	0.02
5	786	34.40019085	-85.21008625	34.30211900	-85.17605900	3.98
6	809	35.99507458	-84.26965674	35.99376600	-84.27094800	0.15
7	810	35.99421141	-84.27132198	35.99376600	-84.27094800	0.04
8	811	36.08076889	-83.89971726	36.03845800	-84.01290900	12.69
9	815	35.89726992	-84.72232821	35.89377100	-84.73433200	1.36
10	823	36.48360324	-85.50974021	36.54534700	-85.49774100	1.47
11	824	36.48288659	-85.50842584	36.54534700	-85.49774100	1.60
12	829	36.10753935	-84.81126197	35.96793200	-85.02001200	23.23
13	831	35.75358247	-83.46154086	35.75767100	-83.44002200	2.45
14	833	35.75155665	-83.43565545	35.74807100	-83.43432800	0.14
15	834	37.91311730	-80.20356886	37.91563700	-80.20648800	0.33
16	835	37.87357399	-80.23048826	37.80723500	-80.28654400	6.40
17	836	39.68206767	-78.45085281	39.64937400	-78.75998300	34.61
18	837	39.73314799	-78.34821548	39.74168800	-78.36296900	1.66
19	846	40.77727832	-77.95599271	40.80756700	-77.94476800	1.42
20	847	40.77983304	-77.95286031	40.80375200	-77.94300900	1.32
21	848	40.77986271	-77.95286341	40.79020300	-77.95093800	0.32
22	853	42.87211130	-73.02005706	42.87450100	-73.05599700	4.02
23	861	42.85327834	-72.83245808	42.85498000	-72.81473700	1.98
24	862	42.86379235	-72.75141246	42.86383100	-72.75202800	0.08
25	868	42.94680496	-72.64105137	42.94685600	-72.64053400	0.06
26	869	42.90591778	-72.41696748	42.90563600	-72.41729100	0.04
27	874	42.86689910	-70.81785793	42.87133100	-70.82089800	0.37
28	875	42.79617604	-70.83998867	42.79726000	-70.82184200	2.09
29	876	41.73779162	-70.38184390	41.73938000	-70.38394400	0.24
30	878	42.06493453	-70.16294179	42.02732800	-70.07644800	9.74
31	882	41.14182464	-72.31767143	41.14090500	-72.31873900	0.13
32	884	35.72895175	-85.38957305	35.65765800	-85.35099600	4.37
33	904	28.75413182	-81.87875482	28.75209600	-81.87955700	0.10
34	905	28.78000927	-81.88790912	28.78009900	-81.88807300	0.02
35	909	34.30832854	-85.85606695	34.26761300	-85.86178100	0.68
36	910	32.80048954	-88.02665172	32.71896000	-88.13067100	11.61
37	911	32.72476114	-88.11519412	32.71416200	-88.13378700	2.06
38	913	32.12513488	-88.87583985	32.32941200	-88.77273300	12.48
39	916	30.11348030	-95.94931872	30.00355200	-95.71436500	26.44
40	918	30.18328588	-95.92658825	30.45230000	-95.94934400	2.53
41	919	29.87434560	-96.19824163	29.86399400	-96.19147200	0.77
42	920	29.86955401	-96.19500001	29.86399400	-96.19147200	0.39
43	922	28.70880333	-97.17322653	28.69630200	-97.23838100	7.28

Record	JSH#	Field_Lat	Field_Lon	Gazetteer_Lat	Gazetteer_Lon	Measured_Distance
44	923	28.70799146	-97.17693309	28.69630200	-97.23838100	6.92
45	927	29.27492084	-98.31989846	29.27385900	-98.32026900	0.05
46	928	29.87053880	-98.40918895	29.70314100	-98.12482100	32.25
47	929	30.00907464	-99.11680455	30.00480300	-99.12717100	1.18
48	932	30.74789825	-99.22717419	30.74911900	-99.23165200	0.52
49	933	30.74635338	-98.41188591	30.75791700	-98.67503800	29.65
50	934	30.75018198	-98.17427102	30.75791700	-98.67503800	56.68
51	935	30.52802660	-97.62081027	30.51886800	-97.67482000	5.69
52	936	30.71951394	-96.98201453	30.71903700	-96.98127100	0.08
53	937	30.71570420	-96.98241301	30.71903700	-96.98127100	0.14
54	938	30.69950759	-96.98281752	30.70731700	-96.98207900	0.14
55	939	30.68202211	-96.94127815	30.68192100	-96.93787200	0.31
56	940	30.72141152	-96.83227041	30.72397800	-96.81913300	1.50
57	941	30.99348793	-96.52475003	30.87620900	-96.59162500	7.32
58	943	32.59272790	-93.80437327	32.51302700	-93.73665100	7.45
59	944	32.77319253	-93.84529363	32.81414200	-93.85061800	0.85
60	945	32.86174449	-93.87356043	32.83923800	-93.85417800	2.18
61	948	34.81959497	-92.33886237	34.82097100	-92.33861700	0.01
62	952	35.78592706	-88.40193601	35.78907000	-88.39157600	1.17
63	953	36.06853123	-83.73866628	36.06586700	-83.73035600	0.94
64	954	36.06250087	-83.68230770	36.06261200	-83.68237800	0.01
65	955	36.02268322	-83.76261505	36.02309300	-83.75991900	0.30
66	892	36.08077291	-83.89953721	36.05819300	-83.92416800	2.73
67	959	34.90872670	-85.10968362	34.92767400	-85.15245600	4.81
68	960	35.72278225	-84.82709111	35.72273400	-84.82790500	0.09
69	962	36.26051151	-82.18652909	36.26111400	-82.18624200	0.04
70	964	36.26042409	-82.21241467	36.26111400	-82.18624200	2.92
71	965	36.27798703	-82.31842194	36.27331000	-82.31844300	0.07
72	967	36.41583455	-82.49037081	36.43469600	-82.48544200	0.62
73	971	36.56387807	-85.52223050	36.56481600	-85.52236700	0.02
74	979	44.56300487	-73.00074868	44.60531200	-73.00848000	1.63
75	993	36.63919936	-84.56031734	36.47541300	-84.68605400	14.01
76	994	36.63351978	-84.55029921	36.47095500	-84.63288600	9.33
77	995	36.63358969	-84.55239334	36.47095500	-84.63288600	9.16
78	996	36.65314752	-84.55757847	36.47095500	-84.63288600	8.62
79	997	36.65165906	-84.55519893	36.47095500	-84.63288600	8.92
80	1004	43.43431262	-89.73230422	43.43441000	-89.72880100	0.41
81	1005	43.43094628	-89.70168823	43.43309000	-89.75515200	5.99
82	1016	33.89400839	-81.04232982	33.89255800	-81.04241400	0.03
83	1018	33.46360395	-80.55038335	33.46317800	-80.53882900	1.28
84	1022	34.22995498	-81.43633107	34.26246100	-81.50355300	7.52
85	966	36.50666000	-82.41638800	36.36758700	-82.41396300	2.07
86	885	36.02198333	-83.77105000	36.02224100	-83.77076400	0.03
87	886	36.02375000	-83.75515000	36.02400300	-83.75200300	0.35
					AVERAGE	**4.87 Km**

Part 3. Creation of GIS infrastructure for SERNEC website

Abstract.

As progress continues toward a uniform databasing standard for collaborating SERNEC institutions, the need for an advanced system of access and delivery for this information is needed in order to expand the use of the data. The work performed to satisfy this need focused on developing a website that hosts all of the currently databased SERNEC records. An important aspect of this pilot website is the use of a GIS-based infrastructure because location and distribution data is of common interest in most ecological plant studies. The newly improved UCHT database (See Part 1) and a preliminary SERNEC dataset were combined in data management software (MS SQL Server 2005). The subsequent integration and online delivery of spatial information was achieved with ESRI's ArcGIS Server. This pilot website, with a GIS-based infrastructure and spatial applications, is intended to serve as a proof-of-concept for the goals of SERNEC as well as a platform from which additional applications can be developed for research-oriented queries on herbarium specimens.

Introduction.

As the number of GIS applications continues to expand in biological research and the use of such software becomes more commonly used, herbaria have begun to incorporate web-based mapping as a way to show the distributions and ranges of plant species. It has been noted that distributing and sharing museum collections over the internet is "...improving science, reducing costs by providing for more efficient and effective biological survey, freeing up scientists to spend more time on research, and leading to a more rapid build-up of knowledge of our environments leading to its improved conservation and sustainable use" (Chapman, 2005). A survey of collaborating SERNEC herbaria shows a wide variety of approaches to presenting county-level distribution maps based upon their collections. Typically, the larger state universities have the most sizable collections and their websites are commonly relied upon for general inquiries regarding species distributions within their home state. For example, the University of Tennessee at Knoxville herbarium (TENN) hosts a website from which county-level distribution of species are shown within the boundary of Tennessee (http://tenn.bio.utk.edu/index.html). Users can query a species and retrieve results based upon an associated, but static, GIF (Graphics Interchange Format) image file (Figure 1). These distribution map images are based upon both specimens at TENN as well as those at Austin Peay State University (APSU) and Vanderbilt University (VDB). Although based upon a substantial number of specimens (over 500,000) and beneficial to many botanists, this system of data delivery has characteristics that limit more useful information from being accessed. Static county-level map images do not allow users to query for collection date, collector, what life stage the specimen was collected in (vegetative, budding, flowering, etc.), or more specific geographic information regarding the collection location (place name, geographic coordinates, etc.). Significant time is also spent by herbarium staff on updating the map images on an individual species basis. Arguments for including data from smaller herbaria within the state can also be made in order to create more accurate distribution maps.

By incorporating GIS technology, the herbarium at the University of North Carolina (NCU) has developed an advanced website that enables users to query a database comprising currently databased specimens from both NCU and North Carolina

State University (NCSC), database records from the North Carolina Natural Heritage Program, the USDA Plants Database, and several literature sources (http://www.herbarium.unc.edu/seflora/firstviewer.htm). User searches based upon taxonomic concepts are queried against the complied database and dynamic results are generated for the southeastern region onto a county-level map. Results of county-level distributions for species are highlighted based upon a color-coded scheme showing what type of data source was referenced (Figure 2).

Work by many other southeastern herbaria has been performed incorporating their own databases with varying degrees of custom search capabilities and result formats. Despite significant progress in individual institutions increasing the access to their collections, there has yet to be an attempt at combining all the current existing SERNEC datasets, each with its varying database standards, into a single standardized, searchable website with server-based geoprocessing by GIS software. The efforts described in this Part detail the standardization of multiple SERNEC datasets through database management software and the design of a GIS-based infrastructure supporting a pilot website for collaborating herbaria.

Materials and Methods.

<u>*Standardization and Merging of Initial SERNEC Datasets.*</u>

A request was sent to SERNEC institutions asking for a copy of any available files containing digitized herbarium records for use in creating a pilot website. In addition to UCHT, six additional herbaria responded: Eastern Kentucky University (EKY), University of Mississippi (MISS), Troy University (TROY), North Carolina State University (NCSC), University of North Carolina, Chapel Hill (NCU), and the University of Tennessee at Knoxville ferns collection (TENN). These seven files consisted of a total of 147,394 specimen records. Although the UCHT database had been previously formatted to the Darwin Core standard (see Part 1), each of the other herbaria had its own method and standard of databasing. For example, some institutions had a single column containing genus, species, and scientific author information while others separated this information into individual columns. Another problem encountered was the diversity of label data entered; few institutions had databased entire locality descriptions from the labels while most entered only place name or county. The format discrepancies between files were corrected by transferring and modifying the initial datasets into the Darwin Core standard for each institution using Microsoft Access 2003. After completion of Darwin Core standardization, the Access files were imported into Microsoft SQL Server 2005 for merging all the files into a single SERNEC database.

Website Development Software.

The Darwin Core–based textual information of the UCHT and SERNEC datasets were loaded onto MS SQL Server 2005. This software was used for its ability to managing, edit, and share large datasets in conjunction with other software programs and applications. This ability to distribute the data into diverse software programs will ensure that the dataset is compatible for enterprise-based scaling. Herbarium datasets were loaded into SQL Server using the DTS Wizard to import the initial Access files into a separate database within the SQL Server Management Studio. These tables were then edited with a table containing geospatial data on all counties within the United States. The counties table contained FIPS (Federal Information Processing Standards) code numbers. This unique 5 digit code number was used to connect the county and state information in the herbarium tables to the county geospatial information contained within the counties table. The purpose of this connection was to allow for the correct county level distribution to be displayed on the website when a specific species query was performed by a user. This "FIPS code connection" was performed within SQL Server Management Studio by creating a new View. Views generate results in a table format dynamically rather than storing data within a permanent table as in SQL Table files. Views within SQL Server are created by scripting a query that joins multiple tables that are joined by at least one common field. For this work several views were created; for the first view, the common fields involved the county and state columns and the view was scripted using a join clause (Figure 3).

Integration of a spatial database engine (SDE) was established using ArcSDE 9.2 through the creation of a new SDE view using command prompt scripting (Figure 4). This was performed in order to connect the SDE, which stores the vector polygons and spatial information for the needed counties and states, to the SQL server which stores all the textual and numeric data of the SERNEC dataset. This method allowed for the standardized SERNEC database to be supplemented with geospatial information for use in the creation of a pilot SERNEC website using ArcGIS Server 9.2 (ESRI, 2006a).

By using an ArcGIS 9.2 (ESRI, 2006b) service based upon a single .mxd file containing the SERNEC dataset, a pilot web mapping service (WMS) was created. The latest ESRI product designed to deliver GIS functions to a web service is ArcGIS Server version 9.2 (ESRI, 2006a) which includes ArcSDE technology. ArcGIS Server 9.2 was used to for the benefit of having a single software platform for the deliverance of GIS applications and services during website development and publishing. The SERNEC dataset in SQL Server, connected to the geospatial data in ArcSDE, was transferred into ESRI ArcMap as an .mxd file. This .mxd file was then published as a service within ArcGIS Server 9.2 through ArcGIS Server Manager. This specified service within ArcGIS Server Manager was then used during application development. Customized application tasks that were established included searches based upon genus, species, keyword, and county. Each search task was directed back to concurrent information column(s) within the SERNEC database and results were displayed on a map of the southeast dynamically, based upon the created FIPS code join, on a county level basis (Figure 5). In addition, results of the actual textual information (in Darwin Core format) for each specimen returned during searches could be accessed by clicking on the "Result Details" tab. To enhance the performance speed of the website, a map cache was created in ArcCatalog (ESRI, 2006c) using the Generate Map Server Cache tool. This allowed for pre-rendered map tiles to be displayed faster for the map service because they are no longer rendered on the fly after a new search is performed.

Multiple tests were performed to ensure each customized search was operating correctly and relevant results were returned. Initial problems were identified during the testing phase and changes were made to applications within ArcGIS Server Manager as needed in order to correct search and/or result issues. After all applications of the site were working correctly, evaluations focusing on user-friendliness were performed. The same techniques described in this section were applied to the UCHT database so that a separate website could be established for the UTC Department of Biological and Environmental Science.

Results & Discussion.

Using the methods described above, a successful pilot SERNEC website was created using a GIS-based infrastructure. Searches on the standardized, combined SERNEC database were able to be performed with county-level distribution results. However, evaluations of the pilot website generated three main concerns about the limitations of "out-of-the-box" editing functionality using ArcGIS Server. The first concern was in regards to the inability of the software to automatically display all county-level results for a search at once; using automated, non-scripted applications of ArcGIS Server created a scenario where each record in the "Results" tab had to be selected before the appropriate county would be highlighted. This creates a cumbersome situation for those interested in a quick distribution map of a species. A second concern was in regards to default result displays where numerous records were returned for commonly collected species; users can be confronted with results showing a long list of several hundred specimens of a species rather than an organized tab which would allow for a table-formatted view of results that could be downloaded or printed in an efficient manner. A third obstacle is that the out-of-the-box editing of search applications within ArcGIS Server are based upon single search criteria. The full benefits of the Darwin Core standardization can only be realized if complex/compound searches are available. For example, researchers will be more likely to use the website if the fields of species, month, life stage, locality, etc. are able to be queried simultaneously to answer ecological questions (e.g. "In what counties does *Cercis canadensis* flower in March as opposed to April").

These three concerns are all the result of the out-of-the-box editing functionality of ArcGIS Server 9.2; without the use of customized editing using Java or .NET scripting the pilot website offers obstacles to user-friendliness and the ability of users to fully benefit from the Darwin Core standardized database. Because the advanced techniques of using programming language was unknown to the author, research was done on the ESRI Developer Network (http://edn.esri.com/) to see if pre-existing documentation or scripts were available to solve our three concerns. Unfortunately, no relevant information was able to be obtained for our customized scripting needs. Due to factors of time and the scope of this thesis, it was not feasible for the author to gain enough experience with Java in order to perform these needed customizations. As of publication of this thesis, efforts are being made for a person with Java experience to assist in the further development of this pilot website.

While concerns have been identified with this pilot website, the goal of developing a GIS-based infrastructure for the collaborations of SERNEC has been accomplished. This will act as a proof-of-concept to the larger herbaria community that large, standardized databases can be combined and allow for the geospatial information of specimens to be accessed within a GIS environment. Future customized programming will ensure that concerns of user-friendliness and the lack of complex queries are addressed so that SERNEC may begin to move out of the pilot stage of its web presence into an official website that scientists and educators can use on a greater scale.

Literature Cited.

Chapman, A.D. (2005) Uses of Primary Species-Occurrence Data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.

ESRI. (2006a) *ArcGIS Server 9.2*. Redlands, California.

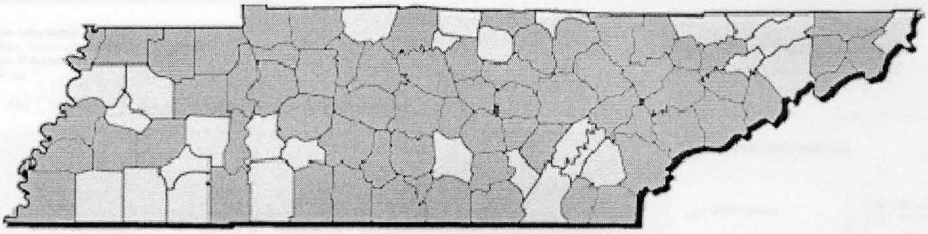
ESRI. (2006b) *ArcGIS 9.2*. Redlands, California.

ESRI. (2006c) *ArcCatalog 9.2*. Redlands, California.

Microsoft. (2005) *Microsoft SQL Server 2005*. Redmond, Washington.

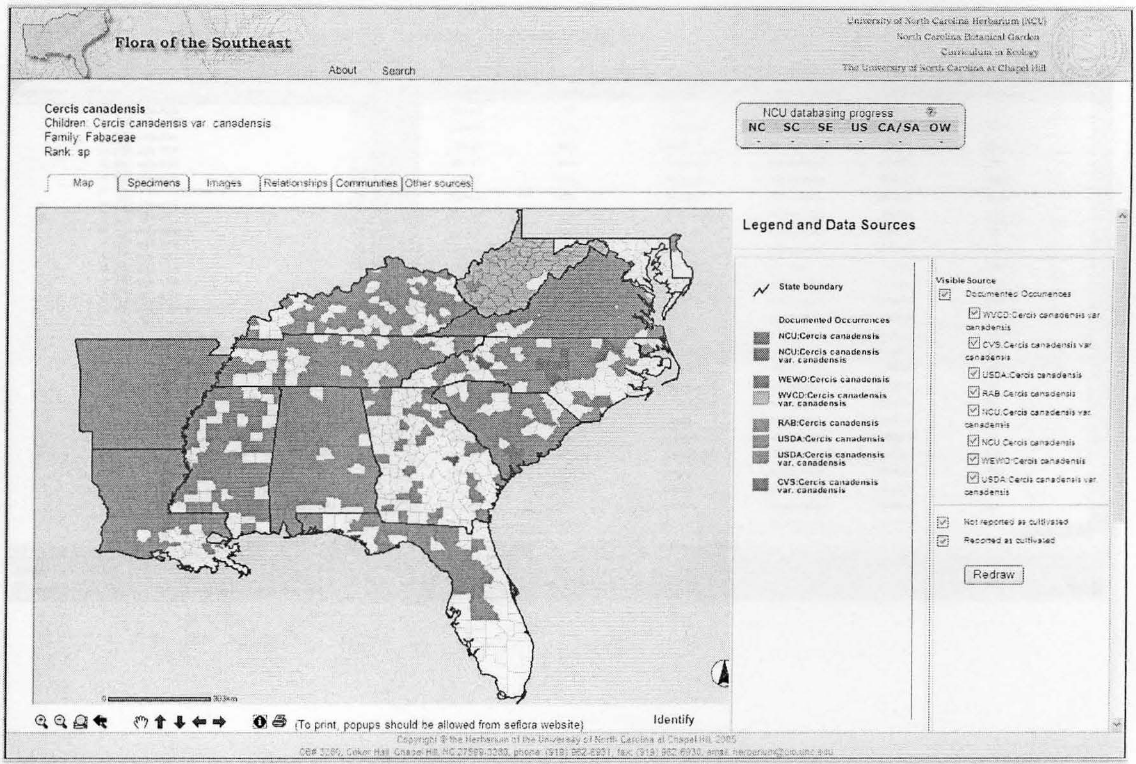
Figures.

Figure 1. Map image (GIF file) of county distribution result for *Cercis canadensis* L. using TENN website.



(http://tenn.bio.utk.edu/vascular/database/vascularmaps.asp?CategoryID=Dicots&FamilyID=Fabaceae&GenusID=Cercis&SpeciesID=canadensis)

Figure 2. Generated map of county distribution result for *Cercis canadensis* L. using NCU website.



(http://www.herbarium.unc.edu/seflora/firstviewer.htm)

Figure 3. The "FIPS code connection" performed within SQL Server Management Studio by creating a new View file.

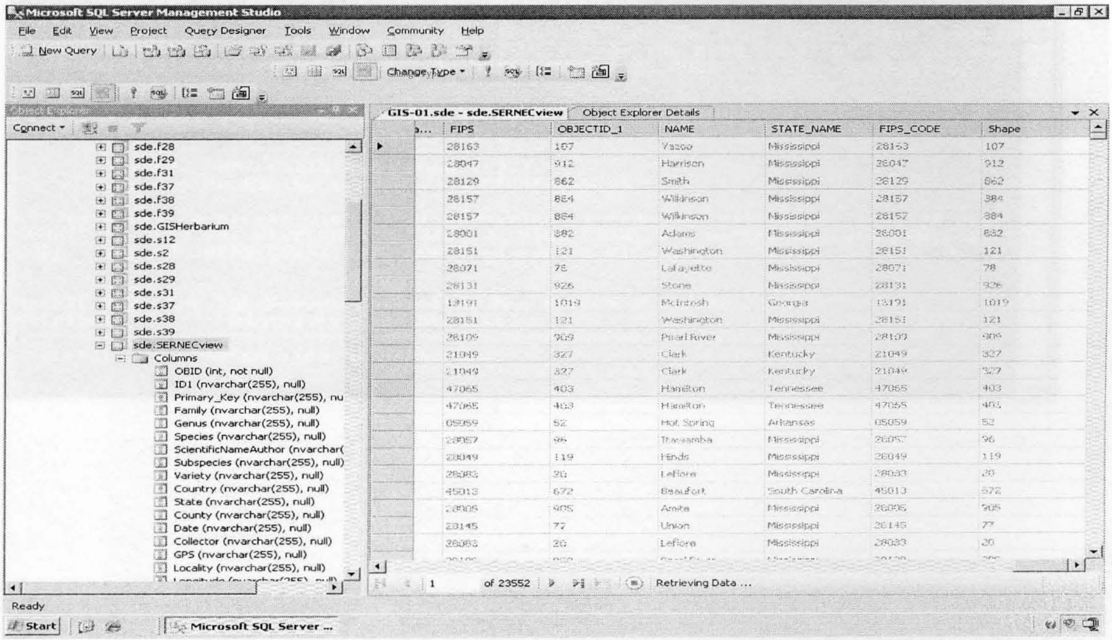


Figure 4. Integration of a spatial database engine (SDE) with SQL database.

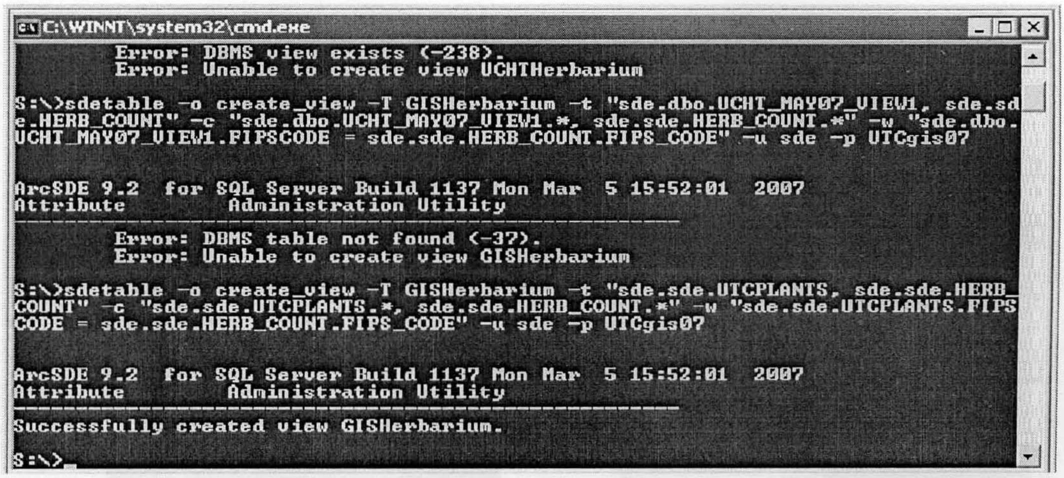


Figure 5. SERNEC website using a FIPS code join to display results on the county level.

PART 4. SPATIOTEMPORAL ANALYSIS OF THREE COMMON WETLAND INVASIVE PLANT SPECIES USING HERBARIUM SPECIMENS AND GIS

This chapter is a revised version of a paper by the same name submitted for publication in the journal *Castanea* in 2008 by Ryan Miller, Andrew Carroll, Thomas Wilson and Joey Shaw.

My primary contributions to this paper include (1) research into topic and method selection, (2) background review and selection of chosen invasive species to examine, (3) data gathering and analysis, and (4) primary writer of the submitted text.

Abstract.

The ecologically harmful impact of invasive plant species has been widely reported in recent years. Herbarium specimens can allow for the study of spatiotemporal spread of invasive species in order to understand distribution characteristics and patterns which can then be utilized for management decisions. In this study, three common wetland invasive plant species were chosen and analyzed via two different methods to determine the historic rate and pattern of spread as well as to identify periods of invasiveness. Performance of these methods varied in scale with a general, area-corrected approach covering the entire southeastern United States and an associate species-corrected method covering smaller, regional areas of dense historical specimen collection. Results showed positive spread over time for these three species in both methods; however, periods of invasiveness did not coincide between the two methods. The use of two different methods and the subsequent comparison of results show the importance of sampling bias correction, scale selection, and adequate sample sizes for spatiotemporal analyses of plant distributions using herbarium records.

Introduction.

Herbarium specimens are a rich source of ecological data that can be used to study a variety of issues, particularly when analyzed in geographic information systems (GIS). Recent studies have focused on the identification of species richness across a landscape to aid in conservation (Funk et al., 1999; Hijmans and Spooner, 2001), modern native species expansion (Lavoie and Saint-Louis, 1999), and in plant genetic resources management (Guarino et al., 2002). In regards to research on invasive plant species, herbarium records by themselves have proven to be an effective data source for the creation and subsequent analysis of distributions (Pysek and Prach, 1993; Stadler et al., 1998; Mihulka et al., 2001; Wu et al., 2005). Secondary sources (e.g. atlases and literature sources based on both herbarium records and field observations) have been used in conjunction with (GIS) to address this same issue (Welk, 2004; Hooftman et al., 2006). Other research on invasive plant species distributions was more focused on coupling the primary data source of herbarium records to GIS for spatiotemporally analyzing the data (Weber, 1998; Saltonstall, 2002; Delisle et al., 2003; Lavoie et al., 2005; Salo, 2005; Vincent, 2005; Barney, 2006; Boylen, 2006; Chauvel, 2006; Lavoie and Lachance, 2006; Lavoie et al., 2007).

This study shows how the expanding ranges of invasive species (*Phragmites australis* (Cav.) Trin. ex Steud. [common reed], *Lythrum salicaria* L. [purple loosestrife], and *Rorippa nasturtium-aquaticum* (L.) Hayek [watercress]) can be addressed with greater understanding by using both the extensive historical information inherent in herbarium specimens and the capability of GIS to efficiently analyze these kinds of data.

Before GIS tools can be applied to herbarium data, the data need to be databased, often following standardized methodology. Efforts to establish a standardized method of databasing and sharing the information of specimens have been attempted by several groups such as the Biodiversity Information Standards (TDWG), Global Biodiversity Information Facility (GBIF), and Science Environment for Ecological Knowledge (SEEK). Members and collaborators of TDWG have accepted the "Darwin Core" as a draft standard for the databasing efforts of biological collections; a recent botanical extension to this proposed standard more clearly delineates elements specific to herbarium data. The standardized entry of herbarium collections allows for large datasets to be compiled from individual herbaria efficiently; once complied, these records can then used within GIS and other ecological modeling platforms for research. Within the southeastern USA, the Southeastern Regional Network of Expertise and Collections (SERNEC) has begun to coordinate databasing efforts that follow these developing global standards and increase access to herbarium data.

Even though SERNEC and other similar efforts are in their infancy, resulting in relatively small datasets, many newly developed methods have focused on various ways to calculate and interpret the spatiotemporal spread of invasive plant species. As seen in these earlier studies, the ways to interpret the spread of invasive plant species are diverse and include various statistical reconstructions on the patterns of dispersion from herbarium records using GIS.

Salo (2005) used a "new occupied areas" method in a study involving the exotic, invasive annual grass *Bromus rubens* L. (red brome) in the southwestern United States. Six hundred and eighty-eight herbarium records were used to map and analyze the introduction and rates of geographic spread over time. Each specimen was given geographic coordinates based upon verbatim label information and each point was then mapped and surrounded by a 50km circular buffer. The increase in new 50 km areas over time was used to describe the "initial lag phase of relatively slow spread, an intermediate period of rapid increase, and a final period of slower spread" (Salo, 2005, p. 173). One weakness of this method is that it does not account for the sampling bias inherent in the background rate of specimen collection.

To separate the actual spread of an invasive plant species from the background rate of specimen collection, a historical collection rate can be determined using associate native species. Delisle et al. (2003) describe the spread of six invasive wetland plant species in Quebec, Canada using procedures that account for sampling biases. The records of five associate native wetland species were used as a control measure; whereby the native records are used to model the rate of specimen collection from wetlands in the general area. By establishing the background rate of specimen collection, to which the invasive species records can then be compared, invasion can be more confidently tested. Proportion curves were also created and analyzed to identify more specific time periods of invasiveness. Through the use of "invasion curves" and "proportion curves", this method allows for calculations of invasion rates and the identification of smaller time periods of invasiveness.

The importance of this topic of study is due to the unique and serious ecological threat that invasive aquatic species pose through their ability to spread quickly via waterways and subsequently lower biodiversity in important wetland habitats. In addition, there is often a high monetary cost involved in their removal and control; it is estimated that purple loosestrife (*Lythrum salicaria*) alone costs $45 million a year to control in the United States (Pimentel et al., 2000). The present availability of data from SERNEC allows for a large database of herbarium records to be used in analyzing the historical spread of invasive plant species.

In this study, we use an improved SERNEC dataset from 29 contributing herbaria and 2 previously developed strategies (Delisle et al., 2003 & Salo, 2005) to perform distribution analyses using herbarium records and GIS on three wetland invasive plant species: *Phragmites australis*, *Lythrum salicaria*, and *Rorippa nasturtium-aquaticum*.

Materials and Methods.

<u>*Herbarium Specimens*</u>.

Specimen label data were collected from herbaria throughout the southeast. Initially 3,317 records were received from 29 herbaria. Specimens lacking year collected or locality information were not used for analysis resulting in 3,202 records. The collection date, county, state and locality/place names were gathered from each record. Only 87 of 3,202 had geographic coordinates on the label. All others were individually georeferenced. The smallest scale of the two analyses described below was 10km; therefore, those specimens described as ≤ 10km within a geologic or cultural place name (e.g. Atlanta, GA, English Mountain, TN) had coordinates obtained through the USGS GNIS website (http://geonames.usgs.gov). The GNIS website allows for advanced searches based upon complex queries on feature/place name, state and county. Those >10km from a place name were georeferenced using various sources including Google Earth software (Google, 2007) & ESRI StreetMap USA (ESRI, 2006) database in ArcGIS. Specimen labels lacking any geographic coordinates and place names were given the centroid coordinates of the given county. The finalized, georeferenced dataset was used for all subsequent analyses.

A total of 3,202 specimens were databased from 29 herbaria; of these specimens, 2,478 were of the natives group (*Typha latifolia* n=897, *Brasenia schreberi* n=270, *Cephalanthus occidentalis* n=1311) and 724 were of the invasive group (*Phragmites australis* n=260, *Lythrum salicaria* n=181, *Rorippa nasturtium-aquaticum* n=283.

Spatiotemporal Analyses.

New Occupied Areas Method. A base map containing the southeastern states and all point locations of the three invasive species was projected using an equidistant conic projection (Clarke 1866 spheroid), as per Salo (2003). An equidistant conic projection acts as a compromise between projections preserving equal areas and those preserving shapes (conformal) (Synder, 1987). For each invasive species record, a 25 km circular area was generated around the point using the Buffer tool in the ArcGIS ArcToolbox (ESRI, 2006). Clusters of overlapping 25km areas were analyzed for each of the three invasive species to identify the earliest record within the clusters. Then all later records within the cluster were omitted as it was presumed that the invasive species has occupied that general area since the earliest record. The elimination of these overlapping records allowed for only geographically unique, whole 25km areas to be studied for each species. The rate of spread for each invasive species was described by plotting the cumulative number of new 25km areas occupied over time.

Associate Species-Corrected Method. The base map containing the southeastern states and all point locations of herbarium records was modified to the North America Albers Equal Area Conic projection; the use of a projected coordinate system allowed for accurate area measurements to be performed during analysis. Layers for analysis consisted of the three individual invasive species records and a combined layer consisting of all of the native species records. In addition, the Create Vector Grid Tool (Hawth's Analysis Tools extension for ArcGIS (Beyer, 2004) was used to create a 10x10km grid which was overlaid on the map extent. Selected regional areas were chosen for each invasive species using the Kernel Density tool in ESRI's Spatial Analyst extension (ESRI, 2006). Kernel density analyses can estimate the probability of finding a specified invasive species within an area using known, recorded locations. Classification of kernel density results was performed using the Natural Breaks (Jenks) method with three classes designated (ESRI, 2006). All break values under 10% were eliminated, leaving a better graphical representation of potential regional study area sites. In each of the kernel density derived study areas occupied grid cells for each invasive species and the combined native species were determined using the Count Points in Polygons Tool in the Hawth's extension. The occupied 10x10km grid cells for each layer were examined for occurrences of multiple records. In cases where more than one record of a layer was in the same grid cell, the record with the earliest year collected was used for the analysis and the others dismissed. Elimination of newer records of a layer in the same cell was done because once an area is occupied by a species of interest it is considered an established presence for the future. For each of the three invasive species and the combined native group, the number of occupied 10x10 km grid cells was recorded for each year. This information was graphed using the log cumulative number of occupied 10x10 km grid cells on a yearly basis. Linear regression lines were calculated to give a slope value for the combined natives group and each of the three invasives. The combined native group's linear regression line was considered the base "invasion curve" – representing the baseline rate of data accumulation regarding plant collecting unique to each kernel density derived study area. The invasion curves of the invasive species were then compared to the combined native group's invasion curve by analyzing the difference in slope. Comparisons of the invasive species' slopes to the combined native species' slope were performed using an ANCOVA test (PROC ANCOVA) in S.A.S. v.9.1.3 (SAS Institute, 2004). Results from the linear regression lines and the subsequent ANCOVA test allowed for an estimation of the invasion rate. In order to account for the sampling bias inherent in the herbarium records used, proportion curves were created to help identify periods of invasiveness while assessing the sampling bias.

Proportion curves were generated for each invasive species through time by dividing the cumulative number of grid cells occupied by the invasive species by the cumulative number of grid cells occupied by the combined native's group for each year.

Proportion curves allow for the identification of fine-scale time periods where potentially rapid range expansion has occurred. Furthermore, according to Delisle et al. (2003), "...if the proportion (of exotic vs. native species) is increasing for a particular time period, this strongly suggests that the area occupied by the exotic species is really expanding, because it is expanding faster than if it was strictly the result of better spatial coverage of the sampling for herbarium specimens" (p. 1035). However, segments of proportion curves that exhibit steady or declining ratios do not necessarily signify reduced or stabilized distributions; rather these segments may be due to expansion that is too slow to be detected or a reduction of sampling efforts.

Results.

New Occupied Areas Method.

The spread of each invasive species was analyzed by creating 25 km buffers around each specimen record using initial sampling sizes of 260, 181 and 284 for *Phragmites australis, Lythrum salicaria* and *Rorippa nasturtium-aquaticum,* respectively. Following the identification of the earliest record in locations of overlap, subsequent later records were eliminated resulting in final sample sizes of 132, 77 and 143 for *Phragmites, Lythrum* and *Rorippa*, respectively. The majority of spread (increased number of occupied 25 km areas) for all three invasive species occurs post 1950 with a general sigmoid shape describing each of the three graphs (Figure 1 a-c).

An average nearest neighbor analysis was used to test the null hypothesis that the invasive records were randomly distributed. The null hypothesis was rejected for each of the three species – the patterns were significantly clustered and not randomly distributed. With a significance level of 0.01 and a critical value of -2.58, all nearest neighbor indexes were between 0.36 – 0.41 (*Phragmites*: index = 0.38, Z score = -19.2 SD; *Lythrum*: index=0.36, Z score=-16.56 SD; *Rorippa*: index=0.41, Z score=-18.9 SD).

Associate Species-Corrected Method.

The same initial invasive species records were used as those in the New Occupied Areas method (*Phragmites* n = 260, *Lythrum* n = 181, *Rorippa* n = 284). Upon analysis of the kernel density results, eight regional study areas were designated with three areas for *Phragmites*, two for *Lythrum* and three for *Rorippa* (Figures 2-4). The size of each regional study area was measured using the XTools Pro version 5.1 extension to ArcGIS (Data East, 2007); area sizes varied from 23,447 km^2 to 56,482 km^2 (Figures 2-4).

Slopes of the invasion curves for each invasive species study area were derived from linear regression analysis and compared to the slope of the combined natives at a 95% confidence interval. For *Phragmites* Area 1 the slopes (b invasive = 0.0165; R^2 = 0.942 and b natives = 0.0159; R^2 = 0.9158) did not show significant difference between each other ($P>0.5$) (Figure 2a). In addition, the proportion curve for *Phragmites* Area 1 supported the ANCOVA results and failed to distinguish any dramatic periods of invasiveness. The slopes of *Phragmites* Area 2 (b invasive = 0.025; R^2 = 0.9219 and b natives = 0.0114; R^2 = 0.9265) did show a significant and distinct difference ($P<0.0001$) (Figure 2b). Results of the proportion curve for *Phragmites* Area 2 identifies a major period of invasiveness as 1996 to 2005 (Figure 5a). The slopes of *Phragmites* Area 3 (b invasive = 0.0194; R^2 = 0.9558 and b natives = 0.0152; R^2 = 0.9738) did show a significant and distinct difference ($P<0.0001$) (Figure 2c). The proportion curve for *Phragmites* Area 3 showed a distinct jump from 1932 and 1937; however, further analysis showed the sample size for this abrupt increase as very low (proportion numbers of 1:7 to 4:13) which negates the implication of this section of the curve as a period of invasiveness (Figure 5b).

For *Lythrum* Area 1 the slopes (b invasive = 0.0148; R^2 = 0.9613 and b natives = 0.0123; R^2 = 0.9656) did show a statistically significant difference between each other ($P = 0.0002$) (Figure 3a). However, the proportion curve for *Lythrum* Area 1 failed to distinguish any dramatic periods of invasiveness. The slopes of *Lythrum* Area 2 (b invasive = 0.0274; R^2 = 0.946 and b native = 0.0136; R^2 = 0.8878) were significantly different ($P<0.0001$; Figure 3b) and the proportion curve showed a consistent trend of increased spread (Figure 5c). The only dramatic spike in the *Lythrum* Area 2 proportion curve was during an period of known increased collecting specifically for *L. salicaria* from 1997 to the present in which specimens were given to West Virginia University and therefore not an actual period of invasiveness (personal communication, Ford-Werntz at WVU 10/22/07).

Slopes for *Rorippa* Area 1 (b invasive = 0.021; R^2 = 0.8955 and b natives = 0.0254; R^2 = 0.9151) were statistically different (P = 0.0052; Figure 4a); however the proportion curve for this study area however failed to show any distinct period of invasiveness. The slopes of *Rorippa* Area 2 (b invasive = 0.0156; R^2 = 0.8587; b native = 0.0169; R^2 = 0.7797) were not significantly different (P>0.3; Figure 4b). In addition, the proportion curve supports the ANCOVA results and fails to show a distinct period of invasiveness. The slopes of Rorippa Area 3 (b invasive = 0.0145; R^2 = 0.9279 and b native = 0.015; R^2 = 0.8094) were not statistically different (P>0.6; Figure 4c) and the proportion curve fails to show any abrupt periods of invasiveness.

Discussion.

In studies like this one, scale is a difficult issue because of 1) the large range at which certain invasive species occur and are spreading (e.g., *Lythrum salicaria* being an issue throughout the United States) and 2) the inherent limitations and potential bias of using herbarium records (e.g., georeferencing vague location information and the botanist effect of collection efforts biased geographically near herbaria and universities (Moerman & Estabrook, 2006). In addition, researchers must use caution when evaluating distributions because spatiotemporal patterns will be measured differently depending on the scale of the analysis (Parker & Pickett, 1998). Because of these issues, our selected study areas and results do not reflect the entire known range of the three invasive species of study; but rather are focused on the southeastern USA, especially where the density of herbarium specimen collection permits reasonable sample sizes.

Within the dataset, the earliest collection date for each invasive species was *P. australis* (1913; Dulac, LA), *R. nasturtium-aquaticum* (1864; Centreville, DE) and *L. salicaria* (1888, Sulphur Springs, VA). These collection dates were remarkably close to the earliest known year of introduction to North America and gave confidence that the temporal scale of our dataset was accurate because the earliest known literary accounts of these species are: 1910 for *P. australis* (invasive spread first noted in literature (Graves et al., 1910)), 1847 for *R. nasturtium-aquaticum* (Mills et al., 1993) and 1814 for *L. salicaria* (Mehrhoff, 2004).

General patterns of spread can be evaluated by identifying periods when large increases of new occupied 25km areas are observed; however the increases may also be due an increase in collecting activities and not necessarily a range expansion. While the new occupied area method allows for interpretations on a broad scale of the increased ranges of the invasive species, it does not account for the sampling bias associated with herbarium records. The associate-species corrected method allows for more detailed periods of invasiveness to be identified while taking into account the inherent sampling bias associated with herbarium specimen collection. The slopes of the invasion curves represent the invasion rate measurement; a higher slope value for an invasive's invasion curve compared to that of the combined native's indicates that the invasive species is spreading. The use of both of these methods allows for the comparison between the two types of analyses and their unique results on the spread of the three invasive wetland species - from a broad geographic perspective using new occupied area methodology to a regionalized, bias-corrected perspective.

Based upon herbarium records throughout the southeast, the new occupied areas method performed on each of the three invasive species show a general sigmoid shape when plotted against time on a yearly basis. Each begins with an initial lag period and shows an abrupt increase in occupied area that is generally sustained throughout the rest of the years. This corresponds with results of Salo (2005) that also exhibited a sigmoid curve trend for the spread of *Bromus rubens* in the southwest. Results of *Phragmites australis* and *Lythrum salicaria* in the new occupied areas method show lag periods ending and the rapid increases in occupied area occurring in the mid-1960s and early 1970s. It appears that the rapid increased occupied area occurs much earlier for *Rorippa nasturtium-aquaticum* with rapid increase in occupied area starting in the early 1930s. While the earliest herbarium records in the dataset had *Rorippa* collected much earlier than the other two invasive species (1864 compared to 1888 for *Lythrum* and 1913 for *Phragmites*), the number of records prior to *Rorippa*'s rapid increase is negligible. There were only nine specimens of *Rorippa*, and 19 specimens for all three invasive species, in the dataset prior to 1930. Because of the low sample size pre-1930, we cannot be sure that *Rorippa* was introduced earlier than the others. However, the significant period of increase in *Rorippa*'s occupied area is truly earlier than the other two species rather than just an artifact of collection in the southeast due to ample specimen numbers post-1930.

This method of new occupied area is more representative of increased spread than using cumulative number of specimens of each species over time. However, while these data show that the three species have increased in occupied area, we cannot separate the potential effect of collection bias over time.

Delisle et al. (2003) argued that one way to separate potential collection bias was to use associate native species records to compare with the invasive species spread. Results of the associate species-corrected method were based upon selected regional study areas chosen by a kernel density analysis. Eight regional study areas were spread throughout the southeast with minimal overlap; this variety of study area locations reflects the diversity of collecting intensities for different species across a broad landscape like the southeastern United States. Of the eight regional study areas used for the associate species-corrected method, five were found to have significant invasive species spread (*Phragmites* areas 2 and 3, *Lythrum* areas 1 and 2, *Rorippa* area 1). In four of these five areas, the slope of the invasive species was found to be significantly greater than that of the combined natives group; this suggests that either the particular invasive species in that area is spreading at an increased rate or that the invasive species was preferentially collected by botanists during that period of time. *Rorippa* area 1 had significantly different slopes but the slope of the combined natives group was found to be greater, this argues that *Rorippa* was either not expanding at a fast enough rate to be detected or was historically under-collected in this area. These slope comparisons of the invasion curves allow only for an overall estimate of rates of invasive species spread; uneven collecting effort over time could cause the invasion curve slopes to be misleading. For example, a general decrease in specimen collecting during more recent years could lead to the slope of the entire time period of the study being underestimated when compared to a situation of consistent collection intensity over time.

In order to identify particular periods of invasiveness and to eliminate the argument of uneven sampling efforts over time, proportion curves were calculated for all eight study areas. Proportion curves were generated by comparing occupied areas of the invasive species to those of the combined natives group for each year collected. A year by year comparison of the proportion of occupied grid areas allows for the identification of true range expansion if the proportion of invasive vs. native species is consistently increasing over a definitive time period. This proportional examination of relatively smaller time periods of course requires a larger sample size to have ample ratios for comparison. Several of the proportion curves generated in this study show periods of significant proportion increases (Figures 5a-c); however, upon further inspection of the raw data it was found that these proportions are based on low sample sizes. For example, while 104 specimens were used to generate the invasion curves in *Phragmites* Area 2, examination of the data underlying the proportion curve for years 1959-1970 shows low sample sizes with a proportion in 1959 of one invasive grid cell to 19 native grid cells and subsequent increases to sample size of five invasive grid cells to 21 native grid cells in 1970. The appearance of significant periods of invasiveness in the other study areas are also mainly due to low sample sizes creating large ratio differences over time.

The average number of records used in the eight study areas was 60 invasive specimens and 118 native specimens. Comparatively, in Delisle et al. (2003) the regional study area was the same for all invasive species studied (southern Quebec, Canada) and larger specimen sample sizes were available for the selected plant species (2668 exotic species records and 2889 native species records). Periods of invasiveness that were identified in Delisle et al. (2003) were, therefore, of greater confidence than the results seen in our study due to higher number of specimen records used.

This study successfully shows the feasibility of using herbarium records with GIS to evaluate spatiotemporal changes of invasive plant species. By using two previously developed methods, estimates on newly occupied areas over time and historical rates of spread were determined. Within the overall rates of spread, the identification of fine-scale time periods of invasiveness using proportion curves was unable to be performed with confidence due to low specimen sample sizes within the regionally selected study areas. The low sample sizes in the proportion curve analyses illustrates the need of regional herbaria to collaborate in combining specimen data into central databases from which specimen information can be distributed for scientific analyses and queries. Zack Murrell (APSU) estimates that there are approximately 10 million herbarium specimens in about 200 southeastern herbaria. We extracted the data for this study from only 29 herbaria; therefore, efforts of SERNEC and similar organizations should continue to support the concept of collective and standardized herbarium databases. Because, as the number of digital specimen datasets grow and the availability of analytical software such as GIS increases, the number and accuracy of studies like this one will certainly add to our overall knowledge of southeastern botany. In turn, this increased knowledge can have major implications for future conservation and management practices. By investigating the historical spatiotemporal dynamics of invasive plant species, land managers can identify areas currently occupied by these harmful species, those areas likely to be under future threat, and the environmental variables associated with their spread.

Literature Cited.

Barney, Jacob N. (2006) North American history of two invasive plant species: phytogeographic distribution, dispersal vectors, and multiple introductions. Biological Invasions 8: 703-717.

Beyer, H.L. (2004) *Hawth's Analysis Tools for ArcGIS*, version 3.26. (http://www.spatialecology.com/htools).

Boylen, C.W., L.W. Eichler, J.S. Bartkowski and S.M. Shaver. (2006) Use of Geographic Information Systems to monitor and predict non-native aquatic plant dispersal through north-western North America. Hydrobiologia 570: 243-248.

Chauvel, Bruno, Fabrice Dessaint, Catherine Cardinal-Legrand and Francois Bretagnolle. (2006) The historical spread of *Ambrosia artemisiifolia* L. in France from herbarium records. Journal of Biogeography 33: 665-673.

Data East, LLC. (2007) *XTools Pro 5.1.* Novosibirsk, Russia.

Delisle, F., C. Lavoie, M. Jean, and D. Lachance. (2003) Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. Journal of Biogeography 30:1033-1042.

Environmental Systems Research Institute, Inc. (ESRI). (2006) *ArcGIS, version 9.2.* Redlands, CA.

Funk, V.A., M. F. Zermoglio and N. Nasir. (1999) Testing the use of specimen collection data and GIS in biodiversity exploration and conservation decision making in Guyana. Biodiversity and Conservation 8:727-751.

Google. (2007) *Google Earth.* Mountain View, CA. (http://earth.google.com/).

Hijmans, R. and D. Spooner. (2001) Geographic distribution of wild potato species. American Journal of Botany 88:2101-2112.

Graves, C.B., E.H. Eames, C.H. Bissell, L. Andrews, E.B. Harger & C.A. Weatherby. (1910) Bulletin of the Connecticut Geological and Natural History Survey No. 14.

Guarino, L., A. Jarvis, R.J. Hijmans, and N. Maxted. (2002) Geographic information systems (GIS) and the conservation and use of plant genetic resources. Managing Plant Genetic Diversity. Eds. J.M.M. Engels, V. Ramanatha Rao, A.H.D. Brown and M.T. Jackson. CABI Publishing. 387-404.

Hooftman, D.A.P., J.G.B. Oostermeijer, J.C.M. den Nijs. (2006) Invasive behaviour of *Lactuca serriola* (Asteraceae) in the Netherlands: Spatial distribution and ecological amplitude. Basic and Applied Ecology 7: 507-519.

Lavoie, Claude and Annie Saint-Louis. (1999) The spread of gray birch (*Betula populifolia*) in eastern Quebec: landscape and historical considerations. Canadian Journal of Botany 77: 859-868.

Lavoie, Claude, Caroline Dufresne and Fanny Delisle. (2005) The spread of reed canarygrass (Phalaris arundinacea) in Quebec: a spatio-temporal perspective. Ecoscience 12 (3): 366-375.

Lavoie, C. and D. Lachance. (2006) A herbarium-based method for reconstructing the phenology of plant species across large areas. American Journal of Botany 93:512-516.

Lavoie, Claude, Yvon Jodoin and Albane Goursaud de Merlis. (2007) How did common ragweed (*Ambrosia artemisiifolia* L.) spread in Quebec? A historical analysis using herbarium records. Journal of Biogeography 34: 1751-1761.

Mehrhoff, L.J. (2004) IPANE: Invasive Plant Atlas of New England. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT (http://www.ipane.org).

Mihulka, Stanislav and Petr Pysek. (2001) Invasion history of *Oenothera* congeners in Europe: a comparative study of spreading rates in the last 200 years. Journal of Biogeography 28: 597-609.

Mills, E.L., J.H. Leach, J.T. Carlton and C.L. Secor. (1993) Exotic species in the Great Lakes: a history of biotic crises and anthropogenic interactions. Journal of Great Lakes Research 19: 1-54.

Moerman, Daniel E. and George F. Estabrook. (2006) The botanist effect: counties with maximal species richness tend to be home to universities and botanists. Journal of Biogeography 33: 1969-1974.

Parker, V. Thomas and Steward T.A. Pickett. (1998) Historical contingency and multiple scales of dynamics within plant communities (Chapter 8). In David L. Peterson a nd V. Thomas Parker, editors, Ecological Scale, pages 171-191, Columbia University Press: New York.

Pimentel, David, Lori Lach, Rodolfo Zuniga and Doug Morrison. (2000) Environmental and economic costs of nonindigenous species in the United States. BioScience 50: 53-65.

Pysek, Petr and Karel Prach. (1993) Plant invasions and the role of riparian habitats: a comparison of four species alien to central Europe. Journal of Biogeography 20: 413-420.

Salo, Lucinda. (2005) Red brome (*Bromus rubens* subsp. *madritensis*) in North America: possible modes for early introductions, subsequent spread. Biological Invasions 7: 165-180.

Saltonstall, Kristin. (2002) Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. Population Biology 99: 2445-2449.

SAS Institute, Inc. (2004) *SAS 9.1.3*. Cary, North Carolina.

Stadler, J., G. Mungai and R. Brandl. (1998) Weed invasion in East Africa: insights from herbarium records. African Journal of Ecology 36: 15-22.

United States Geological Survey (Retrieved 2007) Geographic Names Information System (GNIS). (http://geonames.usgs.gov).

Weber, Ewald. (1998) The dynamics of plant invasions: a case study of three exotic goldenrod species (*Solidago* L.) in Europe. Journal of Biogeography 25: 147-154.

Welk, Erik. (2004) Constraints in range predictions of invasive plant species due to non-equilibrium distribution patterns: Purple loosestrife (*Lythrum salicaria*) in North America. Ecological Modelling 179: 551-567.

Wu, S., M. Rejmanek, E. Grotkopp, and J. DiTomaso. (2005) Herbarium records, actual distribution, and critical attributes of invasive plants: genus *Crotalaria* in Taiwan. Taxon 54:133-138.

Vincent, Michael A. (2005) On the spread and current distribution of *Pyrus calleryana* in the United States. Castanea 70: 20-31.

Figures.

Figure 1. New Occupied Areas Method Results: Graphs display increase in new occupied areas measured using cummulative 25km areas for three wetland invasive plant species (a) Phragmites australis, (b) Lythrum salicaria and (c) Rorippa nasturtium-aquaticum. Maps display 25km areas by time period; 1864-1900: gray circles, 1901-1950: open circles, 1951-present: black circles.

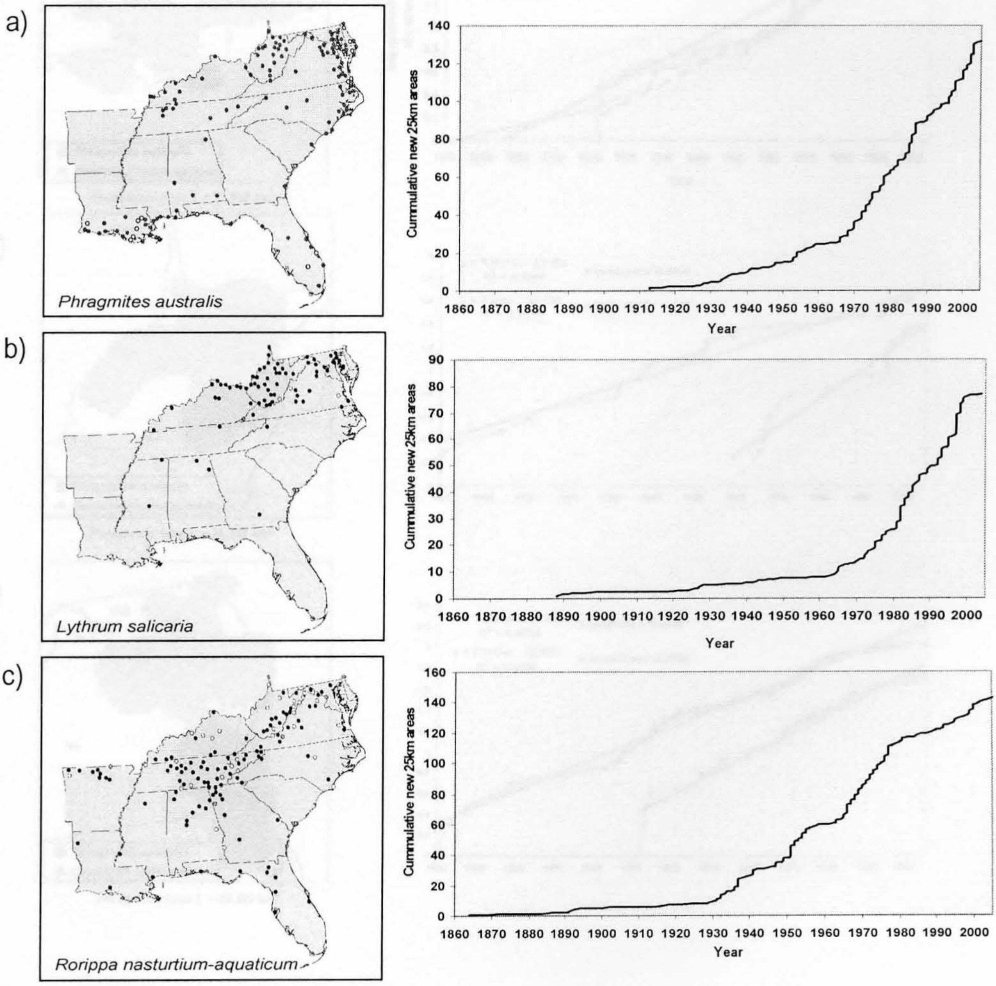


Figure 2. Associate Species-Corrected Method Results for *Phragmites* australis: Maps and results of associate species-corrected method for *Phragmites australis* study areas 1 (a), 2 (b) and 3 (c). Invasion curves were used to compare slopes of combined natives group vs. *Phragmites.*

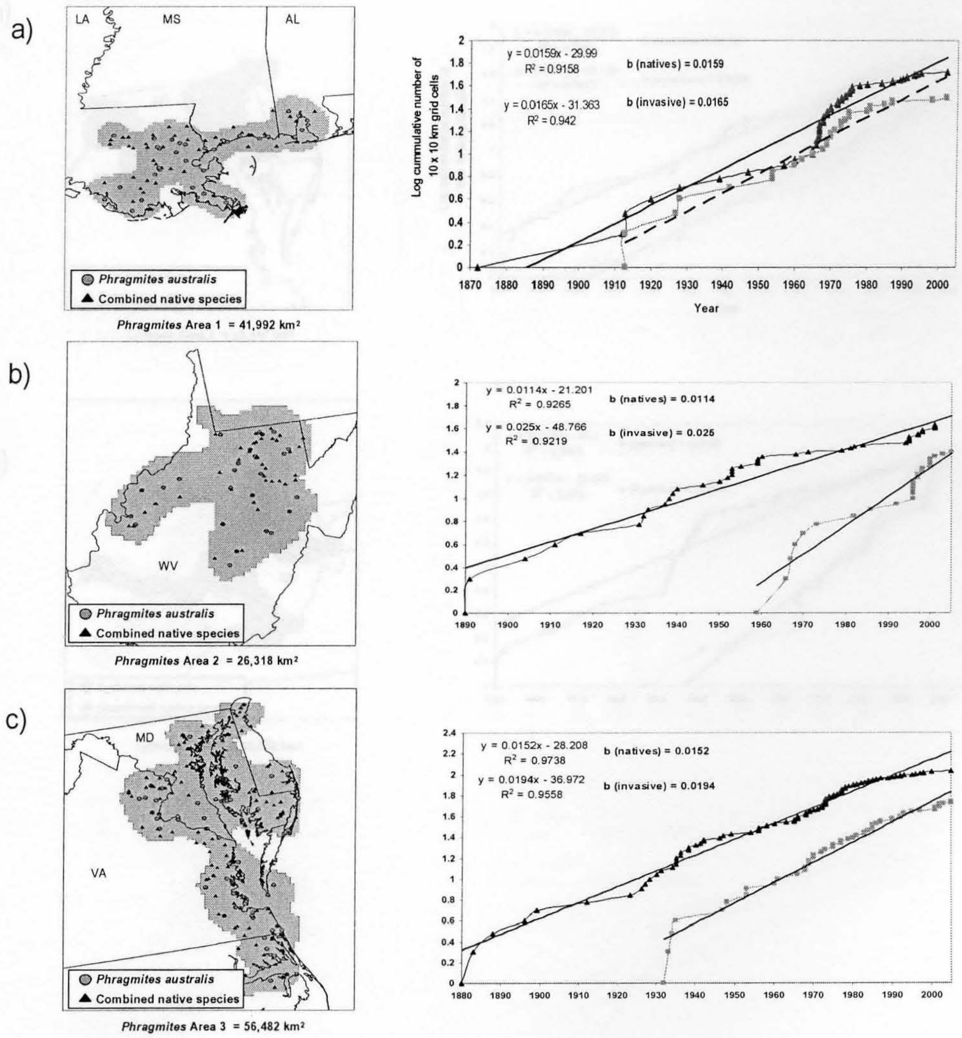


Figure 3. Associate Species-Corrected Method Results for *Lythrum salicaria*: Maps and results of associate species-corrected method for *Lythrum salicaria* study areas 1 (a) and 2 (b). Invasion curves were used to compare slopes of combined natives group vs. *Lythrum*.

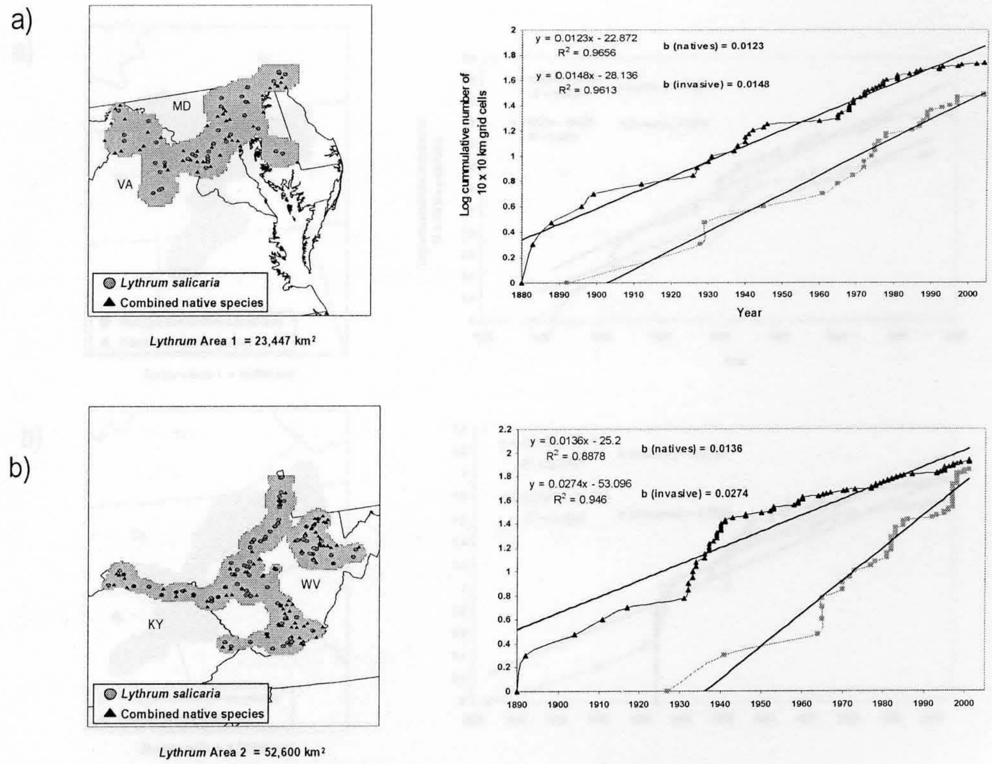


Figure 4. Associate Species-Corrected Method Results for *Rorippa nasturtium-aquaticum*: Maps and results of associate species-corrected method for *Rorippa nasturtium-aquaticum* study areas 1 (a), 2 (b) and 3 (c). Invasion curves were used to compare slopes of combined natives group vs. *Rorippa*.

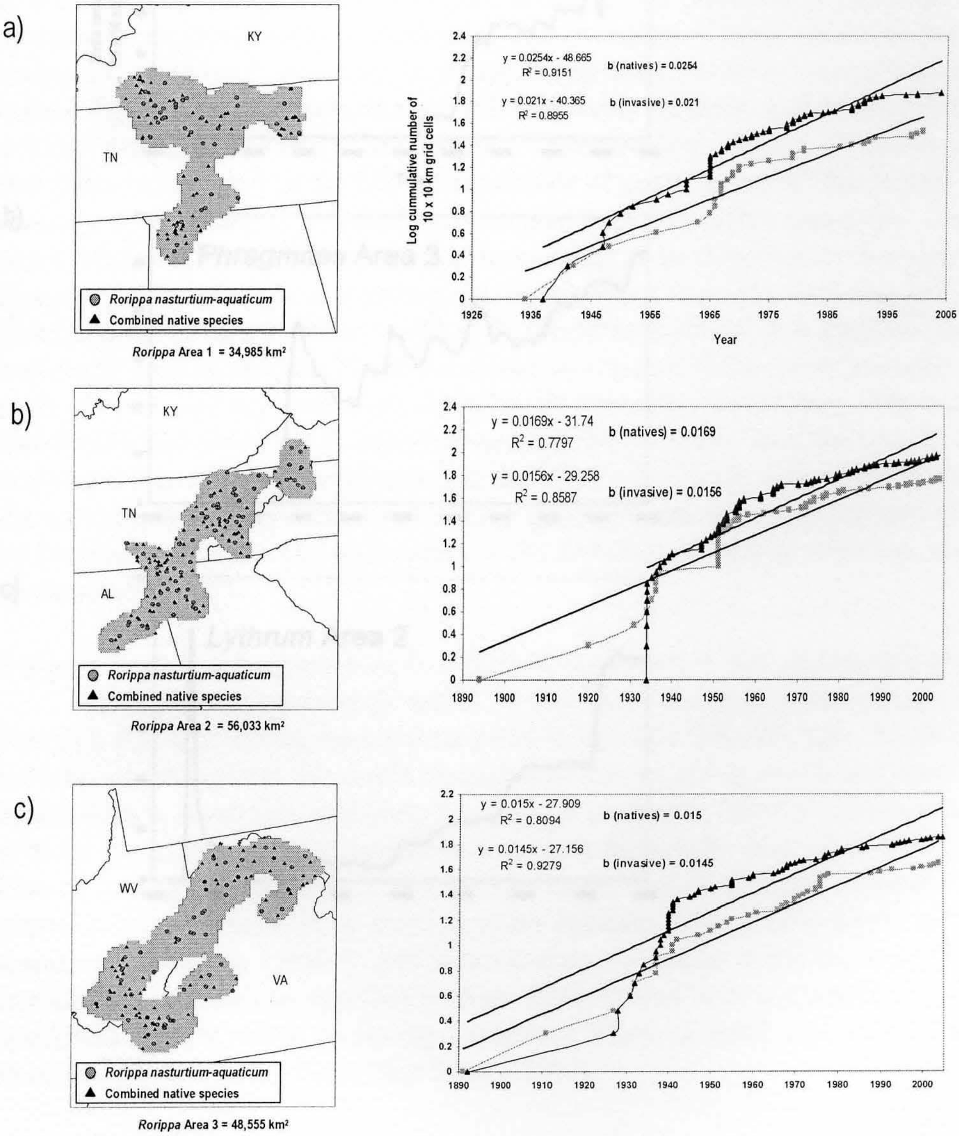
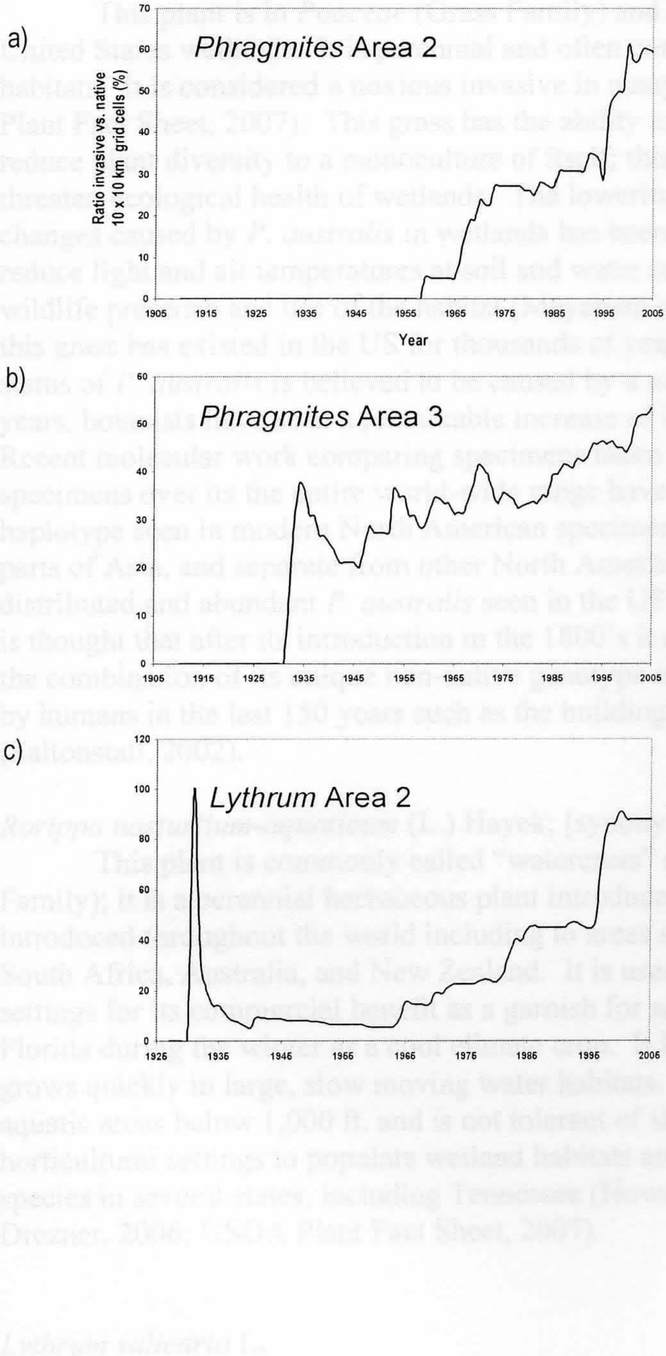


Figure 5. Significant Proportion Curve Results: Three of the eight study areas showed significant proportion curve results: *Phragmites* Areas 2 & 3 and *Lythrum* Area 2.



Appendix.

Appendix 1. Invasive Wetland Plant Species Descriptions.

Phragmites australis (Cav.) Trin. Ex Steud.

This plant is in *Poaceae* (Grass Family) and is the tallest grass in the southeastern United States wetlands. It is perennial and often considered facultative to aquatic habitats. It is considered a noxious invasive in many states, including Tennessee (USDA Plant Fact Sheet, 2007). This grass has the ability to quickly take over a wetland and reduce plant diversity to a monoculture of itself; this reduction of floral diversity can then threaten ecological health of wetlands. The lowering of floral diversity and structural changes caused by *P. australis* in wetlands has been shown to alter nitrogen cycling, reduce light and air temperatures at soil and water surfaces, and subsequently changing wildlife presence and use of the habitat (Meyerson et al., 2000). A native genotype of this grass has existed in the US for thousands of years, however the current invasive status of *P. australis* is believed to be caused by a non-native genotype. Over the last 150 years, botanists have seen a remarkable increase of its distribution and abundance. Recent molecular work comparing specimens taken from the field and herbarium specimens over its the entire world-wide range have shown that the most common haplotype seen in modern North American specimens is the same one seen in Europe and parts of Asia, and separate from other North American haplotypes. The most widely distributed and abundant *P. australis* seen in the US today is of the invasive haplotype. It is thought that after its introduction in the 1800's it quickly expanded its distribution with the combination of its unique non-native genotype and the huge increase of development by humans in the last 150 years such as the building of canals, railroads, and roadways (Saltonstall, 2002).

Rorippa nasturtium-aquaticum (L.) Hayek; [synonym = *Nasturtium officinale* Ait. f.]

This plant is commonly called "watercress" and is in *Brassicaceae* (Mustard Family); it is a perennial herbaceous plant introduced from Europe. It has been introduced throughout the world including to areas such as North and South America, South Africa, Australia, and New Zealand. It is used in North America in horticultural settings for its commercial benefit as a garnish for salads. It is often grown in central Florida during the winter as a cool climate crop. It is an obligate wetland plant that grows quickly in large, slow moving water habitats. It is usually found only in lowland, aquatic areas below 1,000 ft. and is not tolerant of shade. It has escaped from its horticultural settings to populate wetland habitats and is now considered an invasive species in several states, including Tennessee (Howard and Lyon, 1952; Tenorio and Drezner, 2006; USDA Plant Fact Sheet, 2007).

Lythrum salicaria L.

This plant is in *Lythraceae* (Loosestrife Family) and is a widely known as a noxious, highly invasive wetland/aquatic plant. It is native to Eurasia and was originally brought into North America for ornamental and medicinal purposes. It was first reported in North America in 1814 and has since spread throughout most of the US with the majority of its abundance in northeastern wetlands (The Nature Conservancy, 2001). This plant typically occupies weltands and sedge meadows and can quickly out-compete native flora to create a monoculture. This reduces native plant diversity, over-runs open water habitat, and creates declines in animal presence. It is often considered one of the worst invasive plant species in the United States (USDA Plant Fact Sheet, 2007).

Appendix 2. Contributing Herbaria

School	Herbarium Code	# of Specimens	emailed	mailed	visited	online	SERNEC pilot
University of Florida	FLAS	40				x	
Auburn University	AUA	157	x				
Delaware State University	DOV	91		x			
Virginia Polytechnic Institute and State University	VPI	168		x			
Louisiana State University	LSU	219				x	
Old Dominion University	ODU	52		x			
Duke University	DUKE	206		x			
University of North Alabama	UNAF	81	x				
Morehead State University	MDKY	41		x			
Western Carolina University	WCUH	17	x				
Towson University	BALT	82		x			
University of Kentucky	KY	87	x				
University of Southern Mississippi	USMS	44	x				
Emory University	GEO	13	x				
University of Tennessee at Chattanooga	UCHT	44					x
Northern Kentucky University	KNK	74		x			
Wake Forest University	WFU	11		x			
West Virginia University	WVA	366	x	x			
University of Central Arkansas	UCAC	19		x			
Troy University	TROY	40					x
University of Georgia	GA	367			x		
University of Tennessee at Knoxville	TENN	347			x		
Austin Peay State University	APSC	95			x		
George Mason University	GMUF	107		x			
Tennessee Technological University	HTTU	55		x			
University of South Alabama	USAM	9	x				
University of North Carolina	NCU	282				x	
University of the South - Sewanee	UOS	12	x				
University of Mississippi	MISS	191	x				
		3317					

Table 1. Specimen dataset for invasive species analysis.

Typha

10 year period	Total number	Cumulative number
1860-1869	0	0
1870-1879	1	1
1880-1889	1	2
1890-1899	3	5
1900-1909	1	6
1910-1919	2	8
1920-1929	6	14
1930-1939	35	49
1940-1949	40	89
1950-1959	173	262
1960-1969	182	444
1970-1979	194	638
1980-1989	109	747
1990-1999	104	851
2000-Present	46	**897**
	897	

Brasenia

10 year period	Total number	Cumulative number
1860-1869	0	0
1870-1879	1	1
1880-1889	0	1
1890-1899	1	2
1900-1909	1	3
1910-1919	0	3
1920-1929	4	7
1930-1939	13	20
1940-1949	11	31
1950-1959	38	69
1960-1969	39	108
1970-1979	65	173
1980-1989	33	206
1990-1999	52	258
2000-Present	12	**270**
	270	

Cephalanthus

10 year period	Total number	Cumulative number
1860-1869	0	0
1870-1879	0	0
1880-1889	4	4
1890-1899	9	13
1900-1909	1	14
1910-1919	5	19
1920-1929	7	26
1930-1939	92	119
1940-1949	97	216
1950-1959	110	326
1960-1969	326	652
1970-1979	206	858
1980-1989	169	1027
1990-1999	209	1236
2000-Present	76	**1311**
	1311	

Phragmites

10 year period	Total number	Cumulative number
1860-1869	0	
1870-1879	0	0
1880-1889	0	0
1890-1899	0	0
1900-1909	0	0
1910-1919	2	2
1920-1929	3	5
1930-1939	5	10
1940-1949	10	20
1950-1959	11	31
1960-1969	26	57
1970-1979	50	107
1980-1989	49	156
1990-1999	53	209
2000-Present	51	**260**
	260	

Lythrum

10 year period	Total number	Cumulative number
1860-1869	0	0
1870-1879	0	0
1880-1889	1	1
1890-1899	1	2
1900-1909	0	2
1910-1919	0	2
1920-1929	4	6
1930-1939	0	6
1940-1949	4	10
1950-1959	1	11
1960-1969	12	23
1970-1979	24	47
1980-1989	34	81
1990-1999	85	166
2000-Present	15	**181**
	181	

Rorippa

10 year period	Total number	Cumulative number
1860-1869	1	1
1870-1879	0	1
1880-1889	0	1
1890-1899	5	6
1900-1909	1	7
1910-1919	1	8
1920-1929	3	11
1930-1939	26	37
1940-1949	18	55
1950-1959	34	89
1960-1969	46	135
1970-1979	73	208
1980-1989	21	229
1990-1999	32	261
2000-Present	22	**283**
	283	

Table 2. Earliest Records in Dataset.

Invasive Species	Year	Location	Herbarium	Earliest Known Year (North America) in Literature
Phragmites australis (Cav.) Trin. Ex Steud.	1913	Dulac, Louisiana	LSU	1910*
Rorippa nasturtium-aquaticum (L.) Hayek	1864	Centreville, Delaware	DOV	1847**
Lythrum salicaria L.	1888	Sulphur Springs, VA	DOV	1814***

* native genotype has been around U.S. for thousands of years, invasive status refers to introduced exotic genotype (Saltonstall, 2002). 1910 is when expansion of range was first noted in literature: Graves, C.B., Eames, E.H., Bissell, C.H., Andrews, L., Harger, E.B. and Weatherby, C.A. (1910) *Bulletin of the Connecticut Geological and Natural History Survey No. 14* (Case, Lockwood: Brainard, Hartford, CT).

**Earliest record found in literature; believed to be introduced several times since European settlement. *The Sea Grant Nonindigenous Species Site (SGNIS) (http://www.sgnis.org/publicat/papers/19p3.pdf)*

***Mehrhoff, L. J., J. A. Silander, Jr., S. A. Leicht, E. S. Mosher and N. M. Tabak. 2003.IPANE: Invasive Plant Atlas of New England. Department of Ecology & Evolutionary Biology, University of Connecticut, Storrs, CT, USA. (http://www.ipane.org)

Table 3. Sample Sizes for Associate Species-Corrected Method Study Areas.

Regional Study Area	Number of Invasive	Number of Combined Natives
Phragmites Area 1	55	72
Phragmites Area 2	29	75
Phragmites Area 3	69	160
Lythrum Area 1	42	86
Lythrum Area 2	106	138
Rorippa Area 1	50	172
Rorippa Area 2	69	141
Rorippa Area 3	62	102

Part 5. The occupation of new physiographic provinces by Baccharis halimifolia

Abstract.

Baccharis halimifolia (Asteraceae) is a deciduous, dioecious shrub found primarily along Atlantic coast habitats occupying the physiographic province of the Coastal Plain. In the past century, this native coastal species has exhibited a remarkable inland range expansion in the southeast. As a species with well-documented distributions, both in literature and in herbarium collections, *B. halimifolia* offers itself as a prime candidate to use the growing SERNEC data set to examine the modern range expansion of a native species that may be possibly defined as invasive in some areas of the southeast. Using herbarium records and GIS, this study describes the spread of *B. halimifolia* into new physiographic areas of the southeast and compares this to the distribution description of previous literature. From our earliest dataset record in 1863 to 1950 there were two physiographic provinces occupied (CP, Pd); from 1950 to the present an additional number of provinces and province sections have been documented (e.g., Mt and Ouachita). Due to our dataset strongly complimenting the distributions described in the taxonomic literature, and existing discrepancies involving relatively few county records, we believe recent range expansion into previously unoccupied physiographic areas has truly occurred for this species, rather than being an artifact of collection.

Introduction.

Range expansion of non-native and native invasive species.

The establishment of invasive plant species and their subsequent degradation of native habitats has been widely documented (Myers & Bazely, 2003; Mooney et al., 2005). In the southeastern United States, extensive habitat area has been transformed by non-native invasive plants such as *Pueraria lobata* (Willd.) Ohwi (Kudzu), *Lonicera japonica* Thunb. (Japanese honeysuckle) and *Ligustrum sinesne* Lour. (Chinese privet), among many others. The definition of an invasive species has often been elusive with arguments on whether to include native species with rapidly expanding ranges and densities (rather than being exclusive to exotic species) and exactly what factors cause "invasiveness" (Pysek et al., 2004). Native species that quickly expand their range by taking advantage of anthropogenic disturbances, such as land development and the subsequent creation of ruderal habitats, to expand outside of historical distributions can display characteristics associated with invasive species (Lavoie & Saint-Louis, 1999). Recently, members of the Tennessee Exotic Pest Plant Council (TNEPPC) proposed listing the native southeast coastal species *Baccharis halimifolia* L. as invasive due to its expansion into this previously unoccupied state (Estes, 2005).

Invasive plant species often expand their range by dispersing to these newly created ruderal areas and establishing satellite populations, or "nascent foci," which in the future can act as source populations for further expansion into areas previously outside of historic ranges (Moody & Mack, 1998). To study this phenomenon, the use of herbarium records have proven to be an effective data source for the creation and subsequent analysis of invasive distributions (Pysek and Prach, 1993; Stadler et al., 1998; Mihulka et al., 2001; Wu et al., 2005). GIS have also added a new ability to model and predict potential land areas at risk of being occupied by either exotic invasive species or by rapidly expanding ranges of natives species that show invasive characteristics (Peterson et al., 2003; Gillham et al., 2004; Goslee et al., 2006). Other research on invasive plant species distributions has focused on coupling the primary data source of herbarium records to GIS for spatiotemporally analyzing data (Weber, 1998; Saltonstall, 2002; Delisle et al., 2003; Lavoie et al., 2005; Salo, 2005; Vincent, 2005; Lavoie et al., 2007; Miller et al., In review).

Baccharis halimifolia L. species description.

Baccharis halimifolia (Asteraceae) is a deciduous, dioecious shrub found primarily along Atlantic coast habitats occupying the physiographic province of the Coastal Plain. The Coastal Plain is mostly sandy beaches with bordering estuaries and marshes and altitudes less than 500 feet; soils in this province vary with sand and peat along the coasts and alluvium in river floodplains (Hunt, 1967). This species is an early successional woody plant tolerant to a broad range of soil nutrient and salinity conditions along with the ability to survive periodic flooding and drought. Often growing to heights >2m, it can occupy various soil conditions with pH ranging from 3-9 and soil types composed of sand, clay or loam (Nesom, 2001; Westman et al., 1975). Although shade tolerant (Panetta, 1977), *B. halimifolia* has maximum growth and viable seed production (in the form of wind-dispersed achenes) in open, full light conditions (Panetta, 1979a). Buried seeds of this species can become dormant over long periods and still be viable, contributing to its continued presence in regularly disturbed habitats (Panetta, 1979b).

Historical distribution description.

In the past 100 years, *B. halimifolia* has exhibited a considerable expansion of its distribution beyond the Coastal Plain, particularly in the southeastern United States. The historical range of *B. halimifolia* has been described as limited to the coastal plain province of the eastern United States, from Massachusetts to Florida and Texas (Small, 1903) and later "low grounds inland" and rarely in adjacent provinces to the coastal plain (Small, 1933). Additional Arkansas counties were added to its distribution in 1952 (Gleason, 1952). The first significant spread of this species was mentioned in *Manual of the Vascular Flora of the Carolinas* (Radford et al., 1964) stating additional occupation of habitats such as old fields, woodlands and waste ground throughout the coastal plain and piedmont providences, along with the notation "…believed to have been restricted at one time to the outer coastal plain, but now widely spread inland." In the past 40 years, this species has also been collected in northern counties of Mississippi and Alabama. Recent research has documented the occurrence of *B. halimifolia* in Tennessee, particularly in disturbed habitats, and has raised the issue of this species being considered invasive for parts of the southeast (Estes, 2005; Ervin, In review).

Goals.

As a species with well-documented distributions, both in literature and in herbarium collections, *B. halimifolia* offers itself as a prime candidate to use the growing SERNEC data set to examine the modern range expansion of a native species that may be possibly defined as invasive in some areas of the southeast. Using herbarium records and GIS, this study aims to describe the spread of *B. halimifolia* in the last 100 years into new physiographic provinces of the southeast and compare this to the distribution descriptions of previous literature.

Materials and Methods.

Using an improved SERNEC dataset, specimen data of *Baccharis halimifolia* were gathered from 12 southeastern herbaria resulting in a dataset of 1,146 specimens. Contributing herbaria included NCU, USCH, UNAF, LSU, GA, MISS, HTTU, TENN, MO, FLAS, APSC, MISSA. The information gathered from each specimen included the county, state and year collected. This information was analyzed in GIS to determine the earliest recorded presence of *B. halimifolia* for each occupied county in the southeast. A GIS-based map of the physiographic divisions of the eastern U.S. (USGS website, 2008) was downloaded and overlaid on the county information to examine temporal periods of expansion into new physiographic provinces. As structured by Fenneman (1946), the physiographic makeup of the United States is based upon labeled provinces which each have a multitude of "Sections" consisting of more specific descriptions. This paper will focus on the broad physiographic provinces with specific sections of those provinces mentioned as appropriate. Using several methods, we attempted to describe the recent range expansion of *Baccharis halimifolia.* The first method consisted of a 10 year increment examination of the inland "front" distribution. The second method compared the newly occupied physiographic provinces in our data set as compared to distribution descriptions in the taxonomic literature. Interpretations of broad range descriptions in the literature were performed based upon reasonable measurements. For Small's 1903 description of "along coast or rarely inland", an interpretation of "inland" was performed by taking 30 random measurements, from Virginia to Florida, from the coast to the edge of the Coastal Plain (CP) province area. The 30 measurements were averaged and divided by two in order to have a consistent half-width of the CP which allowed for a reasonable "inland" interpretation of Small (1903). Subsequent analyses focused on state-specific, inland expansion over time.

Results and Discussion.

The first modern description of this species' range was that of J.K. Small (1903): "Along the coast or rarely inland, Massachusetts to Florida and Texas." The earliest record in our dataset was from 1863; from 1863 to 1903, our data show only eight occupied counties – all are along the coast with the exception of Columbia County, Florida (Figure 1). However, even though this county is non-coastal it is still within the measured half-width of the CP (87 km), therefore the data reflect Small's 1903 interpretation of the range of *B. halimifolia*.

Publications of Small (1933), Bailey (1949) and Rehder (1940) closely follow this original description with the coastal plain of Texas being first mentioned in 1933 by Small. However, our data show that other physiographic provinces, not mentioned in the literature, were occupied by B. halimifolia from this same time period. Multiple records show that B. halimifolia occupied several Piedmont counties from the mid/late 1930s such as Union and McCormick counties, SC, and Rockingham and Orange counties, NC (Figure 2).

Arkansas is first mentioned as part of this species' range in 1952 by Henry A. Gleason. The data show the earliest AR record occurring within the Coastal Plain in 1940 in Bradley Co., which is in the southern portion of the state. Additional records pre-1952 included Hempstead (1949) and Calhoun (1941) counties. All these counties are in southern Arkansas. In 1958, the range within this state began to expand to the middle and northern portions of the state (Figure 3). During the 1950s and 1960s, new physiographic provinces are occupied with movement outside of the CP in Arkansas into the Ouachita Province (Ouachita Mountain section) and subsequent records showing a more northward spread into the Ouachita Arkansas Valley section in 1966 (Figure 4).

Prior to Radford et al. (1964), the taxonomic literature mentions only the Coastal Plain province as being occupied. Radford et al. (1964), note the occurrence of *B. halimifolia* in the piedmont province for the first time (restricted to the states of VA, GA, FL and MS). However, according to the USGS physiographic map (based on Fenneman, 1946) the piedmont does not extend into FL or MS – this appears to be in error. In addition, there are multiple records in the Piedmont physiographic province from NC and SC well before the 1964 publication date. These discrepancies in Radford (1964) are of note, especially the noticeably absent mention of the Piedmont records in NC and SC in the Radford description (Figure 5).

The first mention of this species being both "far inland" and "weedy" occurs with Godfrey in 1988. Godfrey (1988) states, "Perhaps formerly restricted to near-coastal areas…Now aggressively and noxiously weedy in a wide variety of kinds of disturbed places, even far inland". This range description builds on the initial argument made by Radford et al. (1964) that *B. halimifolia* was spreading inland from the coast. In addition to a broader distribution description, Godfrey also makes the first statement found in the taxonomic literature for *B. halimifolia* being a noxious weed in some areas. Although herbarium specimens cannot easily be used to assist in arguments regarding whether a species is "weedy" or not, the data show collected specimens prior to the publication date of Godfrey as being "far inland" (Figure 6).

In Weakley's (2007) flora he updates the range description for *Baccharis halimifolia.* The occupied piedmont zones are correctly changed to the states of GA, NC, SC, and VA. In addition, our data also show a piedmont province record within Alabama that should be further examined (1966 Clay Co, AL). Of major importance in the Weakley (2007) description is the first mention of the range expansion into mountainous territory, "Mt (NC, SC)"; however, it is described as "rare in Mountains". The data reflect this recent expansion with records from Sevier Co., TN in 1988 (Blue Ridge Province / Valley and Ridge Province) and Floyd Co., GA (Valley and Ridge Province) in 1999 (Figure 7). Nearby to these mountain provinces, Blount County, AL in 1998 is the first site of occupancy of the Appalachian Plateaus Province (Cumberland Plateau section).

Of recent concern among southeastern botanists is the newly found expansion of *B. halimifolia* into Tennessee. The first reported mention of this species in TN was the lone specimen (FLAS #168643) from Sevier Co. in 1988. The next record occurs in 2002 in Giles and Rutherford counties which mark the expansion into the previously unoccupied Interior Low Plateau Province (Highland Rim and Nashville Basin sections, respectively). The western TN counties were surveyed in 2006 (Ervin, In review) producing many county records for this part of the state including Shelby, Fayette, Tipton, Haywood, Hardeman, McNairy and Dyer counties. These new county records collected in 2006 correspond well with the neighboring northward expansion in Mississippi (Figure 8).

Range expansion of this species is highlighted very well in Mississippi due to several factors including heavy collecting and a lot of progress toward specimen databasing of MISS and MISSA leading to a wealth of spatiotemporal data for that state. Analysis of occupied counties within Mississippi, and the earliest year each county was occupied, shows a clear northward expansion of this species within the state. The earliest Mississippi record in our data was in the coastal county of Harrison in 1947 (within the East Gulf Coastal Plain section), which is well within the specific mention of the state of Mississippi in the taxonomic literature (Radford et al., 1964). Prior to 1965, *B. halimifolia* did not occupy the northern half of the state. Furthermore, the nine counties that make up the NE corner of the state were unoccupied until 2006 (Figure 9). It should be noted that a lot of botanical collecting occurred in Mississippi, primarily by S.B. Jones and L.C. Temple, from the mid 1960s through 1970. Because herbarium specimens from Jones and Temple were generally collected from the same period as the observed *B. halimifolia* expansion through the middle portion of the state, the argument can be made that the interpreted spread into northern Mississippi from the mid-1960s is an artifact of collection. Furthermore, our data alone cannot refute the possibility that the northeastern corner of Mississippi was not occupied prior to 2006. However, while collection bias cannot be fully refuted, an examination of the collecting locations of Jones and Temple supports our interpretation. Both collectors have specimens consisting of a wide variety of species in Mississippi. While S.B. Jones collected primarily from the southern half of the state, L.C. Temple collected heavily from all over the state, including a large number of specimens from the northern and northeastern counties. A query of the MISS online database (www.herbarium.olemiss.edu) reveals that Temple collected over 75 *B. halimifolia* specimens throughout the state in addition to hundreds of other specimens consisting of a wide variety of species. Therefore, we believe that had the range of *B.*

halimifolia expanded into northeastern Mississippi prior to 1970, it would likely have been collected.

Conclusions.

This study successfully shows the feasibility of using herbarium records with GIS to evaluate the range change of a native species. Despite the occasional occurrence of county record gaps in the dataset, collaboration through SERNEC allowed for a large quantity of specimen data to be collected and compiled quickly. Therefore, the efforts of SERNEC and similar organizations should continue to support the concept of collective and standardized herbarium databases.

The argument for true range expansion of *B. halimifolia* in the past century, rather than an artifact of more extensive collecting and/or more detailed descriptions in the literature over time, was successfully shown through the use of a large data set from which to test the accuracy of the taxonomic literature. Both the literature and the data available via SERNEC document the range increase of this species. If the inferred range expansion of *B. halimifolia* seen within the distribution descriptions of the literature was false, the data would show consistent and significantly earlier county records within new physiographic provinces and sections as compared to the literature. Because the data strongly complimented the distributions described in the taxonomic literature, and existing discrepancies involved relatively few county records, we believe recent range expansion of this species has truly occurred as reflected through specimen data and subsequent taxonomic literature descriptions.

From the earliest record in 1863 to 1950 there were two physiographic provinces occupied by the dataset: all six sections of the Coastal Plain province and the Piedmont

Province. The data supports the taxonomic literature prior to 1950 which mentions only the Coastal Plain and "rarely adjacent provinces" (Small, 1933). In the past several decades, 1950 to the present, there has been a broad range expansion with five new physiographic provinces being occupied as shown by our data. Since 1950, *B. halimifolia* has spread into the following physiographic provinces: Ouachita, Interior Low Plateaus, Valley and Ridge, Blue Ridge, and the Appalachian Plateaus (Cumberland Plateau section). This large-scale expansion into new habitat types by a native species is unusual and warrants more research into the ecology of *B. halimifolia.*

The first mention of *B. halimifolia* being a "weedy" species was by Godfrey (1988). Since that time, several prominent southeastern botanists have noted personal observations of this species occupying ruderal habitats, specifically along roadways and newly logged or cleared land (L. Dwayne Estes, Austin Peay State University, personal communication 2007; A. Weakley, University of North Carolina, Chapel Hill, personal. communication 2008). Observers have described the occupation of these disturbed sites by *B. halimifolia* as occurring rapidly with populations being more dense than typically seen in non-disturbed sites. Although research is needed to corroborate these observations, the recent spread of *B. halimfolia* into new physiographic provinces and states, as seen in the results, suggest that this species possesses the characteristics of a potential invasive species for several states, including Tennessee.

Literature Cited.

Bailey, L.H. (1949) Manual of Cultivated Plants. Macmillan: New York.

Delisle, F., C. Lavoie, M. Jean, and D. Lachance. (2003) Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. Journal of Biogeography 30:1033-1042.

Ervin, Gary N. In review. Distribution, habitat characteristics, and new county-level records of *Baccharis halimifolia* L. on a portion of its present US range boundary. Southeastern Naturalist.

Estes, L. Dwayne. (2005) The vascular flora of Giles County, Tennessee. Sida 21: 2343-2388.

Fenneman, N.M. and D.W. Johnson. (1946) Physiographic divisions of the conterminous U.S.. USGS: Washington, D.C.

Gillham, J.H., A.L. Hild, J.H. Johnson, E.R. Hunt, D.T. Whitson. (2004) Weed invasion susceptibility prediction (WISP) model for use with geographic information systems. Arid Land Research and Management 18: 1-12.

Gleason, Henry A. (1952) Illustrated Flora of the Northeastern United States and Adjacent Canada: Volume 3. New York Botanical Garden Press: Bronx.

Gleason, Henry A. and Arthur Cronquist. (1963) Manual of Vascular Plants of Northeastern United States and Adjacent Canada. Van Nostrand: Princeton, New Jersey.

Godfrey, Robert K. and Jean W. Wooten. (1981) Aquatic and Wetland Plants of Southeastern United States: Dicotyledons. The University of Georgia Press: Athens, Georgia.

Godfrey, Robert K. (1988) Trees, Shrubs, and Woody Vines of Northern Florida and Adjacent Georgia and Alabama. The University of Georgia Press: Athens, Georgia.

Goslee, S.C., P.C. Peters, G.K. Beck. (2006) Spatial prediction of invasion success across heterogeneous landscapes using an individual-based model. Biological Invasions 8: 193-200.

Hunt, Charles B. (1967) Physiography of the United States. W.H. Freeman and Company: San Francisco.

Lavoie, Claude and Annie Saint-Louis. (1999) The spread of gray birch (*Betula populifolia*) in eastern Quebec: landscape and historical considerations. Canadian Journal of Botany 77: 859-868).

Lavoie, Claude, Caroline Dufresne and Fanny Delisle. (2005) The spread of reed canarygrass (Phalaris arundinacea) in Quebec: a spatio-temporal perspective. Ecoscience 12 (3): 366-375.

Lavoie, Claude, Yvon Jodoin and Albane Goursaud de Merlis. (2007) How did common ragweed (*Ambrosia artemisiifolia* L.) spread in Quebec? A historical analysis using herbarium records. Journal of Biogeography 34: 1751-1761.

Mooney, H.A., R.N. Mack, J.A. McNeely, L.E. Neville, P.J. Schei, J.K. Waage, eds. (2005) Invasive Alien Species. Island Press: Washington, D.C..

Mihulka, Stanislav and Petr Pysek. (2001) Invasion history of *Oenothera* congeners in Europe: a comparative study of spreading rates in the last 200 years. Journal of Biogeography 28: 597-609.

Miller, Ryan, Andrew Carroll, Thomas Wilson, and Joey Shaw. (In Review) Spatiotemporal analysis of three common wetland invasive plant species using herbarium specimens and geographic information systems. Castanea.

Myers, Judith H. and Dawn R. Bazely. (2003) Ecology and Control of Introduced Plants. Cambridge University Press: Cambridge.

Nesom, Guy. (2001) USDA Plant Guide: Groundsel Tree *Baccharis halimifolia* L. (http://plants.usda.gov/plantguide/pdf/pg_baha.pdf).

Panetta, F.D. (1977) The effects of shade upon seedling growth in groundsel bush (*Baccharis halimifolia* L.) Australian Journal of Agricultural Research 28: 681-90.

Panetta, F.D. (1979a) The effects of vegetation development upon achene production in the woody weed, groundsel bush (*Baccharis halimifolia* L.). Australian Journal of Agricultural Research 30: 1053-65.

Panetta, F.D. (1979b) Germination and seed survival in the woody weed, groundsel bush (*Baccharis halimifolia* L.). Australian Journal of Agricultural Research 30: 1067-77.

Peterson, A.T., M. Papes, D.A. Kluza. (2003) Predicting the potential invasive distributions of four alien plant species in North America. Weed Science 51: 863-868.

Pysek, Petr and Karel Prach. (1993) Plant invasions and the role of riparian habitats: a comparison of four species alien to central Europe. Journal of Biogeography 20: 413-420.

Pysek, P., D.M. Richardson, M. Rejmanek, G.L. Webster, M. Williamson, and J. Kirschner. (2004) Alien plants in checklists and floras: towards better communication between taxonomists and ecologists. Taxon 53: 131-143.

Radford, Albert E., Harry E. Ahles, and C. Ritchie Bell. (1964) Manual of the Vascular Flora of the Carolinas. The University of North Carolina Press: Chapel Hill, North Carolina.

Rehder, Alfred. (1940) Manual of Cultivated Trees and Shrubs. Macmillan: New York.

Salo, Lucinda. (2005) Red brome (*Bromus rubens* subsp. *madritensis*) in North America: possible modes for early introductions, subsequent spread. Biological Invasions 7: 165-180.

Saltonstall, Kristin. (2002) Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. Population Biology 99: 2445-2449.

Small, John Kunkel. (1903) Flora of the Southeastern United States. Published by the author.

Small, John Kunkel. (1933) Manual of the Southeastern Flora. The University of North Carolina Press: Chapel Hill, North Carolina.

Stadler, J., G. Mungai and R. Brandl. (1998) Weed invasion in East Africa: insights from herbarium records. African Journal of Ecology 36: 15-22.

United States Geological Survey (USGS). (2008) "Physiographic divisions of the conterminousU.S." (http://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml).

Vincent, Michael A. (2005) On the spread and current distribution of *Pyrus calleryana* in the United States. Castanea 70: 20-31.

Weakley, Alan S. (2007) "Flora of the Carolinas, Virginia, Georgia, and surrounding areas: working draft of 11 January 2007". (http://www.herbarium.unc.edu/WeakleysFlora.pdf).

Weber, Ewald. (1998) The dynamics of plant invasions: a case study of three exotic goldenrod species (*Solidago* L.) in Europe. Journal of Biogeography 25: 147-154.

Westman, W.E., F.D. Panetta, and T.D. Stanley. (1975) Ecological studies on reproduction and establishment of the woody weed, groundsel bush (*Baccharis halimifolia* L.: Asteraceae).

Wu, S., M. Rejmanek, E. Grotkopp, and J. DiTomaso. (2005) Herbarium records, actual distribution, and critical attributes of invasive plants: genus *Crotalaria* in Taiwan. Taxon 54:133-138.

Figures.

Figure 1. *Baccharis halimifolia* Records 1863-1903. all records located in counties along coast or slightly inland, as per Small 1903.

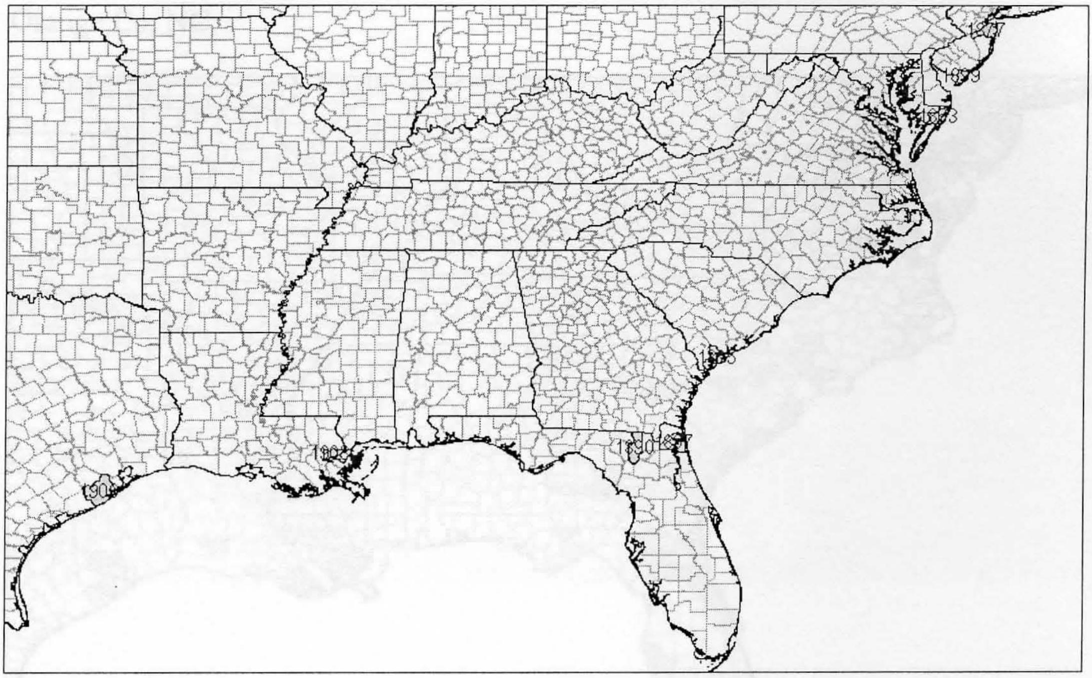


Figure 2. Inland county records of note. Several occupied inland counties are found in our dataset of note prior to mention in taxonomic literature, mostly occurring in the mid/late 1930s including: Greenville Co., SC, Wilkes Co., GA, Rockingham Co, NC, and Orange Co, NC.

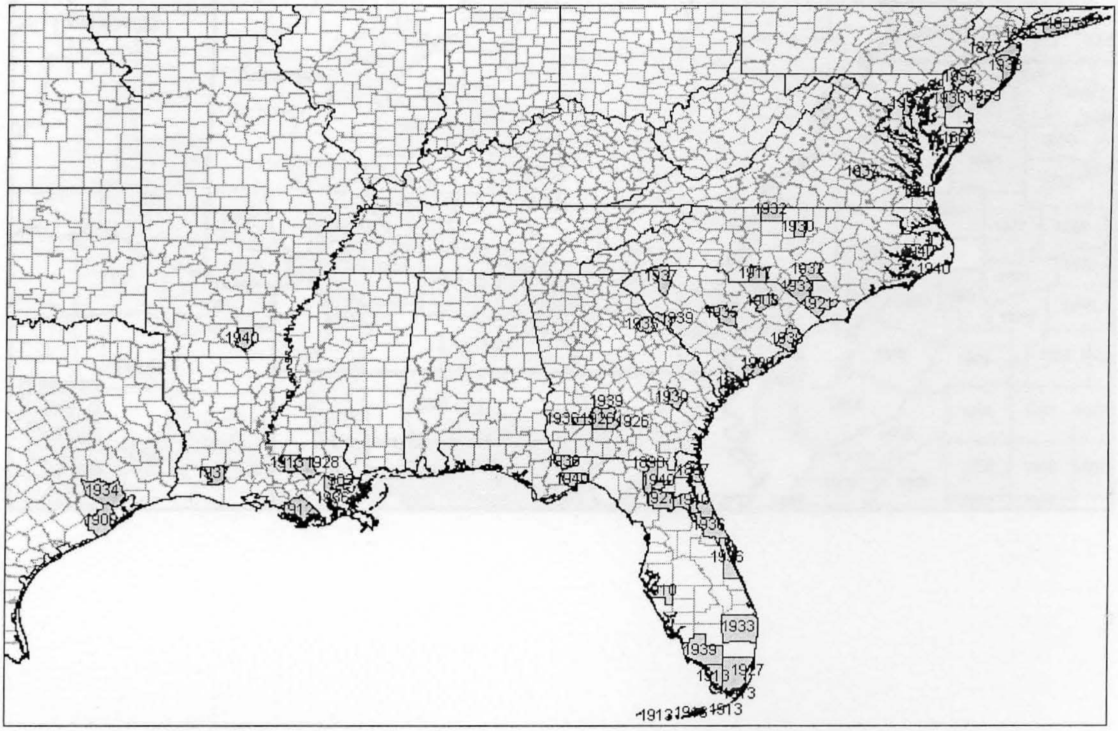


Figure 3. Expansion of *B. halimifolia* in Arkansas.

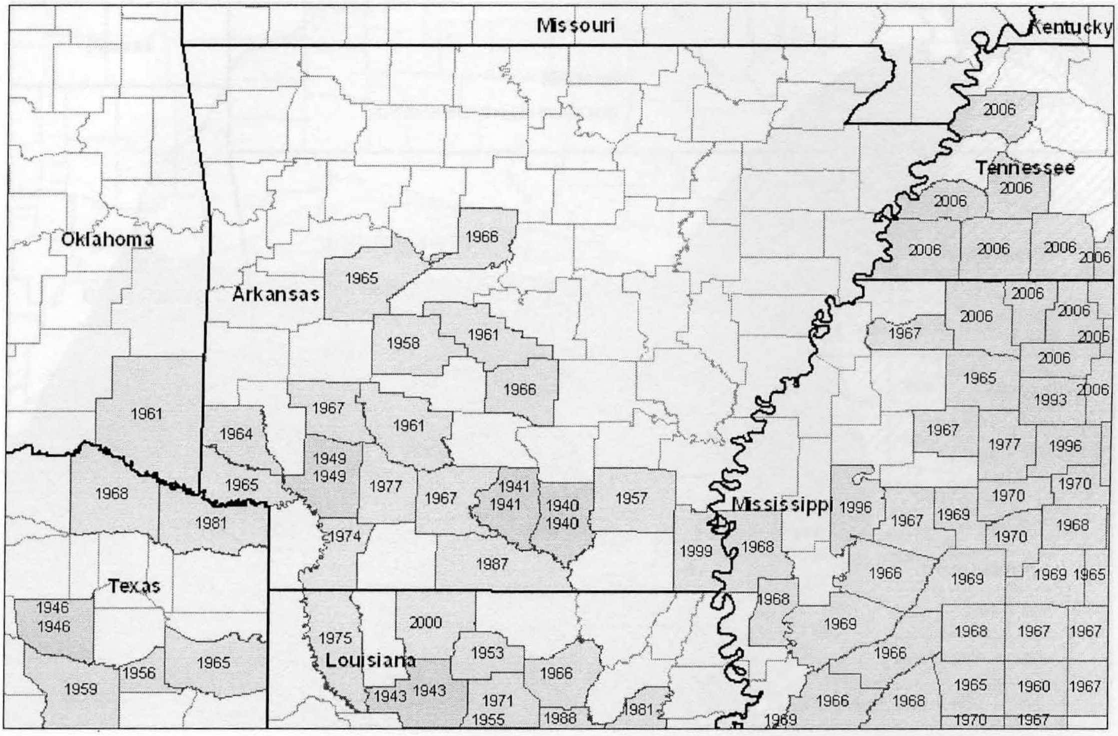


Figure 4. New occupied physiographic regions in Arkansas.

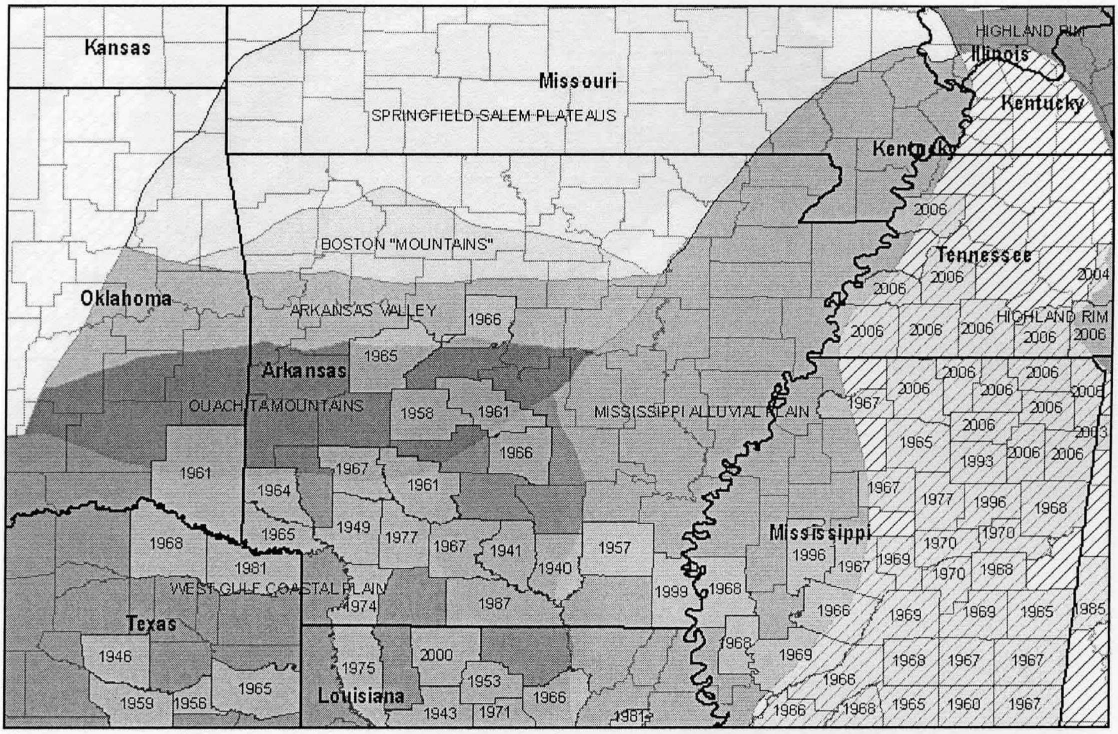


Figure 5. Discrepancies in piedmont provinces from Radford et al. 1964.

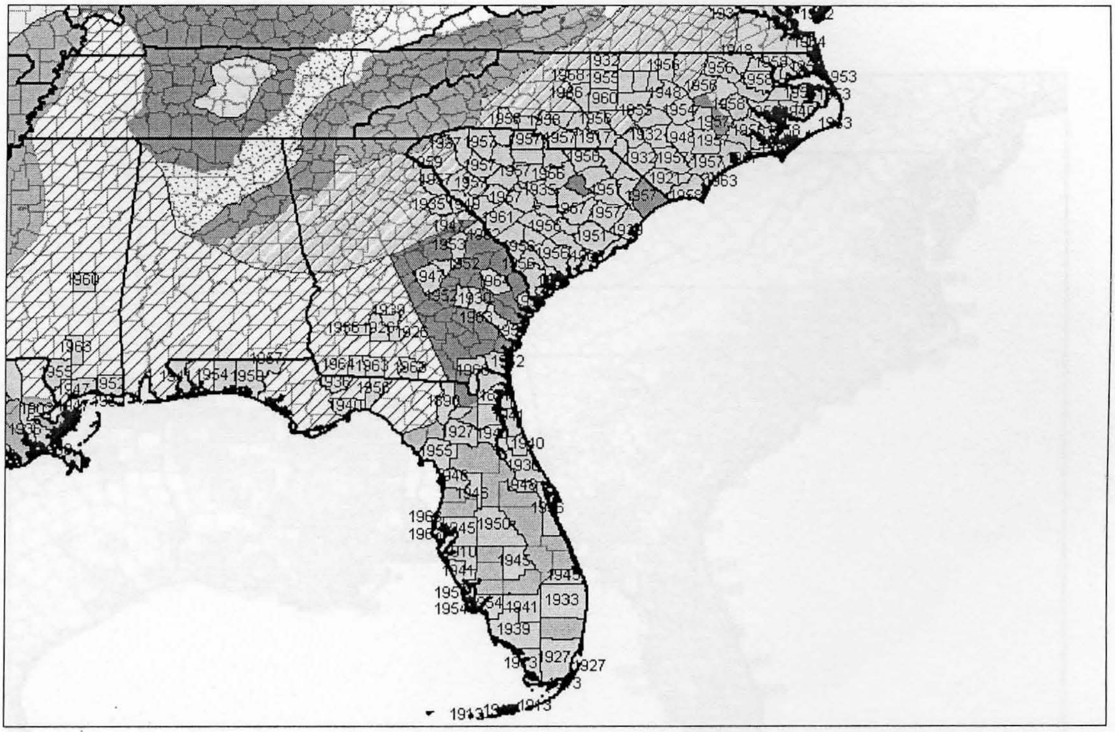


Figure 6. "Far inland" records supporting Godfrey (1988).

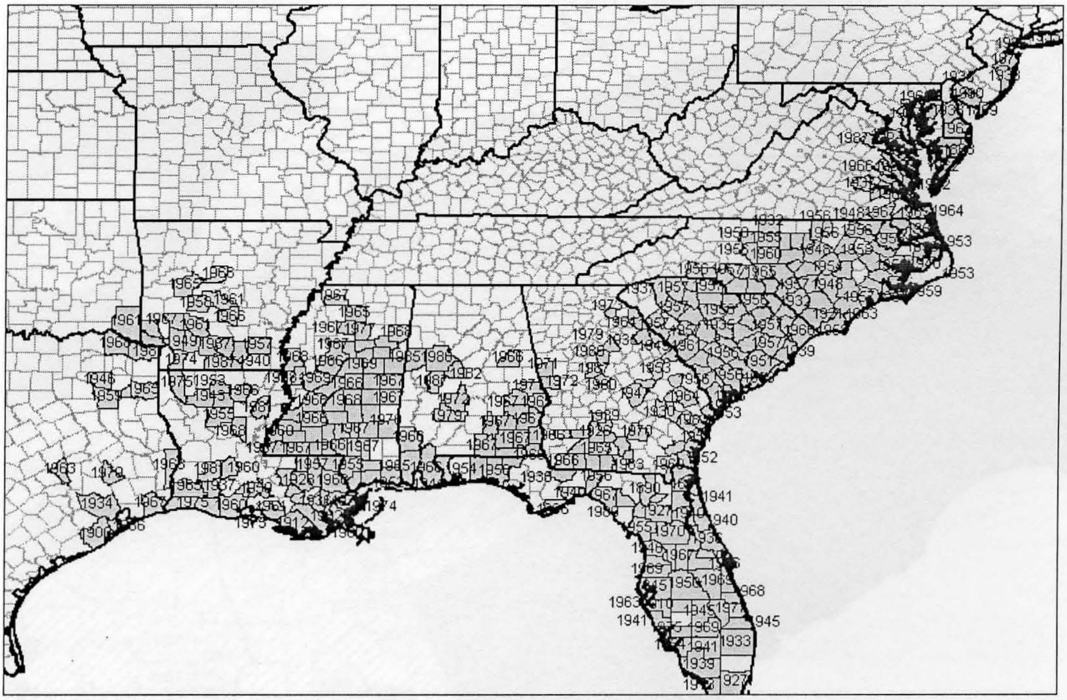


Figure 7. Weakley's correct description of NC/SC piedmont and new expansion into the Mountains (Blue Ridge) province.

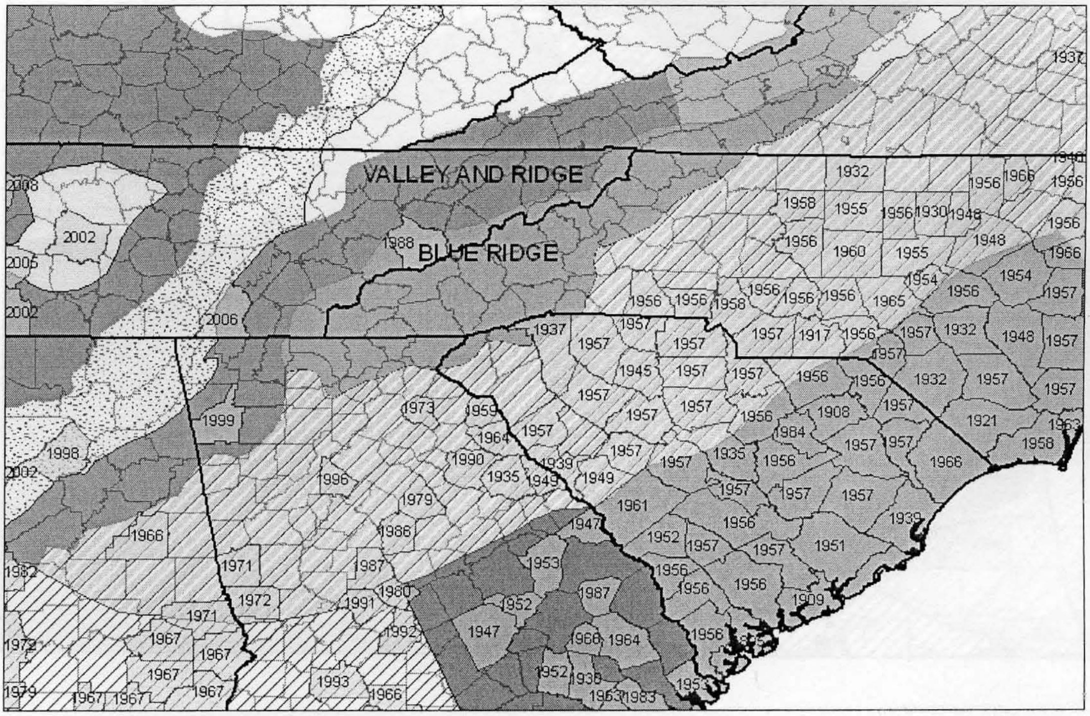


Figure 8. Expansion into Tennessee and unique physiographic provinces.

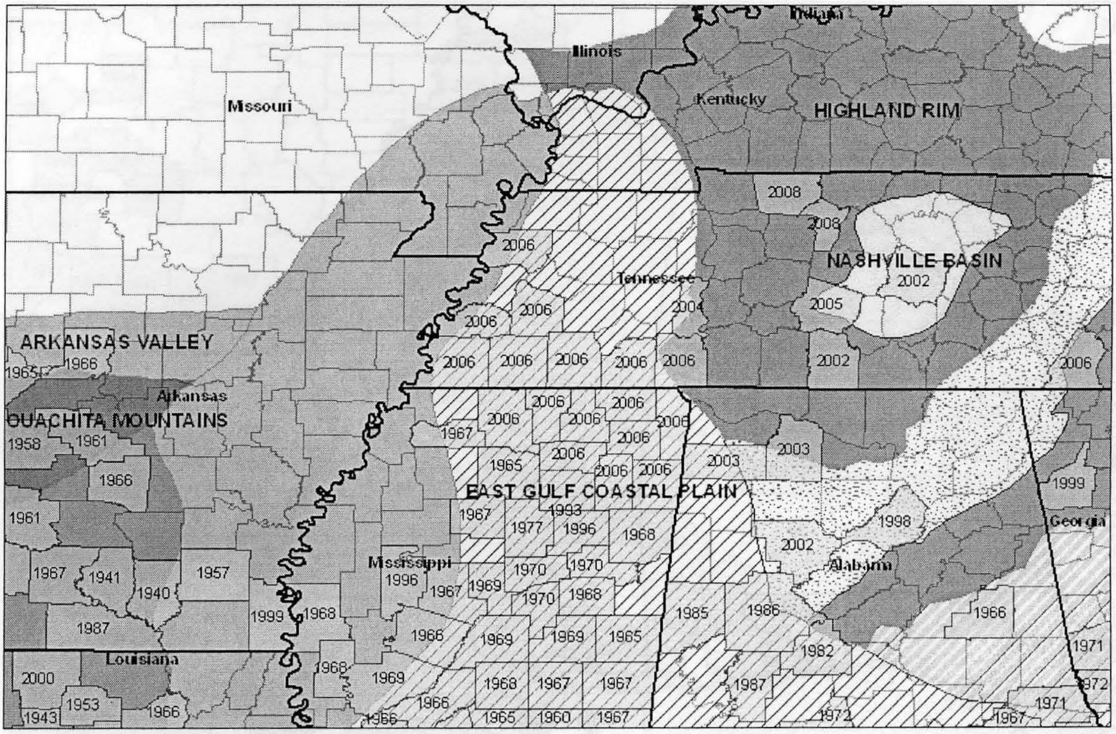
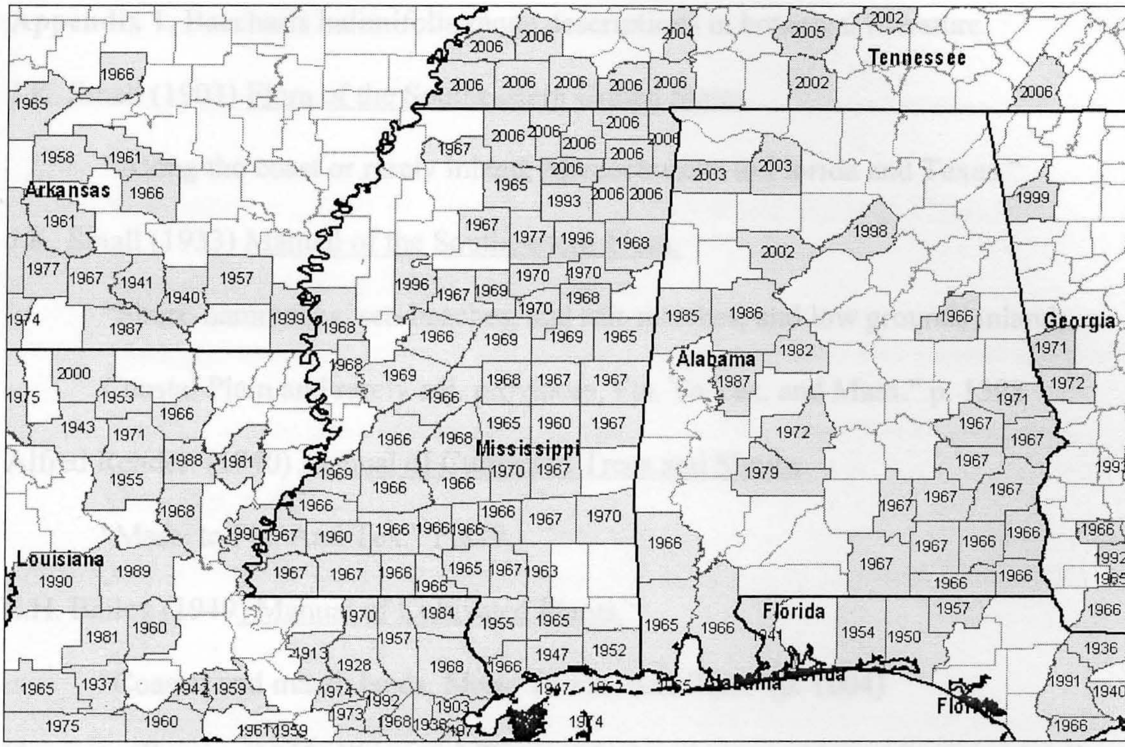


Figure 9. Northward expansion in Mississippi.



Appendix.

Appendix 1. Baccharis halimifolia range descriptions in botanical literature.

J.K. Small (1903) Flora of the Southeastern United States

"Along the coast or rarely inland, Massachusetts to Florida and Texas."

J.K. Small (1933) Manual of the Southeastern Flora.

"Shore-hammocks, sea-beaches, and salt-marshes, and low grounds inland, Coastal Plain and rarely adj. provinces, Fla. To Tex. and Mass." p. 1398

Alfred Rehder. (1940) Manual of Cultivated Trees and Shrubs.

"Mass. to Fla. And Tex." P.879

L.H. Bailey (1949) Manual of Cultivated Plants.

"Coastal and marsh lands, Mass. To Fla. And Tex." (p. 1004)

Henry A. Gleason (1952) Illustrated Flora of the Northeastern United States and Adjacent Canada. Volume 3.

"Marshes and beaches, especially near seashore, Mass. To Fla., Ark., and Tex.; also in the W.I." (p. 477).

Henry A. Gleason and Arthur Cronquist (1963) Manual of Vascular Plants of Northeastern United States and Adjacent Canada.

"Marshes and beaches, especially near the seashore; Mass. To Fla., Ark. and Tex.; W.I."

Albert E. Radford, Harry E. Ahles, and C. Ritchie Bell (1964) Manual of the Vascular Flora of the Carolinas.

"Old fields, woodlands and waste ground; throughout cp. And pied. [Va., Ga., Fla., Miss.] Believed to have been restricted at one time to the outer cp., but now widely spread inland." (p. 1067)

Robert K. Godfrey and Jean W. Wooten (1981) Aquatic and Wetland Plants of Southeastern United States: Dicotyledons.

"Marshes and banks of marshes, shores, swales, old fields and various disturbed places. Coastal plain and piedmont, Mass. to s. Fla., westward to Tex., Ark., Okla." (p. 839)

Robert K. Godfrey (1988) Trees, Shrubs, and Woody Vines of Northern Florida and Adjacent Georgia and Alabama.

"Perhaps formerly restricted to near-coastal areas, marshes, shores, swales, and the like. Now aggressively and noxiously weedy in a wide variety of kinds of disturbed places, even far inland; throughout our area. (Coastal plain and piedmont, Mass. to s. pen. Fla., westward to Tex., Ark.; W.I.)" p. 206

Alan S. Weakley (2006) Flora of the Carolinas, Virginia, Georgia, and surrounding areas.

"Cp, Pd (GA, NC, SC, VA), Mt (NC, SC): fresh and brackish marshes, marsh borders, hammocks, moist abused land, roadsides, ditches, old fields, and a wide variety of disturbed areas; common (rare in Mountains and VA Piedmont). September-October. Se. MA south to s. FL, west to TX, AR, and OK; West Indies."

PART 6. GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

The results of this research will be of interest to the botanical community both in the academic and public setting. As the amount of human development rises and natural plant populations decline, there will be a necessity for the understanding of plant distributions and the identification of areas in need of conservation. By developing a digital format in which herbarium data are freely accessible and made available to researchers in the fields of GIS, plant ecology, biogeography and conservation, a greater amount of data can be utilized to more accurately reflect the natural world. This ongoing research will allow UTC to be on the forefront of a major current interest in the field of botany as efforts of herbaria across the United States begin to digitize, georeference, and distribute specimen data for research and education (Appendix 1).

In the public arena, herbarium data that contain information relevant to status as species of special concern, flowering phenology, species distribution, or conservation needs could also be used by such community groups as garden clubs, botanical societies, and schools as a method to help plan field trips for when and where to go for specific viewing of certain plants or to aid in educating younger students about the plants of their community and the need to protect them.

Using GIS applications to take advantage of the large amount of raw data inherent in herbarium specimens is a novel, innovative technique which is just beginning to be tested. This research will act as a framework for future endeavors in using GIS to extract information from raw herbarium data for use in various categories of botanical research.

Future directions for this work, in relation to the goals of SERNEC and NBII, include additional evaluations and feedback of both the GIS infrastructure and pilot website from SERNEC collaborators. Recent suggestions regarding additional data to host on the website for analysis have included soil, precipitation, elevation, and physiographic layers. It is anticipated that along with the inclusion of additional ecological data, customized result displays, and the ever-growing SERNEC dataset, the GIS-enabled website produced in this thesis work will develop into a prime source for southeastern botanists from which research ideas can be developed and explored. In addition, it is hoped that the two ecological analyses discussed will advance the knowledge and understanding of these selected invasive plant species and also encourage more research using similar techniques.

Appendix.

Appendix 1. Description of Collaborative Agencies/Organization Hierarchy

The thesis work involved collaboration with several federal agencies and state universities and has been funded by NBII-SAIN. The goals set forward by NBII will be directed through the Southern Appalachian Information Node (SAIN). Coordination for this project outside of the UTC Graduate Committee involves Jean Freeney (SAIN Manager) and Fred Rascoe (SAIN Point of Contact/Knowledge Manager) as well as Dr. Zack Murrell (Appalachian State University, Department of Biology) who is the president of SERNEC and SHC.

USGS (United States Geological Survey)

"The USGS serves the Nation by providing reliable scientific information to describe and understand the Earth; minimize loss of life and property from natural disasters; manage water, biological, energy, and mineral resources; and enhance and protect our quality of life" (http://www.usgs.gov/aboutusgs/, 2006).

NBII (National Biological Information Infrastructure)

"The National Biological Information Infrastructure (NBII) is a broad, collaborative program to provide increased access to data and information on the nation's biological resources. The NBII links diverse, high-quality biological databases, information products, and analytical tools maintained by NBII partners and other contributors in government agencies, academic institutions, non-government organizations, and private industry. NBII partners and collaborators also work on new standards, tools, and technologies that make it easier to find, integrate, and apply biological resources information. Resource managers, scientists, educators, and the general public use the NBII to answer a wide range of questions related to the management, use, or conservation of this nation's biological resources" (http://www.nbii.gov/about/, 2006).

SAIN (Southern Appalachian Information Node)

"The Southern Appalachian Information Node (SAIN) is a consortium of public and private partners who work together to build and maintain an integrated information system that serves as the gateway to regional biological information to enable the use of science and information technologies for decision-making, sustainable development, research, education, and outreach. SAIN provides web-based access to regional information resources, data, expert lists, maps, and educational tools to ensure that any development or use of the area is approached in a thoughtful and informed way without disrupting the essence of human life in Appalachia or compromising the region's biodiversity" (http://sain.nbii.org/about.shtml, 2006).

SERNEC (Southeast Regional Network of Expertise and Collections)

"SERNEC is an organization devoted to making the resources of these nearly 150 regional Herbaria of the Southeast available online, in concurrence with developing global standards, so that all available data can then be studied regionally or globally as one virtual, researchable collection. *SERNEC* will improve access to specimen data of a richly biodiverse ecological environment, and provide a platform for herbarium curators and plant scientists to exchange ideas, share expertise, and benefit from the value of information shared across institutions" (http://lugh.sunsite.utk.edu/drupalsites/SERNEC/, 2006).

SHC (Society of Herbarium Curators)

"The purpose of the society shall be to promote and expand the role of herbaria in botanical research, teaching, and service to the community at large, to provide a forum for discussion and action on all issues confronting herbaria, and to extend its efforts and interject its influence toward the protection and preservation of endangered herbaria. In particular, regional networks will be used to reach out to groups that have been historically underrepresented in the botanical and conservation communities, to land managers and state and federal agencies, and to the K-12 students and teachers" (http://serenity.sunsite.utk.edu/shc/?q=node/2, 2007).

VITA

Ryan Miller was born in Dayton, Ohio on March 21, 1980. He graduated from Archbishop Alter High School (Kettering, Ohio) in 1998. From there he attended the University of Dayton and graduated cum laude with a Bachelor of Science degree in Environmental Biology. After graduation in 2002, he moved to Tucson, Arizona and worked as a Research Assistant for two years in the School of Natural Resources at the University of Arizona. In 2004, he moved back to Ohio and accepted a position as a Lab Technician at Data Chem Laboratories in Cincinnati. In 2005, he was accepted to graduate school at the University of Tennessee at Chattanooga and expanded his knowledge and interests in botany and geographic information systems. Currently, he is employed as a Senior Geospatial Analyst with Wiser Company in Murfreesboro, Tennessee.