12-2014

Applications of a New Geodata Crawler for Landscape Ecology: From Mapping Natural Stream Hydrology to Monitoring Endangered Beetles

Douglas Ryan Leasure
University of Arkansas, Fayetteville

Applications of a New Geodata Crawler for Landscape Ecology:
From Mapping Natural Stream Hydrology to Monitoring Endangered Beetles

Applications of a New Geodata Crawler for Landscape Ecology:
From Mapping Natural Stream Hydrology to Monitoring Endangered Beetles

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Biology

by

Douglas Ryan Leasure
University of Arkansas – Fort Smith
Bachelor of Science in Biology, 2009

December 2014
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

Dr. Daniel D. Magoulick
Dissertation Director

Dr. Gary R. Huxel
Committee Member

Dr. Steven L. Stephenson
Committee Member

Dr. Jason A. Tullis
Committee Member

Abstract

An investigation of endangered American burying beetle (*Nicrophorus americanus*) ecology led to development of a Geodata Crawler with applications in eco-hydrology. Geodata Crawler includes a national GIS (geospatial information systems) database with layers that quantify climate, land cover, soils, human development, and other attributes of the biosphere. For user-locations in the continental United States, Geodata Crawler can rapidly tabulate site-specific statistics within automatically delineated sample areas: points, site radii, watersheds, and riparian zones, among others. Geodata Crawler supported a multi-scale analysis of *N. americanus* habitat at a military installation in western Arkansas to produce a Landsat-based monitoring tool. Royle's N-mixture model was used to simultaneously account for 1) the detection process associated with baited pitfall traps, and 2) the ecological processes driving spatial patterns of abundance. Detection rates of *N. americanus* averaged 20% and were optimized at about 29º C on nights with high humidity and slight wind. Effective sample radii assessed using marked beetles released at known locations were no more than 800 m, and detection rates dropped below 5% beyond 400 m. *Nicrophorus americanus* abundance was associated with native grasslands and open-canopy oak woodlands with rolling topography, sandy loam soils, and moderate disturbances from wildfire. Habitat measured within 800 m site radii produced best fitting models compared to 100 or 1600 m radii. A new above-ground bucket trap was evaluated in comparison to standard pitfall traps. Compared to standard pitfall traps, above-ground bucket traps were safer for beetles, more resistant to scavengers, and more time-efficient for workers to install. The first application of Geodata Crawler for aquatic ecology was an eco-hydrology project that identified seven natural flow regimes of the Ozark-Ouachita Interior Highlands based on daily hydrological data from 64 reference streams.

Geodata Crawler quantified climate and catchment characteristics necessary to predict natural flow regimes of 24,557 un-gaged stream segments. This dissertation demonstrated the utility of Geodata Crawler for eco-hydrology and species distribution modeling. Development will continue to expand potential applications to include landscape genetics and climate change, and also to support web-based project submission, cluster computing, and FTP-based data retrieval.

Acknowledgements

This dissertation would not have been possible without contributions from many people. My graduate committee including Dan Magoulick, Gary Huxel, Jason Tullis, and Steven Stephenson provided constructive criticism throughout. Dr. Magoulick first introduced me to eco-hydrology and provided me a valuable foundation in experimental design and statistical analysis. Dr. Huxel provided me a broad foundation in ecology by encouraging exploration of the literature, and he sparked my interest in quantitative ecology. Dr. Tullis provided valuable training in GIS and remote sensing and first exposed me to Python programming. Dr. Stephenson trained me in forest ecology and field-based vegetation assessment. Scott Longing and Pablo Bacon provided initial motivation to develop Geodata Crawler with their excitement for aquatic insects like the Sulphur Springs diving beetle (*Heterosternuta sulphuria*). None of my work with *N. americanus* would have been possible without support from employees at Fort Chaffee Maneuver Training Center, particularly Dan Farrer, Beth Phillips, Dustin Opine, and Doug Cobb at the Environmental Branch, and military personnel at the Range Control office, particularly Sergeants Swad, Richmond, and Mickelson. *N. americanus* field work was also supported by FTN Associates, Ltd., particularly David Rupe, Jeremy Rigsby, Blake Griffis, and Robert Gregory. Funding for *N. americanus* field work was provided by the Arkansas Military Department through Fort Chaffee Environmental Branch. Arkansas Game and Fish Commission funded our eco-hydrology work with two State Wildlife Grants. Throughout graduate school, my family provided constant support and encouragement that kept me focused.

Table of Contents

List of Tables

are Groundwater Stable (GS), Groundwater (G), Groundwater Flashy (GF), Perennial Runoff (PR), Runoff Flashy (RF), Intermittent Runoff (IR), and Intermittent Flashy (IF).

Chapter IV: Geodata Crawler: A centralized national geodatabase and automated multi-scale data crawler

Table 1. These spatial scale designations are found at the beginning of column headings in Geodata Crawler's output. # represents the user option LOCAL_RADIUS, and * represents the user option RIPARIAN_BUFFER_WIDTH (see section User Options).

Conclusion:

Table 1. Sets of non-redundant flow metrics selected to represent ecologically important components of each natural flow regime. Lists contain a flow metric to represent each of the top three principal components with eigenvalues greater than one. Underlined metrics were selected for cluster analysis. Metrics in bold were considered the best descriptors (i.e. highest component loading on the 1st principal component) for nine ecologically important aspects of each flow regime. Metrics in parentheses had equal component loadings.

List of Figures

Figure 4. The local-watershed spatial scale includes all land areas that drain into a user-defined location, but that are also within some user-defined radius *x* of sites.

Figure 5. The riparian spatial scale includes all that areas within a site's watershed, but that are also within some user-defined distance from streams (see Fig. 6).

Figure 6. The riparian spatial scale is delineated based on a user-defined stream buffer distance *x*.

Figure 7. The local-riparian spatial scale is the intersection of the local and riparian scales.

Figure 8. The local-riparian spatial scale includes all areas within a site's watershed that are within a user-defined radius from sites *x* and also within a user-defined distance from streams *y*.

Conclusion:

Figure 1. Predicted recapture probabilities as a function of release distance. Sample ranges appeared to be similar between trap methods, but sample sizes were too small to support robust conclusions. Recapture rates were less than 25% for both methods, even when beetles were released nearby.

Figure 2. Modeled relationships between observation covariates and detection probabilities. Thick lines are model predictions when other covariates are held to their means and thin lines are 95% confidence intervals. Scatter plots show model predictions using site specific values of all covariates each year. When site-specific model predictions deviate from the line, it is due to the influence of other covariates in the model, not model error. The gray regions indicate the range of optimal flight temperatures reported by Merrick and Smith (2004) for Nicrophorus hybridus.

Figure 3. Layout of a standard pitfall trap transect showing an 800 m trap sample range (USFWS estimate) and a 200 m trap sample range in comparison to 20 m trap spacing to illustrate the lack of independence among traps in a transect.

Figure 4. Average differences in N. americanus abundance estimates between methods using various sample effort conversion rates (n = 20). Normalized average differences are mean differences divided by their standard deviations to provide a standardized scale for comparisons because abundance estimates—and therefore differences between them—are inherently smaller when trap-nights are artificially increased.

Figure 5. Spatio-temporal dynamics of three site covariates and abundance model predictions at Fort Chaffee. The habitat model holds the YEAR factor constant at "2007"

while the abundance model allows YEAR to vary. Red corresponds to low values and blue corresponds to high values in all maps.

Figure 6. Suggested improvements to above-ground bucket trap design using a wooden rain cover and landing pad to reduce disturbances from scavengers, reduce maintenance, and increase bait-life.

Figure 7. Natural flow regimes of 64 reference gages were identified using mixture-model cluster analysis based on 10 flow metrics. Natural flow regimes of all stream segments were predicted based on climate and catchment characteristics using a random forest model.

Figure 8. Distribution of measurement uncertainties (bias, precision, and accuracy) among 170 flow metrics for increasing period of record lengths from 1 to 30 years. Each data point represents average uncertainty for a given flow metric among nine streams analyzed.

Figure 9. Measurement uncertainties in terms of bias, precision, and accuracy for 170 flow metrics based on a 15 year period of record. Bars represent average values ± one standard deviation among the nine streams analyzed.

List of Papers

Leasure DR. 2014. Landsat-based monitoring of an endangered beetle: Addressing issues of high mobility, annual life history, and imperfect detection. *In review* for *Landscape Ecology*.

Leasure DR, Magoulick DD, Longing SD. 2014. Natural flow regimes of the Ozark-Ouachita Interior Highlands region. *In press* for *River Research and Applications*.

Leasure DR, Rupe DM, Phillips EA, Opine DR, Huxel GR. 2012. Efficient new above-ground bucket traps produce comparable data to that of standard transects for endangered American burying beetles (Silphidae: *Nicrophorus americanus* Olivier). *The Coleopterists Bulletin* **66**(3):209-218.

Introduction

Simon Levin (1992) argued that "the problem of pattern and scale is the central problem in ecology, unifying population biology and ecosystems science, marrying basic and applied ecology." Measurements of any patterns are dependent on the spatial and temporal scales of observation, and ecological processes often have multiple components operating at different scales. A focus on how multi-scale landscape patterns affect population and community processes has a rich tradition in ecological research (MacArthur & Wilson 1967, MacArthur 1972, Pickett & White 1987, Hanski 1999, Turner *et al.* 2001, Hubbell 2001, Manel *et al.* 2003). Technological advances in geographic information systems (GIS) and remote sensing are now providing unprecedented amounts of high quality data that can be used to study ecological phenomena at multiple spatial and temporal scales. This, along with increased accessibility to powerful analytical techniques like machine learning and Bayesian statistics, that can accommodate these often high-dimensional datasets, has resulted in a flurry of research activity in several ecological sub-disciplines like climate change, landscape ecology, eco-hydrology, and landscape genetics. These are inherently scale-dependent areas of study, and spatial scale is often the primary difference in how they utilize GIS and remote sensing data. For example, projects in all disciplines may require information about forest cover on the landscape, but eco-hydrology may be interested in forest cover within watersheds, landscape ecology may be interested in various site radii, and landscape genetics may be in interested in paths connecting sites. Although availability of GIS and remote sensing data has drastically increased in recent years, it can be difficult to acquire and process geodata to generate site-specific tabular data at an appropriate spatial scale(s) for the question being addressed. This not only limits the number of sites and variety of GIS data used in many studies, but it can also discourage interdisciplinary collaboration.

Geodata Crawler is a centralized national geodatabase and automated multi-scale data crawler that can rapidly build project-specific geodatabases, delineate multi-scale sample areas at user-locations anywhere in the continental United States, and tabulate data from within these sample areas (see Chapter IV). Geodata Crawler's national geodatabase currently includes datasets such as land cover, soils, topography, hydrology, and climate, and new datasets are regularly added. It can delineate site-specific sample areas using several spatial scales: point, local (site radius), watershed, riparian, local-watershed, local-riparian, and stream paths or linear paths connecting sites. This new tool was initially developed in support of data collection required for this dissertation, but it now provides a template for a broader GIS data serving system that could provide rapid access to customized site-specific data at multiple spatial scales for user's with little or no GIS experience. Geodata Crawler's development began with an investigation of American burying beetle spatial ecology that required data collection using multiple site radii (see Chapter II), and as research interests broadened to include aquatic beetles, eco-hydrology, and gene flow, Geodata Crawler development continued with improved processing efficiency, inclusion of additional national GIS datasets, and development of new spatial scales for data collection (*e.g.* watersheds, riparian zones, and stream paths).

American Burying Beetle

The American burying beetle (Silphidae: *Nicrophorus americanus* Olivier) was placed on the endangered species list in 1989 due to a drastic range contraction in the late 19th and early 20th centuries (Raithel 1991, Sikes & Raithel 2002). Although once found throughout most of the United States east of the Rocky Mountains, *N. americanus* are now known from only three regions: Oklahoma (and Arkansas), Nebraska, and Block Island, Rhode Island (Fig. 1). *N. americanus* is a highly mobile annual species that must fly in search of rat- or quail-sized carcasses small enough to be buried or moved into underground brood chambers, and large enough to adequately provision developing larvae (Scott 1998, Kozol *et al.* 1988). Adult beetles can feed on carcasses of any size, so reproductive carcasses are assumed to be the most limited resource on the landscape (Raithel 1991, Sikes & Raithel 2002). The federal strategy to conserve remaining *N. americanus* populations historically relied on trap and relocation efforts prior to large habitat disturbances in counties where the species was known to occur (*e.g.* USFWS 2007). This strategy was adopted *in lieu* of a habitat-based conservation strategy because the most-important habitat feature, availability of suitably-sized carcasses for reproduction, is difficult to assess and manage. Efforts to identify vegetation communities related to beetle abundance have produced conflicting results (see Sikes & Raithel 2002 for a review). Despite these difficulties, recent conservation efforts have shifted towards a habitat-based strategy (USFWS 2014a). This dissertation (Chapters I and II) will support these efforts by evaluating standard trap methods and assessing habitat associations with particular emphasis on spatial scale and detection probabilities to provide methodological recommendations, habitat descriptions for an *N. americanus* population in Arkansas, and a remote sensing-based monitoring tool.

Adult burying beetles fly in search of carrion at night when temperatures are 15 - 35^{o} C, and optimal flight temperatures are around 25^{o} C (USFWS 2014b, Bedick 1999, Raithel 1991, Merrick & Smith 2004). Suitable reproductive carcasses for *N. americanus* are 80 to 200 g animals that have not yet been colonized by fly larvae (Kozol *et al.* 1988, Raithel 1991). Larger carcasses are associated with increased fecundity (Kozol *et al.* 1988), but larger carcasses are more difficult to bury and to defend from flies, fungi, and bacteria (Scott 1998, Wilson & Fudge 1984, Scott & Traniello 1990).

When a suitable reproductive carcass has been located, beetles release pheromones to attract potential mates, but competitors are also attracted (Trumbo & Bloch 2002, Müller & Eggert 1987). Outcomes of direct competitive interactions among burying beetles often depend on body size, and larger beetles generally win (Otronen 1988). The American burying beetle is the largest *Nicrophorus* species in North America, so it should generally prevail in direct competitions for reproductive carcasses with other burying beetles. However, it may lose resources to smaller species through exploitative competition, when smaller beetles locate and bury carcasses before *N. americanus* arrives (Mathews 1995). Exploitative competition may be facilitated by inter-specific differences in phenology and diel patterns of flight activity based on temperature and light conditions (Wilson *et al.* 1984). Further niche-segregation among *Nicrophorus spp.* may occur based on carcass size such that larger species are better able to bury and preserve larger carcasses, and smaller species can have higher reproductive output on small carcasses (Kozol *et al.* 1988, Trumbo 1990). Parents must often defend their brood chambers from being usurped by larger burying beetles that would kill their offspring and utilize brood chambers for their own reproduction (Scott 1990).

There is also intense competition from vertebrate scavengers, flies, ants, and other scavenging insects (Scott *et al.* 1987, Scott 1994, Trumbo 1990). Competition from vertebrate scavengers like raccoons (*Procyon lotor*), skunks (*Mephitis mephitis*), and opossums (*Didelphis virginianus*) pressure *N. americanus* to quickly bury carcasses. In addition to competing for resources, opossums have been recorded preferentially eating adult *N. americanus* at a carcass (W. Hoback *unpublished data*). Even after burial, resources must continue to be defended from other scavenging insects like rove beetles (Staphlynidae) and ants. Ants can render carcasses unusable for burying beetles, and they can kill burying beetles in traps (*personal observation*). Burying beetles avoid carcasses that have already been colonized by fly larvae, and they actively remove fly eggs from brood chambers. Adult beetles even have a mutualistic relationship with phoretic mites (*i.e. Poecilochirus necrophori* Vitz.) that eat fly eggs in the brood chamber (Springett 1968, Wilson & Knollenberg 1987).

After a group of burying beetles have arrived at a suitable carcass, direct physical competitions will occur among individuals of each sex. The pair of beetles that emerge victorious, and sometimes a few remaining intra-specific competitors, will cooperate to bury the carcass or move it into an existing animal burrow nearby (Trumbo *et al.* 1994, Wilson & Fudge 1984, Smith *et al.* 2000). As a mating pair of beetles bury a carcass, they strip it of fur or feathers and preserve it using hindgut secretions (Scott 1998, Hoback *et al.* 2004). Eggs will be deposited in the soil nearby and when larvae eclose they will feed on the carcass for about a week before pupation. Throughout larval development, one or both parents may provide extended care by returning to the brood chamber with regurgitated meals (Scott 1998). Parents regularly clean the brood chamber of fly eggs, maggots, and fungi, the primary competitors with larval *N. americanus* (Wilson 1983, Scott, 1994, Scott 1998). Parental care, particularly bi-parental care that often occurs in burying beetles, is extremely rare in non-social insects (Scott 1998). About 45 to 60 days after brood initiation, larvae emerge from pupation as adult beetles capable of flight and reproductively active. They must locate a reproductive carcass and secure a mate within one year to complete their life cycle, and they can only search at night when temperatures are about 15 to 35^{o} C (USFWS 2014b, Bedick 1999, Raithel 1991, Merrick & Smith 2004).

Many hypotheses have been proposed to explain the decline of *N. americanus* including DDT/pesticide use, light pollution, pathogens, competition, or habitat loss. Perhaps the most likely cause was declining availability of reproductive carcasses on the landscape either due to loss of suitable vegetation communities for carcass-producing species, an increase in competition from vertebrate scavengers, or both (Sikes & Raithel 2002). There has been debate over which vegetation communities are associated with *N. americanus* abundance, and hypotheses have been proposed claiming the species is a forest specialist (Anderson 1982, Walker 1952, Lomolino & Creighton 1996), a prairie specialist (Kozol *et al.* 1988, Bedick *et al.* 1999), or a generalist (Lomolino *et al.* 1995). As pointed out by Sikes and Raithel (2002), we should not expect this species to be specialized on any particular vegetation community since it selects habitat based on availability of appropriately-sized carrion. Unfortunately, carrion availability is not amenable to direct management and is difficult to quantify at large spatial scales relevant to *N. americanus* habitat selection. For these reasons, it is important to identify vegetation communities, or other manageable site characteristics, associated with *N. americanus* abundance to increase effectiveness of habitat conservation and restoration efforts. Although suitable vegetation communities may not be consistent throughout the range of *N. americanus*, some regional consistency is expected from vegetation communities associated with species capable of producing adequate reproductive carcasses.

There are a few confounding factors that may have contributed to difficulties describing *N. americanus* habitat that will be addressed here (Chapters I and II):

1. The large spatial scale of *N. americanus* habitat selection associated with its strong dispersal ability makes it difficult to measure habitat in the field;

2. *N. americanus* has potential to respond to ephemeral landscape characteristics like wildfires that are difficult to quantify in a single year and that are not well represented by temporally-static land cover maps;
3. Potentially inconsistent sampling efficiency associated with baited pitfall traps due to their large sample ranges and low detection probabilities may confound site-abundance estimates; and
4. Measurement error may be introduced by standard *N. americanus* data handling procedures that do not acknowledge large sample radii associated with counts from baited pitfall traps.

Field studies have generally focused on habitat in the immediate vicinity of trap locations, but individual *N. americanus* can move several kilometers in a single night (*personal observation*, Creighton & Schnell 1998, Bedick *et al.* 2004, Creighton *et al.* 1993). This may necessitate habitat assessments at larger spatial scales than are possible in the field, but optimal spatial scales are currently unknown. GIS data and satellite-imagery provide ideal data for assessing *N. americanus* habitat at large spatial scales to identify important habitat features and appropriate spatial scales for conservation. Satellite images provide the ability to measure vegetation conditions annually so that ephemeral effects of wildfires and other disturbances can be quantified.

Factors that may affect detection of *N. americanus* with baited pitfall traps are not well understood. Baited pitfall traps rely on *N. americanus* aerial foraging behavior, and since flight in burying beetles is temperature dependent (Merrick and Smith 2004), it is reasonable to assume that detection rates may vary with temperature. Wind may also affect detection rates because it is known to affect insects' abilities to track odor plumes (Elkinton *et al.* 1987, Murlis *et al.*

1992), and it likely increases cooling rates of burying beetles during flight (Merrick & Smith 2004). Failure to account for differences in detection rates achieved on different sampling occasion may bias site-abundance estimates and confound investigations of habitat associations. Recent advances in occupancy modeling, particularly Royle's N-mixture models of abundance and detection, provide an ideal modeling framework to address potentially biased site-abundance estimates due to low and inconsistent detection probabilities, and to identify important factors that may affect detection rates.

A better understanding is needed of the spatial scale associated with count data from baited pitfall traps to ensure that data from various trap methods are comparable and to minimize measurement error associated with how trap data are handled. There are several baited pitfall trap designs approved for *N. americanus* trapping. All traps are baited with rotten chicken or other meat, and they are set for at least three nights (USFWS 2014b). Some methods use multiple traps spaced 20 m along a transect at each site, and other methods use only a single trap. The most common trap method in Oklahoma and Arkansas has been a transect of eight pitfall traps made of 32 fl. oz. cups, and the most common method in Nebraska is a single pitfall trap made from a five gallon bucket. Beetle trap success is often standardized by trap effort (*i.e.* beetles per trap-night) before being used as the response variable for *N. americanus* habitat studies. A single night of trapping with a transect of eight traps spaced 20 m is usually counted as eight trap-nights. If four traps were disturbed by scavengers, sample effort would be reduced by four. However, since an individual trap may have a sample radius up to 800 m, traps spaced 20 m may not represent independent sample units. The standard approach for quantifying sample effort ignores the spatial scale of measurement and may introduce significant bias when some traps are disturbed or when data are compared among trap methods. The effective sample radius of baited pitfall traps has been roughly estimated to be 800 m based on overnight flight distances reported for *N. americanus* (USFWS 2014). We will directly assess sample ranges using beetles released at known distances from single traps and transects.

Eco-hydrology

A collaborative side project investigated multi-scale habitat selection of the Sulphur Springs diving beetle (Dytiscidae: *Heterosternuta sulphuria*), and this led to development of additional spatial scales for Geodata Crawler relevant to stream ecology like watershed- and riparian-scale sample areas. These new spatial scales, combined with Geodata Crawler's existing infrastructure and national geodatabases, broadened its applicability to include research disciplines like eco-hydrology.

Eco-hydrology is an interdisciplinary field that studies the interaction of hydrologic regimes and ecosystems. It has recently become a vibrant area of research due to pressing issues like climate change, water shortages, and large-scale hydrologic alterations from reservoirs, water withdrawals, agriculture, and urbanization. Eco-hydrology has benefited considerably from the recent boom in availability of GIS and remote sensing data and this has fueled development of new GIS-based methods including a widely used hydrologic disturbance index (Falcone *et al.* 2010), machine learning methods for assessing hydrologic alteration and predicting effects of climate change (Carlisle *et al.* 2010, Liermann *et al.* 2011), and a new risk-based framework for conservation of water resources and aquatic ecosystems that has been widely adopted by the water management community (Poff *et al.* 2010).

Eco-hydrology is an inherently multi-scale field of research in which landscape data often need to be collected from watersheds, riparian zones, and point-locations. The large-scale nature of eco-hydrology lends itself to GIS-based analyses, and a substantial amount of hydrology-related national GIS datasets have been developed to support this area of research. Access to these data is often limited as previously discussed for GIS and remote sensing data in general. Data are available from many different sources, so they can be difficult to locate or even to be aware of. A significant amount of storage capacity and processing effort may be required to generate site specific data collected from unique sample areas, like watersheds and riparian zones. This requires specialized software, expert knowledge, and time. Incorporating hydrology datasets and stream-related spatial scales into Geodata Crawler provides efficient data collection in support of recently developed eco-hydrology methods allowing their application at tens-of-thousands of stream locations, rather than only hundreds.

The first application of Geodata Crawler in eco-hydrology will involve identifying and mapping natural flow regimes in the Ozark-Ouachita Interior Highlands region (Chapter III). This project will work within the framework of Poff *et al.* (2010) to provide a foundation for regional risk-based water management by identifying the natural flow regimes of the region, quantifying their hydrologic attributes, and mapping their geographic distributions. Poff *et al.* (2010) recommended classifying relatively undisturbed reference streams based on a suite of ecologically-relevant hydrologic attributes to identify natural flow regimes that are expected to have unique ecological attributes (Poff *et al.* 1997, Olden *et al.* 2011). Hydrologic alteration and ecological responses should be assessed separately for each natural flow regime to identify potentially unique vulnerabilities (Poff *et al.* 2010). Predicting stream hydrology expected under natural conditions based on GIS-based landscape and climate data allows site assessments of hydrologic alteration, even in the absence of pre-disturbance hydrologic data from a stream gage (Carlisle *et al.* 2010).

We will produce an interactive Google Earth map document in which 24,557 individual stream segments of the Interior Highlands region can be clicked to display dozens of custom multi-scale landscape and climate characteristics for that stream segment, predicted values for a suite of ecologically-relevant hydrologic attributes, and predicted probabilities of membership in each natural flow regime. This will provide a foundation for assessing ecological responses to hydrologic alteration for each of the region's natural flow regimes. This project will demonstrate Geodata Crawler's ability to provide flexible and efficient data collection to support new methods in eco-hydrology.

Works Cited

Anderson RS. 1982. On the decreasing abundance of *Nicrophorus americanus* Olivier (Coleoptera: Silphidae) in eastern North America. *Coleopterists Bulletin* 36:362-365.

Bedick JC, Ratcliffe BC, Hoback WW, Higley LG. 1999. Distribution, ecology, and population dynamics of the American burying beetle [*Nicrophorus americanus* Olivier (Coleoptera, Silphidae)] in south-central Nebraska, USA. *Journal of Insect Conservation* 3:171-181.

Bedick J, Ratcliffe B, Higley L. 2004. A new sampling protocol for the endangered American burying beetle, *Nicrophorus americanus* Olivier (Coleoptera: Silphidae). *The Coleopterists Bulletin* **58**(1):57-70.

Carlisle DM, Falcone J, Wolock DM, Meador RM, Norris RH. 2010. Predicting the natural flow regime: Models for assessing hydrological alteration in streams. *River Research and Applications* **26**:118-136.

Creighton JC, Lomolino MV, Schnell GD. 1993. Survey methods for the ABB (*Nicrophorus americanus*) in Oklahoma and Arkansas. Oklahoma Biological Survey: Norman, OK.Creighton JC, Schnell GD. 1998. Short-term movement patterns of the endangered American burying beetle Nicrophorus americanus. *Biological Conservation* **86**:281-287.

Elkinton JS, Schal C, Ono T, Carde RT. 1987. Pheromone puff trajectory and upwind flight of male gypsy moths in a forest. *Physiological Entomology* **12**:399-406.

Falcone JA, Carlisle DM, Weber LC. 2010. Quantifying human disturbance in watersheds: variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams. *Ecological Indicators* **10**:264-273.

Hanski I. 1999. *Metapopulation Ecology.* Oxford University Press: Oxford, UK.

Hoback WW, Bishop AA, Kroemer J, Scalzitti J, Shaffer JJ. 2004. Differences among antimicrobial properties of carrion beetle secretions reflect phylogeny and ecology. *Journal of Chemical Ecology* **30**:719-729.

Hubbell SP. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography.* Princeton University Press: Princeton, NJ.

Levin SA. 1992. The problem of pattern and scale in ecology. *Ecology* **73**:1943-1967.

Liermann CAR, Olden JD, Beechie TJ, Kennard MJ, Skidmore PB, Konrad CP, Imaki H. 2011. Hydrogeomorphic classification of Washington State rivers to support emerging environmental flow strategies. *River Research and Applications* **28**:1340-1358.

Lomolino MV, Creighton JC, Schnell GD, Certain DL. 1995. Ecology and conservation of the endangered American burying beetle (*Nicrophorus americanus*). *Conservation Biology* 9(3):605-614.

Lomolino MV, Creighton JC. 1996. Habitat selection, breeding success and conservation of the endangered American burying beetle *Nicrophorus americanus*. *Biological Conservation* 77:235-241.

MacArthur RA. 1972. *Geographical Ecology*. Princeton University Press: Princeton, NJ.

MacArther RH, Wilson EO. 1967. *The Theory of Island Biogeography*. Princeton University Press: Princeton, NJ.

Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* **18**:189-197.

Mathews CY. 1995. Interspecific competition between burying beetles *Nicrophorus americanus* and *N. orbicollis*. MS thesis, University of Oklahoma.

Merrick JM, Smith RJ. 2004. Temperature regulation in burying beetles (*Nicrophorus spp.*: Coleoptera: Silphidae): effects of body size, morphology and environmental temperature. *The Journal of Experimental Biology* **207**:723-733.

Müller JK, Eggert AK. 1987. Effects of carrion-independent pheromone emission by male burying beetles (Silphidae: *Necrophorus*). *Ethology* **76**:297-304.

Murlis J, Elkinton JS, Carde RT. 1992. Odor plumes and how insects use them. *Annual Review of Entomology* **37**:505-32.

Olden JD, Kennard MJ, Pusey BJ. 2011. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Ecohydrology* **5**:503-518.

Otrenen M. 1988. The effect of body size on the outcome of fights in burying beetles. *Annales Zoologici Fennici* **25**:191-201.

Pickett STA, White PS. 1987. *The Ecology of Natural Disturbance and Patch Dynamics*. Academic Press: Waltham, MA.

Poff NL, Allan JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, Sparks RE, Stromberg JC. 1997. The natural flow regime. *BioScience* **47**:769-784.

Poff NL, Richter B, Arthington AH, Bunn SE, Naiman RJ, Kendy E, Acreman M, Apse C, Beldsoe BP, Freeman M, Henriksen J, Jacobson RB, Kennen J, Merritt DM, O'Keefe J, Olden J, Rogers K, Tharme RE, Warner A. 2010. The Ecological Limits of Hydrologic Alteration (ELOHA): A new framework for developing regional environmental flow standards. *Freshwater Biology* **55**:147-170.

Python. 2012. Python v2.7.3. Python Software Foundation. Available at: http://www.python.org

Scott MP, Traniello JFA, Fetherston IA. 1987. Competition for prey between ants and burying beetles (*Nicrophorus spp.*): differences between northern and southern temperate sites. *Psyche* **94**:325-332.

Scott MP. 1990. Brood guarding and the evolution of male parental care in burying beetles. *Behavioral Ecology and Sociobiology* **26**:31-39.

Scott MP. 1994. Competition with flies promotes communal breeding in the burying beetle, *Nicrophorus tomentosus*. *Behavioral Ecology and Sociobiology* **34**:367-373.

Scott MP. 1998. The ecology and behavior of burying beetles. *Annual Review of Entomology* 43:595-618.

Scott MP, Traniello JFA. 1990. Behavioural and ecological correlates of male and female parental care and reproductive success in burying beetles (*Nicrophorus spp.*). *Animal Behavior* **39**:274-283.

Smith RJ, Bonilla M, Calahan C. 2000. Comparison of reproductive success of *in-situ* burial versus the use of abandoned burrows for carcass interment by *Nicrophorus investigator* (Coleoptera: Silphidae). *Journal of the Kansas Entomological Society* **73**:148-154.

Springett BP. 1968. Aspects of the relationship between burying beetles, *Nicrophorus spp.* and the mite *Poecilochirus necrophori* Vitz. *Journal of Animal Ecology* 37:417-24.

Trumbo ST. 1990. Interference competition among burying beetles. *Ecological Entomology* **15**:347-355.

Trumbo ST. 1994. Interspecific competition and the evolution of communal breeding in burying beetles. *American Micland Naturalist* **131**:169-174.

Trumbo ST. Bloch PL. 2002. Competition between *Nicrophorus orbicollis* and *N. defodiens*: resource locating efficiency and temporal partitioning. *Northeastern Naturalist* **9**:13-26.

Turner MG, Gardner RH, O'Neill RV. 2001. *Landscape Ecology in Theory and Practice: Pattern and Process*. Springer Science+Business Media, Inc.: New York, NY.

[USFWS] United States Fish and Wildlife Service. 2004. *Current and reported historical range of the American burying beetle.* U.S. Fish and Wildlife Service, Oklahoma Ecological Services Field Office: Tulsa, OK.

[USFWS] United States Fish and Wildlife Service. 2007. *American burying beetle Nicrophorus americanus trapping and relocating guidance.* U.S. Fish and Wildlife Service, Arkansas Ecological Field Office: Conway, AR.

[USFWS] United States Fish and Wildlife Service. 2014a. *American Burying Beetle Conservation Strategy for the Establishment, Management, and Operations of Mitigation Lands.* U.S. Fish and Wildlife Service, Southwest Region: Albuquerque, NM. Available at http://www.fws.gov/southwest/es/oklahoma/ABB_Add_Info.htm

[USFWS] United States Fish and Wildlife Service. 2014b. *American Burying Beetle Nicrophorus americanus Oklahoma Presence/Absence Live-trapping Survey Guidance*. U.S. Fish and Wildlife Service, Ecological Services Field Office: Tulsa, OK. Available at http://www.fws.gov/southwest/es/oklahoma/ABB_Add_Info.htm

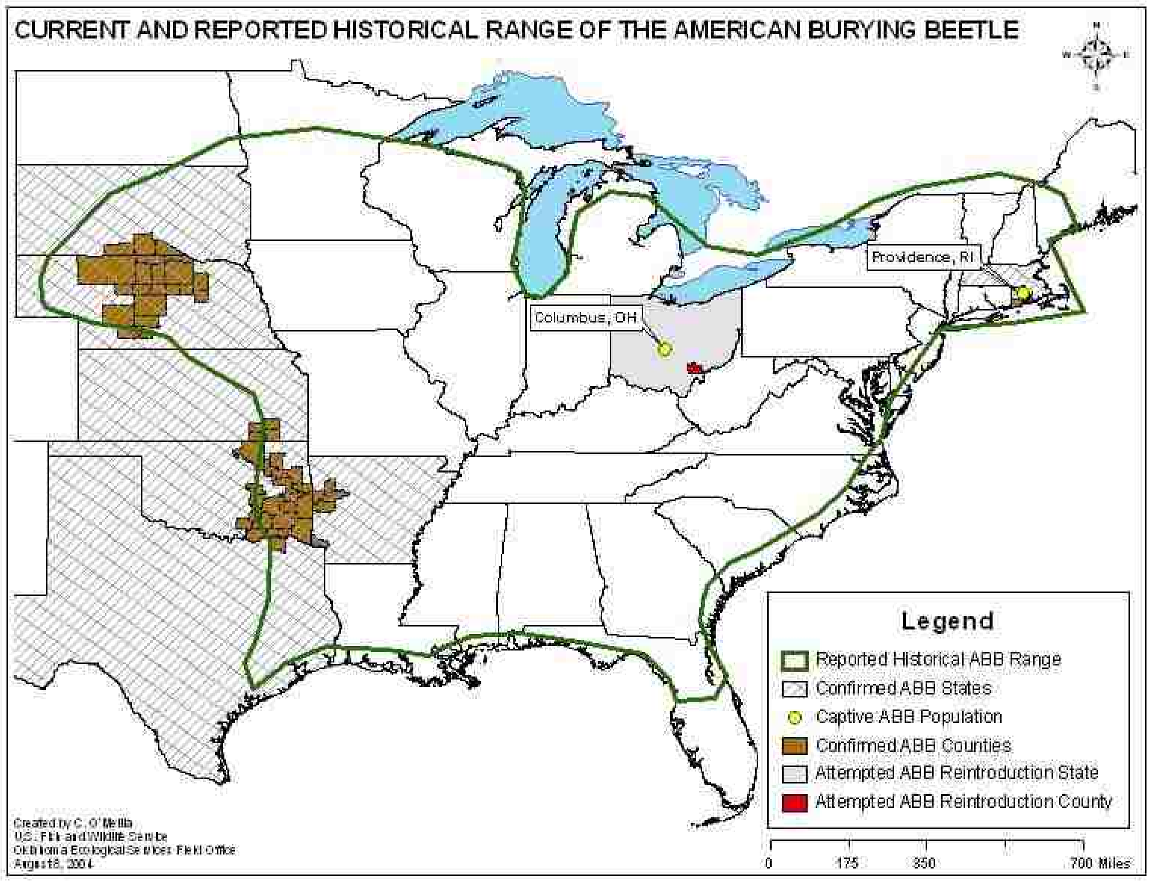
Wilson DS, Fudge J. 1984. Burying beetles: intraspecific interactions and reproductive success in the field. *Ecological Entomology* **9**:195-203.

Wilson DS, Knollenberg WG. 1987. Adaptive indirect effects: the fitness of burying beetles with and without their phoretic mites. *Evolutionary Ecology* **1**:139-159.

Wilson DS, Knollenberg WG, Fudge J. 1984. Species packing and temperature dependent competition among burying beetles (Siliphidae, *Nicrophorus*). *Ecological Entomology* 9:205-216.

Figures

Figure 1.
The historic and current distribution of *N. americanus* (USFWS 2004). Since the creation of this map, a captive ABB population has been established at the St. Louis Zoo in St. Louis, MO, and a non-essential experimental population has been introduced at the Wah'kon-tah Prairie in southwestern Missouri (Cedar & St. Claire counties). Reintroductions attempted at the Wayne National Forest in southeast Ohio were not successful.



Chapter I:

Efficient new above-ground bucket traps produce comparable data to that of standard transects for endangered American burying beetles (Silphidae: *Nicrophorus americanus* Olivier)

The Coleopterists Bulletin 66(3):209-218. 2012.

Douglas R. Leasure[1,2,3], David M. Rupe[2], Elizabeth A. Phillips[3], Dustin R. Opine[3], Gary R. Huxel[1]

[1] Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701

[2] FTN Associates, Ltd., Fayetteville, AR 72703

[3] Environmental Branch, Chaffee Maneuver Training Center, Fort Chaffee, AR 72905

Abstract

Federal sampling guidelines for the endangered American burying beetle (Silphidae: *Nicrophorus americanus* Olivier) have historically recommended transects of eight baited pitfall traps spaced 20 m. We compared a new above-ground bucket trap sampling method to standard transects in terms of capture rates, time efficiency, trap mortality, disturbance, and sample range. A single bucket trap was set for three consecutive nights at each site (three bucket-nights) rather than a transect of eight traps set for three nights like standard pitfall traps (24 trap-nights). To facilitate comparisons between methods, an appropriate sample effort conversion was determined to convert bucket-nights to trap-nights. Bucket traps were 75% more time efficient than standard transects and were more resistant to disturbances from scavengers. *Nicrophorus americanus* abundance estimates were significantly different between methods when a bucket-night was treated as equivalent to a trap-night. The most appropriate sample effort conversion was one bucket-night equals eight trap-nights. For both trap types, the probability of recapture was less than 25% for beetles released directly adjacent to traps and dropped below 5% for beetles greater than 300 m from traps. No trap mortalities resulted from either method in this study, but bucket traps were designed to reduce risks from the most common causes of trap mortality: drowning, heat stress, and predation. Bucket traps had rain covers and allowed for drainage, increased ventilation, and excluded some common predators found in standard pitfall traps. We recommend exclusive use of above-ground bucket traps in future *N. americanus* surveys due to increased time-efficiency, comparability with standard transects, decreased susceptibility to disturbance, larger bait size, and likely decrease in trap mortality.

Keywords: Coleoptera, trap methodology, pitfall trap, sample range, sample effort conversion

Introduction

The American burying beetle (Silphidae: *Nicrophorus americanus* Olivier) was placed on the endangered species list in 1989 due to a drastic range contraction in the late 19th and early 20th centuries (Raithel 1991). These beetles historically ranged throughout much of the United States east of the Rocky Mountains but are now limited to portions of Oklahoma, Arkansas, Nebraska, South Dakota, and Rhode Island. *Nicrophorus americanus* bury appropriately sized carrion in underground brood chambers for oviposition (preferred carrion mass is 80-100 g; Kozol *et al.* 1988). After removing feathers or hair from buried carrion, adults use oral and anal secretions to preserve the food source for later consumption by larvae, which also receive regurgitated supplemental feedings from one or both parents (Raithel 1991).

The U.S. Fish and Wildlife Service's standardized trap protocols have traditionally called for transects of eight baited pitfall traps set for three consecutive nights (USFWS 2010, Kozol 1990, Creighton *et al.* 1993, Bedick *et al.* 2004). *Nicrophorus americanus* abundances are reported as beetles per trap-night where a transect represents 24 trap-nights (8 traps x 3 nights). Federal guidelines have traditionally favored pitfall traps made from 24 fl. oz. cups buried to ground level (*i.e.* USFWS 2010). We designed and evaluated new above-ground bucket traps that required only a single trap at each site for three consecutive nights (three bucket-nights). Standard pitfall traps were compared to new above-ground bucket traps in terms of *N. americanus* capture rates, ability to detect presence/absence of a population, mortality risk, resistance to scavengers, effective sample range, and setup/maintenance time.

Bedick *et al.* (2004) advocated pitfall traps made of five-gallon buckets because they (1) accommodated larger pieces of bait, (2) provided larger areas for trapped beetles, and (3) improved ventilation. Several disadvantages were mentioned that included difficulty burying a five-gallon bucket in rocky soils and mortality risk due to trap inundation despite the use of rain covers and soil berms. The bucket traps evaluated in the present study address these problems because they are installed above the ground and allow for drainage.

Recent federal trapping guidelines (USFWS 2011) allowed bucket pitfall traps using either a single five-gallon bucket or five one-gallon buckets dug into the ground and set for three consecutive nights (three trap-nights or 15 trap-nights, respectively). However, there is no established basis for quantifying sample effort for bucket traps so that *N. americanus* abundance estimates (beetles per trap-night) are comparable across methods. If the sample effort associated with a bucket-night is not equivalent to that of a standard trap-night, bucket traps will produce incompatible *N. americanus* abundance estimates without an appropriate sample-effort conversion.

Sample effort metrics for standard transects have treated individual traps in transects as independent sample units (eight trap-nights) and subtracted a trap-night for each trap disturbed by scavengers. However, a transect's effectiveness may not be significantly diminished by several disturbed traps because bait remains in nearby undisturbed traps. It seems likely that nearby baited traps reliably attract beetles considering the 800 m estimated sample range of the traps (USFWS 2010) in comparison to the 20 m trap spacing (Fig. 1). Subtracting trap-nights for disturbed traps could artificially inflate beetle abundance estimates (beetles per trap-night) if a transect's effectiveness was not significantly diminished by disturbances. We evaluated an alternative method based on transect-nights that required at least four undisturbed traps for a transect-night to be valid, but otherwise ignored disturbances. Sample effort metrics based on transect-nights rather than trap-nights would simplify comparisons among various methods.

Methods

This field study was conducted as part of the 2011 annual *N. americanus* survey at Chaffee Maneuver Training Center (Fort Chaffee) in the Arkansas River Valley of western Arkansas. Fort Chaffee is a 26,000 hectare military training installation that hosts one of the largest remaining *N. americanus* populations. Wildfires and ground disturbances associated with military training have resulted in a patchwork of successional vegetation communities including native grasslands, shrublands, oak woodlands, and climax oak-hickory forests. Dominant flora in these communities include broomsedge bluestem (*Andropogon virginicus*), little bluestem (*Schizachyrium scoparium*), big bluestem (*Andropogon gerardii*), winged sumac (*Rhus copalinum*), winged elm (*Ulmus alata*), post oak (*Quercus stellata*), blackjack oak (*Quercus marilandica*), mockernut hickory (*Carya tomentosus*), and black hickory (*Carya texana*). Species potentially serving as carrion resources for *N. americanus* reproduction—based on expected fecundities, mortality rates, and reported *N. americanus* carrion size preference (Kozol *et al.* 1988)—are abundant in these successional communities (*personal observation*). These species include, but are not limited to, northern bobwhite (*Colinus virginianus*), hispid cotton rats (*Sigmodon hispidus*), and eastern cotton tail kits (*Sylvilagus floridanus*).

Above-ground bucket traps were designed and tested for use at Fort Chaffee where digging restrictions prohibited standard pitfall traps. Bucket traps (Fig. 2; Leasure *et al.* 2012) were made from five gallon buckets with small drain holes in the bottom and vent holes in the sides. A 15 cm hole was cut in the lid and a nine fl. oz. bait cup was suspended above the hole with wire and baited with a chicken leg drumstick that had previously been placed outdoors in a sealed container for 24 – 36 hours. A 20 cm funnel was attached to the inside of the lid beneath the hole. Erosion control matting was zip-tied to a 30 x 60 cm piece of welded wire fencing which was then attached to the bucket with wire to serve as a landing pad. An inverted plastic bowl was attached to the wire fence to serve as a rain cover and provide shade. Two 5 x 10 cm holes were cut in the rain cover to allow beetles access to the trap. Two eye-bolts were installed in the side of the bucket which allowed the bucket to be attached to a rebar stake in the ground. A two foot length of rebar was hammered into the ground and then used to secure the trap by sliding the eye bolts protruding from the side of the bucket over the rebar. A wet sponge was placed inside as a water source. A single bucket trap was placed on the surface of the ground at each sample location for three consecutive nights.

Standard pitfall trap transects consisted of eight baited pitfall traps (Fig. 3) spaced 20 meters and set for three consecutive nights. Each trap was made of two 32 fl. oz. plastic cups placed one inside the other in an excavation so that a 1.5 cm lip remained above the soil line. Bait cups were made from two fl. oz. plastic condiment cups suspended above the pitfall traps with wire. Split chicken breasts with skin and bone were cut into 15 – 20 g pieces and placed outdoors in a sealed container for 24 – 36 hours prior to being used as bait. Rain covers were made by attaching an inverted plastic bowl to a 30 x 30 cm piece of wire fence with a minimum 2.5 cm mesh size. Rain covers were secured over traps using landscape pins to prevent trap inundation and to discourage scavengers.

Twenty eight (28) sample sites were selected from CMTC's established *N. americanus* sample sites so that no sites were within 1.6 km of each other (Fig. 4). This buffer was intended to maintain independence of adjacent samples and was based on the estimated effective sample range of baited pitfall traps for *N. americanus* (800 m; USFWS 2010). Between 9 July and 7 August, each site was sampled during two separate three night periods with at least a three night lag between sampling periods. 10 sites were randomly chosen to receive bucket traps during the first sampling period and another 10 sites received standard transects. The alternate method was used at these sites during the second sampling period. Eight control sites were sampled with standard transects during both sampling periods. The order in which methods were assigned was randomized to limit temporal confounding effects and a lag period was used to limit sequence effects where prior trapping may influence future efforts. The average lag period was longer for control sites (17 days) than for experimental sites (four days). Results are presented with and without control sites and this discrepancy in lag periods did not alter conclusions.

The difference in *N. americanus* abundance estimates (Δ beetles per trap-night) between buckets and transects was calculated for every site using various sample effort conversions for bucket data (*i.e* 1 bucket-night = 1, 2, 3, …16 trap-nights). For each sample effort conversion rate, a single-sample two-tailed t-test (SAS 2010) was used to test the null hypothesis that the differences in abundance estimates between methods were not significantly different than zero. For each sample effort conversion rate, a one-way ANOVA (SAS 2010) was used to test that disparities in relative abundance estimates between methods were no different than disparities between sampling periods at control sites. Data were graphically analyzed for normality and ANOVA residual plots were checked for homogeneous variances.

We expected significantly different relative abundance estimates between methods when bucket-nights were treated as equivalent to trap-nights. We expected no difference in relative abundance estimates when bucket-nights were treated as being roughly equivalent to full transects (eight trap-nights). "Transect-nights" were also evaluated as units of sample-effort where a transect-night was only valid if at least four traps remained undisturbed.

The time required for a team of two researchers to install traps was recorded with a stopwatch. All traps were checked daily before 10:00 am and the numbers of all *Nicrophorus* species captured and any trap disturbances were recorded. Traps were re-baited daily or every other day to prevent desiccation of bait. Time required to service traps was recorded each day, not including time processing captured beetles. Number of *N. americanus* was recorded daily for each trap as well as any trap disturbances (*i.e.* missing bait, damaged rain cover, trap removed, etc.). Disturbance rates (*e.g.* disturbed bucket-traps vs. standard transects with at least one disturbed trap) were compared using an equality of proportions test. Captured *N. americanus* were sexed, aged (teneral or adult), and marked with paint pens to codify capture date and trap site.

Marked beetles were released at various distances within 500 m of traps and release locations were recorded with a Trimble GeoXT GPS unit. Flight distances were determined for marked beetles recaptured the following day using ArcGIS 10.0 (Esri 2010). Butler *et al.* (2012) showed that about 92% of burying beetles marked with enamel paint retained their mark for two days and the present study required only a single day of mark retention. Logistic regression (Systat 2007) was used to model the probability of recapturing marked beetles as a function of release distances from traps. Effective sample radii of both trap types were assessed graphically based on how recapture probabilities decreased as release distances increased. Trap method (bucket or transect) was included as a factor to test for differences in sample radii between methods. Conclusions drawn from this analysis were tentative due to small sample sizes (n = 65, six recaptures).

Program PRESENCE (Hines, J. E. 2006) was used to estimate the ability of each trap method to detect an *N. americanus* population in a presence-absence survey. Covariates that were thought to affect detectability such as average overnight wind speed and average temperature the previous day (influencing bait desiccation) were assessed, but these covariates did not improve detectability estimates. The simplest models are reported here: no variation in detection probabilities across sampling occasions and no sample- or site-covariates.

The current practice of subtracting trap-nights to adjust for trap disturbances is based on the assumption that trap disturbances significantly reduce sampling efficiency. If true, a negative relationship should exist between numbers of disturbed traps and numbers of beetles captured across a large number of transects with various levels of disturbance. To assess this, we analyzed historic data from CMTC that were collected using standard pitfall trap transects to produce 594 records of *N. americanus* abundances and trap disturbances (FTN 2007, 2009, 2010, 2011). Sites without *N. americanus* were not included because capture rates cannot vary at sites without a detectable population. Due to abundance and disturbance data having negative exponential distributions, Spearman's rank correlation coefficients were used to characterize correlations and 1000 bootstrap randomizations were used to estimate P-values (two-tailed) and 95% confidence intervals (bias-corrected and accelerated method; Systat 2007). Due to concern about potential lack of independence among three consecutive sample nights at a single site, data were separated by sample night (*i.e.* 1, 2, or 3) and correlations were estimated separately for each group as well as with groups pooled.

If disturbances reduced sampling efficiency, the largest number of beetles would be expected on nights with the least number of disturbed traps at sites where a gradient of disturbances occurred across three sample nights. 32 sites were identified where a gradient of disturbances had occurred across three sample nights and where *N. americanus* were detected. Sample nights were ranked at each site based on *N. americanus* abundances (1, 2, or 3; ties averaged) to provide within-site ranked abundances that were normally distributed. Within-site ranked abundance was the response variable in an ANOVA model (Systat 2007) with number of disturbed traps as a six level factor. Histograms and residuals plots were graphically analyzed to assess normality and homogeneity of variances. This dataset was not balanced (*i.e.* unequal number of trap-nights for each level of disturbance) and did not have adequate samples for transects with more than four disturbed traps. For these reasons, interpretation of these ANOVA results should be conservative and only considered valid for transects with less than five disturbed traps.

Results

Table 1 compares bucket traps to standard transects in terms of capture rates, mortality rates, disturbance rates, recapture rates, and time required to set and check traps. Bucket traps outperformed standard transects in all categories although no mortalities were recorded with either method. It took a team of two experienced field workers about 4 times longer to run a standard transect for three days compared to a bucket trap. Bucket traps were better able to prevent scavengers from disturbing traps, evidenced by an 18.3% lower disturbance rate compared to standard transects with at least one trap disturbed ($P = 0.001$, $CI_{95\%}$ = 7.7 – 29.0%).

Nicrophorus americanus abundance estimates from bucket traps and standard transects were significantly different when a bucket-night was treated as equivalent to a trap-night (P = 0.0351, n = 20), reiterating the need for a sample effort conversion. Abundance estimates were not significantly different between methods when a bucket-night was treated as equivalent to eight trap-nights (P = 0.8864, n = 20, *power* = 0.8, *effect size* = 0.125 beetles/trap-night). Abundance estimates showed insignificant differences (P > 0.2) between methods when buckets were treated as five to 13 transect-nights, but the least difference was found when a bucket-night was treated as equivalent to eight trap-nights (Fig. 5). At control sites, *N. americanus* abundance estimates differed by an average of 0.129 beetles per trap-night between sampling periods. This was no different than the average disparity in abundance estimates between methods at experimental sites (0.115 beetles per trap-night, $P = 0.6681$, $n = 28$).

Trap method did not significantly affect recapture probability ($P = 0.739$) suggesting that both trap methods have similar sampling efficiency. A graphical comparison also showed that both methods had similar effective sample radii (Fig. 6). The probability of recapture never exceeded 25% for beetles released directly adjacent to traps and was below 5% for beetles greater than 300 m from traps. The small number of recaptures made conclusions based on these data tentative.

Detection probabilities were similar between methods when presence-absence data were analyzed in program PRESENCE. Bucket traps had a detection probability of $0.475 \pm$ SE 0.105 and standard pitfall trap transects had a detection probability of 0.510 ± 0.092. The probability of detecting an *N. americanus* population at a site was virtually equivalent between methods. Note that this analysis is based on presence or absence of *N. americanus* at a site rather than the probability of detecting individual beetles as in the logistic regression analysis above.

No significant correlation was found between *N. americanus* capture rates and the number of disturbed traps which suggested that a few disturbed traps in a transect did not reduce the sampling efficiency. Correlation estimates, bootstrapped confidence intervals, and hypothesis test results are presented in Table 2 with data grouped by sample night and pooled.

An ANOVA was used to test for differences in ranked beetle abundances (ranked within sites) across a gradient of disturbance rates at each site, but no significant effect was found ($P = 0.166$). Tukey's pairwise comparisons showed no significant differences among levels of disturbance although there was a slight increase in trap success with four disturbed traps and a slight decrease with five disturbances (Fig. 7). These results suggest that transects with at least four undisturbed traps are likely as effective as full transects. Graphical examination of residuals showed no evidence of heteroskedasticity.

The lack of association between disturbed traps and capture rates suggests that subtracting trap-nights for disturbed traps may result in overestimated *N. americanus* abundances (beetles per trap-night). Note that few records existed with a high number of trap disturbances, so conclusions are limited to situations with zero to four disturbed traps. Transect-nights may be a more appropriate unit of sample effort for pitfall trap transects with a minimum number of undisturbed traps required for a transect-night to be valid. A single-sample t-test showed no significant difference between relative abundances derived using bucket-nights vs. transect-nights ($P = 0.8997$, $n = 20$). This is similar to counting a bucket-night as 8 trap-nights except that it ignores disturbed traps in standard transects.

Discussion

No *N. americanus* died in traps used for this comparison, but two mortalities were recorded at nearby sites sampled with standard transects as part of concurrent research. These deaths were the result of wildfire and predation while trapped. Flightless predators such as predaceous ground beetles (Carabidae) that were commonly captured along with *N. americanus* in standard pitfall traps were generally excluded from above-ground bucket traps. We expected above-ground bucket traps to prevent drowning mortality because they allowed rain water to drain, unlike standard pitfall traps, but drought conditions during this study prevented this assessment. In a separate study, Leasure (unpublished data 2011) conducted 126 bucket-nights of trapping with above-ground bucket traps and captured 547 *Nicrophorus spp.* individuals with only a single trap mortality, which is a testament to the low mortality rate associated with above-ground bucket traps.

Bucket traps were more successful in preventing scavengers from stealing bait. Only a single bucket trap was disturbed throughout the study where a scavenger had broken the plastic rain cover to gain access to the bait (2% of bucket-nights vs. 20% of standard transects with disturbances). Similar disturbances with this trap design were periodically found in other research, but were not a major problem (Leasure 2011, unpublished data). We designed and tested a new landing pad and rain cover (Fig. 8; Leasure *et al.* 2012) made from plywood with a bait container fixed to the underside of the rain cover. This study design should deny scavengers access to bait and prevent these types of disturbances as well as making trap servicing quicker and cleaner. The more enclosed bait container should delay desiccation of bait while providing adequate dispersal of bait odor.

In addition to the benefits of above-ground bucket traps, we have shown *N. americanus* abundance data from buckets to be comparable to standard pitfall trap transects when bucket-nights are treated as eight trap-nights. Abundance data were not comparable when a bucket-night was treated as equivalent to a trap-night, emphasizing the need for a sample effort conversion to maintain the integrity of comparisons with historic datasets. We suggest that a transect of closely spaced traps (*i.e.* pitfall cups, pitfall buckets, and above ground buckets spaced 20 m) should be considered a single sample unit measured in transect-nights with some minimum number of undisturbed traps required. To support this recommendation, we presented rank-correlations and ANOVA results that showed *N. americanus* capture rates were not reduced with up to four disturbed traps. However, additional field work is required to make conclusions about the effects of five to eight disturbed traps and should be designed so that numbers of disturbed traps are experimentally controlled to provide a balanced dataset (equal numbers of trap-nights carried out for each level of disturbance).

Both trap types had very low probabilities of capturing beetles known to be in the vicinity. For any release distance within 500 meters of traps, the probability of recapture never exceeded 25% and dropped below 5% at 300 meters. This suggests that bait away (USFWS 2007a) and relocation efforts (USFWS 2007b) may be marginally effective at best. Even if we generously estimate traps to detect 40% of individuals present, 60% of the local *N. americanus* population likely remains undetected in trap and relocation efforts. We suggest that mark-recapture and release distance data be recorded in other *N. americanus* surveys to improve understanding of effective sample range and detectability. Appropriate data would include the original capture locations and release locations of all marked beetles released at various distances from traps (*i.e.* 100, 200, 300, … 800 m) and recapture locations of recaptured beetles.

Compared to standard pitfall trap transects, above-ground bucket traps were found to be time-efficient, safe for trapped beetles, and resistant to disturbances. *Nicrophorus americanus* abundance estimates from bucket traps were comparable to standard transects when a bucket-night was treated as equivalent to eight trap-nights or a transect-night. The sample effort metric for bucket traps is more robust than standard transects because only a single trap is used and a larger piece of bait can be accommodated which delays desiccation resulting in more consistent effectiveness. Based on results presented here, we recommend exclusive use of above-ground bucket traps for future *N. americanus* surveys.

Acknowledgments

This project was funded by the Arkansas State Military Department through the Environmental Branch at Chaffee Maneuver Training Center. Field assistants included Robert Gregory, Jeremy Rigsby, and Blake Griffis from FTN Associates, Ltd. Dr. Daniel Magoulick from the Department of Biological Sciences at the University of Arkansas provided helpful experimental design advice. Robert Gregory and James McKeever from FTN Associates, Ltd. provided CAD drawings. This work would not have been possible without the cooperation and support of the U.S. Fish and Wildlife Service, particularly Erin Leone and Chris Davidson at the Arkansas Ecological Services Field Office.

Works Cited

Bedick, J. C., and B. C. Ratcliffe, L. G. Higley. 2004. A new sampling protocol for the endangered American burying beetle, *Nicrophorus americanus* Olivier (Coleoptera: Silphidae). *The Coleopterists Bulletin* **58**(1):57-70.

Butler, S. R., J. Jurzenski, W. W. Hoback. 2012. Evaluation of marking techniques, mark retention, and mortality due to marking of burying beetles (Coleoptera: Silphidae). *The Coleopterists Bulletin* **66**(2):149-154.

Creighton, C., M. Lomolino, and G. Schnell. 1993. Survey methods for the American burying beetle (*Nicrophorus americanus*) in Oklahoma and Arkansas. Unpublished report.

Esri. 2010. ArcMap 10.0 Evaluation Edition, Build 2414. Redlands, CA.

FTN. 2007. American burying beetle (*Nicrophorus americanus*) survey at Fort Chaffee Maneuver Training Center, Arkansas. FTN Associates Ltd.: Fayetteville, AR.

FTN. 2009. American burying beetle (*Nicrophorus americanus*) survey, Fort Chaffee Maneuver Training Center, Fort Chaffee, Arkansas. FTN Associates Ltd.: Fayetteville, AR.

FTN. 2010. American burying beetle (*Nicrophorus americanus*) survey, Chaffee Maneuver Training Center, Fort Chaffee, Arkansas. FTN Associates Ltd.: Fayetteville, AR.

FTN. 2011. American burying beetle (*Nicrophorus americanus*) survey, Chaffee Maneuver Training Center, Fort Chaffee, Arkansas. FTN Associates Ltd.: Little Rock, AR.

Kozol, A. J., M. P. Scott, J. F. A. Traniello. 1988. The American burying beetle *Nicrophorus americanus:* Studies on the natural history of a declining species. *Psyche* **95**(4):167-176.

Kozol, A. J. 1990. Suggested survey protocol for *Nicrophorus americanus*, the American burying beetle. New England Field Office of Endangered Species.

Leasure, D. R., D. M. Rupe, E. A. Phillips, D. R. Opine, G. R. Huxel. 2012. Instructions for building above-ground bucket traps for American burying beetles (*Nicrophorus americanus*) [Appendix X] *In:* American burying beetle (*Nicrophorus americanus*) rangewide survey guidance. U.S. Fish and Wildlife Service, Tulsa, OK. Available from: http://www.fws.gov/southwest/es/Oklahoma/ABB_Add_Info.htm (Accessed on 28 June 2012).

Raithel, C. 1991. American burying beetle (*Nicrophorus americanus*) recovery plan. Rhode Island Division of Fish and Wildlife, West Kingston, RI.

SAS. 2010. JMP 8.0. SAS Institute Inc., Cary, NC.

Systat. 2007. Systat 12 for Windows. Version 12.00.08. SYSTAT Software, Inc. Chicago, IL.

[USFWS] United States Fish and Wildlife Service. 2007a. American burying beetle *Nicrophorus americanus* baiting away guidance for projects in Oklahoma. USFWS Division of Ecological Services: Tulsa, Oklahoma. 13 June 2007. Available from http://www.fws.gov/southwest/es/oklahoma/beetle1.htm (Accessed on 19 Aug 2011).

[USFWS] United States Fish and Wildlife Service. 2007b. American burying beetle *Nicrophorus americanus* trapping and relocating guidance in Oklahoma. USFWS Division of Ecological Services: Tulsa, Oklahoma. 13 June 2007. Available from http://www.fws.gov/southwest/es/oklahoma/beetle1.htm (Accessed on 19 Aug 2011).

[USFWS] United States Fish and Wildlife Service. 2010. American burying beetle *Nicrophorus americanus* survey guidance. USFWS Division of Ecological Services: Conway, Arkansas. 20 April 2010.

[USFWS] United States Fish and Wildlife Service. 2011. American burying beetle *Nicrophorus americanus* rangewide survey guidance. USFWS Division of Ecological Services. 20 April 2011.

Tables

Table 1.
Comparisons of above-ground bucket traps to standard pitfall trap transects using a variety of important metrics.

	Bucket Traps	n	Standard Transects	n
Average *N. americanus* per trap-night †	0.108	20	0.102	20
Total ABBs captured	48	20	45	20
Time to install traps (hh:mm)	3:52	13	19:32	9
Time to check traps daily	2:26	54	8:10	48
Average total time	11:13	20	44:02	20
ABB Mortality Rate ‡	0.00%	60	0.00%	60
Disturbance Rate	1.70%	60	20.00% *	60
ABB Recapture Rate	14.71%	34	6.67%	30

† 1 bucket-night = 8 trap-nights

‡ Per bucket-night or transect-night

* Proportion of standard transects with at least one disturbed trap

Table 2.
Spearman rank correlation coefficients between the number of *N. americanus* captured and the number of disturbed traps in standard transects. Confidence intervals and 2-tailed p-values were estimated using 1000 bootstrap randomizations.

Sample Night	Sample size	Spearman Correlation	95% Confidence Interval		Bootstrapped P-value	Bootstrapped correlations -0.1 to 0.1
			Minimum	Maximum		
1	198	0.053	-0.09	0.206	0.553	71.70%
2	198	-0.016	-0.15	0.113	0.802	83.90%
3	198	-0.096	-0.252	0.042	0.493	52.60%
All	594	-0.023	-0.109	0.056	0.619	96.50%

Figures

Figure 1.
Layout of a standard pitfall trap transect showing an 800 m trap sample range (USFWS estimate) and a 200 m trap sample range in comparison to 20 m trap spacing to illustrate the lack of independence among traps in a transect.

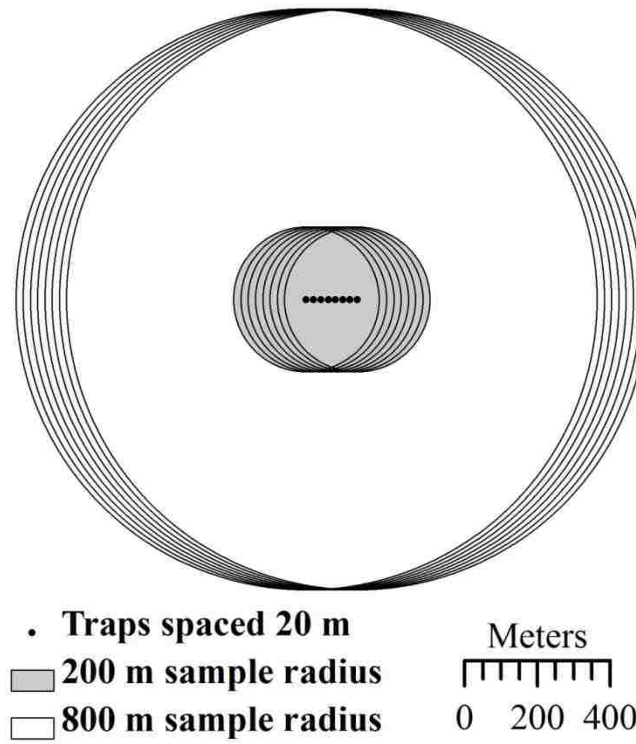


Figure 2.
New above-ground bucket trap design with wire trap cover that used a single trap at each site.

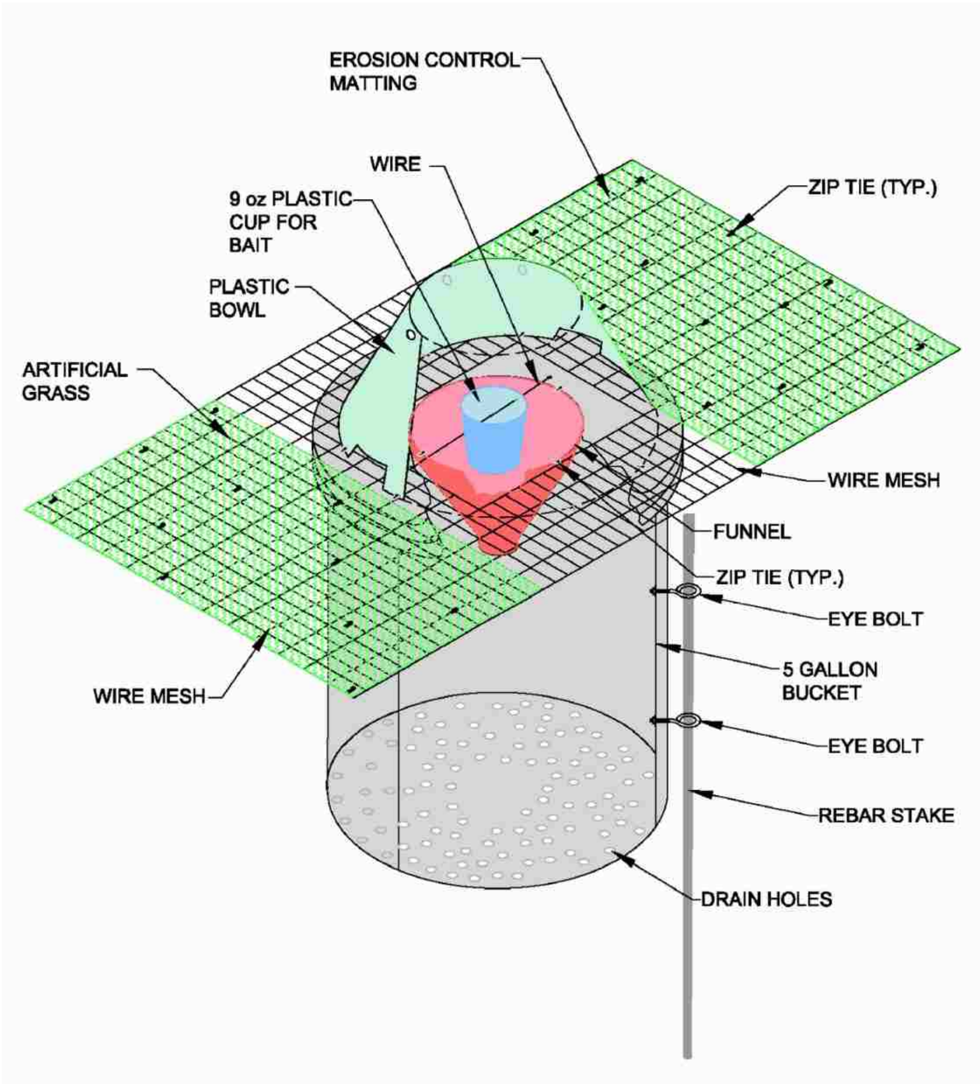


Figure 3.
Standard pitfall trap design. A standard transect consisted of eight traps spaced 20 m.

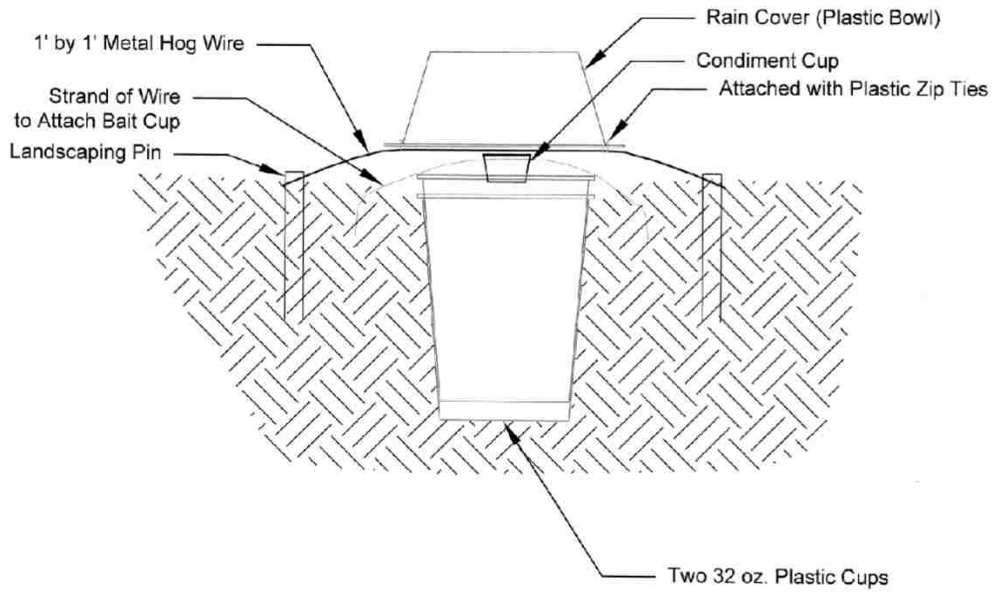


Figure 4.
Sample site layout (n=28) showing 800 m site buffers (open circles). The order in which trap methods were used during the two sample periods at each site included buckets then transects (circles), transects then buckets (squares), and transects during both sample periods at control sites (triangles).

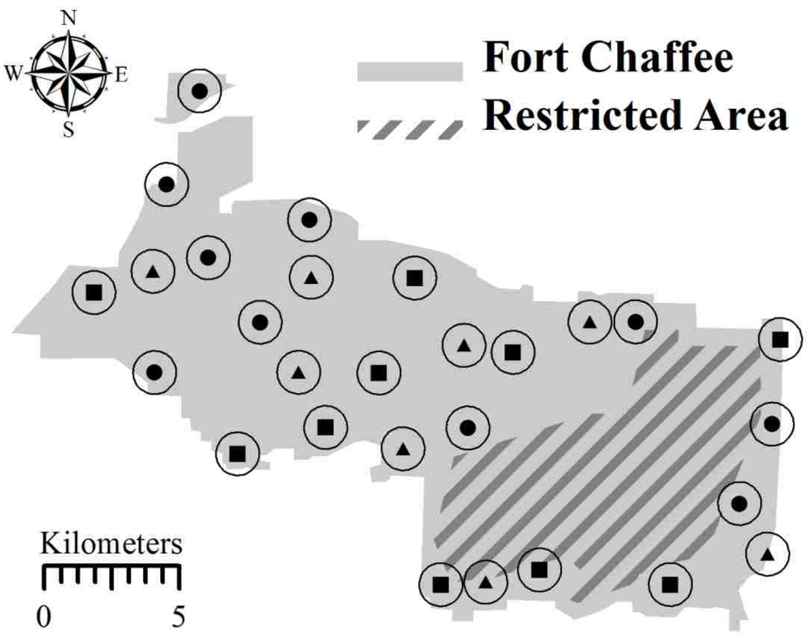


Figure 5.
Average differences in *N. americanus* abundance estimates between methods using various sample effort conversion rates (n = 20). Normalized average differences are mean differences divided by their standard deviations to provide a standardized scale for comparisons because abundance estimates—and therefore differences between them—are inherently smaller when trap-nights are artificially increased.

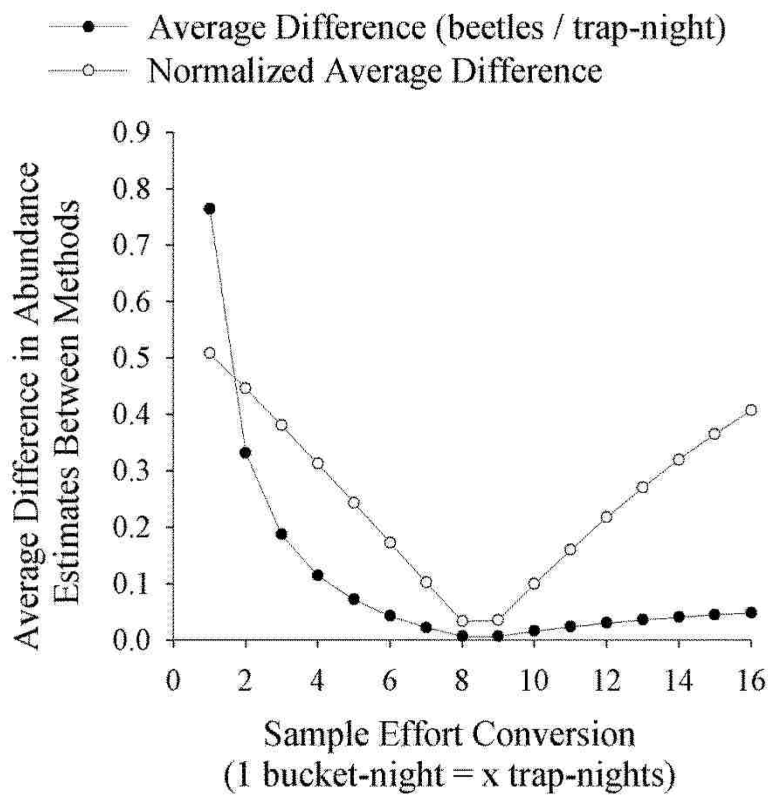


Figure 6.
Predicted recapture probabilities as a function of release distance. Sample ranges appeared to be similar between trap methods, but sample sizes were too small to support robust conclusions. Recapture rates were less than 25% for both methods, even when beetles were released nearby.

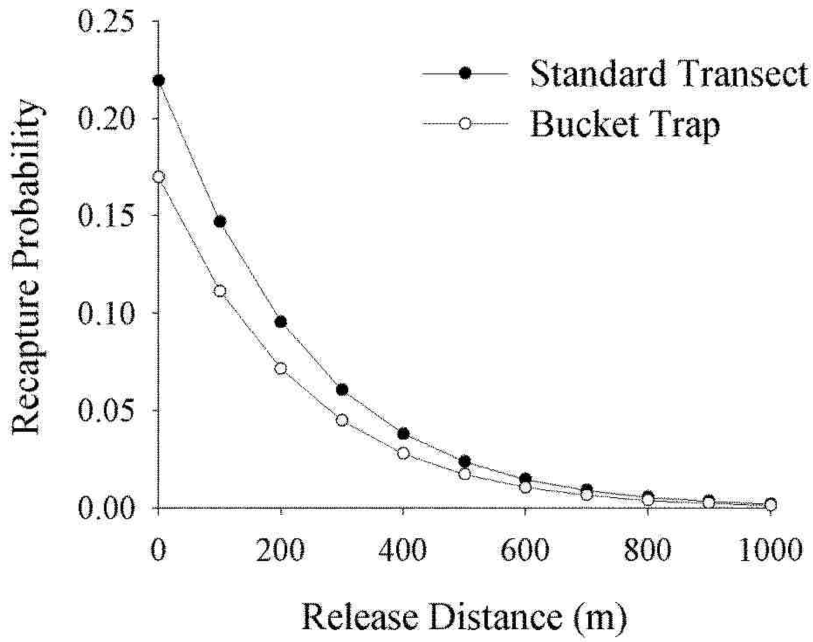


Figure 7.
ANOVA group means showing the relationship between beetle captures and the numbers of disturbed traps. Beetle abundances were ranked for the three sample nights at each site such that increasing ranks represented increasing abundances. Precision was low in groups 4 and 5 due to small sample sizes (n = 6 and n = 4 respectively).

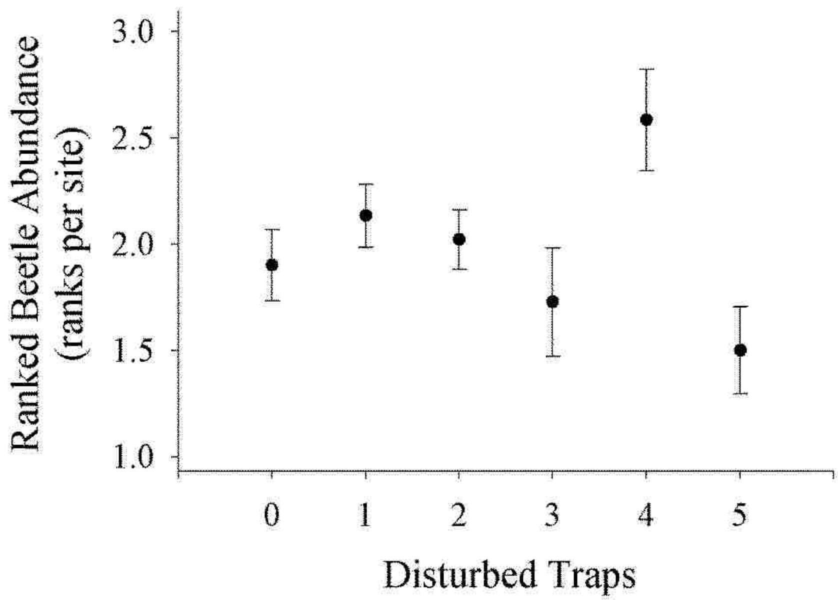
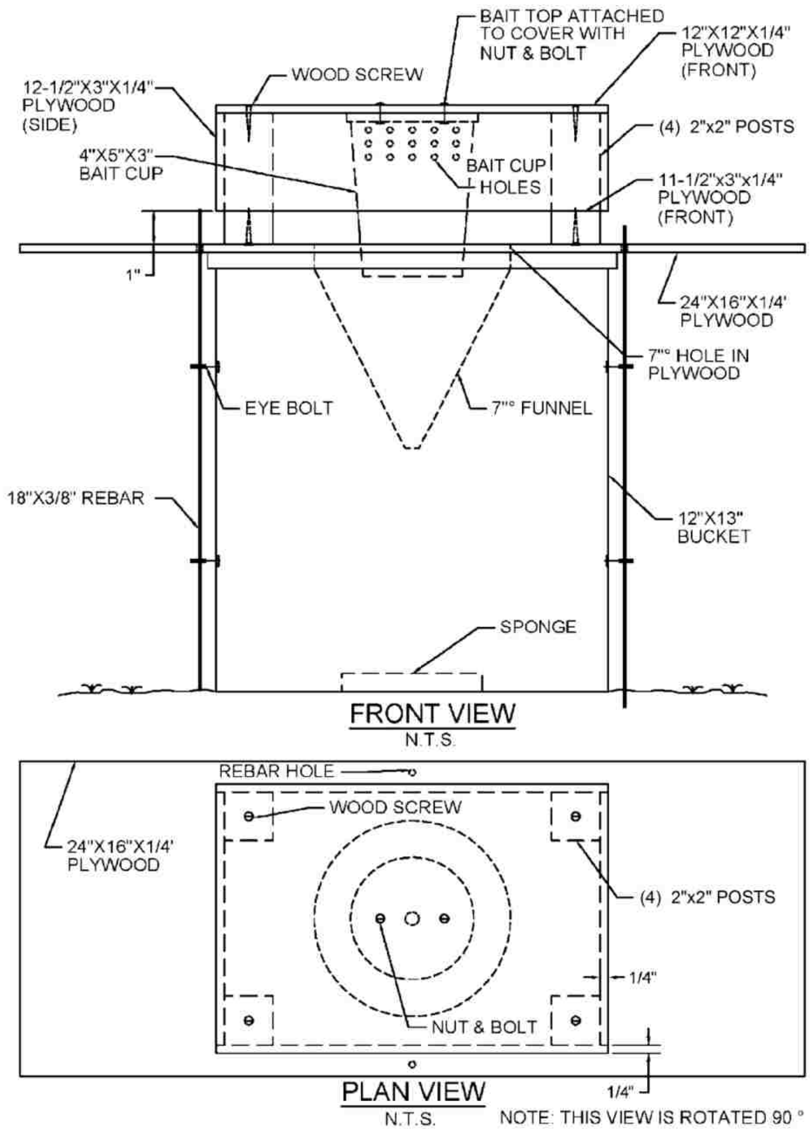


Figure 8.
Suggested improvements to above-ground bucket trap design using a wooden rain cover and landing pad to reduce disturbances from scavengers, reduce maintenance, and increase bait-life.



Appendixes

Appendix 1: Lead Author Confirmation Letter

J. William Fulbright College of Arts and Sciences
Department of Biological Sciences

Chapter 1, titled "Efficient new above-ground bucket traps produce comparable data to that of standard transects for endangered American burying beetles (Silphidae: Nicrophorus americanus Olivier)" of D. R. Leasure's dissertation was published in *The Coleopterists Bulletin* in 2012 with coauthors: David M. Rupe, Elizabeth A. Phillips, Dustin R. Opine, and Gary R. Huxel.

I, Dr. Daniel D. Magoulick, advisor of Douglas Ryan Leasure, confirm Douglas Ryan Leasure was first author and completed at least 51% of the work for this manuscript.

Daniel D. Magoulick Assistant Unit Leader U. S. Geological Survey Arkansas Cooperative Fish and Wildlife Research Unit	Date

Science Engineering, Room 601 • Fayetteville, AR 72701-1201 • 479-575-3251 • Fax: 575-4010 • www.uark.edu

Chapter II:

Landsat-based monitoring of an endangered beetle: Addressing issues of high mobility, annual life history, and imperfect detection

In review for *Landscape Ecology*

Douglas R. Leasure

Department of Biological Sciences, University of Arkansas,
Fayetteville, AR 72701 USA

Abstract

A conservation priority for the endangered American burying beetle (Nicrophorus americanus) has been to implement habitat-based conservation, but its annual life history, strong dispersal ability, and low detectability have contributed to difficulties identifying manageable habitat characteristics and mapping the species' current distribution. I assessed habitat within three site radii (100, 800, and 1600 m) to determine an appropriate spatial scale for habitat assessment of this mobile species. A Landsat time-series was used to quantify successional dynamics likely to be important for an annual species. Royle's N-mixture model accounted for imperfect detection with baited pitfall traps, and was used to assess competing hypotheses that explained patterns of beetle abundance in western Arkansas. Factors hypothesized to affect beetle detection were temperature, dew point, wind speed, topographic position, and forest cover. Factors hypothesized to affect beetle abundance were vegetation structure, disturbance history, soil texture, and topographic wetness. Detection rates of *N. americanus* during our sampling periods averaged 0.20 ± 0.108 (± SD), and were dependent on overnight temperature, dew point, and wind speed. Results suggested upper and lower temperature thresholds beyond which detection was reduced. Habitat assessments were most effective within 800 m site radii, the estimated sample range of traps. *Nicrophorus americanus* abundance was associated with grasslands and open-canopy woodlands with rolling topography, sandy loam soils, and moderate patchy disturbances from wildfires or troop maneuvers. Large annual fluctuations in *N. americanus* population sizes were apparent, but availability of suitable vegetation communities appeared fairly stable. Results were consistent with the hypothesis that *N. americanus* populations have suffered from widespread losses of early successional communities. This project provided important conservation recommendations for an endangered species,

demonstrated efficacy of Landsat-based monitoring, and provided a framework for assessing habitat of mobile annual species that are difficult to detect.

Introduction

Top conservation priorities for the endangered American burying beetle (*Nicrophorus americanus*) have been to identify manageable habitat characteristics and to map its current distribution, but its annual life history, strong dispersal ability, and low detectability have made this difficult. Highly mobile species often require habitat assessments at fairly large spatial scales, but it may be unclear which spatial scales are most appropriate. Habitat assessments at the landscape scale usually rely on geographic information systems (GIS), but land cover maps do not adequately represent temporal dynamics of vegetation condition and disturbance history that may be key factors influencing patterns of abundance for annual species. Sampling protocols, in this case baited pitfall traps, usually do not detect all individuals at a site, and detection rates may vary among sites, confounding abundance estimates. For non-homeothermic animals, like insects, detection may be influenced by temperature, and burying beetle flight is known to be temperature-dependent (Merrick and Smith 2004). These challenges have been an impediment to understanding the decline of *N. americanus* and developing effective habitat conservation strategies.

Many hypotheses have been proposed to explain the decline of *N. americanus*, but perhaps the most plausible explanation is based on a decline in availability of optimally sized carrion for reproduction (Sikes and Raithel 2002). Adult burying beetles (*Nicrophorus* spp.) can feed on carcasses of any size, but they must fly in search of suitably sized vertebrate carrion small enough to be buried into underground brood chambers, yet big enough to adequately provision larvae. *Nicrophorus americanus* appears to prefer 80 to 200 g avian and mammalian carcasses for reproduction (Kozol *et al.* 1988), and larger carcasses are associated with larger brood sizes (Wilson and Fudge 1984, Scott 1998). Intense competition for carrion resources occurs among burying beetle species and the outcome often depends on relative body sizes, search and burying efficiencies, carrion mass, and air temperature (Otronen 1988, Wilson and Fudge 1984, Wilson et al. 1984, Matthews 1995, Kozol *et al.* 1988). Other important competitors include microbes, flies, ants, and vertebrate scavengers (USFWS 1991, Sikes and Raithel 2002). Carrion availability for reproduction is likely a key factor determining spatial patterns of abundance for *N. americanus* throughout its range, but carrion availability is difficult to assess directly.

Most habitat studies have focused on indirect relationships of *N. americanus* abundance with vegetation and soils, assuming that these factors are associated with carrion availability and site suitability for brood rearing. Anderson (1982) originally suggested that *N. americanus* may be associated with mature closed canopy forests with deep soils, based on habitat of a similar European species, *N. germanicus*. Field studies of *N. americanus* habitat in North America have produced conflicting claims that *N. americanus* is a generalist (Lomolino *et al.* 1995), a forest specialist (Anderson 1982, Walker 1952, Lomolino and Creighton 1996), or a grassland specialist (Kozol *et al.* 1988, Bedick *et al.* 1999). In an attempt to more directly assess carrion availability, Holloway and Schnell (1997) estimated biomass and species richness of birds and mammals documented in the vicinity of *N. americanus* traps, but no strong relationships with *N. americanus* abundance were identified. Soil texture and moisture conditions have been found to be associated with *N. americanus* abundance, and this is likely related to substrate suitability for brood chamber construction (Creighton *et al.* 1993, Lomolino *et al.* 1995, Hoback unpublished data). All of these field studies assessed habitat in the areas directly adjacent to *N. americanus* trap locations.

Relying on *in situ* measurement of habitat characteristics in the immediate vicinity of trap locations ignores the likelihood of attracting beetles from several hundred meters away where habitat characteristics may differ significantly. *Nicrophorus americanus* has strong flight ability and searches frequently for food and reproductive carcasses (Creighton and Schnell 1998, Bedick *et al.* 2004, Creighton *et al.* 1993). An individual beetle was documented moving 6.7 kilometers in a single night at Fort Chaffee (personal observation). Standard baited pitfall traps may attract beetles from up to 800 m away, a sample area of about 2 km^2 (Leasure *et al.* 2012, USFWS 2014). Habitat assessments are usually conducted at much smaller spatial scales (*i.e.* within 100 m of traps), and this may be one factor that has contributed to difficulties detecting habitat associations. There has been some success using GIS to assess *N. americanus* habitat at larger spatial scales (Crawford and Hoagland 2010, McPherron *et al.* 2012, Jurzenski et al. 2014), but land cover maps do not account for vegetation dynamics and disturbance histories that may be important for this annual species. Satellite images from Landsat (USGS 2013) can be used to measure these vegetation dynamics at large spatial scales.

Detection of *N. americanus* using baited pitfall traps is dependent on flight activity of individual beetles and their ability to track the bait's odor plume. Standard trap protocols generally detect less than 25% of individuals within the sample area, and detection rates decline rapidly with increasing distance from traps (Leasure *et al.* 2012, Backlund *et al*. 2008). The effect of weather on the detection process is not well understood. Temperature is likely the most important determinant of flight activity with upper and lower bounds beyond which flight activity is reduced or physiologically impossible (Taylor 1963, Merrick and Smith 2004). Due to the susceptibility of burying beetles to desiccation (Bedick *et al.* 2006, personal observation), humidity may also be an important determinant of flight activity. Wind speeds greater than 16 km/h have been suspected of discouraging flight in *N. americanus* (USFWS 2014), but empirical evidence in support of this hypothesis is lacking. In fact, moderate winds are thought to improve the ability of many flying insects to track odor plumes because searching animals can simply fly upwind within the odor plume to locate the source (Murlis *et al.* 1992). Forest structure and topography can distort odor plumes and make them more difficult for flying insects to follow (Murlis *et al.* 1992, Elkinton *et al.* 1987). The confounding effect of imperfect detection on estimated site abundances may have contributed to difficulties characterizing suitable habitat for *N. americanus*.

Working at a military training installation in western Arkansas that supports one of the largest remaining *N. americanus* populations, this project's goals were to 1) assess factors affecting detection of beetles with baited pitfall traps, 2) identify important habitat characteristics, and 3) map spatio-temporal patterns of habitat quality and beetle abundance throughout the study area. Factors hypothesized to explain variability in detection of *N. americanus* included temperature, wind, humidity, topographic position, and forest cover. Factors hypothesized to affect abundance *of N. americanus* included vegetation structure, disturbance history, soil texture, and topographic wetness. A multi-scale model comparison approach with Royle's (2004) N-mixture models was used to assess competing hypotheses and to identify an appropriate spatial scale for habitat assessment. Despite our inability to assess carrion availability directly, I expected *N. americanus* abundance to be indirectly related to vegetation communities associated with those species producing suitable carcasses for *N. americanus* reproduction. I expected these habitat associations to be more evident when controlling effects of imperfect detection and when assessing habitat at larger spatial scales in a way that captures inter-annual vegetation dynamics. Although suitable vegetation communities may not be consistent throughout the range of *N. americanus*, some regional consistency should be expected.

Methods

Study Site

This study was conducted over five years at Chaffee Maneuver Training Center, a 26,000 hectare military training installation in the Arkansas River Valley of western Arkansas. Fort Chaffee hosts one of the largest remaining *N. americanus* populations which has been monitored annually since 1992. Chaffee vegetation communities are distributed as a mosaic of successional stages ranging from closed canopy oak-hickory forests typical of the region to fire-disturbed native prairies that were historically widespread in the region but are now rare. Fire disturbance associated with military training and an active prescribed fire program have maintained the patchwork of successional communities including native prairies, shrublands, and woodlands of various sizes. Open-canopy post oak (*Quercus stellata*) woodlands are fairly common on the landscape with scattered oak trees (*e.g.* basal area ≈ 1.7 m^2/hectare) and an understory of native grassland plants. Dominant flora at Fort Chaffee include broomsedge bluestem (*Andropogon virginicus*), little bluestem (*Schizachyrium scoparium*), big bluestem (*Andropogon gerardii*), winged sumac (*Rhus copalinum*), winged elm (*Ulmus alata*), post oak (*Quercus stellata*), blackjack oak (*Quercus marilandica*), mockernut hickory (*Carya tomentosus*), and black hickory (*Carya texana*).

Field Methods

We sampled *N. americanus* abundance at about 50 sites per year from 2007 to 2011 using consistent trapping protocols and fairly consistent sample locations among years (n = 257 including all years). Sites were sometimes moved slightly between years due to conflicts with military training, and to increase site spacing when possible. All sites used for this study were at least 1 km from any neighboring sites sampled the same year. Previous work suggested that the probability of detecting beetles more than 500 meters from traps was less than 5% (Leasure *et al.* 2012), so the 1 km buffer between sample sites should adequately maintain site independence.

Trap design followed United States Fish and Wildlife Service guidelines (*e.g.* USFWS 2011) with eight baited pitfall traps spaced 20 m and set for three nights. Each trap was made of two 32 fl. oz. plastic cups placed one inside the other in an excavation so that a 1.5 cm lip remained above the soil line. A moist piece of sponge was placed in each trap to hydrate trapped beetles. Bait cups were made from two fl. oz. plastic condiment cups suspended above the pitfall traps with wire. Chicken breasts with skin and bone were cut into 15 – 20 g pieces and allowed to rot outdoors in a sealed container for 24 to 36 hours prior to being used as bait. Rain covers were made by attaching an inverted plastic bowl to a 30 by 30 cm piece of wire fence with a minimum 2.5 cm mesh size. Rain covers were secured over traps using landscape pins to prevent trap inundation and to discourage scavengers. This trap design remained fairly consistent throughout the study although minor changes were made to the trap cover between years to better discourage scavengers. Traps were checked each morning before 1000 CST and *N. americanus* were counted. A trap-night was only considered valid if at least 4 traps remained undisturbed by scavengers, overnight temperatures remained above 15.5^{o} C, and no significant storms occurred. An additional night of trapping was done if any of these conditions were not met.

Digital Data Collection

Weather attributes thought to affect abundance and detection of *N. americanus* were quantified using hourly weather records from an airport 8 km from the study area (WeatherUnderground, Inc. 2013; Airport Code = FSM). All weather attributes were measured for 12 hour periods from 1900 to 0700 CST each night to correspond with nocturnal foraging activity of *N. americanus* (Bedick *et al.* 1999). Six weather covariates thought to affect *N. americanus* detection were measured for each trap night: average temperature (TMP), number of hours 24 to 33^{o} C (TMP_{2433}), average dew point (DEW), average dew point during hours when temperatures were 24 to 33^{o} C (DEW_{2433}), average wind speed (WND), and average wind speed when temperatures were 24 to 33^{o}C (WND_{2433}). The temperature range of 24 to 33^{o} C is the range of thoracic temperatures recorded during flight of another fairly large North American burying beetle, *N. hybridus* (Merrick and Smith 2004).

Several GIS datasets were used to quantify landscape characteristics thought to be associated with *N. americanus* abundance or detection. Landscape characteristics were quantified within 100, 800, and 1600 m radii around trap locations. Sample radii used for data collection are denoted as subscripts to variable names. A custom Python script (Python 2012) was used to automate the process of delineating sample areas and tabulating multi-scale GIS data for each site using ArcGIS 10.2 software (Esri 2013).

Based on findings that *N. americanus* prefer soils with greater than 40% sand (Lomolino *et al.* 1995), county soil survey data from Sebastian, Crawford, Franklin, and Logan counties in Arkansas (NRCS 2013) were used to quantify percent coverage of soils with greater than 40%

sand within three site radii to derive predictor variables $SNDY_{100}$, $SNDY_{800}$, and $SNDY_{1600}$. Lab studies have suggested that *N. americanus* prefer moist soils (Hoback *unpublished data*), and other members of the genus *Nicrophorus* can be extremely susceptible to dessication (Bedick *et al.* 2006). Although they are hypothesized to be associated with moist soils, burying beetles must avoid flood prone areas to prevent inundation of their underground brood chambers. This suggests that intermediate topographic wetness may be optimal. A topographic wetness index was calculated as $\frac{\ln(drainage\_area)}{\tan(slope)}$ using a digital elevation model with 30 m spatial resolution (Beven and Kirkby 1979, Sørenson *et al.* 2006, USEPA and USGS 2012). Topographic wetness values were averaged within three site radii to create predictor variables $TWET_{100}$, $TWET_{800}$, and $TWET_{1600}$.

Fort Chaffee's map of vegetation communities (Emrick and Dorr 2004) was used to quantify vegetation communities despite being several years outdated because it was the most detailed vegetation map available, it was field validated at over 1,000 locations, and it recognized differences in vegetation structure that are particularly important at Fort Chaffee (*i.e.* forest versus woodland). Percent coverage of four types of vegetation structure were quantified within three site radii to create predictor variables for grassland (GRS_{100}, GRS_{800}, GRS_{1600}), shrubland (SHB_{100}, SHB_{800}, SHB_{1600}), woodland (WDL_{100}, WDL_{800}, WDL_{1600}), and forest (FOR_{100}, FOR_{800}, FOR_{1600}). Emrick and Dorr's (2004) vegetation classification system defined forests as communities with greater than 60% canopy cover, whereas woodlands had 10-60% canopy cover. Shrublands and grasslands had less than 10% canopy cover. Woody vegetation made up more than 25% of ground cover in shrublands, but not grasslands. Emrick and Dorr's (2004) vegetation map did not extend beyond the Fort Chaffee property boundary, so vegetation communities outside the property boundary were classified using the National Land Cover

Dataset (USGS 2010). Land surrounding Fort Chaffee was mostly urban, forest, agriculture, or hayfield, and contained no shrubland or woodland communities. This made it relatively easy to manually reclassify pixels as either forest or grassland (not including hayfields or agriculture) to match the four types of vegetation structure currently being assessed.

A Landsat-5 Thematic Mapper image was obtained from each year 2006 to 2011 with acquisition dates ranging from 19 July to 6 September (USGS 2013). Images were selected from each year to minimize cloud cover over the study area and to minimize differences among acquisition dates. Images were atmospherically corrected using ATCOR2 for ERDAS IMAGINE (Intergraph 2013, Geosystems 2013) to obtain surface reflectance values. Radiometric correction used revised calibration parameters for images after May 5, 2003 (Chander and Markham 2003).

The normalized difference water index (NDWI) is a Landsat-derived index that quantifies vegetation biomass and moisture conditions (Gao 1996). Compared to the normalized difference vegetation index (NDVI), a widely used index of vegetation biomass, NDWI is less sensitive to atmospheric effects, it more consistently incorporates data from the sub-canopy layer, and it is more sensitive to vegetation moisture content (Gao 1996). NDWI was calculated from Landsat surface reflectance in near-infrared band (NIR, band 4) and mid-infrared band (MIR, band 5) as $\frac{NIR - MIR}{NIR + MIR}$. NDWI rasters were zero-centered by subtracting mean NDWI within the study area for each year. This was done to minimize effects of phenological differences among years, and to focus on site characteristics relative to available habitat each year. NDWI values were then averaged within three site radii for each year ($NDWI_{100}$, $NDWI_{800}$, $NDWI_{1600}$). If *N. americanus* abundance was associated with forested habitat, a positive relationship with NDWI was expected, whereas if *N. americanus* abundance was associated with open-canopy or grassland habitats, a negative relationship was expected.

Change in NDWI was calculated each year by subtracting the previous year's zero-centered NDWI raster from the current year's raster. Change in NDWI was then averaged within three site radii for each year ($\Delta NDWI_{100}$, $\Delta NDWI_{800}$, $\Delta NDWI_{1600}$). *Nicrophorus americanus* may be associated with vegetation disturbed by wildfires or troop maneuvers ($\Delta NDWI < 0$), vegetation recovering from disturbance the previous year ($\Delta NDWI > 0$), or undisturbed vegetation ($\Delta NDWI = 0$).

Detection of *N. americanus* may be associated with topographic position because of varying wind exposure and dispersal of odor plumes from baited pitfall traps (Murlis *et al.* 1992). Topographic position (*TPOS*) was measured as the elevation of a given point divided by the average elevation within 400 m of that point using a digital elevation model with 30 m resolution (USEPA and USGS 2012). Forest cover may also be associated with detection of *N. americanus* due to its potential influence on dispersal of the bait's odor plume (Murlis *et al.* 1992, Elkinton *et al.* 1987). For this purpose, forest cover was quantified within 100 m of trap sites (FOR_{100}) using the land cover map.

Statistical Model

Royle's (2004) N-mixture model was implemented using the R package *Unmarked* (Fiske *et al.* 2011, Royle and Dorazio 2008) to simultaneously model the abundance and detection of *N. americanus*. Two variations of Akaike information criterion (AIC) were used to compare models, AIC corrected for small sample sizes (AICc), and quasi-likelihood AICc to accommodate over-dispersion (QAICc; Burnham and Anderson 2002). All covariates were centered and scaled by subtracting their mean and dividing by their standard deviation prior to analysis to stabilize the numerical optimization algorithm (Fiske and Chandler 2014).

Beetle count data included 257 samples that were collected across five years from 2007 to 2011 (about 50 sites per year). Each sample included three observations, usually from three consecutive nights. Samples from different years at the same site were treated as independent samples. *Nicrophorus americanus* has a lifespan of one year so it is unlikely that an individual beetle could survive to be counted in two consecutive years. Because of strong dispersal ability, individual beetles could easily disperse throughout Fort Chaffee within a year. An individual beetle has been documented flying 6.7 km in a single night at Fort Chaffee (*personal observation*) and the entire study area is only about 30 km at its widest. Therefore, areas where *N. americanus* is consistently abundant from year to year are the result of independent habitat selection by new generations of beetles each year.

Global models were designed with the goal of estimating less than one parameter for every 10 samples. Sample size is somewhat ambiguous for Royle's (2004) N-mixture models because the abundance model treats each site as a sample (*i.e.* n = 257), while the detection model treats each observation period as a sample (*i.e.* n = 771). For the purposes of limiting parameters estimated in the global model, I assumed a sample size of 257 and limited the number of parameters to 25, including zero-inflation and over-dispersion parameters.

Twelve global models were estimated initially to compare model fit between Poisson and zero-inflated Poisson models, and among different spatial scales for measuring site covariates. A negative binomial model was assessed initially but it was not used because it had unstable parameter estimates at different values of *K*, an argument of the *pcount* function that sets the upper limit of integration (R package *Unmarked*; Fiske *et al.* 2011). The zero-inflated Poisson model was selected based on AICc, and K was set to 200 resulting in stable parameter estimates. The zero-inflated Poisson model estimated a zero-inflation parameter Ψ and used it to adjust site abundance estimates λ_i as $\lambda_i(1-\Psi)$.

Global models were built for each of three site radii used to measure site covariates (100, 800, and 1600 m). Two global models were built for each spatial scale. One used the vegetation map to quantify vegetation characteristics, and the other used satellite images. Global models were compared based on AICc. All subsequent models were built using the spatial scale and probability density function of the best global model.

Over-dispersion (*i.e.* extra-Poisson variation) was evident in histograms of the raw counts, and this may have resulted from pheromones being released by trapped beetles that attracted additional beetles at a greater rate than bait alone. Model fit and over-dispersion of the global model were assessed based on chi-square χ^2 statistics from the model and from 500 parametric bootstrap simulations. A measure of over-dispersion, c-hat, was calculated as $\frac{\chi^2_{model}}{mean(\chi^2_{bootstrap})}$. C-hat values greater than one were interpreted as over-dispersion and c-hat values greater than four were interpreted as lack of model fit (Burnham and Anderson 2002). Due to overdispersion, QAICc was the information criterion used to compare models (Burnham and Anderson 2002). Models that improved QAICc by at least two were selected, and models within two QAICc of the best model were selected if they had fewer parameters. Standard errors for parameter estimates were calculated using nonparametric bootstrap simulations with the R-package *Unmarked* (Fiske 2011). Standard errors for model predictions were calculated by the delta method using the R-package *AICcmodavg* (Oehlert 1992, Mazerolle 2013).

Nine covariates were assessed in relation to detection of *N. americanus* (Table 1): $YEAR$, TMP, TMP_{2433}, DEW, DEW_{2433}, WND, WND_{2433}, FOR_{100}, and $TPOS$. Thirty-nine detection models were identified *a priori* that represented competing hypotheses to be compared. All models of detection contained a *YEAR* factor and used the global abundance model during model selection. Each covariate was assessed initially to identify temporal scales of measurement that achieved the best model fit. For example, either *TMP* or TMP_{2433} would be included in multiple variable models, but not both. This was done to reduce the number of multiple variable models being compared, and to eliminate multi-collinearity. No multiple variable models contained covariates with Spearman correlation coefficients greater than 0.7. Except for a null model with only a *YEAR* factor, temperature was included in all models of detection. All combinations of the other variables were compared. The detection covariates selected by this process were used as the detection model when comparing models of abundance.

Eight covariates of abundance were assessed and three of these covariates were modeled with and without a quadratic term to represent hump-shaped and linear relationships (Table 1). A total of 72 competing hypotheses were identified *a priori* that were based on topographic features, soil attributes, coverage of vegetation communities, and normalized difference water index (NDWI) values. Models used either vegetation covariates derived from satellite images or vegetation covariates derived from the vegetation map, but not both. Eighteen hypotheses were based on the vegetation map, 48 hypotheses were based on Landsat-derived covariates, and six hypotheses included no measures of vegetation.

The best-fitting model was used to estimate detection rates achieved during our observation periods, and to estimate site-specific beetle abundances. A one-way ANOVA was used to test for differences in predicted detection probabilities among years. Site covariates from the selected model were sampled at 4,224 points on a 250 m grid covering the study area, and model predictions were made at each grid point. The grid of points was rasterized and the number of beetles occupying each 250 m raster cell was estimated assuming that predicted site abundances represented the number of beetles within either a 400 m or an 800 m site radius. It is important to note that the appropriate site radius is unknown. Total beetle population estimates are provided only as rough estimates and to illustrate the uncertainty in population estimates introduced by ambiguous sample areas associated with baited pitfall traps. The *trends* in estimated beetle populations among years should be consistent regardless of appropriate site radius, as long as the site radius is constant among years.

Inter-annual trends in total *N. americanus* population size at Fort Chaffee were estimated three ways: (1) by summing predicted beetle densities for all cells of the 250 m grid covering the study area, (2) by summing predicted beetle abundances each year across 29 sites that were consistently sampled all five years of the study, and (3) by taking the maximum observed beetle count among three sampling occasions at each site and summing them among the 29 consistently sampled sites each year. The first and second approach directly model the detection process, but the second approach is spatially limited to only 29 sample locations whereas the first approach assesses the entire study area. The third approach ignores the detection process, relying on observed counts, and is also spatially limited. Calculating beetle density per site (approach 1) requires an estimate of the area sampled by each trap. The area sampled by each trap is not known exactly, so results were compared assuming 400 m and 800 m sample radii.

Results

A zero-inflated Poisson distribution fit the data better than a Poisson distribution (ΔQAICc = 113). Model fit of the zero-inflated global model was adequate, but a c-hat value of 3.62 suggested over-dispersion. Using an 800 m site radius to measure site covariates for the global model resulted in best model fit (ΔQAICc = 137 and 131 compared to 100 m and 1600 m site radii, respectively). The zero-inflated Poisson model was used for all subsequent models of abundance and 800 m site radii were used to measure site covariates.

The model building process resulted in a final model with four covariates of detection, four covariates of abundance, and 21 estimated parameters including the zero-inflation parameter and c-hat. C-hat for the final selected model was 3.65 indicating model fit and over-dispersion similar to the global model. Model residuals had a median of -0.42 beetles, and 95% of residuals were between -6.4 and 10.3 beetles (Fig. 1). Some extreme residuals (max = 58.4) occurred at sites with very high abundances that were underestimated by the model. Parameter estimates and standard errors for the final selected model are reported in Table 2.

The final model of detection p (Fig. 2, Table 2) was:

$$p \sim YEAR + TMP + TMP^2 + DEW_{2433} + WND_{2433} \,.$$

Detection rates estimated for trapping periods at our sample sites averaged 0.201 ± 0.108 (± 1 SD), and detection rates varied significantly among years ($P < 0.01$; Fig. 3). Over-night average temperature *TMP* was an important predictor of detection probability and it was best modeled using a quadratic term (ΔQAICc = 90 without a quadratic term) indicating upper and lower temperature thresholds beyond which detection probabilities of *N. americanus* were reduced (Fig. 2). The optimal overnight average temperature for detection of *N. americanus* was approximately 29^{o} C. Temperature appeared to be the primary factor reducing predicted detection probabilities when overnight average temperatures were outside the range of optimal flight temperatures. However, when overnight average temperatures were within that range of optimal flight temperatures, increased dew points DEW_{2433} often increased predicted detection probabilities well above that expected based on temperature alone (Fig. 2). Wind speeds WND_{2433} were also positively related to detection probabilities, although to a lesser degree than dew point or temperature (Fig. 2, Table 2). Model fit was best when dew points and wind speeds were measured only during overnight hours with optimal flight temperatures, as opposed to analyzing entire overnight periods (ΔQAICc = 104 and 10.0 for dew point and wind, respectively). Forest cover FOR_{100} and topographic position *TPOS* were not selected for the detection model.

The final model of abundance λ (Fig. 4, Table 2) was:

$$\lambda \sim (1-\varphi)(YEAR + TWET_{800} + {TWET_{800}}^2 + NDWI_{800} + \Delta NDWI_{800} + {\Delta NDWI_{800}}^2)\ ,$$

where φ is the zero-inflation parameter. Topographic wetness $TWET_{800}$ was selected for the final model with a quadratic (hump-shaped) relationship with abundance (ΔQAICc = 3.6 without a quadratic term). Peak abundances were predicted at moderately low topographic wetness values, with a slight decrease at extremely dry ridge-top sites, and a strong decrease at wetter bottomland sites (Fig. 4). Normalized difference water index $NDWI_{800}$ was selected for the final model and it was modeled as a negative linear relationship (ΔQAICc = 1.9 with a quadratic term). Maximum abundances were predicted at grassland and woodland sites (low NDWI), and low abundances were predicted at forested sites, particularly bottomland forests (high NDWI; Fig. 4).

Annual change in normalized difference water index $\Delta NDWI_{800}$ was included in the final model with a quadratic term (ΔQAICc = 40 without quadratic term). Peak abundances were predicted at moderate or moderately high $\Delta NDWI_{800}$ values (*i.e.* 0 to 0.15) indicative of areas that either remained undisturbed between years or were recovering from small or patchy disturbances that occurred the *previous* year (usually wildfire or troop maneuvers). Low values of $\Delta NDWI_{800}$ (*i.e.* < 0.05) indicative of disturbances during the *current* year showed a strong negative association with *N. americanus* abundance. Beetle abundances were also negatively associated with extremely high values of $\Delta NDWI_{800}$ (*i.e.* > 0.15) indicative of areas recovering from exceptionally large or intense disturbances the previous year.

The best abundance model based on the vegetation community map fit the data much worse than the model based on Landsat-derived vegetation metrics (ΔQAICc = 145). The vegetation community map provided no information about temporal dynamics of vegetation condition, such as effects of wildfire and drought that were quantifiable using year-specific Landsat images. The best model based on the vegetation map included a hump-shaped relationship with topographic wetness, similar to the model based on NDWI. However, in the model based on the vegetation map, three additional predictors of *N. americanus* abundance were important: cover of soils with greater than 40% sand $SNDY_{800}$ (β = 0.395, *SE* = 0.127), cover of grasslands (β = 0.320, *SE* = 0.276), and cover of woodlands (β = 0.269, *SE* = 0.243). Cover of shrublands and forests were also included in this model, but parameter estimates for both were within one standard error of zero.

Spatial extrapolation of model predictions across a 250 m grid covering the study area suggested fairly constant habitat availability among years with some fluctuations due to wildfires, but the core area with good habitat was consistently in the central-western part of the study area (Fig. 5, habitat model). This area consisted mostly of interspersed bluestem grasslands and open-canopy oak woodlands with sandy loam soils and gently rolling topography that experienced patchy wildfires every few years. Model predictions showed major fluctuations in beetle abundances among years, but the core habitat area always supported higher abundances than other areas. A secondary area of suitable habitat in the east, particularly the southeast, supported high abundances in good years (Fig. 5, abundance model).

Strong annual fluctuations of the *N. americanus* population at Fort Chaffee were apparent with all three indicators of population trends (Fig. 6). The population trend indicated by detection-corrected abundance estimates for 29 consistently sampled sites at Chaffee were very similar to the trend indicated by spatially-extrapolated model predictions. Ignoring the detection process and instead using the maximum count among three sample periods at each site to estimate site abundances also provided a reasonable estimate of overall population trends, but this approach appeared to underestimate the population in good years and overestimate in bad years (Fig. 6).

Discussion

Over-dispersion in beetle count data may have been the result of trapped beetles releasing pheromones and attracting additional beetles at a greater rate than bait alone. Model residuals were greatest at sites with high beetle abundance and the largest residuals were always the result of underestimation by the model, as would be expected in the case of trapped beetles releasing pheromones. The zero-inflation term may have contributed to underestimation because it reduced predicted abundances most drastically at high abundance sites, although Poisson models without the zero-inflation term had much poorer fits to the data.

The effective sample range of baited pitfall traps for *N. americanus* has been estimated to be 800 m (USFWS 2014), and model fit was best when an 800 m site radius was used to measure site covariates compared to 100 or 1600 m radii. With the ability to fly over fairly large distances, beetles foraging in good habitat may be attracted into nearby poor habitat by a baited trap. Failure to account for this movement when choosing the spatial scale for habitat assessments may obscure significant habitat associations.

Three weather factors were associated with detection of *N. americanus* with baited pitfall traps: temperature, dew point, and wind. A weak positive relationship was found with overnight wind speeds, but it is important to note that this relationship might have been altered (strengthened or nullified) with site-specific wind data, as opposed to weather data obtained from a nearby airport. Temperature was best modeled as a quadratic relationship indicating lower and upper temperature thresholds beyond which detection probabilities were reduced. Maximum detection was achieved when overnight average temperatures were about 29^{o} C (Fig. 2), and this agreed with previous research that analyzed thermal tolerances during flight in three *Nicrophorus* species (Merrick and Smith 2004).

Current regulations impose a lower temperature threshold beyond which presence/absence surveys are not valid (*i.e.* 15.5^{o} C; USFWS 2014), but no upper threshold has been established. Our results suggest that false negative survey results may be more likely during hot summer months. Failure to detect a population due to weather conditions during pre-construction surveys for this endangered species could result in development projects occurring in occupied habitat without appropriate conservation measures. This may be particularly important in the southern range of *N. americanus* (*i.e.* Arkansas and Oklahoma) where temperatures regularly exceed optimal flight temperatures for burying beetles.

Abundance of *N. americanus* was negatively associated with the normalized difference water index $NDWI_{800}$, indicating a negative association with forested habitats at Fort Chaffee, particularly bottomland forests. This contradicts some previous research suggesting that *N. americanus* is a forest specialist (see Sikes and Raithel 2002 for a review). Our results indicated that *N. americanus* at Fort Chaffee were associated with grasslands and open-canopy woodlands with moderate patchy disturbances from wildfires and troop maneuvers. This is similar to habitat occupied by *N. americanus* populations throughout Oklahoma, particularly at places like the Tallgrass Prairie Preserve, where prescribed fire and free range bison maintain grassland vegetation, and Camp Gruber Training Center, where prescribed fire and military training maintain a patchwork of successional vegetation communities similar to Fort Chaffee. Based on the hump-shaped response of *N. americanus* abundance to $\Delta NDWI_{800}$ (Fig. 2), moderate disturbances likely had a positive effect after a year of recovery, while very large or intense disturbances may have had a negative effect even after a year of recovery. All disturbances appeared to have a short-term negative effect during the year they occurred.

Strong inter-annual fluctuations in *N. americanus* abundance at Fort Chaffee were apparent with all three measures of total population that were assessed (Fig. 6). The similarity in trends indicated by model predictions extrapolated throughout the study area versus model predictions at only 29 consistently sampled sites suggested that these 29 sites provided adequate spatial coverage for population monitoring at Fort Chaffee. Ignoring the detection process and summing raw counts at those 29 sites also tracked the overall population trend among years fairly well, suggesting that modeling the detection process may not be required to monitor overall population trends, although some bias was apparent. It is important to note that assessing population trends among years assumed that sample radii of traps were constant among years. Population estimates were given assuming two different site radii to illustrate potential error in trend estimates if, for example, sample radii varied between 400 and 800 m among years (Fig. 6). Sample radii and detection probabilities were likely influenced similarly by weather as it relates to flight behavior and dispersal of odor from bait. For this reason, it is fairly reasonable to assume that detection-corrected abundance estimates represented relatively consistent site radii among sites and years, but sample radii were not directly assessed here.

Although carrion availability could not be assessed directly, patchy successional vegetation communities at Fort Chaffee were assumed to be associated with increased availability of preferred carrion for *N. americanus* reproduction. In a *post hoc* literature review, previous surveys of avian and small mammal communities at Fort Chaffee were used to identify several species that may contribute carcasses for *N. americanus* reproduction (Murray 2004, Nupp 2007, *unpublished reports*). Likely candidates included hispid cotton rat (*Sigmodon hispidus*), eastern cotton tail (*Sylvilagus floridanus*), and northern bobwhite (*Colinus virginianus*). All of these species occupy open-canopy woodlands and grasslands at Fort Chaffee, can be found in high abundances, have the preferred body size for *N. americanus* reproductive carcasses, and are known to undergo boom-and-bust population dynamics that could produce a substantial number of carcasses on the landscape (Hernandez and Peterson 2007, Sealander and Heidt 1990). Boom-and-bust population dynamics of species producing suitable reproductive carcasses for *N. americanus* may be related to inter-annual population fluctuations that were observed for *N. americanus* at Fort Chaffee.

If habitat associations of *N. americanus* at Fort Chaffee are indicative of historical habitat—they may or may not be—then the decline of *N. americanus* throughout the 20th century may have been associated with losses of early-mid successional vegetation communities and reductions in carcasses produced by the boom-and-bust mammalian and avian species associated with these communities. Forest and wetland communities have been the major focus of landscape-scale conservation in the United States, while early-successional communities have received much less attention (Askins 2001). Oak savannas, like those utilized by *N. americanus* at Fort Chaffee, have been reduced by 99.9% throughout the Midwest (Noss *et al.* 1995). In fact, most of the ecosystems in eastern North America that have declined by greater than 98% are grassland, savanna, or shrubland communities (Noss *et al.* 1995, Askins 2001). Declines of these early successional ecosystems have been accompanied by declines in many disturbance-dependent animal species that rely on the structure and diversity of these communities (Hunter *et al.* 2001). For example, the decline of northern bobwhite (*Colinus virginianus*), a species likely to provide important carrion resources for *N. americanus* at Fort Chaffee, has been largely attributed to agricultural intensification and the loss of patchy fallow fields with herbaceous diversity and hedgerows (Herbert Stoddard 1978). A focus on forest conservation and fire suppression throughout the 20th century also resulted in widespread losses of grassland and woodland communities due to forest encroachment (Noss *et al.* 1995, Taft 1997, Foti 2004). The current distribution of *N. americanus* throughout Arkansas and Oklahoma includes abundance hotspots on military installations and managed prairie preserves that experience regular disturbances from wildfires, bison (*Bison bison*), or troop maneuvers. The annual life history of *N. americanus* and its strong dispersal ability makes it well adapted to exploit ephemeral resources that may result from these patchy disturbances.

N. americanus is an annual species with strong dispersal ability, and it is fairly difficult to detect using standard survey protocols. These characteristics pose challenges for *N. americanus* habitat assessments, and also for many other species with similar characteristics. Landsat allowed ephemeral effects of wildfire to be quantified annually, which is important for species capable of tracking patchy ephemeral resources on the landscape. The multi-scale approach provided flexibility necessary to identify an appropriate spatial scale for habitat assessments which is often difficult to determine *a priori*. The detection model not only reduced bias of site abundance estimates, but it also identified weather conditions that may severely reduce detectability with standard trap methods. Understanding the detection process associated with endangered species survey methods is particularly important because conservation planning often relies on pre-development surveys to assess conservation needs. These methodological tools were selected to address specific difficulties encountered assessing habitat for *N. americanus*, but the overall approach has broad applications for conservation biogeography of species presenting similar challenges.

Acknowledgments

This project was funded and coordinated by Fort Chaffee Maneuver Training Center, Environmental Branch, particularly Beth Phillips, Dustin Opine, and Doug Cobb. Field work was completed in cooperation with FTN Associates, Ltd. project managers David Rupe and David Williams, and field crews that included Jeremy Rigsby, Blake Griffis, Robert Gregory, and Robin Verble. Valuable comments and data analysis guidance were provided by my dissertation committee that included Dan Magoulick, Gary Huxel, Jason Tullis, and Steve Stephenson. This project required significant pre-planning and daily coordination with military personnel at Fort Chaffee Range Control, particularly Sergeants Swad, Richardson, and Mickelson. We owe them many thanks for being flexible and keeping us safe. Technical support and software licensing were provided by University of Arkansas, Center for Advanced Spatial Technologies (CAST).

Work Cited

Anderson RS. 1982. On the decreasing abundance of *Nicrophorus americanus* Olivier (Coleoptera: Silphidae) in eastern North America. *Coleopterists Bulletin* **36**:362-365.

Askins RA. 2001. Sustaining biological diversity in early successional communities: The challenge of managing unpopular habitats. *Wildlife Society Bulletin* **29**:407-412.

Backlund DC, Marrone GM, Williams CK, Tilmon K. 2008. Population estimate of the endangered American burying beetle, *Nicrophorus americanus* Olivier (Coleoptera: Silphidae) in South Dakota. *The Coleopterists Bulletin* **62**(1):9-15.

Bedick JC, Ratcliffe BC, Hoback WW, Higley LG. 1999. Distribution, ecology, and population dynamics of the American burying beetle [*Nicrophorus americanus* Olivier (Coleoptera, Silphidae)] in south-central Nebraska, USA. *Journal of Insect Conservation* **3**:171-181.

Bedick J, Ratcliffe B, Higley L. 2004. A new sampling protocol for the endangered American burying beetle, *Nicrophorus americanus* Olivier (Coleoptera: Silphidae). *The Coleopterists Bulletin* **58**(1):57-70.

Bedick JC, Hoback WW, Albrecht MC. 2006. High water-loss rates and rapid dehydration in the burying beetle, *Nicrophorus marginatus*. *Physiological Entomology* **31**:23-29.

Beven KJ, Kirkby MJ. 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences—Bulletin* **24**:43-69

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: A practical information-theoretic approach. 2nd Edition. Springer Science+Business Media, Inc.: New York, NY. 488 pp.

Chander G, Markham B. 2003. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges. *Geoscience and Remote Sensing* **41**(11): 2674-2677.

Crawford PHC, Hoagland BW. 2010. Using species distribution models to guide conservation at the state level: the endangered American burying beetle (*Nicrophorus americanus*) in Oklahoma. *Journal of Insect Conservation* **14**:511-521.

Creighton JC, Lomolino MV, Schnell GD. 1993. Survey methods for the ABB (*Nicrophorus americanus*) in Oklahoma and Arkansas. Oklahoma Biological Survey: Norman, OK.

Creighton JC, Schnell GD. 1998. Short-term movement patterns of the endangered American burying beetle *Nicrophorus americanus*. *Biological Conservation* **86**:281-287.

Elkinton JS, Schal C, Ono T, Carde RT. 1987. Pheromone puff trajectory and upwind flight of male gypsy moths in a forest. *Physiological Entomology* **12**:399-406.

Emrick VR and Dorr JL. 2004. Fort Chaffee-Maneuver Training Center Vegetative Communities Map and Classification System. Conservation Management Institute-Military Lands Division, College of Natural Resources, Virginia Polytechnic Institute and State University.

Esri. 2013. ArcGIS 10.2 for Desktop. Version 10.2.0.3348. Esri Inc.: Redlands, CA.

Fiske I, Chandler R, Miler D, Royle A, Kery M. 2011. Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software* **43**(10):1-23.

Fiske I, Chandler R. 2014. Overview of Unmarked: An R package for the analysis of data from unmarked animals. Available at http://cran.at.r-project.org/web/packages/unmarked/vignettes/unmarked.pdf

Foti TL. 2004. Upland hardwood forests and related communities of the Arkansas Ozarks in the early 19th century. Proceedings of the Upland Oak Ecology Symposium: History, Current conditions, and Sustainability. USDA Forest Service, Southern Research Station: Asheville, NC. General Technical Report SRS-73.

Gao, BC. 1996. NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of Environment* **58**:257-266.

Geosystems. 2013. ATCOR for IMAGINE 2013. GEOSYSTEMS GmbH: Germering, Germany.

Hernandez F, Peterson MJ. 2007. Northern bobwhite ecology and life history. Chapter in *Texas Quails: Ecology and Management*, 1st edition. Editor Leonard A. Brennan. Texas A&M University Press: College Station, TX.

Holloway AK, Schnell GD. 1997. Relationship between numbers of the endangered American burying beetle *Nicrophorus americanus* Olivier (Coleoptera: Silphidae) and available food resources. *Biological Conservation* **81**:145-152.

Hunter WC, Buehler DA, Canterbury RA, Confer JL, Hamel PB. 2001. Conservation of disturbance-dependent birds in eastern North America. *Wildlife Society Bulletin* **29**:440-455.

Intergraph. 2013. Erdas Imagine 2013. Version 13.00.00. Build 281. Intergraph Corporation: Hunstville, Alabama, USA.

Jurzenski JD, Jorgensen CF, Bishop A, Grosse R, Riens J, Hoback WW. 2014. Identifying priority conservation areas for the American burying beetle, *Nicrophorus americanus* (Coleoptera: Silphidae), a habitat generalist. *Systematics and Biodiversity* **12**(2):149-162.

Kozol AJ, Scott MP, Traniello JFA. 1988. The American burying beetle, *Nicrophorus americanus*: Studies on the natural history of a declining species. *Psyche* **95**:167-176.

Leasure DR, Rupe DM, Phillips BA, Opine DR, Huxel GR. 2012. Efficient new above-ground bucket traps produce comparable data to that of standard transects for endangered American burying beetles (Silphidae: *Nicrophorus americanus* Olivier). *Coleopterists Bulletin* **66**(3):209-218.

Lomolino MV, Creighton JC, Schnell GD, Certain DL. 1995. Ecology and conservation of the endangered American burying beetle (*Nicrophorus americanus*). *Conservation Biology* **9**(3):605-614.

Lomolino MV, Creighton JC. 1996. Habitat selection, breeding success and conservation of the endangered American burying beetle *Nicrophorus americanus*. *Biological Conservation* **77**:235-241.

Mathews CY. 1995. Interspecific competition between burying beetles *Nicrophorus americanus* and *N. orbicollis*. MS thesis, University of Oklahoma.

Mazerolle M. 2013. AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c). R package version 1.35. Available at http://CRAN.R-project.org/package=AICcmodavg.

McPherron MM, Jurzenski J, Albrecht MC, Farnsworth-Hoback KM, Walker TL, Hoback WW. 2012. Model of habitat suitability for American burying beetles in Nebraska's Loess Cantons ecosystem. *Trends in Entomology* **8**:27-36.

Merrick JM, Smith RJ. 2004. Temperature regulation in burying beetles (*Nicrophorus spp.*: Coleoptera: Silphidae): effects of body size, morphology and environmental temperature. *The Journal of Experimental Biology* **207**:723-733.

Murlis J, Elkinton JS, Carde RT. 1992. Odor plumes and how insects use them. *Annual Review of Entomology* **37**:505-32.

Noss RF, LaRoe ET III, Scott JM. 1995. Endangered ecosystems of the United States: A preliminary assessment of loss and degradation. Biological Report 28, National Biological Service: Washington, D.C.

[NRCS] Natural Resources Conservation Service. 2013. Soil Survey Geographic Database (SSURGO 2.2). U.S. Department of Agriculture, Natural Resources Conservation Service. Available at http://datagateway.nrcs.usda.gov/

Oehlert GW. 1992. A note on the delta method. *American Statistician* **46**:27–29.

Otronen M. 1988. The effect of body size on the outcome of fights in burying beetles (*Nicrophorus*). *Annales Zoologici Fennici* **25**:191-201.

Python. 2012. Python v2.7.3. Python Software Foundation. Available at http://www.python.org

Royle JA. 2004. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60**:108-115.

Royle JA, Dorazio RM. 2008. *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities.* Academic Press: London, UK.

Scott MP. 1998. The ecology and behavior of burying beetles. *Annual Review of Entomology* **43**:595-618.

Sealander JA, Heidt GA. 1990. *Arkansas Mammals: Their Natural History, Classification, and Distribution.* The University of Arkansas Press: Fayetteville, AR.

Sikes DS, Raithel CJ. 2002. A review of hypotheses of decline of the endangered American burying beetle (Silphidae: *Nicrophorus americanus* Olivier). *Journal of Insect Conservation* **6**: 103-113.

Sørenson R, Zinko U, Seibert J. 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences* **10**:101-112.

Stoddard HL. 1978. *Bobwhite Quail: Its Habitats, Preservation, and Increase.* Scribner's: New York, New York.

Taft JB. 1997. Savanna and Open-woodland Communities. In *Conservation in Highly Fragmented Landscapes.* Springer: New York, New York. pp. 24-54.

Taylor LR. 1963. Analysis of the effect of temperature on insects in flight. *Journal of Animal Ecology* **32**(1):99-117.

[USEPA] United States Environmental Protection Agency, [USGS] United States Geological Survey. 2012. National Hydrography Dataset Plus – NHDPlus. Available at http://www.horizon-systems.com/NHDPlus/

[USFWS] United States Fish and Wildlife Service. 1991. American Burying Beetle (*Nicrophorus americanus*) Recovery Plan. United States Department of Interior, Fish and Wildlife Service: Newton Corner, MA. 80 pp. Available at http://www.fws.gov/southdakotafieldoffice/abbrecoveryplan.pdf

[USFWS] United States Fish and Wildlife Service. 2011. American Burying Beetle *Nicrophorus americanus* Rangewide Survey Guidance. United States Department of Interior, Fish and Wildlife Service, Division of Ecological Services: Conway, AR. 20 April, 2011. Available at http://www.fws.gov/arkansas-es/docs/ABB%20Rangewide%20Surveys%20Protocol%204-20-2011%20final.pdf

[USFWS] United States Fish and Wildlife Service. 2014. American Burying Beetle *Nicrophorus americanus* Oklahoma Presence/Absence Live-trapping Survey Guidance. United States Department of Interior, Fish and Wildlife Service, Division of Ecological Services: Tulsa, OK. May, 2014. Available at http://www.fws.gov/southwest/es/oklahoma/ABB_Add_Info.htm

[USGS] United States Geological Survey. 2010. National Land Cover Version 1. National Biological Information Infrastructure, Gap Analysis Program (GAP). Available at: http://www.gap.uidaho.edu/Portal/DataDownload.html

[USGS] United States Geological Survey. 2013. USGS Global Visualization Viewer. United States Department of Interior, United States Geological Survey, Earth Resources Observation Center (EROS). Available at http://glovis.usgs.gov/

Walker JT Jr. 1952. Ecological studies of the arthropods associated with certain decaying materials in four habitats. *Ecology* **38**:262-276.

Wilson DS, Fudge J. 1984. Burying beetles: intraspecific interactions and reproductive success in the field. *Ecological Entomology* **9**:195-203.

Wilson DS, Knollenberg WG, Fudge J. 1984. Species packing and temperature dependent competition among burying beetles (Siliphidae, *Nicrophorus*) *Ecological Entomology* **9**:205-216.

Weather Underground, Inc. 2013. Weather History for Fort Smith, AR. Available at http://www.wunderground.com/history/airport/KFSM/. Accessed: 19 July, 2013.

Tables

Table 1.
Abbreviations and brief descriptions for all covariates assessed in competing models.

Covariate	Description
Detection Model	
TMP	Mean temperature
$TMP + TMP^2$	Mean temperature (quadratic)
TMP_{2433}	Number of hours from 24-33°C
DEW	Mean dew point
DEW_{2433}	Mean dew point when temp. 24-33°C
WND	Mean wind speed
WND_{2433}	Mean wind speed when temp. 24-33°C
TPOS	Topographic position
FOR	Coverage of forests within 100 m
Abundance Model	
TWET	Topographic wetness average
$TWET + TWET^2$	Topographic wetness average (quadratic)
SNDY	Coverage of soils with greater than 40% sand
GRS	Coverage of grasslands
SHB	Coverage of shrublands
WDL	Coverage of woodlands
FOR	Coverage of forests
NDWI	Normalized difference water index average
$NDWI + NDWI^2$	Normalized difference water index average (quadratic)
$\Delta NDWI$	Normalized difference water index average
$\Delta NDWI + \Delta NDWI^2$	Normalized difference water index average (quadratic)

Table 2.

Parameter estimates from final selected model. Parameter estimates represent linear increases in detection (logit-scale) and abundance (log-scale) expected to result from a one standard deviation increase in each covariate. Standard errors were calculated from 500 nonparametric bootstrap simulations.

Covariate	Estimate	SE
C-hat	3.651	
Detection Model		
Intercept	-1.921	0.186
YEAR-2008	1.030	0.334
YEAR-2009	-0.478	0.172
YEAR-2010	0.214	0.284
YEAR-2011	2.512	0.439
TMP	-0.337	0.064
TMP^2	-0.226	0.032
DEW_{2433}	0.710	0.069
WND_{2433}	0.276	0.099
Abundance Model		
Zero-inflation	-1.628	0.216
Intercept	3.222	0.377
YEAR - 2008	-0.949	0.586
YEAR - 2009	0.335	0.430
YEAR - 2010	-1.206	0.566
YEAR - 2011	-2.234	0.447
TWET	-0.440	0.194
$TWET^2$	-0.099	0.117
NDWI	-0.704	0.117
$\Delta NDWI$	0.293	0.099
$\Delta NDWI^2$	-0.105	0.063

Figures

Figure 1.
Model residuals for each year showing the difference between observed beetle counts and predicted beetle counts (site abundance x detection probability) for each sampling occasion. Positive residuals represent underestimation by the model and negative residuals represent overestimation.

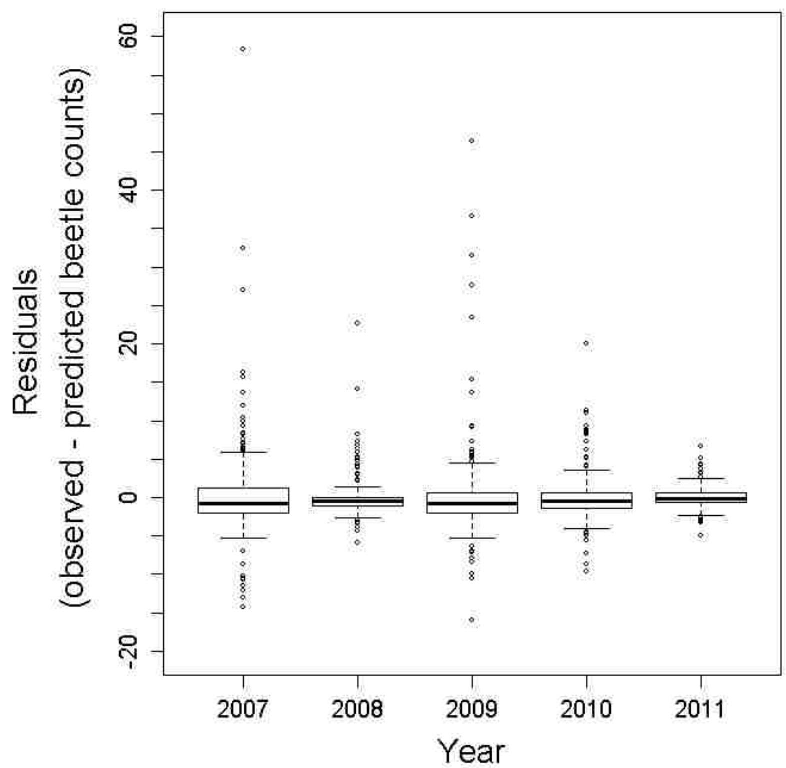


Figure 2.
Modeled relationships between observation covariates and detection probabilities. Thick lines are model predictions when other covariates are held to their means and thin lines are 95% confidence intervals. Scatterplots show model predictions, not observed data, for observation periods at each of our sites each year. When scatterplot points deviate from the line, it is due to additive effects of other covariates in the model, not model error. The gray regions indicate the range of thoracic temperatures during flight reported for *Nicrophorus hybridus* (Merrick & Smith 2004).

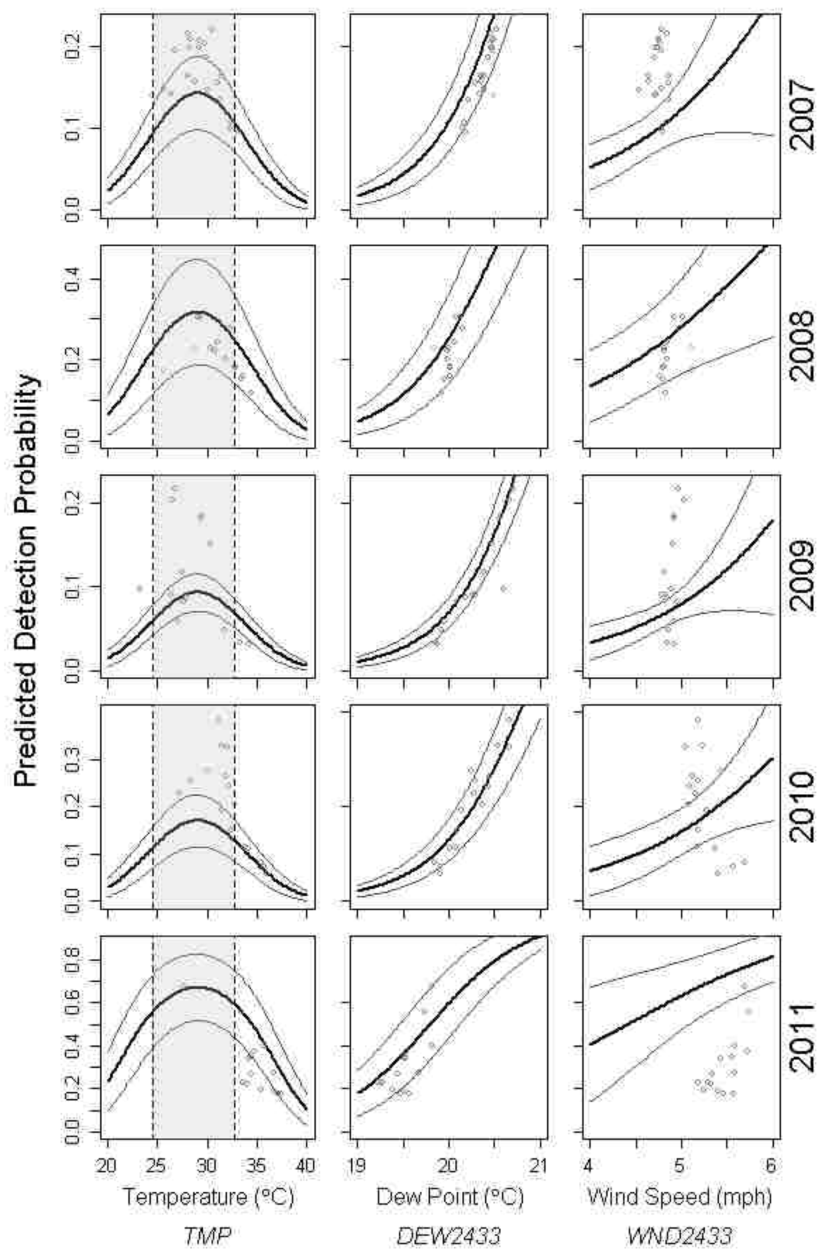


Figure 3.
Annual trend in predicted detection probabilities for observation periods at our sample sites. Fluctuating detection rates among years could obscure or exaggerate population trend estimates.

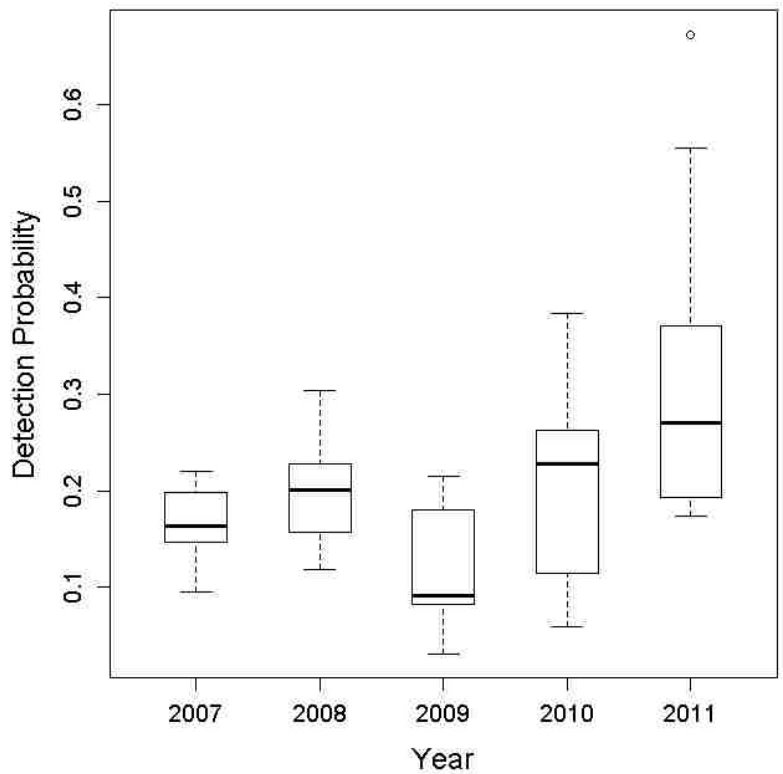


Figure 4.
Relationships between site covariates and abundance of *N. americanus.* Plots for continuous covariates assumed *YEAR* = 2007, and these relationships were representative of other years. Bold lines represent model predictions while other covariates were held to their means, and thin lines represent 95% confidence intervals. $NDWI_{800}$ values to the right of dotted lines generally represent forest communities and values to the left represent woodland, shrubland, or grassland communities. $dNDWI_{800}$ values between the dotted lines represent no major changes in vegetation, values above this range represent re-vegetation following disturbances from the previous year (*i.e.* fire), and values below this range represent disturbances that occurred during the current year.

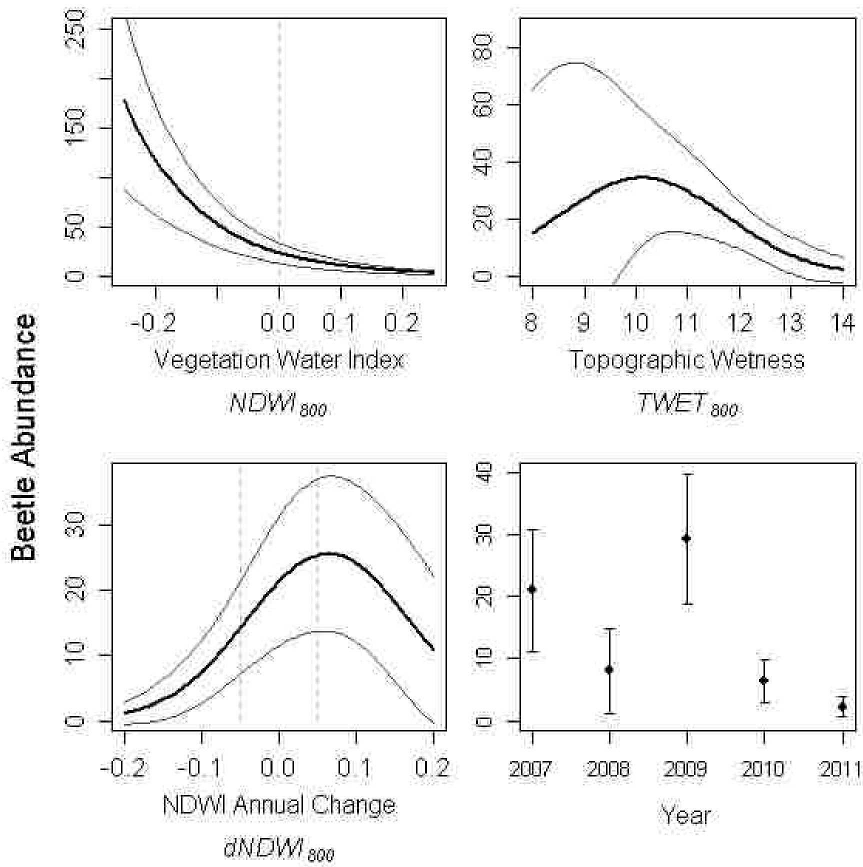


Figure 5.

Spatio-temporal dynamics of three site covariates and abundance model predictions at Fort Chaffee. The habitat model holds the *YEAR* factor constant at “2007” while the abundance model allows *YEAR* to vary. Red corresponds to low values and blue corresponds to high values in all maps.

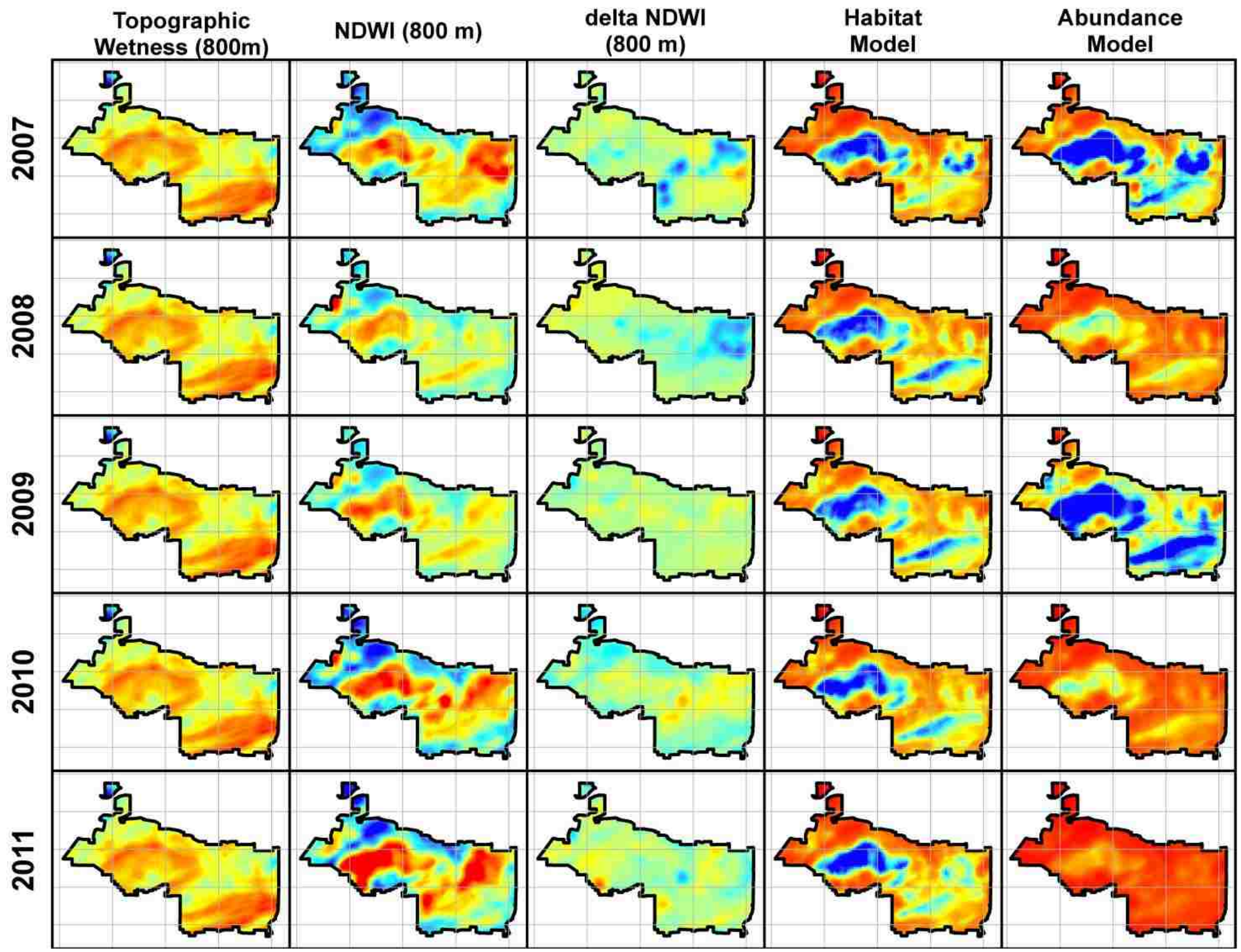
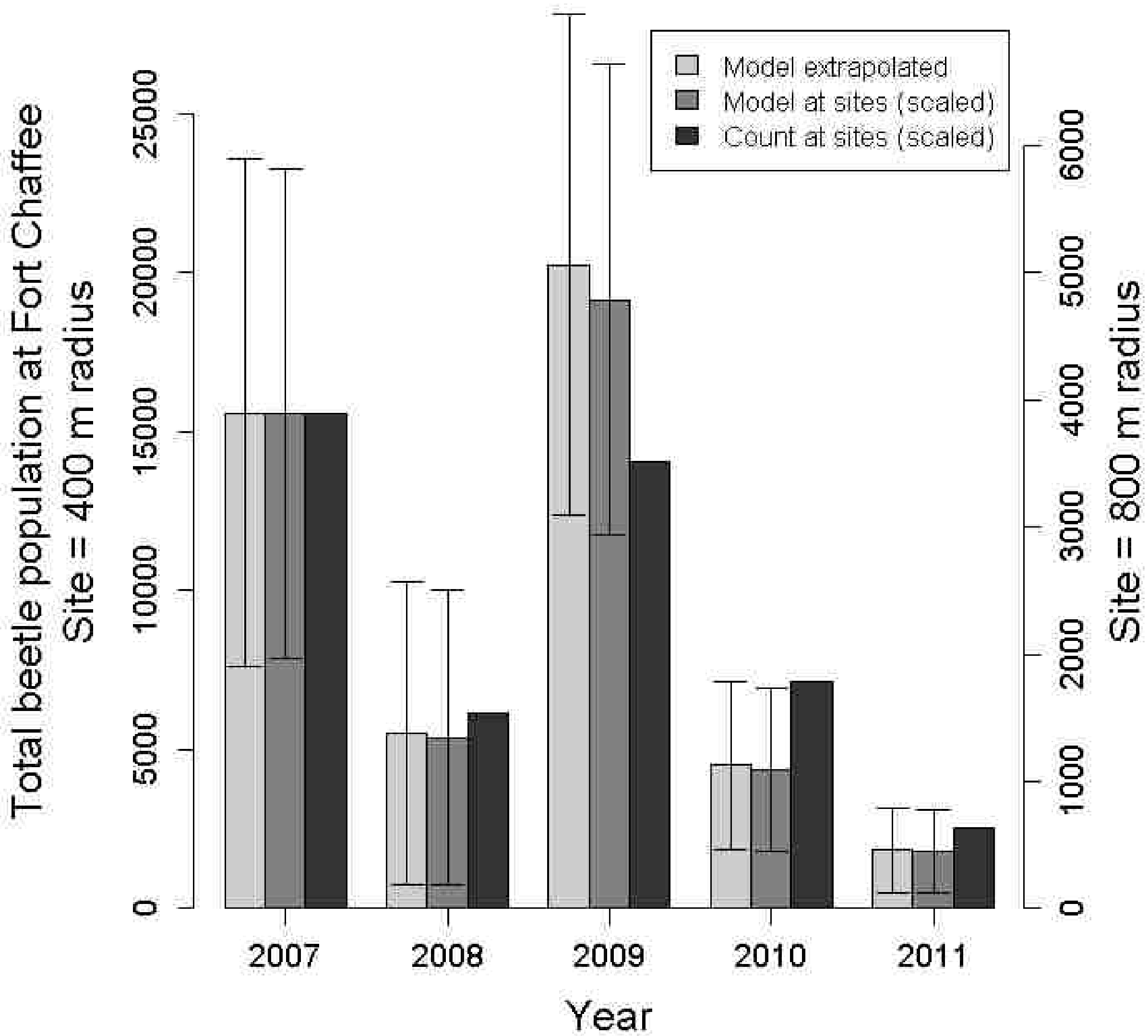


Figure 6.
Overall population trend of *N. americanus* at Fort Chaffee across 5 years estimated using three methods: model extrapolated across a 250 m grid, summing model predictions at 29 consistently sampled sites (re-scaled multiplying by 18.7), or summing maxima of raw counts among three sample periods at the same sites (re-scaled multiplying by 55). The y-axis on the left estimates the total population at Fort Chaffee using the extrapolated model assuming each sample site includes a 400 m radius around traps, and the y-axis on the right assumes sites include an 800 m radius around traps. Error bars represent 95% confidence intervals calculated using the delta method (Oehlert 1992, Mazerolle 2013) and a c-hat of 3.65.



Appendixes

Appendix 1: Lead Author Confirmation Letter

J. William Fulbright College of Arts and Sciences
Department of Biological Sciences

Chapter 2, "Landsat-based monitoring of an endangered beetle: Addressing issues of high mobility, annual life history, and imperfect detection" of D. R. Leasure's dissertation is intended for submission for publication.

I, Dr. Daniel D. Magoulick, advisor of Douglas Ryan Leasure, confirm Douglas Ryan Leasure will be first author and completed at least 51% of the work for this manuscript.

______________________________ ______________________________

Daniel D. Magoulick
Assistant Unit Leader
U. S. Geological Survey
Arkansas Cooperative Fish and Wildlife
Research Unit

Date

Science Engineering, Room 601 • Fayetteville, AR 72701-1201 • 479-575-3251 • Fax: 575-4010 • www.uark.edu

Chapter III:

Natural flow regimes of the Ozark-Ouachita Interior Highlands region

In press for *River Research and Applications*

Douglas R. Leasure[1], Daniel D. Magoulick[2], & Scott D. Longing[1,3]

[1] Arkansas Cooperative Fish and Wildlife Research Unit, Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701 USA

[2] U.S. Geological Survey, Arkansas Cooperative Fish and Wildlife Research Unit, Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701 USA

[3] Department of Plant and Soil Sciences, Texas Tech University, Lubbock, TX 79409 USA

Abstract

Natural flow regimes represent the hydrologic conditions to which native aquatic organisms are best adapted. We completed a regional river classification and quantitative descriptions of each natural flow regime for the Ozark-Ouachita Interior Highlands region of Arkansas, Missouri, and Oklahoma. On the basis of daily flow records from 64 reference streams, seven natural flow regimes were identified with mixture model cluster analysis: Groundwater Stable, Groundwater, Groundwater Flashy, Perennial Runoff, Runoff Flashy, Intermittent Runoff, and Intermittent Flashy. Sets of flow metrics were selected that best quantified nine ecologically important components of these natural flow regimes. An uncertainty analysis was performed to avoid selecting metrics strongly affected by measurement uncertainty that can result from short periods of record. Measurement uncertainties (bias, precision, and accuracy) were assessed for 170 commonly used flow metrics. The ranges of variability expected for select flow metrics under natural conditions were quantified for each flow regime to provide a reference for future assessments of hydrologic alteration. A random forest model was used to predict the natural flow regimes of all stream segments in the study area based on climate and catchment characteristics and a map was produced. The geographic distribution of flow regimes suggested distinct eco-hydrological regions that may be useful for conservation planning. This project provides a hydrologic foundation for future examination of flow-ecology relationships in the region.

Keywords: hydrologic classification, hydrologic alteration, eco-hydrology, flow ecology, ELOHA, measurement uncertainty, hydrologic index tool

Introduction

The natural flow regimes of stream and river ecosystems represent the undisturbed hydrologic conditions to which native aquatic organisms are best adapted (Poff *et al.*, 1997; Bunn & Arthington, 2002; Lytle & Poff, 2004). Several ecologically important components of natural flow regimes are susceptible to human alterations such as average flow, flow variability, flood frequency, and duration of low flow events (Poff *et al.*, 1997). Anthropogenic alterations of natural flow regimes from dams, agriculture, forest loss, and urbanization are widespread throughout North America, often restructuring aquatic communities in favour of species best adapted to disturbed hydrologic conditions (Poff *et al.*, 1997; Bunn & Arthington, 2002; Carlisle *et al.*, 2010a). Ecological consequences of flow alteration may differ among natural flow regimes, each with unique species assemblages, flow-ecology relationships, and susceptibility to particular forms of flow alteration.

Hydrologic classification has been widely adopted in eco-hydrology, often with the goal of characterizing flow-ecology relationships and crafting water management policies for individual types of rivers or streams. Olden *et al.* (2011) gave a recent review and methodological framework for hydrologic classification that we largely adopted here. This process has been used to develop several state, regional, and national river classification systems (Poff, 1996; Kennard *et al.*, 2010b; Liermann *et al.*, 2011; McManamay *et al.*, 2011).

Poff *et al.* (2010) provided a framework for assessing ecological limits of hydrologic alteration that included four steps: 1) identification of reference hydrographs and flow metrics, 2) hydrologic classification of natural flow regimes, 3) assessment of flow alteration, and 4) establishing relationships between flow alteration and ecological responses. We adopted this framework for the current study and completed the first two steps of the process. In this framework, flow regimes are characterized using at least five ecologically important components of natural flow regimes (Richter *et al.*, 1996; Poff *et al.*, 1997; Poff *et al.*, 2010): magnitude, duration, frequency, timing, and rate of change of flows.

We identified natural flow regimes of the Ozark-Ouachita Interior Highlands region of Arkansas, Missouri, and Oklahoma and mapped the geographic distribution of natural flow regimes across all stream segments in the study area. Measurement uncertainties were assessed for 170 commonly used flow metrics and important metrics were identified that best quantified the natural range of variation observed for several ecologically important components of these flow regimes. This will provide a foundation for future work characterizing flow alteration-ecological response relationships in the region.

Methods

The boundaries of the study area were delineated based on ecoregions and hydrologic units in Arkansas, Missouri, and Oklahoma that included the Ozark Highlands, Boston Mountains, Arkansas Valley, Ouachita Mountains, and Arkansas's South Central Plains (USEPA 2010, USEPA & USGS 2012). USGS stream gaging stations within the project area (n = 491; Fig. 1) were screened to identify least-disturbed reference streams with adequate daily flow records to best represent the natural flow regimes of the region. Streams with less than 15 years of daily flow data or drainage areas more than 10,000 km^2 were excluded. Hydrologic alteration was assessed initially using the hydrologic disturbance index (Falcone *et al.*, 2010a; Falcone *et al.*, 2010b) which was a composite index based on quantity of water withdrawals, density of major dams, change in dam storage from 1950 to 2009, percent canals in the watershed, water discharge locations, road density, and land cover fragmentation. The hydrologic disturbance index ranged nationally from 1 to 42 with a national median of 15. All streams used in this analysis had hydrologic disturbance indices less than the median among all gaged streams in the study area (HDI $\leq$ 13). Arkansas' Mississippi alluvial plain was excluded from this analysis because no streams had appropriate HDI values. Additional reference gage screening included reviews of annual water data reports available for 75% of gages (USGS, 2012), visual inspection of satellite imagery using Google Earth (Google Inc., 2013), and inspection of GIS data from the National Inventory of Dams (USACE, 2012), National Pollution Discharge Elimination System (USEPA, 2013), National Hydrography Dataset Plus (USEPA & USGS, 2012), and the 1992 National Land Cover Database (Vogelmann *et al.*, 2001). Sixty-six gaged streams in reference condition were retained for further consideration.

Average daily flow and peak annual flow data for the 66 remaining gages were obtained from the National Water Information System (USGS, 2012). An initial attempt was made to minimize measurement uncertainties associated with flow metrics by requiring a minimum of 15 complete years of flow data for each gage and a minimum 50% overlap among flow records (Kennard *et al.*, 2010a). To meet these requirements and to maximize the number of gages retained, a temporal window was selected beginning in water year 1955 and ending in water year 2010. Water years were from October 1 to September 30 and were named for the year in which they ended. The temporal window from 1955 to 2010 included 66 reference streams with at least 15 years of daily flow data, except one of these streams had only 14 complete years of data. Two strongly intermittent streams were eliminated from further analysis because their 25th percentiles of daily flow were zero resulting in many undefined or anomalous flow metrics. There were 64 reference gages retained for further analysis (Fig. 1), and their flow records averaged 63% overlap within our temporal window.

Average daily flow and peak flow data from 64 reference streams were processed using the Hydrologic Index Tool (Henriksen *et al.*, 2006a, 2006b) to produce 171 hydrologic indices describing the flow regimes of each stream. Flow metric names used here followed Olden & Poff's (2003) alpha-numeric designations that categorized metrics into nine categories representing ecologically important components of flow regimes (Appendix 1). These categories (and their alpha-designations) are magnitude of average flow (MA), magnitude of high flow (MH), magnitude of low flow (ML), frequency of high flow (FH), frequency of low flow (FL), duration of high flow (DH), duration of low flow (DL), timing (T), and rate of change (R).

Flow metrics describing relatively rare events can have high measurement uncertainty when period of record is short. We selected nine reference streams with at least 50 years of daily flow data to perform an uncertainty analysis following the approach of Kennard *et al.* (2010a; Fig. 1). Subsets of consecutive years of data were extracted from each flow record using random start years to provide 20 shortened flow records for each flow record length of 1, 2, 3, …, 30 years. The Hydrologic Index Tool was used to calculate 170 flow metrics for each new flow record. One metric (DL19) was excluded from the uncertainty analysis due to undefined values. For calculation of bias, precision, and accuracy of each flow metric associated with each period of record length, the flow metric values calculated from the complete records (>50 years) were treated as the expected values, and the flow metric values calculated from the shortened records (1, 2, 3, …, 30 consecutive years) were treated as the observed values. Bias was calculated as the absolute percentage difference between observed *o* and expected *e* values $\frac{|o-e|}{0.5(o+e)}$ averaged across the 20 replicates. Precision was calculated as the coefficient of variation (standard deviation/mean) generated across the 20 replicates. Accuracy (mean squared error; MSE) was calculated as the mean of the squared differences between observed and expected values. Flow metrics were excluded from further analysis if mean+SD of bias or precision were greater than 0.5, or if mean+SD of accuracy was greater than 0.25. These thresholds were selected subjectively to exclude metrics with the greatest measurement uncertainties based on a 15 year period of record.

To reduce the effect of river size on the flow classification, all magnitude metrics were divided by mean daily flow, median daily flow, or catchment area prior to variable selection and cluster analysis. Many magnitude metrics were automatically standardized by the hydrologic index tool, but we manually standardized additional magnitude metrics by dividing by mean daily flow (MA12-23, ML1-12, MH1-12, DL1-5, DH1-5, RA1, and RA3). All metrics were ln(x+1) transformed, except MH19 which was already normally distributed and centered on zero. Although choice of transformation can affect clustering results, we felt the log transform was justified to better meet the assumption of multivariate normality associated with Gaussian mixture models, and the precedent has been well established in the flow classification and statistical clustering literatures (Cheeseman & Stutz, 1996; Kennard *et al.*, 2010b; Liermann *et al.*, 2011; MacManamay *et al.*, 2011).

Olden and Poff's (2003) variable selection procedure based on principal components analyses was used to identify flow metrics from each of nine categories representing ecologically important components of flow regimes that best accounted for dominant patterns of variation while minimizing redundancy among retained flow metrics. Eight metrics were excluded *a priori* due to undefined values for intermittent streams (MA6-7, ML18, ML21, DL6-8, DL19). Metrics with unacceptable measurement uncertainty were also excluded. Principal components analyses (SAS, 2012) were conducted separately for flow metrics within each of the nine flow metric categories. From each category, a flow metric was selected from among the metrics with highest component loadings on the first principal component (*i.e.* within 0.05 of the highest loading). Measurement uncertainty, univariate cluster structure (*e.g.* bimodality of histograms), and ecological relevance were considered when selecting metrics. Despite its high measurement uncertainty, *no-flow days (DL18)* was added to the final list of selected metrics due to its central role in defining intermittent streams (*e.g.* Poff, 1996; Olden & Poff, 2003; McManamay *et al.*, 2011) and its ecological importance in the Interior Highlands where stream drying can play a major role in structuring ecological communities (Flinders & Magoulick, 2003; Dekar & Magoulick, 2007; Ludlam & Magoulick, 2010).

Gaussian mixture model clustering was used to classify 64 reference streams based on the 10 selected flow metrics using the R package MCLUST v4.0 (Fraley *et al.*, 2012; Fraley & Raftery, 2002; RStudio, 2012; R Core Team, 2012). Gaussian mixture models in MCLUST defined clusters as multivariate normal distributions that could be allowed to vary in terms of volume, shape, and orientation. Cluster attributes were parameterized as covariance matrices estimated using eigenvalue decomposition (Fraley & Raftery, 2002). MCLUST included 10 different parameterizations ranging in complexity from a simple model where all clusters were spherical with equal volume (similar to k-means clustering), to a complex model where volume, shape, and orientation were allowed to vary among clusters. With this model-based clustering approach, the number of clusters and their parameterizations were selected using Bayesian Information Criterion (BIC) to identify the most parsimonious and best fitting models. Our models used MCLUST default priors to minimize issues resulting from singularities in covariance matrices (Fraley & Raftery, 2007).

After all reference streams had been classified, the variable selection procedure (Olden & Poff, 2003) was repeated for each stream-type individually to identify custom sets of flow metrics that best described variation among streams within each natural flow regime. Flow metrics with unacceptable measurement uncertainty were excluded. The flow metric with the highest component loading was selected for each of the top three principal components with eigenvalues of one or greater. To provide a quantitative characterization of each natural flow regime, percentiles were calculated for these custom sets of flow metrics and also for the ten metrics used for clustering.

A random forest model was used to predict the natural flow regimes of all stream segments in our study area based on their landscape and climate characteristics using the R package RANDOMFOREST (Liaw & Wiener, 2002; Breiman, 2001). Decision tree classifiers like random forest have been used successfully in recent years to model a variety of stream flow characteristics and anthropogenic impacts (Carlisle *et al.*, 2010b; Kennard *et al.*, 2010b; Liermann *et al.*, 2011). Random forest models use an ensemble of classification trees where each tree is trained on a bootstrap sample from the original data. This ensemble approach produces pseudo-probabilities of class membership based on the proportion of trees voting for membership in each class. The class with the highest proportion of votes wins, and we calculated prediction uncertainty as the proportion of votes for all but the winning class. Bootstrap sampling helps prevent over-fitting and it also allows an out-of-bag error rate to be calculated by classifying each site using only trees that did not contain that site in their bootstrapped training data. Compared to error rates based on predictions from the full model, out-of-bag error rates more accurately reflect misclassification rates when applying the model at new sites. Over-fitting is also reduced by randomly selecting a small number of candidate variables to be assessed for discriminating among classes at each node in the classification trees, as opposed to assessing all variables at each node. Our random forest model assessed 10 variables at each node. Random forest provides a measure of variable importance based on decrease in classification accuracy caused by random permutations of each variable. We used this measure of variable importance to identify predictor variables to be included in a reduced final model. To balance class error rates among natural flow regimes that had imbalanced class frequencies, our random forest model used stratified bootstrap samples with per class sample sizes equal to class frequencies, and weighted penalties for class error rates that were the inverse of class frequencies (Chen *et al.*, 2004). Our goal was to achieve an overall out-of-bag classification error rate less than 25% and per class error rates less than 50%, similar to Liermann *et al.* (2011).

Landscape and climate characteristics were quantified at our reference gages using a custom Python script and the *ArcPy* package (PSF, 2010; Esri, 2013) to collect data for 249 GIS-based variables describing climate (Hijmans *et al.*, 2005; Diluzio *et al.* 2008; McCabe & Wolock 2009), topography (USEPA & USGS, 2012; Wolock & McCabe 1995), soil characteristics (Wolock 1997), geology (King & Beikman, 1974; Schruben *et al.*, 1997; Hunt, 1979), groundwater (Wolock, 2003a, Wolock, 2003b; USGS & USEPA, 1999; USGS, 2003), hydrology (Wolock & McCabe, 1999; USEPA & USGS, 2012), and land cover (*i.e.* coverage of forests & wetlands; Vogelmann *et al.*, 2001). A random forest model was built using all 249 variables and then reduced to include only the 30 most important variables (Table 1). Those 30 variables were collected at 24,557 stream segments throughout the study area using a custom Python script. A stream segment was defined as a section of stream between two confluences and the downstream-most point of a stream segment was used to delineate its catchment area. Landscape characteristics and climate were quantified within the entire catchment of each stream segment and also at the downstream-most point. Streams draining less than 5 km^2 or more than

10,000 km^2 were excluded. Two variables had undefined values at some stream segments: point-based percent clay, and catchment-based maximum stream slope. Dropping these two variables had no effect on classification error rates and so the reduced 28-variable random forest model was used to classify all stream segments.

Results

Uncertainty analysis suggested that a minimum period of record of 15 years was adequate to minimize measurement uncertainty for most flow metrics (Fig. 2). Twenty flow metrics were excluded from further analysis due to unacceptable measurement uncertainty based on a 15 year period of record (Fig. 3). The variable selection procedure identified nine flow metrics to be used in cluster analysis that best represented natural variation among reference streams for nine ecologically important aspects of flow regimes (Table 2, column 1): MA4, ML17, MH14, FL3, FH7, DL4, DH4, TA1, and RA3. Despite its high measurement uncertainty, we added *no-flow days* (DL18) to this list of selected metrics due to its ecological and hydrological importance.

The seven natural flow regimes identified with cluster analysis were: Groundwater Stable (GS), Groundwater (G), Groundwater Flashy (GF), Perennial Runoff (PR), Runoff Flashy (RF), Intermittent Runoff (IR), and Intermittent Flashy (IF). This seven class model maintained equal volume, shape, and orientation among clusters (model name = "EEE", G = 7). Classification uncertainties for the seven class model suggested strong model support (mean = 0.026%, max = 1.4%), and BIC clearly selected this model over the second best model (ΔBIC = 39). An interactive Google Earth map is provided that reports classification uncertainties, probabilities of class membership, and values for the 10 flow metrics used in cluster analysis for all reference gages (Appendix 2).

Each of the seven natural flow regimes occupied a distinct region of multivariate space defined by the 10 flow metrics used in cluster analysis (Fig. 4). Three broad flow regimes (groundwater, runoff, and intermittent) could be identified from just two flow metrics: *frequency of low flow spells (FL3)* and *no-flow days (DL18)*. Groundwater-influenced streams averaged less than two low flow spells per year (*i.e.* flow less than 5% mean daily flow; FL3), while other streams averaged more. Intermittent streams averaged more than 15 no-flow days per year, while other streams averaged less. Percentiles were reported for all flow metrics used in cluster analysis to provide guidelines for comparison with other streams (Appendix 3).

The *Groundwater Stable* flow regime was found in large rivers with extremely low flow variability and high constancy associated with significant groundwater recharge. These rivers had the greatest baseflows and they never had flow below 5% of mean daily flow. They had large drainage areas ranging from 1453 to 5278 km^2, and mean daily flows ranged from 728 to 2801 cfs. This flow regime was mostly restricted to the eastern end of the Ozark Plateaus aquifer where significant groundwater recharge occurred (Fig. 5).

The *Groundwater* flow regime was found in large rivers that had more daily flow variability and more frequent flooding than Groundwater Stable streams. Groundwater streams had decreased baseflow compared to Groundwater Stable streams, but flow never dropped below 5% of mean daily flow. Mean daily flow in this flow regime ranged from 474 to 3047 cfs, and drainage areas ranged from 1030 to 8236 km^2. This flow regime was associated with mainstem rivers of the Ozark Highlands with tributaries that were dominated by the Groundwater Flashy flow regime.

The *Groundwater Flashy* flow regime was found in streams with a range of drainage areas from 11 to 3237 km^2 that included some very small drainage areas usually associated with intermittent streams, but Groundwater Flashy streams had less daily flow variability than any runoff-dominated streams. Mean daily flow in Groundwater Flashy streams ranged from 4.3 to 905 cfs. These streams never dried up completely, although their flow was sometimes less than 5% mean daily flow, unlike Groundwater or Groundwater Stable streams. This flow regime had less constancy and lower baseflows than any of the other groundwater-influenced flow regimes. This was the most common flow regime in the Ozark Highlands ecoregion.

The *Perennial Runoff* flow regime averaged more low flow spells and had lower baseflows than any of the groundwater-influenced flow regimes, but these streams were rarely if ever reduced to zero flow. Perennial Runoff streams were most common along the edges of the Ozark Plateau aquifer where there was likely some minor groundwater recharge (Fig. 5). Compared to other runoff-dominated streams, this flow regime had higher baseflows and more constancy. Drainage areas of Runoff streams ranged from 257 to 2839 km^2, and mean daily flow ranged from 85 to 1157 cfs. Tributaries of these streams often consisted of a mixture of Runoff Flashy and Groundwater Flashy streams.

The *Runoff Flashy* flow regime was similar to the Perennial Runoff flow regime, except that these streams averaged 2 to 15 days of no flow per year. Streams with Runoff Flashy flow regimes tended to have slightly smaller drainage areas than Perennial Runoff streams, ranging from 105 to 1088 km^2, and mean daily flows ranged from 46 to 567 cfs. Baseflows were similar to Intermittent Runoff streams although Runoff Flashy streams receded slower, had fewer days of no flow, and less daily flow variability. This was the most common flow regime in the Boston Mountains ecoregion.

The *Intermittent Runoff* flow regime was found in streams with relatively small drainage areas (70 to 622 km^2) that averaged 14 to 50 days of no-flow per year. Mean daily flows ranged from 40 to 264 cfs. Compared to the Intermittent Flashy flow regime, these streams receded slower, had fewer no-flow days, and fewer low flow events. The Intermittent Runoff flow regime was common in the Ouachita Mountains, South Central Plains, and portions of the Arkansas Valley south of the Arkansas River.

The *Intermittent Flashy* flow regime was found in streams with very small drainage areas (8 to 22 km^2) that were dry for one to three months each year. Flow in these streams receded much faster than in any other flow regime. These streams averaged at least six low flow spells per year and mean daily flows ranged from 2.8 to 8.7 cfs. Constancy was high compared to Intermittent Runoff streams, probably due to longer periods of no flow. This flow regime was found in headwater streams throughout our study area (Fig. 5).

After the seven natural flow regimes had been identified, a custom set of nine flow metrics was selected for each flow regime separately (Table 2, columns 2-7). These custom sets of flow metrics best captured variation *within* flow regimes, as opposed to the ten flow metrics used in cluster analysis that were selected to best capture variation *among* flow regimes. For each custom set of flow metrics, percentiles were reported to quantify the bounds of variation expected for each flow regime under natural conditions (Appendix 4).

The random forest model to predict natural flow regimes at ungaged or disturbed streams adequately met our goals for out-of-bag classification error rates, even though the overall error rate of 26.6% was slightly higher than our goal of 25% (Table 3, Fig. 5). The flow regimes with the worst error rates were Intermittent Runoff and Groundwater Flashy that had out-of-bag error rates of 43% and 42%, respectively (Table 3). Misclassification of reference streams never occurred with predictions from the full model (*i.e.* 0% "in-bag" error rates). An interactive Google Earth map is provided (Appendix 2) with "probabilities" of belonging to each natural flow regime for all stream segments in the study area, as well as segment-specific values for the 28 landscape and climate variables used to classify streams (Table 1).

Discussion

The seven natural flow regimes that we identified in the Ozark-Ouachita Interior Highlands region can be used to guide future research and management. Our seven flow regimes could be interpreted as sub-classifications of Poff's (1996) three flow regimes from this region, similar to the hierarchical interpretation from another regional hydrologic classification in the southeastern United States (McManamay *et al.*, 2011). Graphical examination of our clustering results suggested that three broad flow regimes could be easily identified from only two flow metrics (FL3 & DL18; Fig. 4). These three groups corresponded with Poff's (1996) three flow regimes from this region: groundwater, perennial runoff, and intermittent runoff.

Our uncertainty analysis suggested that a minimum 15 year period of record was adequate to maintain acceptable measurement uncertainty for most metrics. Kennard *et al.* (2010a) analyzed measurement uncertainty using a different set of flow metrics describing Australian streams and also recommended a minimum 15 year period of record. The nine streams used in our uncertainty analysis included streams that were classified as Groundwater Stable (2), Groundwater (2), Groundwater Flashy (1), Perennial Runoff (1), and Runoff Flashy (3), but no intermittent streams met the criteria of at least 50 years period of record. Measurement uncertainties for some metrics could be more extreme for intermittent streams than our analysis suggested.

Some metrics with unacceptable measurement uncertainty may be ecologically or hydrologically important, such as *number of no-flow days (DL18)*. This metric can be difficult to quantify accurately with a 15 year period of record due to its sensitivity to rare events and potential bias when the period of record includes unusual drought like the most recent decade. The risk of including such a metric in cluster analysis is the identification of spurious hydrologic classes based on differences in period of record. We included this metric in our cluster analysis due to its ecological and hydrological importance in the Interior Highlands, although we also included another metric of low flow duration that had acceptable measurement uncertainty (DL4). Measurement uncertainty associated with *number of no-flow days (DL18)* may have had some effect on our cluster analysis, as evidenced by the Intermittent Flashy flow regime consisting of only streams with relatively short flow records (20-22 years; Fig. 4). However, we did not view this flow regime as a spurious class because it could also be distinguished based on other metrics with lower measurement uncertainty, such as fall rate and constancy (RA3 and TA1; Fig. 4).

The geographic distribution of natural flow regimes within our study area suggested distinct eco-hydrological regions that may be useful for conservation planning (Fig. 5). The Ozark Highlands were dominated by Groundwater Flashy streams with an area of Groundwater Stable rivers at the east end of the Ozark Plateaus aquifer. Perennial Runoff streams were mostly found along the edges of the Ozark Plateaus aquifer where minor groundwater influence likely reduced the intensity of stream drying. The Boston Mountains and Arkansas Valley north of the Arkansas River were dominated by Runoff Flashy streams. Areas south of the Arkansas River contained a mixture of Runoff Flashy and Intermittent Runoff streams. A small region of Groundwater Flashy streams occurred in the Ouachita Mountains near the Caddo and Cossatot Rivers. Native aquatic species are likely best adapted to the natural flow regimes that are dominant on the landscape and so we would expect regional differences in species composition to be associated with these eco-hydrological regions.

We identified important flow metrics (Table 2) and quantified their ranges of variability for each natural flow regime (Fig. 4, Appendices 2 & 3). These results provide quantitative guidelines that may be used to assess flow alteration in streams with adequate daily flow records (*i.e.* $\geq$15 years) that can be clearly assigned to a particular natural flow regime. However, *ad hoc* stream classification based on flow characteristics presented here may be problematic for gaged streams with inadequate flow records or with significant hydrologic alteration. For example, if flow variability were increased in a Groundwater stream due to urbanization, *ad hoc* stream classification could erroneously identify it as a Groundwater Flashy stream and its hydrology may appear normal for that type of stream despite being outside the normal conditions experienced by ecological communities of Groundwater streams. Our map of natural flow regimes (Figure 5, Appendix 2) can be used to identify the natural flow regimes of ungaged and disturbed streams within our study area. We encourage evaluation of multiple stream segments in the vicinity of the segment-of-interest, and consideration of class probabilities and prediction uncertainties.

Our study provides the hydrologic foundation and classification of natural flow regimes for the Interior Highlands region of Arkansas, Oklahoma, and Missouri. Native aquatic organisms of this region are adapted to hydrologic conditions associated with at least one of these natural flow regimes (Poff *et al.*, 1997; Bunn & Arthington, 2002). Regional water management should reflect hydrological differences among natural flow regimes and address their unique ecological sensitivities to flow alterations (Arthington *et al.*, 2006, Poff *et al.*, 2010). Important next steps will include assessing flow alteration, predicting ecological responses to flow alteration in each natural flow regime, and evaluating potential effects of climate change on regional hydrogeography. River conservation and water management in the Interior Highlands region will benefit from adopting the concept of natural flow regimes and a risk-based water management framework.

Acknowledgments

This project was funded by a State Wildlife Grant from the Arkansas Game and Fish Commission (grant no. T-30-12). Facilities were provided by the Arkansas Cooperative Fish and Wildlife Research Unit and the University of Arkansas. Dr. James Falcone provided his database of Hydrologic Disturbance Index values for USGS stream gages which helped with reference site selection, and he also provided several important GIS datasets that were used to predict the natural flow regimes of ungaged streams. Melissa Riskin from the New Jersey Water Science Center generously helped troubleshoot issues with the Hydrologic Index Tool even in the aftermath of hurricane Sandy. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Work Cited

Arthington AH, Bunn SE, Poff NL, Naiman RJ. 2006. The challenge of providing environmental flow rules to sustain river ecosystems. *Ecological Application* **16**:1311-1318.

Breiman L. 2001. Random Forests. *Machine Learning* **45**:5-32.

Bunn SE, Arthington AH. 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. *Environmental Management* **30**:492-507.

Carlisle DM, Wolock DM, Meador MR. 2010a. Alteration of streamflow magnitudes and potential ecological consequences: a multiregional assessment. *Frontiers in Ecology and the Environment* **9**:264-270.

Carlisle DM, Falcone J, Wolock DM, Meador RM, Norris RH. 2010b. Predicting the natural flow regime: Models for assessing hydrological alteration in streams. *River Research and Applications* **26**:118-136.

Chen C, Liaw A, Breiman L. 2004. *Using random forest to learn imbalanced data*. Department of Statistics, University of California, Berkeley: Berkely, CA. July, 2004. Tech report number 666.

Cheeseman P, Stutz J. 1996. Bayesian classification (AutoClass): theory and results. In *Advances in Knowledge Discovery and Data Mining* Fayyad U (ed). AAAI Press and MIT Press: Cambridge, MA; pp. 153-180.

Colwell RK. 1974. Predictability, constancy, and contingency of periodic phenomena. *Ecology* **55**:1148–1153.

Dekar MP, Magoulick DD. 2007. Factors affecting fish assemblage structure during seasonal stream drying. *Ecology of Freshwater Fish* 16: 335-342.

DiLuzio M, Johnson GL, Daly C, Eischeid JK, Arnold JG. 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *International Journal of Climatology* **28**:2031-2064.

[Esri] Environmental Systems Research Institute. 2013. ArcMap 10.2.0.3348. Earth Systems Research Institute (Esri): Redlands, CA.

Falcone JA, Carlisle DM, Weber LC. 2010a. Quantifying human disturbance in watersheds: variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams. *Ecological Indicators* **10**:264-273.

Falcone JA, Carlisle DM, Wolock DM, Meador MR. 2010b. GAGES: a stream gage database for evaluating natural and altered flow conditions in the conterminous United States. *Ecology* **91**: 621. (Ecol. Arch. E091-045-D1).

Flinders CA, Magoulick DD. 2003. Effects of stream permanence on crayfish community structure. *American Midland Naturalist* **149**:134-147.

Fraley C, Raftery A. 2002. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97**:611-631.

Fraley C, Raftery A. 2007. Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classification* **24**:155-181.

Fraley C, Raftery A, Scrucca L. 2012. Package ‘mclust’ v4.0: Normal mixture modeling for model-based clustering, classification, and density estimation. August 9, 2012.

Google, Inc. 2013. Google Earth ver. 7.1.2.2041. Google Inc.: Mountain View, CA. Available from http://earth.google.com

Henriksen JA, Heasley J, Kennen JG, Nieswand S. 2006a. Hydrologic Index Tool version 1.4. U.S. Geological Survey. Available from https://www.fort.usgs.gov/products/22798

Henriksen JA, Heasley J, Kennen JG, Nieswand S. 2006b. Users' manual for the Hydroecological Integrity Assessment Process software (including the New Jersey Assessment Tools): U.S. Geological Survey Open-File Report 2006-1093. 72 p.

Hijmans RF, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**:1965-1978.

Hughes JMR, James B. 1989. A hydrological regionalization of streams in Victoria, Australia, with implication for stream ecology. *Australian Journal of Marine and Freshwater Research* **40**:303–326.

Hunt CB. 1979. Surficial geology of the conterminous United States: U.S. Geological Survey, National Atlas of the United States of America.

Kennard MJ, Mackay SJ, Pusey BJ, Olden JD, Marsh N. 2010a. Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. *River Research Applications* **26**:137-156.

Kennard MJ, Pusey BJ, Olden JD, Mackay SJ, Stein JL, Marsh N. 2010b. Classification of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology* **55**:171-193.

King PB, Beikman HM. 1974. Explanatory text to accompany the geologic map of the United States. U.S. Geological Survey Professional Paper 901.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. *R News* **2**: 18-22.

Liermann CAR, Olden JD, Beechie TJ, Kennard MJ, Skidmore PB, Konrad CP, Imaki H. 2011. Hydrogeomorphic classification of Washington State rivers to support emerging environmental flow strategies. *River Research and Applications* **28**:1340-1358.

Ludlam JP, Magoulick DD. 2010. Environmental conditions and biotic interactions influence ecosystem structure and function in a drying stream. *Hydrobiologia* **644**:127-137.

Lytle DA, Poff NL. 2004. Adaptation to natural flow regimes. *Trends in Ecology and Evolution* **19**:94-100.

McCabe GJ, Wolock DM. 2009. Recent declines in western US snowpack in the context of 20th century climate variability. *Earth Interactions* **13**:1-15.

McManamay RA, Orth DJ, Dolloff CA, Frimpong EA. 2011. A regional classification of unregulated stream flows: spatial resolution and hierarchical framework. *River Research and Applications* **28**:1019-1033.

Olden JD, Poff NL. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research Applications* **19**:101-121.

Olden JD, Kennard MJ, Pusey BJ. 2011. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Ecohydrology* **5**:503-518.

Poff NL. 1996. A hydrogeography of unregulated streams in the United States and an examination of scale-dependence in some hydrologic descriptors. *Freshwater Biology* **36**:71-79.

Poff NL, Allen JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, Sparks RE, Stromberg JC. 1997. The natural flow regime. *Bioscience* **47**:769-784.

Poff NL, Richter B, Arthington AH, Bunn SE, Naiman RJ, Kendy E, Acreman M, Apse C, Beldsoe BP, Freeman M, Henriksen J, Jacobson RB, Kennen J, Merritt DM, O'Keefe J, Olden J, Rogers K, Tharme RE, Warner A. 2010. The Ecological Limits of Hydrologic Alteration (ELOHA): A new framework for developing regional environmental flow standards. *Freshwater Biology* **55**:147-170.

[PSF] Python Software Foundation. 2010. *Python 2.7.3*. Available at http://www.python.org/download/releases/2.7.3/. Accessed 4 October, 2012.

R Core Team. 2012. *R v2.15.3: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. Available at http://www.R-project.org. Accessed 1 March, 2013.

Richter BD, Baumgartner JV, Powell J, Braun DP. 1996. A method for assessing hydrologic alteration within ecosystems. *Conservation Biology* **10**:1163-1174.

RStudio. 2012. *RStudio v3.0.97.237*. RStudio, Inc.: Boston, MA. Available at http://www.rstudio.org. Accessed 11 September, 2012.

SAS. 2012. JMP Pro 10.0.0. 64-bit Edition. SAS Institute, Inc.

Schreve RL. 1966. Statistical law of stream numbers. *The Journal of Geology* **74**:17-37.

Schruben PG, Arndt RE, Bawiec WJ. 1997. Geology of the conterminous United States at 1:2,500,000 scale – A digital representation of the 1974 P.B. King and H.M. Beikman map. U.S. Geological Survey Digital Data Series 11, release 2.

[USACE] United States Army Corps of Engineers. 2012. National Inventory of Dams. Available at http://geo.usace.army.mil/pgis/f?p=397:1:0::NO. Accessed 6 August, 2012.

[USEPA] United States Environmental Protection Agency. 2010. *Level III Ecoregions of Arkansas*. USEPA Office of Research and Development – National Health and Environmental Effects Research Laboratory (NHEERL): Corvallis, OR. Available at http://www.epa.gov/wed/pages/ecoregions.htm. Accessed 28 May 2013.

[USEPA] United States Environmental Protection Agency. 2013. National Pollution Discharge Elimination System (NPDES). Available at http://www.epa.gov/enviro/geo_data.html. Accessed 21 November, 2013.

[USEPA] United States Environmental Protection Agency, [USGS] United States Geological Survey. 2012. National Hydrography Dataset Plus, version 2.10. Available at http://www.horizon-systems.com/NHDPlus/NHDPlusV2_data.php. Accessed May, 2013.

[USGS] United States Geological Survey. 2003. *Principal Aquifers of the 48 Conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands,Version 1.0.* U.S. Geological Survey: Reston, VA. Available at http://water.usgs.gov/lookup/getspatial?aquifers_us. Accessed 4 May, 2013.

[USGS] United States Geological Survey. 2012. National Water Information System: Web Interface. Available at http://waterdata.usgs.gov/nwis. Accessed 14 December, 2012.

[USGS] United States Geological Survey, [USEPA] United States Environmental Protection Agency. 1999. *National Hydrography Dataset*. U.S. Geological Survey: Reston, VA. Available at http://nhd.usgs.gov.

Vogelmann JE, Howard SM, Yang L, Larson CR, Wylie BK, Van Driel JN. 2001. Completion of the 1990's national land cover data set for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* **67**:650-662.

Wolock DM. 1997. STATSGO soil characteristics for the conterminous United States. U.S. Geological Survey Open-File Report 656.

Wolock DM, McCabe GJ Jr. 1995. Comparison of single and multiple flow-direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research* **31**:1315-1324.

Wolock DM, McCabe GJ Jr. 1999. Explaining spatial variability in mean annual runoff in the conterminous United States. *Climate Research* **11**:149-159.

Wolock DM. 2003a. Estimated mean annual natural ground-water recharge in the conterminous United States. U.S. Geological Survey Open-File Report 03-311.

Wolock DM. 2003b. Base-flow index grid for the conterminous United States. U.S. Geological Survey Open-File Report 03-263.

Tables

Table 1.
GIS-based variables describing site characteristics used by the random forest model to classify natural flow regimes of un-gaged streams. Two spatial extents were used to quantify site characteristics: point-based (P) and catchment-wide (C).

Site Characteristic	Spatial Extent	Resolution	Source
Climate			
Average precipitation wettest quarter	C	30 arcsec	Hijmans *et al.* 2005
Average precipitation in October	C	30 arcsec	Hijmans *et al.* 2005
Average temperature annual range	C	30 arcsec	Hijmans *et al.* 2005
Topography			
Maximum terrain slope	C	30 m	USEPA & USGS 2012
Soil Characteristics			
Average percent clay	C	1 km	Wolock 1997*
Average percent sand	C	1 km	Wolock 1997*
Average soil fractions less than 2mm	C	1 km	Wolock 1997*
Average soil fractions less than 5mm	C	1 km	Wolock 1997*
Coverage of soil hydrologic group D	C	1 km	Wolock 1997*
Average rainfall and runoff factor (R-factor)	P, C	1 km	Wolock 1997*
Average erodability (K-factor)	C	1 km	Wolock 1997*
Average rock depth	C	1 km	Wolock 1997*
Average organic matter content	C	1 km	Wolock 1997*
Average bulk density	C	1 km	Wolock 1997*
Geology			
Coverage of floodplain and alluvium gravel terraces	C	1:7,500,000	Hunt 1979*
Coverage of sandstone	C	1:2,500,000	King and Beikman 1974, Schruben *et al.* 1997
Coverage of dolostone	C	1:2,500,000	King and Beikman 1974, Schruben *et al.* 1997
Groundwater			
Coverage of Ozark Plateau aquifer	C	1:2,500,000	USGS 2003
Average baseflow index	P, C	1 km	Wolock 2003a
Average groundwater recharge index	P	1 km	Wolock 2003a
Number of springs	C	point feature	USGS & USEPA 1999
Density of springs	C	point feature	USGS & USEPA 1999
Hydrology			
Catchment area	C	30 m	USEPA & USGS 2012
Total length of streams	C	30 m	USEPA & USGS 2012
Average annual runoff	C	1 km	Wolock & McCabe 1999*
Stream order (Shreve method)	C	30 m	USEPA & USGS 2012, Shreve 1966

* Data provided by James Falcone, *personal communication*

Table 2.

Sets of non-redundant flow metrics selected to represent ecologically important components of each natural flow regime. Lists contain a flow metric to represent each of the top three principal components with eigenvalues greater than one. Underlined metrics were selected for cluster analysis. Metrics in bold were considered the best descriptors (*i.e.* highest component loading on the 1st principal component) for nine ecologically important aspects of each flow regime. Metrics in parentheses had equal component loadings.

	All Streams (n=64)	**Stable Groundwater** (n=5)	**Groundwater** (n=6)	**Groundwater Flashy** (n=12)	**Perennial Runoff** (n=13)	**Runoff Flashy** (n=17)	**Intermittent Runoff** (n=7)	**Intermittent Flashy** (n=4)
Magnitude:								
Average Flow	**MA4**, MA41, MA13	**MA19**, MA8, MA2	**MA37**, MA1,MA17	**MA44**, MA26, MA8	**MA29**, MA40, MA13	**MA5**, MA26, MA13	**MA44**, MA18, MA2	**MA34**, MA1, MA43
Low Flow	**ML17**	**ML8**, ML22	**ML7**, ML3	**ML10**, ML13, ML14	**ML19**, ML1, ML13	**ML19**, ML2, ML1	**ML8**, ML12, ML13	**ML10**, ML9, ML4
High Flow	**MH14**, MH13, MH18	**MH27**, MH18, MH23	**MH27**, MH17, MH18	**MH25**, MH20, MH6	**MH27**, MH17, MH18	**MH27**, MH20, MH13	**MH21**, MH4, MH5	**MH27**, MH1, MH14
Frequency:								
Low Flow	**FL3**, FL1	**FL1 (FL2)**, FL3	**FL1 (FL2)**, FL3	**FL2**	**FL1**	**FL3**, FL2	**FL1**	**FL1**
High Flow	**FH7**, FH9, FH11	**FH9**, FH11	**FH3**, FH2, FH10	**FH6**, FH4, FH10	**FH1**, FH3, FH2	**FH1**, FH4, FH2	**FH6**, FH10	**FH1**, FH11, FH9
Duration:								
Low Flow	**DL4,** DL16, DL18*	**DL14**, DL16	**DL11**, DL17	**DL3**, DL5, DL9	**DL3**, DL5, DL16	**DL12**, DL10, DL5	**DL3**, DL10, DL9	**DL2**, DL10, DL16
High Flow	**DH4,** DH8, DH7	**DH12**, DH14, DH19	**DH1**, DH15, DH24	**DH5**, DH1, DH7	**DH8**, DH11, DH23	**DH12**, DH15, DH6	**DH18**, DH2, DH14	**DH2**, DH18, DH21
Timing:								
Average, Low, & High Flow	**TA1**, TL2, TH1	**TA1**, TH1	**TA2**, TH2	**TA2**, TA3, TL1	**TA1**, TH1, TA3	**TA2**, TH2, TL1	**TH1,** TA1, TL2	**TA1**, TH2
Rate of Change:								
Average Flow	**RA3**, RA4	**RA9**, RA2	**RA1**, RA9	**RA3**, RA2	**RA1**, RA4	**RA1**, RA4, RA9	**RA1**, RA9	**RA3**, RA9

* DL18 (no-flow days) was selected based on ecological and hydrological relevance even though it was excluded from the metric selection process due to high measurement uncertainty.

Table 3.
Confusion matrix showing out-of-bag classification error for the random forest model that was used to predict the natural flow regimes of ungaged and disturbed stream segments throughout our study area. The overall out-of-bag error rate for the model was 26.56%. Rows represent the actual reference stream classifications and columns represent the out-of-bag predictions from the random forest model. Natural flow regimes are Groundwater Stable (GS), Groundwater (G), Groundwater Flashy (GF), Perennial Runoff (PR), Runoff Flashy (RF), Intermittent Runoff (IR), and Intermittent Flashy (IF).

	GS	G	GF	PR	RF	IR	IF	Class error rate (%)
GS	5	0	0	0	0	0	0	0
G	0	5	1	0	0	0	0	16.7
GF	0	1	7	2	1	0	1	41.7
PR	0	1	2	8	2	0	0	38.5
RF	0	0	0	2	14	1	0	17.6
IR	0	0	0	0	3	4	0	42.9
IF	0	0	0	0	0	0	4	0

Figures

Figure 1.
Geographic distribution of all available USGS stream gages, our 64 reference gages, and the nine gages used for uncertainty analysis.

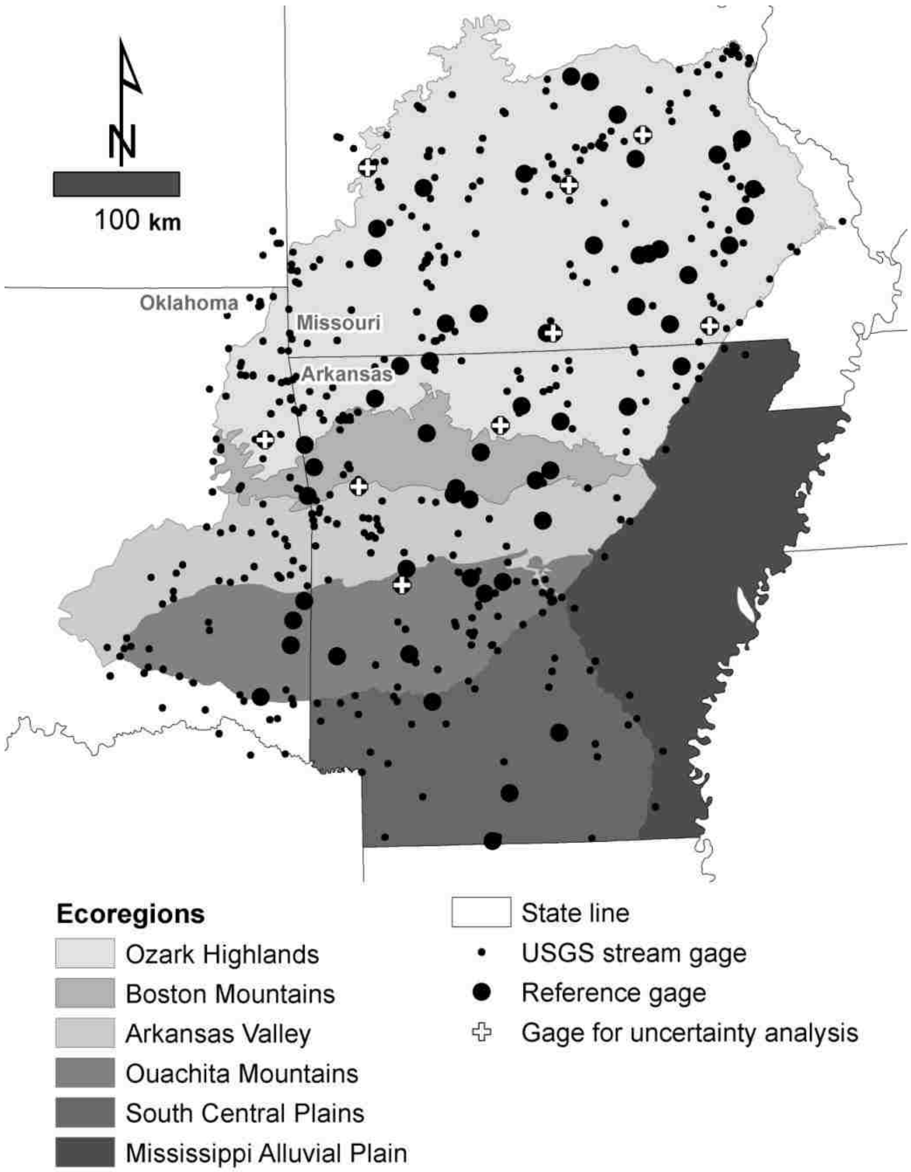


Figure 2.
Distribution of measurement uncertainties (bias, precision, and accuracy) among 170 flow metrics for increasing period of record lengths from 1 to 30 years. Each data point represents average uncertainty for a given flow metric among nine streams analyzed.

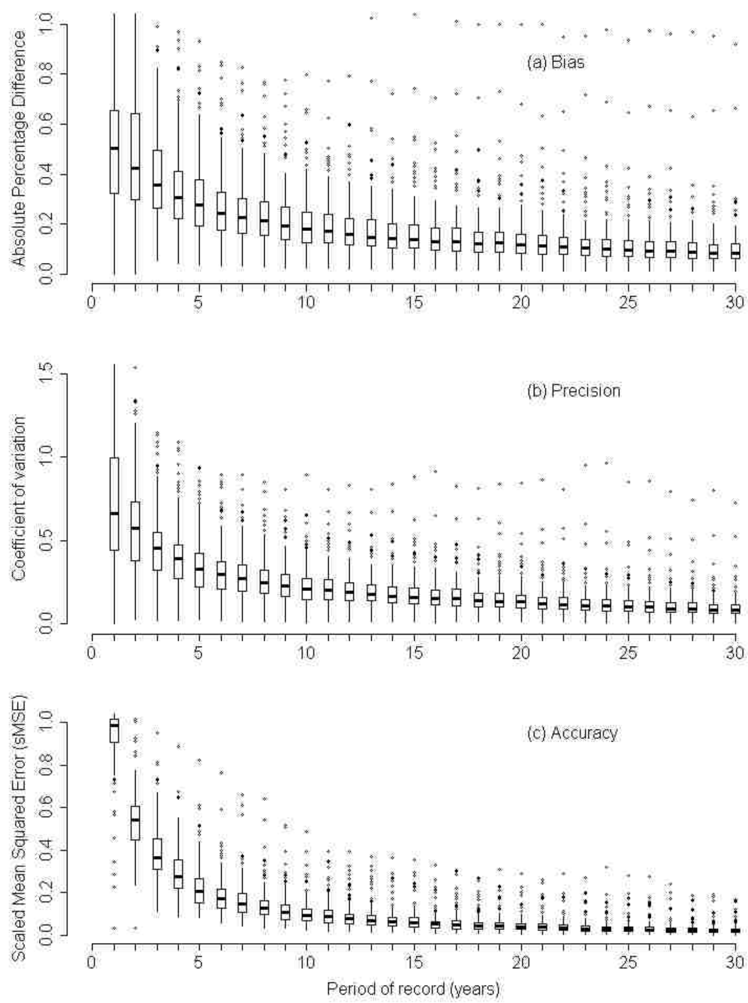


Figure 3.
Measurement uncertainties in terms of bias, precision, and accuracy for 170 flow metrics based on a 15 year period of record. Bars represent average values ± one standard deviation among the nine streams analyzed.

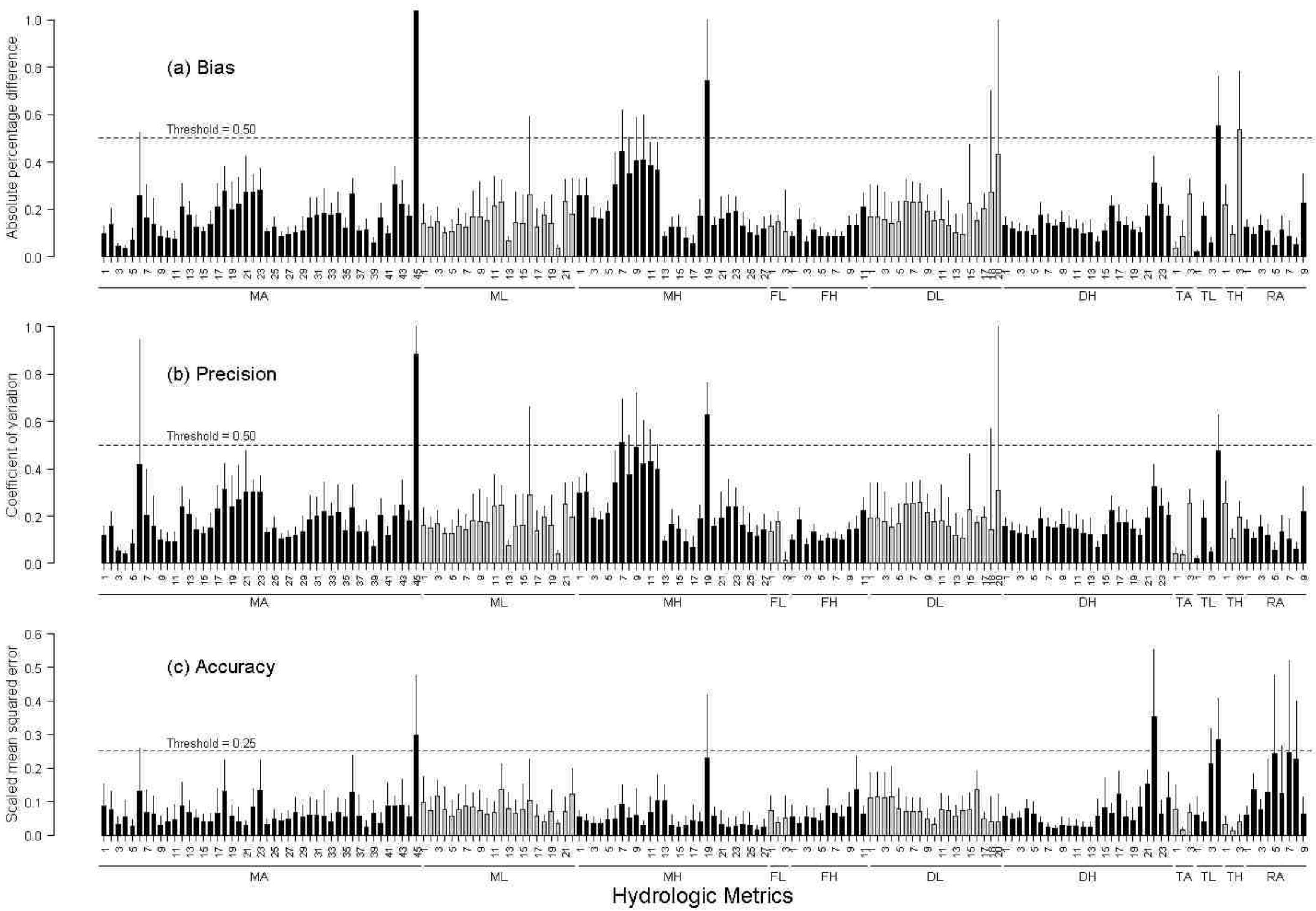


Figure 4.

Flow metrics used in cluster analysis compared among natural flow regimes: Groundwater Stable (GS), Groundwater (G), Groundwater Flashy (GF), Perennial Runoff (PR), Runoff Flashy (RF), Intermittent Runoff (IR), and Intermittent Flashy (IF).

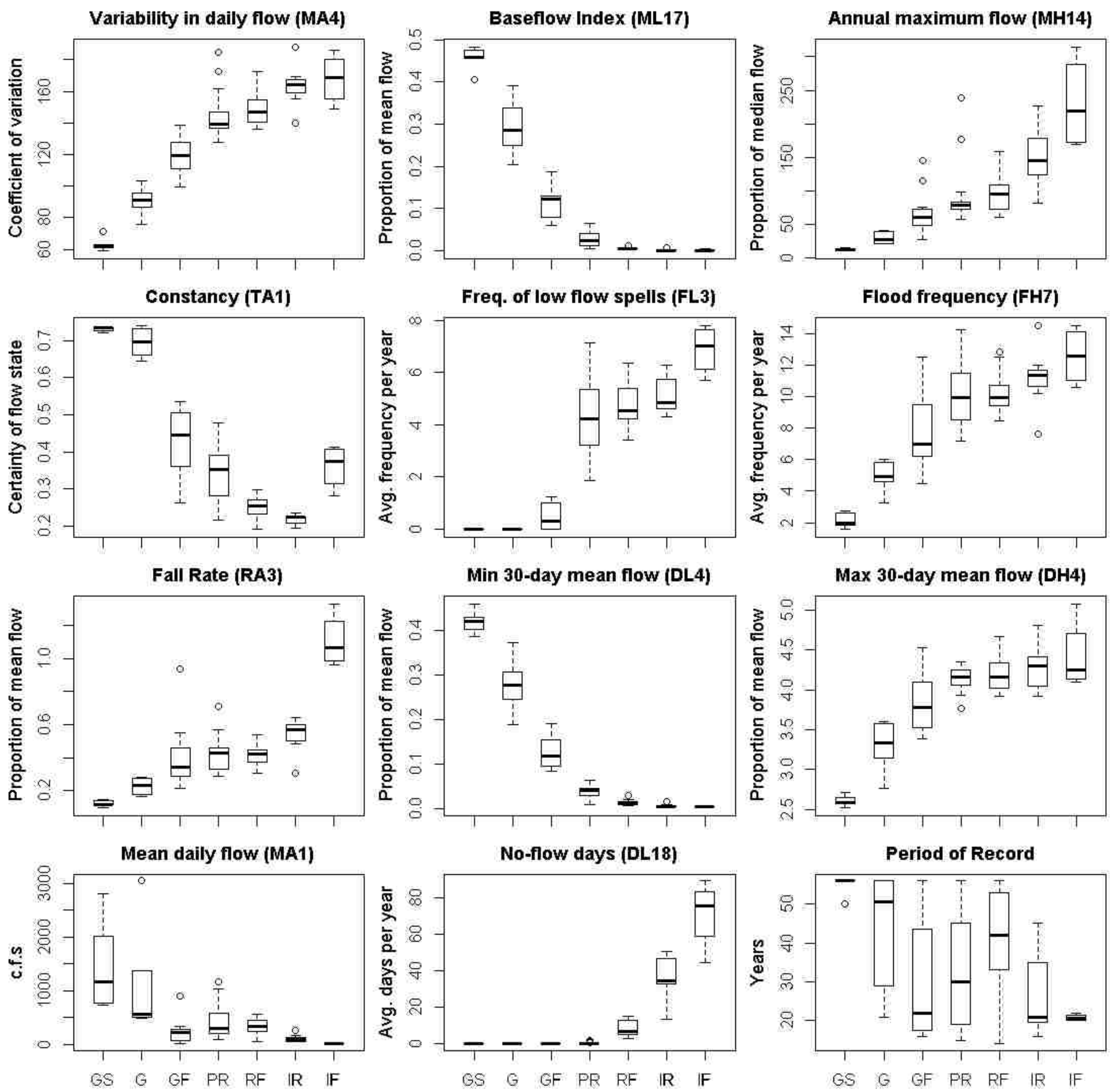
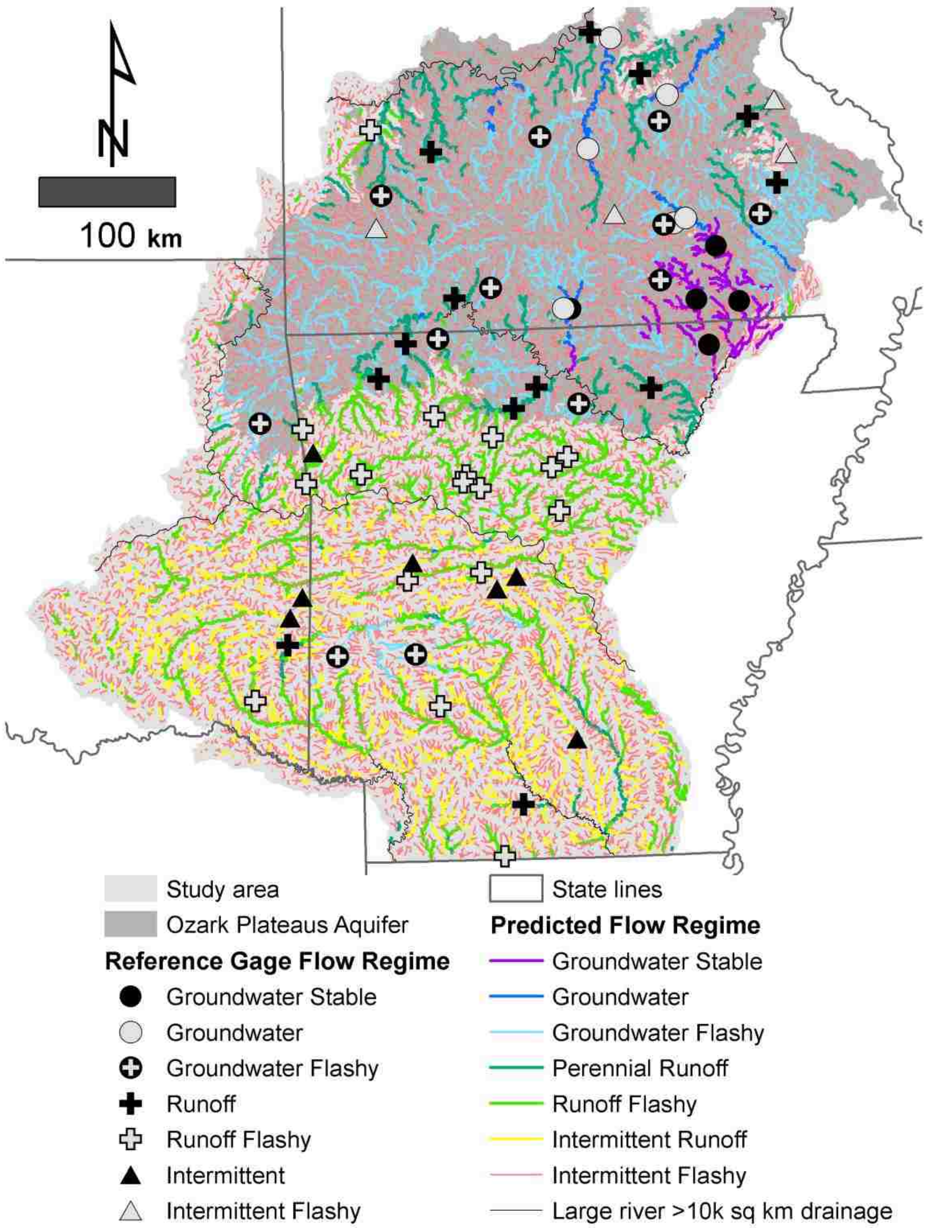


Figure 5.
Natural flow regimes of 64 reference gages were identified using mixture- model cluster analysis based on 10 flow metrics. Natural flow regimes of all stream segments were predicted based on climate and catchment characteristics using a random forest model.



Appendixes

Appendix 1.

Flow metric definitions and alpha-numeric designations. Table adapted from Olden & Poff (2003). Metrics are based on daily flow records that used cubic feet per second (cfs) as the measurement unit for flow. Detailed explanations of metric calculations are given in Appendix 5 of Henrikson *et al.* (2006b)

Code	Flow Metric Name	Description
Magnitude of flow events		
Average flow conditions		
MA1	Mean daily flows	Mean daily flow
MA2	Median daily flows	Median daily flow
MA3	Variability in daily flows 1	Coefficient of variation in daily flows
MA4	Variability in daily flows 2	Coefficient of variation of the logs in daily flows corresponding to the {5th, 10th, 15th,…,85th, 90th, 95th} percentiles
MA5	Skewness in daily flows	Mean daily flows divided by median daily flows
MA6-8	Ranges in daily flows	Ratio of 10th/90th, 20th/80th, and 25th/75th percentiles in daily flows over all years
MA9-11	Spreads in daily flows	Ranges in daily flows (MA6-8) divided by median daily flows
MA12-23	Mean monthly flows	Mean monthly flow for all months
MA24-35	Variability in monthly flows	Coefficient of variation in monthly flows for all months
MA36-38	Variability across monthly flows 1	Variability in monthly flows divided by median monthly flows, where variability is calculated as range, interquartile, and 90th-10th percentile.
MA39	Variability across monthly flows 2	Coefficient of variation in mean monthly flows
MA40	Skewness in monthly flows	(Mean monthly flow - median monthly flow)/median monthly flow
MA41	Mean annual runoff	Mean annual flow divided by catchment area
MA42-44	Variability across annual flows	Variability in annual flows divided by median annual flows, where variability is calculated as range, interquartile, and 90th-10th percentile.
MA45	Skewness in annual flows	(Mean annual flow - median annual flow)/median annual flow
Low flow conditions		
ML1-12	Mean minimum monthly flows	Mean minimum monthly flow for all months
ML13	Variability across minimum monthly flows	Coefficient of variation in minimum monthly flows
ML14	Mean of annual minimum flows	Mean of the lowest annual daily flow divided by median annual daily flow averaged across all years
ML15	Low flow index	Mean of the lowest annual daily flow divided by mean annual daily flow averaged across all years
ML16	Median of annual minimum flows 2	Median of the lowest annual daily flows divided by median annual daily flows averaged across all years
ML17	Baseflow index 1	Seven-day minimum flow divided by mean annual daily flows

ML18	Variability in Baseflow Index 1	Coefficient of variation in ML17
ML19	Baseflow index 2	Mean of the ratio of the lowest annual daily flow to the mean annual daily flow times 100 averaged across all years
ML20	Baseflow index 3	Ratio of baseflow volume to total flow volume
ML21	Variability across annual minimum flows	Coefficient of variation in annual minimum flows averaged across all years
ML22	Specific mean annual minimum flows	Mean annual minimum flows divided by catchment area
High flow conditions		
MH1-12	Mean maximum monthly flows	Mean of the maximum monthly flows for all months
MH13	Variability across maximum monthly flows	Coefficient of variation in mean maximum monthly flows
MH14	Median of annual maximum flows	Median of the highest annual daily flow divided by the median annual daily flow averaged across all years.
MH15-17	High flow discharge	Mean of the 1st, 10th, and 25th percentile from the flow duration curve divided by median daily flow across all years
MH18	Variability across annual maximum flows	Coefficient of variation of logarithmic annual maximum flows
MH19	Skewness in annual maximum flows	See Hughes and James (1989)
MH20	Specific mean annual maximum flows	Mean annual maximum flows divided by catchment area
MH21-23	High flow volume	Mean of the high flow volume (calculated as the area between the hydrograph and the upper threshold during the high flow event) divided by median annual daily flow across all years. The upper threshold is defined as 1, 3, and 7 times median annual flow
MH24-26	High peak flow 1	Mean of the high peak flow during the high flow event (defined by the upper threshold) divided by median annual daily flow. The upper threshold is defined as 1, 3, and 7 times median annual flow
MH27	High peak flow 2	See MH24-26, where the upper threshold is defined as the 25th percentile from the flow duration curve
Frequency of flow events		
Low flow conditions		
FL1	Low flood pulse count	Number of annual occurrences during which the magnitude of flow remains below a lower threshold. Hydrologic pulses are defined as those periods within a year in which the flow drops below the 25th percentile (low pulse) of all daily values for the time period.
FL2	Variability in low flood pulse count	Coefficient of variation in FL1
FL3	Frequency of low flow spells	Total number of low flow spells (threshold equal to 5% of mean daily flow) divided by the record length in years
High flow conditions		
FH1	High flood pulse count 1	See FL1, where the high pulse is defined as the 75th percentile
FH2	Variability in high flood pulse count 1	Coefficient of variation in FH1
FH3-4	High flood pulse count2	See FH1, where the upper threshold is defined as 3 and 7 times median daily flow, and the value is represented as an average instead of a tabulated count
FH5-7	Flood frequency 1	Mean number of high flow events per year using an upper threshold of 1, 3, and 7 times median flow over all years
FH8-9	Flood frequency 2	See FH5-7, where the 25th and 75th percentile are used as the upper threshold

FH10	Flood frequency 3	See FH5-7, where the median of the annual minima is used as the upper threshold
FH11	Flood frequency 4	Mean number of discrete flood events per year
Duration of flow events		
Low flow conditions		
DL1-5	Annual minima of 1-/3-/7-/30-/90-day means of daily discharge	Magnitude of minimum annual flow of various duration, ranging from daily to seasonal
DL6-10	Variability in annual minima of one-/3-/7-/30-/90-day means of daily discharge	Coefficient of variation in DL1-5
DL11-13	Means of 1-/7-/30-day minima of daily discharge	Mean annual 1-day/7-day/30-day minimum, respectively, divided by median flow
DL14-15	Low exceedence flows	Mean magnitude of flows exceeded 75% and 90% of the time (calculated from the flow duration curve) divided by median daily flow, respectively, over all years
DL16	Low flow pulse duration	Mean duration of FL1
DL17	Variability in low flow pulse duration	Coefficient of variation in DL16
DL18	Number of zero-flow days	Mean annual number of days having zero daily flow
DL19	Variability in number of zero-flow days	Coefficient of variation in DL18
DL20	Percent of zero-flow months	Percentage of all months with zero flow
High flow conditions		
DH1-5	Annual maxima of 1-/3-/7-/30-/90-day means of daily discharge	Magnitude of maximum annual flow of various duration, ranging from daily to seasonal
DH6-10	Variability in annual maxima of 1-/3-/7-/30-/90-day means of daily discharge	Coefficient of variation in DH1-5
DH11-13	Means of 1-/7-/30-day maxima of daily discharge	Mean annual 1-day/7-day/30-day maximum, respectively, divided by median flow
DH14	Flood duration 1	Monthly flow equalled or exceeded 95% of the time divided by mean monthly flow
DH15	High flow pulse duration	Mean duration of FH1
DH16	Variability in high flow pulse duration	Coefficient of variation in DH15
DH17-19	High flow duration 1	See DH15, where the upper threshold is defined as 1, 3, and 7 times median flows, and the value is represented as an average instead of a tabulated count
DH20-21	High flow duration 2	See DH17-19, where the upper threshold is defined as the 25th and 75th percentile of median flows
DH22	Flood interval	Mean annual median interval in days between floods over all years
DH23	Flood duration 2	Mean annual number of days that flows remain above the flood threshold averaged across all years
DH24	Flood free days	Mean annual maximum number of 365 days over all water years during which no floods occurred over all years
Timing of flow events		
Average flow conditions		
TA1	Constancy	See Colwell (1974), Henriksen *et al.* (2006b, Appendix 5)

TA2	Predictability of flow	Composed of two independent, additive components: constancy (a measure of temporal invariance) and contingency (a measure of periodicity)
TA3	Seasonal predictability of flooding	Maximum proportion of all floods over the period of record that fall in any one of six 60-day 'seasonal' windows
Low flow conditions		
TL1	Julian date of annual minimum	The mean Julian date of the 1-day annual minimum flow over all years
TL2	Variability in Julian date of annual minimum	Coefficient of variation in TL1
TL3	Seasonal predictability of low flow	Proportion of low-flow events $\geq$ 5-year magnitude falling in a 60-day 'seasonal' window
TL4	Seasonal predictability of non-low flow	Maximum proportion of the year (number of days/365) during which no 5-year + low flows have ever occurred over the entire period of record
High flow conditions		
TH1	Julian date of annual maximum	The mean Julian date of the 1-day annual maximum flow over all years
TH2	Variability in Julian date of annual maximum	Coefficient of variation in TH1
TH3	Seasonal predictability of non-flooding	Maximum proportion of the year (number of days/365) during which no floods have ever occurred over the period of record
Rate of change in flow events		
Average flow conditions		
RA1	Rise rate	Mean rate of positive changes in flow from one day to the next
RA2	Variability in rise rate	Coefficient of variation in RA1
RA3	Fall rate	Mean rate of negative changes in flow from one day to the next
RA4	Variability in fall rate	Coefficient of variation in RA3
RA5	No day rises	Ratio of days where the flow is higher than the previous day
RA6-7	Change of flow	Median of difference between natural logarithm of flows between two consecutive days with increasing/decreasing flow
RA8	Reversals	Number of negative and positive changes in water conditions from one day to the next
RA9	Variability in reversals	Coefficient of variation in RA8

Appendix 2.

An interactive Google Earth map document with natural flow regimes of reference streams and predicted natural flow regimes of all stream segments in the study area (n = 24,557). Reference gages can be clicked to report probabilities of class membership from the cluster model, cluster uncertainty, and values for the 10 flow metrics used in cluster analysis. Stream segments can be clicked to report probabilities of class membership from the random forest classifier, classification uncertainties, and values for 28 landscape and climate variables used by the random forest model. Probabilities of class membership from random forest are not true probabilities. They represent the proportion of classification trees in the random forest model that voted for membership in each class.

Appendix 3.

For each stream type, percentiles are given for catchment area, mean daily flow (MA1), and all flow metrics used in cluster analysis. Flow metrics are defined in Appendix 1. Hydrologic regimes are: Groundwater Stable (GS), Groundwater (G), Groundwater Flashy (GF), Perennial Runoff (PR), Runoff Flashy (RF), Intermittent Runoff (IR), and Intermittent Flashy (IF).

	Percentiles														
Hydrologic Regime	**5th**	**25th**	**50th**	**75th**	**95th**	**5th**	**25th**	**50th**	**75th**	**95th**	**5th**	**25th**	**50th**	**75th**	**95th**
	Daily flow variability (MA4)					**Baseflow Index (ML17)**					**Annual maximum flow (MH14)**				
GS	59.4	60.9	62.4	62.9	69.5	0.415	0.456	0.458	0.475	0.481	11.3	11.9	12.2	14.7	15.3
G	78.7	86.7	91.0	95.4	101.3	0.215	0.254	0.286	0.331	0.378	22.0	23.6	28.4	37.3	41.1
GF	102.8	112.8	119.1	124.3	137.9	0.070	0.081	0.122	0.130	0.160	29.8	49.2	60.9	70.7	128.5
PR	131.6	136.6	139.5	146.8	177.7	0.006	0.011	0.025	0.041	0.052	59.5	72.7	78.3	83.6	202.1
RF	137.1	140.3	146.8	154.7	170.3	0.003	0.004	0.005	0.006	0.009	64.2	73.3	96.0	108.7	126.6
IR	144.5	159.3	163.9	167.7	182.5	0.000	0.001	0.001	0.002	0.006	92.5	123.7	145.0	178.9	215.5
IF	151.0	158.7	168.7	177.9	184.2	0.000	0.001	0.001	0.002	0.004	170.6	174.7	219.7	275.8	306.7
	Constancy (TA1)					**Frequency of low flow spells (FL3)**					**Flood frequency (FH7)**				
GS	0.722	0.728	0.734	0.736	0.738	0.0	0.0	0.0	0.0	0.0	1.6	1.9	2.0	2.6	2.7
G	0.649	0.670	0.696	0.724	0.739	0.0	0.0	0.0	0.0	0.0	3.6	4.6	4.9	5.7	6.0
GF	0.293	0.362	0.447	0.502	0.526	0.0	0.0	0.3	1.0	1.2	4.8	6.3	7.0	9.5	10.9
PR	0.250	0.283	0.354	0.390	0.458	2.3	3.2	4.2	5.4	6.5	7.8	8.5	9.9	11.5	13.6
RF	0.209	0.234	0.256	0.271	0.292	3.9	4.2	4.5	5.4	6.4	9.0	9.4	9.9	10.7	12.6
IR	0.200	0.210	0.225	0.226	0.234	4.4	4.6	4.8	5.7	6.1	8.4	10.7	11.3	11.7	13.7
IF	0.293	0.331	0.375	0.405	0.411	5.8	6.3	7.0	7.6	7.8	10.7	11.3	12.6	13.9	14.4
	Fall Rate (RA3)					**Min 30-day mean flow (DL4)**					**Max 30-day mean (DH4)**				
GS	0.101	0.116	0.117	0.143	0.147	0.391	0.404	0.421	0.432	0.455	2.53	2.57	2.58	2.64	2.70
G	0.170	0.188	0.230	0.267	0.278	0.203	0.248	0.277	0.305	0.357	2.86	3.17	3.33	3.54	3.60
GF	0.240	0.291	0.344	0.449	0.726	0.087	0.098	0.119	0.153	0.173	3.40	3.56	3.78	4.06	4.44
PR	0.307	0.330	0.430	0.461	0.624	0.016	0.029	0.040	0.045	0.060	3.87	4.06	4.16	4.25	4.35
RF	0.331	0.374	0.420	0.446	0.538	0.008	0.010	0.013	0.016	0.022	3.93	4.02	4.15	4.34	4.62
IR	0.359	0.500	0.567	0.602	0.640	0.003	0.004	0.005	0.007	0.013	3.92	4.04	4.29	4.41	4.70
IF	0.968	1.003	1.067	1.171	1.298	0.002	0.004	0.005	0.005	0.005	4.11	4.15	4.25	4.52	4.97
	Mean daily flow (MA1)					**No-flow days (DL18)**					**Catchment area (km^2)**				
GS	737	773	1162	2018	2645	0	0	0	0	0	1573	2054	2937	4318	5086
G	480	504	560	1181	2630	0	0	0	0	0	1136	1457	1750	2977	7001
GF	28	87	226	274	592	0	0	0	0	0	80	212	584	778	1971
PR	120	209	301	581	1086	0	0	0	0	1.6	313	495	715	1365	2424
RF	95	238	335	455	547	2.9	5.3	6.9	12.8	14.5	140	438	624	816	1088
IR	43	58	84	128	233	19.0	32.9	34.7	46.7	50.2	76	98	119	202	498
IF	2.9	3.0	4.0	5.8	8.1	48.9	66.8	75.6	80.1	87.6	8.9	9.7	10.2	13.2	20.0

Appendix 4.

Percentiles of custom sets of flow metrics for each stream type, selected to best represent variation in each of nine ecologically important components of each natural flow regime (Table 2). Flow metrics are defined in Appendix 1.

Flow Metric	Percentiles 5th	25th	50th	75th	95th
			Groundwater Stable		
MA19	0.517	0.548	0.574	0.576	0.597
ML8	0.424	0.471	0.492	0.496	0.525
MH27	5.30	5.36	5.87	6.73	6.80
FL1	3.50	4.13	4.20	4.52	4.83
FH9	3.63	4.29	4.34	4.48	4.85
DL14	0.653	0.655	0.670	0.721	0.738
DH12	6.50	6.96	6.97	7.65	8.29
TA1	0.722	0.728	0.734	0.736	0.738
RA9	12.80	13.49	14.23	14.56	14.62
			Groundwater Flashy		
MA44	0.81	1.22	1.39	1.52	2.02
ML10	0.081	0.119	0.140	0.183	0.199
MH25	14.5	17.5	20.1	27.8	39.3
FL2	38.2	48.4	55.6	59.4	71.6
FH6	6.23	7.87	8.61	10.34	14.78
DL3	0.062	0.069	0.097	0.116	0.142
DH5	2.048	2.146	2.224	2.341	2.425
TA2	44.7	47.4	54.4	62.1	64.1
RA3	0.240	0.291	0.344	0.449	0.726
			Runoff Flashy		
MA5	3.04	3.23	3.79	4.08	5.61
ML19	0.166	0.260	0.304	0.403	0.579
MH27	21.9	25.4	31.1	34.5	44.4
FL3	3.88	4.22	4.53	5.40	6.36
FH1	9.13	9.71	10.00	10.96	12.65
DL12	0.0098	0.0120	0.0160	0.0210	0.0420
DH12	28.1	31.9	37.9	44.2	65.9
TA2	33.3	39.6	43.0	44.4	45.0
RA1	0.95	1.12	1.31	1.39	1.65
			Intermittent Flashy		
MA34	122.7	126.9	128.6	144.5	174.0
ML10	0.009	0.013	0.018	0.028	0.032
MH27	30.7	41.3	52.1	63.4	82.8
FL1	3.61	3.80	3.89	4.83	5.38
FH1	8.05	10.17	11.89	12.27	14.02
DL2	0.0002	0.0004	0.0010	0.0012	0.0043
DH2	17.13	17.41	18.21	19.80	21.59
TA1	0.200	0.210	0.225	0.226	0.234
RA3	0.359	0.500	0.567	0.602	0.640

Flow Metric	Percentiles 5th	25th	50th	75th	95th
			Groundwater		
MA37	1.28	1.41	1.50	1.62	1.63
ML7	0.267	0.288	0.328	0.378	0.439
MH27	8.24	9.35	11.24	12.81	13.58
FL1	5.12	5.21	5.42	5.51	6.06
FH3	42.7	48.3	52.6	56.4	59.2
DL11	0.366	0.459	0.494	0.516	0.557
DH1	13.3	17.7	20.4	22.1	22.8
TA2	71.0	72.2	75.3	77.8	79.8
RA1	0.434	0.538	0.641	0.730	0.760
			Perennial Runoff		
MA29	69.4	80.9	94.8	117.0	176.5
ML19	0.50	0.82	2.07	3.21	4.17
MH27	21.9	24.2	25.5	27.1	53.0
FL1	4.18	4.95	5.50	5.88	6.57
FH1	8.59	10.00	10.64	12.58	14.12
DL3	0.0060	0.0111	0.0234	0.0305	0.0422
DH8	42.7	52.6	55.3	58.6	68.6
TA1	0.250	0.283	0.354	0.390	0.458
RA1	0.79	1.03	1.28	1.42	1.90
			Intermittent Runoff		
MA44	0.99	1.06	1.19	1.59	2.19
ML8	0.001	0.003	0.004	0.006	0.010
MH21	132.7	187.8	227.4	282.9	652.3
FL1	3.61	3.80	3.89	4.83	5.38
FH6	7.42	9.63	11.22	12.12	13.13
DL3	0.0003	0.0006	0.0012	0.0015	0.0050
DH18	8.00	8.95	10.24	12.09	19.51
TH1	6.2	17.4	30.7	84.1	97.7
RA1	0.924	1.657	1.747	2.054	2.326

Appendix 5: Lead Author Confirmation Letter

J. William Fulbright College of Arts and Sciences
Department of Biological Sciences

Chapter 3, "Natural flow regimes of the Ozark-Ouachita Interior Highlands region" of D. R. Leasure's dissertation was published in *River Research and Applications* with coauthors: Daniel D. Magoulick and Scott D. Longing.

I, Dr. Daniel D. Magoulick, advisor of Douglas Ryan Leasure, confirm Douglas Ryan Leasure was first author and completed at least 51% of the work for this manuscript.

Daniel D. Magoulick Assistant Unit Leader U. S. Geological Survey Arkansas Cooperative Fish and Wildlife Research Unit	Date

Science Engineering, Room 601 • Fayetteville, AR 72701-1201 • 479-575-3251 • Fax: 575-4010 • www.uark.edu

Chapter IV:

Geodata Crawler: A centralized national geodatabase and automated multi-scale data crawler

Douglas R. Leasure

Department of Biological Sciences, University of Arkansas,
Fayetteville, AR 72701 USA

Abstract

There is now an unprecedented availability of GIS and remote sensing data that provides a powerful new tool for scientific research, but it is often difficult to acquire and process these data to generate tabular datasets that quantify landscapes at specific research sites using appropriate spatial scales for the questions being investigated. This can limit the number of sample locations and the variety of GIS data and spatial scales included in studies, and it may even prevent some researchers from utilizing GIS resources at all. Geodata Crawler contains a centralized national geodatabase with dozens of ecologically-relevant datasets including land cover types, soils, climate characteristics, hydrology, and human populations. The automated multi-scale data crawler delineates customized sample areas for user-locations anywhere in the continental United States and tabulates summary statistics from national geodatabases within those sample areas. Six spatial scales are available for delineating sample areas: point, local, watershed, riparian, local-watershed, and local-riparian. User options allow customization of these spatial scales by adjusting, for example, the site radius used by the local scale, or the stream buffer size used by the riparian scale. Geodata Crawler output includes 1) a project-specific geodatabase with all GIS layers required to collect user-requested data, 2) polygons representing sample areas delineated at each site, and 3) tabular data appropriate for most statistical analyses. Geodata Crawler can run on a single local machine or on a server allowing remote access by multiple users. Several time-saving features are available that include simultaneous processing of multiple projects on multiple processing cores, data archiving for rapid retrieval by other projects, and simultaneous processing of subsets of user-locations from a single project. Future directions for the Geodata Crawler project are discussed including web-based project submission and cluster computing.

Introduction

The unprecedented availability of GIS (*i.e.* geographic information systems) and remotely sensed data in recent years provides a rich resource that supports scientific research in many disciplines including landscape ecology, eco-hydrology, climate change, and landscape genetics. Increased availability of spatial data has been coupled with increased access to powerful analytical techniques like machine learning and Bayesian statistics that can accommodate these high dimensional datasets. Although significant progress has been made leveraging these tools, several impediments remain to efficient acquisition and processing of large GIS datasets to generate site-specific tabular data appropriate for particular research questions and analytical methods.

Many national and global GIS datasets are freely available for download, but access is scattered among various websites and FTP servers, and data are often stored in disparate file formats and spatial resolutions. Processing these data requires expert knowledge, specialized software, and significant computing power. Samples must then be extracted from national datasets for specific study locations at spatial scales that are appropriate for the research question being addressed. For example, two research projects, one studying endangered beetle habitat and the other studying natural stream hydrology, both required information from national land cover datasets about forest cover, but one was interested in forest cover within multiple site radii, while the other was interested in forest cover within watersheds of study sites (Chapters 2 & 3). The time-consuming and difficult process of delineating multi-scale sample areas for particular research locations and then tabulating appropriate data can severely limit the number of sample locations and the variety of GIS data included in a study.

Geodata Crawler was developed to address these issues by providing a centralized national geodatabase and an automated multi-scale data crawler that can rapidly build project specific GIS databases, delineate customized sample areas for user-defined locations, and tabulate requested data that can be exported in spreadsheet format suitable for most statistical analyses. Only minimal GIS skills are required from the user, and no GIS data are required other than sample locations and project boundaries. Geodata Crawler was coded using Python programming language using the *ArcPy* package which provides command-line access to Esri's ArcGIS toolbox (Esri 2013a, Python 2012). It requires ArcGIS for Desktop 10.2 with the advanced license and the Spatial Analyst extension. Exact system requirements have not been explored thoroughly, but known issues can occur on systems with less than 8 GB RAM or single core processors. Two primary computer platforms were used in the development of Geodata Crawler:

1. Windows 7 Professional (64-bit) with an Intel Core i7-2600 CPU (3.40 GHz, 16 GB RAM) and an external RAID 1 hard drive array (3 TB actual storage, USB 3.0), and
2. Windows Server 2008 R2 Enterprise (64-bit) with Intel Xeon CPUs (2 processors @ 2.13GHz, 24 GB RAM) and an internal RAID ?? SCSI hard drive array (4 TB actual storage).

Geodata Crawler can be setup to run on a local machine, or it can be setup to run on a server so that it can accept simultaneous data requests from remote users. Remote users can create a mapped drive to the Geodata Crawler server to submit jobs and retrieve output. This enables them to execute projects remotely without ArcGIS licenses on their local computers. Geodata Crawler can utilize multiple processors (or cores) to process several jobs simultaneously. Multi-processing functionality was implemented using Python package *multiprocessing*. Geodata Crawler can also be configured to archive site-specific data in a system geodatabase for rapid retrieval by future projects. This can drastically improve performance when a project requests a large amount of data that were previously archived by another project, but it can also decrease processing efficiency when previously archived data do not exist. The archiving function can be toggled on or off for individual projects (see *User Options* section). The archiving function is supported by Geodata Crawler's site identification system that provides a unique 11-digit numeric identifier for points along a 30 m grid covering the continental United States. All user-provided locations are snapped to this grid prior to analysis.

Input/Output

User-input generally consists of three items: 1) a Geodata Crawler user-options worksheet, 2) an ArcGIS polygon shapefile to identify the project boundary, and 3) an ArcGIS point shapefile to identify locations of-interest (optional). The user-options worksheet enables customization of a Geodata Crawler project to best meet specific research goals and to optimize processing efficiency. It contains a list of variables and spatial scales that can be toggled on or off. Geodata Crawler automatically builds project-specific geodatabases that contain all user-requested GIS layers clipped to their project boundaries. All data requested for a single project must use the same project boundary. User-provided locations will be used as data collection sites. If no sample sites are provided, random locations can be generated by Geodata Crawler. Users may also provide their own rasters or Landsat images for automated processing. All user-inputs must be projected in the Albers NAD 1983 datum.

Geodata Crawler output includes: 1) a project-specific geodatabase, 2) a point shapefile with an attribute table containing requested data for all sample sites, and 3) a polygon shapefile for each requested spatial scale containing boundaries for sample areas delineated at each site.

The project specific geodatabase will contain all of the original GIS layers as well as intermediate products required to collect requested data. It is named *PROJECT_NAME*/GEODATABASE_#dm.gdb, where # is the spatial resolution in decimeters of rasters in the geodatabase. This geodatabase will be retained by Geodata Crawler to use in subsequent data requests for the same project. A separate geodatabase will be created for site-specific output data and it will be named *PROJECT_NAME*/OUTPUT.gdb. Output data will include a point shapefile with an attribute table containing requested data. The attribute table can be exported to Microsoft Excel using the “TableToExcel” tool in ArcGIS. The exported table will contain a row for each sample site and a column for each requested variable. Column names begin with a designation for the spatial scale of data collection (Table 1). Appendix 3 provides detailed descriptions of output data. OUTPUT.gdb will also contain polygon shapefiles representing sample areas for each site, and their attribute tables will contain all data ever collected at that spatial scale for a project. These polygon features can be used to produce maps and to support custom GIS analyses. All output data is projected in the Albers NAD 1983 datum.

Variables

Geodata Crawler’s user options worksheet (Appendix 1) contains a row for each variable with a column for user-input that allows custom variable selection. A Geodata Crawler variable is a combination of:

1. A spatial scale for delineating sample areas (see *Spatial Scales* section),
2. A GIS layer from the national geodatabase (see *National Geodatabase* section), and
3. A statistic to summarize data within each location’s sample area (*e.g.* mean, mode, standard deviation, coverage of a particular map unit; see *Appendix 2*).

Some variables can be toggled with a simple true/false response in the user-options worksheet, but other variables require selection of individual map units for data collection. For example, *average elevation within watersheds* can be toggled with a true/false response, but *percent coverage of pine forest within watersheds* requires the user to specify pine forest as the land cover of-interest. Variables that require specific map units to be identified accept a list of map unit codes so that data can be collected for multiple map units during a single Geodata Crawler run. User options may be available to customize the way in which data are collected for some variables (see *User Options* section and Appendix 2). Appendix 2 provides descriptions of all Geodata Crawler variables and map unit codes, and it can be used as a reference when filling out the user-options worksheet. Appendix 3 provides descriptions of all column names that will be output for each variable, and it can be used as a reference when interpreting Geodata Crawler output.

Spatial Scales

Geodata Crawler automates the process of delineating sample areas associated with each user-defined location. It includes six flexible spatial scales used to delineate sample areas: point, local, watershed, riparian, local-watershed, and local-riparian. Several options are available to customize spatial scales to meet specific research goals (see *User Options* section).

The *point* spatial scale is the simplest because no sample area needs to be delineated for each user-defined location. The point coordinates themselves serve as the “sample area”. Variables collected at this spatial scale generally either extract a value from a GIS layer at each point, or they measure the distance from the point to a landscape feature (*e.g.* nearest forest patch greater than 10 hectares). Variables that measure the distance to a landscape feature can be customized using the user options *DIST_TO_RADIUS* and *DIST_TO_MIN_PATCH_SIZE*.

Local spatial scale (Fig. 1) is a circular sample area centered at a user-defined location. Its size is determined by a user-defined site radius that can be customized using the user option *LOCAL_RADIUS*.

Watershed spatial scale (Fig. 2) contains all land areas that drain into a user-defined location. This is delineated using the ArcGIS watershed tool (Esri 2013) based on the flow accumulation and flow direction rasters of the National Hydrography Dataset Plus (USEPA & USGS 2012). Watersheds can be delineated for terrestrial and stream sites. It is important to use the *MOVE_TO_STREAM* user option to ensure that locations meant to be on a stream are exactly located on the GIS representation of the stream channel.

Local-watershed spatial scale (Figs. 3 & 4) is the area where the local and watershed scales intersect. This contains land areas that drain into a user-defined location, but that are also within a user-defined radius of the site. This spatial scale can be customized using the user option *LOCAL_RADIUS*.

Riparian spatial scale (Figs. 5 & 6) contains areas within a site's watershed that are also within some user-defined distance of a stream. Riparian zones cannot be delineated for sites that are not on streams. Non-stream sites will be skipped in riparian-scale data requests. Sites can be snapped to streams using the *MOVE_TO_STREAM* user option, and this will guarantee that riparian-scale data are collected for all user-locations. Streams are delineated based on the flow accumulation raster of the National Hydrography Dataset Plus (USEPA & USGS 2012). The riparian spatial scale can be customized using the user options *RIPARIAN_BUFFER_WIDTH* and *MIN_SHED*.

Local-riparian spatial scale (Figs. 7 & 8) is the area where the local and riparian scales intersect. This contains land areas that drain into a user-defined location, but that are also within a user-defined radius of the site and within a user-defined distance from a stream. Local-riparian zones cannot be delineated for sites that are not on streams. This spatial scale can be customized using the user options *RIPARIAN_BUFFER_WIDTH*, *LOCAL_RADIUS*, and *MIN_SHED*.

National Geodatabase

Geodata Crawler includes a set of national GIS datasets used to derive over 1,000 variables available in Geodata Crawler. These data describe landscape characteristics like land cover, topography, hydrology, and climate, and they are all publicly available free-of-charge. New datasets are regularly added to Geodata Crawler, and it was designed to streamline this process. Data are projected to the Albers NAD 1983 datum and clipped to include only the continental United States before being stored in Geodata Crawler's system databases (ArcGIS file geodatabases). Both vector and raster data can be utilized, and rasters are stored using their original spatial resolutions. Some of these datasets may be used to derive additional datasets when project geodatabases are built. For example, flow accumulation rasters from the National Hydrography Dataset (NHD) are used to delineate streams for each project individually, rather than relying on NHD flow lines because this increases flexibility in how streams are delineated for each project and it also reduces the overall disk space required to store Geodata Crawler's system databases (*i.e.* currently 68.2 GB). Geodata Crawler resamples rasters to a uniform spatial resolution using bilinear interpolation as it builds project geodatabases, and it can be configured to either detect the highest spatial resolution among requested datasets or to simply resample all rasters to a system-defined resolution (*i.e.* 30 m). Rasters are never resampled to a lower resolution.

Geodata Crawler currently includes the following national geodatabases:

National Land Cover (USGS 2010, Fry *et al.* 2011, Homer *et al.* 2007, Vogelmann *et al.* 2001, Price *et al.* 2007):

These digital maps depict the geographic distribution of land cover classes for five time periods: 1970-80s, 1990s, 2001, 2006, and 2010. All of these data sets use hierarchical land cover classifications with two or three "levels". Each specific-level classification (*e.g.* pine forest) is nested within at least one broad-level classification (*e.g.* forest). The 1970-80s data have 37 classes and two levels. The 1992 data have 21 classes and two levels. The 2001 and 2006 data have 16 classes and two levels. The 2010 data have 583 classes and three levels. The 2001 and 2006 data include rasters that quantify percent impervious surfaces, and the 2001 data include a raster that quantifies percent canopy cover. See Appendix 2 for descriptions of all land cover classes and levels. These data are stored in raster format with 30 m spatial resolutions.

U.S. General Soil Map (USDA 2006):

This is a digital representation of the U.S. general soil map and it depicts the geographic distributions of 9,193 soil types. These data were originally stored in vector format with accompanying tabular data describing soil attributes. Soil data were originally mapped in 1-by 2-degree quadrangles. See Appendix 2 for names and some attributes of all soil types. Original data were converted to raster format for the Geodata Crawler system database and stored with 30 m spatial resolution.

Geologic Rock Type (Schruben *et al.* 1994, King & Beikman 1974):

This is a digital representation of a geologic map depicting the geographic distributions of 175 rock types. See Appendix 2 names of all geologic rock types. It was originally stored in vector format and map units were mapped at the 1:2,500,000 scale. It was converted to raster format for the Geodata Crawler system database and stored with a 30 m spatial resolution.

Aquifers (USGS 2003):

This dataset identifies the boundaries of the principal aquifers in the conterminous United States mapped at 1:2,500,000 scale. Aquifer names and their rock types are included in the attribute table, and they are listed in Appendix 2. These vector data are stored as a polygon feature in an ArcGIS file geodatabase.

Baseflow Index & Groundwater Recharge (Wolock 2003a, 2003b):

The base-flow index raster was created by interpolating base-flow index values estimated at USGS stream gage locations. Base-flow is the component of stream flow that can be attributed to ground-water discharge into streams. The groundwater recharge raster provides an index of mean annual natural ground-water recharge calculated by multiplying the base-flow index raster by a raster of mean annual runoff values. The mean annual runoff data used for this calculation were long-term averages (1951-1980) of stream flow divided by drainage area. These data are all stored in raster format with 1 km resolution.

Soil attributes for hydrological modeling (Carlisle *et al.* 2010):

This is a set of 35 layers describing characteristics of soil and geology that were originally processed and used by Carlisle *et al.* (2010) for flow alteration modeling. These data were obtained from James Falcone, a USGS GIS specialist and co-author of the hydrological modeling project for which the data were developed (Carlisle *et al.* 2010). The data were originally derived from the U.S. general soil map (USDA 2006) or the geologic map of Reed and Bush (2005). All layers are stored in raster format with 1 km spatial resolution.

National Hydrography Dataset Plus (USEPA & USGS 2012):

This dataset includes rasters that quantify elevation, flow accumulation, and flow direction that are essential for delineating streams, watersheds, and riparian zones. All rasters have 30 m spatial resolution and are stored as raster mosaics.

National Hydrography Dataset (Smiley *et al.* 2009):

Two products from the National Hydrography Dataset are currently used by Geodata Crawler: The flow lines that are classified as artificial paths or canals, and the "spring seeps" points that identify geographic locations of known natural springs. Both are stored as vector data (line and point features) in an ArcGIS file geodatabase.

WorldClim (Hijmans *et al.* 2005):

This geodatabase contains a set of 67 global climate layers describing current climate conditions (*i.e.* 1950-2000). It includes measures of precipitation and temperature (annual and monthly mean, minimum, and maximum), and it includes 19 BioClim variables that were designed to measure biologically relevant aspects of climate like mean diurnal range in temperature and precipitation during the driest quarter each year. All layers are stored in raster format with 30 arc-second spatial resolution (~800 m).

PRISM U.S. Climate Normals (Daly *et al.* 2008):

This geodatabase contains a set of 39 climate layers for the conterminous United States describing current climate conditions (*i.e.* 1971-2000). It includes measures of annual and monthly precipitation and temperature (minimum and maximum). All layers are stored in raster format with 30 arc-second spatial resolutions (~800 m).

U.S. Census Grids (Seirup and Yetman 2006):

These raster datasets describe human demographics based on U.S. Census data from 2000, including population density, household density, education levels, income, and ethnicities. All rasters have 30 arc-second spatial resolution (~800 m).

Nighttime Lights (NOAA 2012):

These data provide an index of intensity of nighttime lights viewed from space. The files are cloud-free composites made using data from the Operational Linescan System (OLS) of the Defense Meteorological Satellite Program (DMSP). The data were screened to minimize bias due to sunlit areas, glare, clouds, or the aurora. Two time periods have been incorporated into Geodata Crawler (2000 and 2010), but data are available for all years 1992 to 2012 (NOAA 2012). Several related datasets are available such as average visible light, average stable lights (discarding ephemeral lights such as wildfire), and average light intensity scaled by the frequency of light detections in each pixel. Scaling data by frequency of detections normalizes data to account for variations in the persistence of lighting. These raster datasets have 30 arc-second spatial resolutions (~800 m).

Agricultural Pesticides (Nakagaki 2007a, 2007b):

These data quantify pesticide applications (kg/km^2) in 1992 and 1997. The 1992 data include a raster for each of 199 pesticides and the 1997 data include a raster for each of 219 pesticides. Data for both years include a raster with application rates summed across all pesticides. Non-agricultural uses of pesticides are not included in these datasets. All rasters have 1 km spatial resolution.

Roads 2011 (USDC 2011):

This national roads layer was created by merging TIGER/Line road layers for primary and secondary roads in each state. Roads are categorized as primary, secondary, local/rural, or four-wheel drive roads (see Appendix 2 for all road types). These vector data are stored as a line feature with an attribute table in a file geodatabase.

Forest Service Active Fire Mapping (USDA 2011):

These data were obtained as a point shapefile that depicted active wildfires detected throughout each year 2000-2011 by the MODIS satellite. MODIS images have 1 km spatial resolution, 36 spectral bands, and they are acquired for every location on earth every 1 to 2 days. Points in the active fire shapefile represent the center of a MODIS image's pixel where a fire was detected. These data were converted to raster format with 1 km spatial resolution for use with Geodata Crawler. Each pixel was classified as either burned (1) or unburned (0) depending on presence of any active fire points for a particular year were within a pixel. The dataset was also used to create a national fire frequency raster that measured the number of fires per decade (2000-2011).

U.S. Historical Oil & Gas Development (Biewick 2008):

These data depict oil and gas development throughout the continental United States for each decade of the 20th century, and also for the pre-1900 and 2005-2006 time periods. Due to the proprietary nature of exact well locations, the U.S. was divided into grid cells ¼ square mile and oil/gas development was categorized as absent, oil only, gas only, oil and gas, or unknown well type. These raster data have 804.7 m (0.5 mile) spatial resolutions.

Arkansas Oil & Gas Development (AOGC 2013):

These data contain point locations of 55,126 oil and gas wells in the state of Arkansas. Attribute information identifies each well as gas, oil, or oil and gas, and identifies its status as active, inactive, permitted, plugged, or spud. These data were obtained as a Google Earth map document and converted to a point feature in an ArcGIS file geodatabase.

User Options

The user-options worksheet (Appendix 1) begins with 24 user options that allow users to customize each project. User-options can be manipulated to balance project needs with processing efficiency. User options require either true/false, text, or numerical responses, and default values are provided for all user-options. In addition to the user-options listed below, the user options worksheet contains a list of all variables that can be collected (see *Variables* section).

PROJECT_NAME (text):

A name for the Geodata Crawler system folder that will contain all data for a project. Data for all Geodata Crawler runs associated with a single project will be saved into this folder. All subsequent data collection runs for a single project must have identical project boundaries. The default value is “PROJECT”.

OUTPUT_FILE_NAME (text):

A name for point shapefile that will contain requested data for each sample location. The output file will be *PROJECT_NAME*/OUTPUT.gdb/*OUTPUT_FILE_NAME*. If a file with this name already exists in the project folder, it will be overwritten. The default value is “OUTPUT”.

USE_RANDOM_POINTS_YN (T/F):

If "T", a set of random points will be created within the project boundary and user-provided locations will be ignored. Two additional user-options control the number of random points and their spacing. The default value is "F".

RANDOM_POINT_COUNT (numeric):

The number of random points created. The default value is 10.

RANDOM_POINT_SPACING (numeric):

The minimum distance (in meters) allowed between random points. The default value is 50 m.

MIN_SHED (numeric):

The minimum drainage area (km^2) used to define the smallest stream-of-interest for the current data collection run. This value will be noted in output spreadsheets for any variables that may be affected by its value, and separate columns will be maintained in output spreadsheets if the same variable is collected using multiple values of this user-option. Separate stream layers will be created in the projects geodatabase for each value that is used for this option. The default value is 3 km^2.

MAX_SHED (numeric):

The maximum drainage area (km^2) used to define the largest stream-of-interest for the current data collection run. This value will be noted in output spreadsheets for any variables that may be affected by its value, and separate columns will be maintained in output spreadsheets if the same variable is collected using multiple values of this user-option. Separate stream layers will be created in the projects geodatabase for each value that is used for this option. The default value is *null* which will not impose an upper-limit of stream size.

MOVE_TO_STREAM_YN (T/F):

If "T", all sample locations will be moved to the nearest stream, centered on a raster cell of the digital elevation model used to delineate streams. This guarantees that points intended to be stream locations are not several meters away from the digital representation of the stream which can severely alter results, particularly for data collected from watersheds or riparian zones. This can also be used with random points to identify random stream sites. The default value is "F".

MOVE_TO_STREAM_SEGMENT_YN (T/F):

If "T", all sample locations will be moved to the nearest stream, and then moved to the downstream-most point of that stream segment. A stream segment is defined as a section of stream between two confluences. The default value is "F".

FIND_STREAM_SEGMENT_POINTS_YN (T/F):

If "T", all stream segments in the study area will be identified and saved as a GIS layer. Then, the downstream-most point of each segment will be identified and saved as a separate GIS layer. This will be automatically toggled on with *MOVE_TO_STREAM_SEGMENT_YN*. The default value is "F".

REMOVE_DUPLICATES_YN (T/F):

This option allows duplicate locations to be removed prior to sample area delineation and data collection. Duplicates will be removed after random points are created, points have been snapped to the nearest Geodata Crawler grid point (always within 30 m), and points have been moved to streams or stream segments. This will improve processing efficiency, and all duplicate rows will be retained in the output spreadsheet even though only one of them was actually processed. The default value is “T”.

RIPARIAN_BUFFER_WIDTH (numeric):

Defines the distance (in meters) away from streams used to delineate sample areas for upstream riparian zones. Very large values may result in sample areas that are identical to the watershed of a point because riparian zones cannot go beyond the watershed boundary. This value will affect delineation of sample areas at the riparian and local-riparian spatial scales. The default value is 1000 m.

LOCAL_RADIUS (numeric):

Defines the distance (in meters) used as a site radius to delineate local-scale sample areas. Sample locations that are within this distance from the project boundary may have incomplete data for local-scale variables. This value will affect delineation of sample areas at the local, local-watershed, and local-riparian spatial scales. The default value is 1000 m.

DIST_TO_RADIUS (numeric):

Some variables will measure the distance from user-locations to some landscape feature like a road or a patch of forest. This user-option sets the upper limit to search for the landscape feature and setting this value to a reasonably small distance (in meters) can drastically improve processing efficiency, particularly for projects with a large project area. The default value is 1000 m.

DIST_TO_MIN_PATCH_SIZE (numeric):

This option sets the minimum patch size (in hectares) to be recognized when measuring the distance from a user-location to a landscape feature like the nearest patch of forest.

Any patches smaller than this value will not be considered when making measurements. The default value is 0 which will not impose a lower limit on patch sizes being assessed.

LANDSAT_FILE_NAME (text):

Some variables allow a user to provide a Landsat image for automated processing by Geodata Crawler. The image must be provided as a multi-band composite .tif file. This option allows the user to specify the filename of the Landsat image. The default value is "LANDSAT.tif".

USER_RASTER_FILE_NAME (text):

Some variables allow a user to provide a raster dataset for automated processing by Geodata Crawler. The raster dataset must be provided as a single band .tif file. This option allows the user to specify the filename of this raster dataset. The default value is "USER_RASTER.tif".

QUERY_TYPE (0, 0.5, 1, 2, or 3):

0: Normal Geodata Crawler operation with automated sample area delineation for user-defined point features and data tabulation from within those sample areas based on user-defined variables.

0.5: Similar to query type 0, except user-defined points will be ignored if their sample areas do not already exist in the project's output geodatabase. Processing efficiency is improved because no sample areas will be delineated.

1: All data from Geodata Crawler's archives within the project boundary will be returned. User-defined points will be ignored, no sample areas will be delineated and user's variable selection will be ignored.

2: Data from Geodata Crawler's archives within the project boundary will be returned for sites that contain data for *any* of the variables selected by the user. User-defined points will be ignored and no sample areas will be delineated.

3: Data from Geodata Crawler's archives within the project boundary will be returned for sites that contain data for all of the variables selected by the user. User-defined points will be ignored and no sample areas will be delineated.

CREATE_SAMPLE_AREAS_ONLY_YN (T/F):

If "T", sample areas will be delineated for all user sites, but no data will be collected. The default value is "F".

CLONES (numeric):

This can be an integer from zero to four that will allow user-locations to be sub-divided (up to four times) to accommodate simultaneous processing of subsets of locations. This will drastically increase processing efficiency with a multi-core or multi-processing computer, but provides no advantage on a single core computer. Each cloned project will contain a subset of user locations and replicates of project geodatabases. Although replicating project geodatabases requires additional disk space (*e.g.* dozens of GB per clone for large projects), it facilitates multi-processing without requiring multi-user editing capability which is not supported by file geodatabases. Multi-user editing is supported by ArcSDE geodatabases (unavailable with ArcGIS for Desktop), and this would circumvent the need to clone projects. The default value is zero, resulting in no cloned projects being created.

TOGGLE_IMPORT_ARCHIVES (T/F):

If "T", Geodata Crawlers system archives will be searched for pre-existing data matching the data requested for the current project's locations. Performing this search can drastically increase processing time for projects with many locations and large data requests that are not in the archives. Using the archive function has resulted in read/write conflicts when processing multiple projects simultaneously. The default value is "F".

TestMode (T/F):

If "T", archived data will be written to a geodatabase separate from Geodata Crawler system archives to prevent potentially erroneous test data from being archived. Console windows will also be left open at the end of each project's run to allow manual manipulation of Python objects at the end of a run, or in the event of an error. The default value is "F".

ALL_VARIABLES (T/F):

If "T", all possible variables will be collected at user-defined locations. This is included for testing purposes and should never be used with more than a few locations. The default value is "F".

Future Directions

Geodata Crawler is a powerful new tool that can help to overcome GIS bottlenecks in data analysis work flows across many different research disciplines. Five important next steps are envisioned to broaden research applicability, improve processing efficiency, and increase ease-of-use:

1. Develop a web interface for setting up projects and submitting them to a server,
2. Adapt to the Linux operating system for cluster computing,

3. Shift to ArcSDE geodatabases with multi-user editing,
4. Add path-based spatial scales including stream paths, linear paths, and least-cost paths between sample locations,
5. Continue to incorporate new national GIS datasets, particularly future climate data for climate change research.

The Geodata Crawler website (http://www.geodatacrawler.com) gives a general overview of Geodata Crawler, its national geodatabases, and its spatial scales, but it does not currently support online setup or submission of projects. Adding this important feature would broaden the reach of Geodata Crawler. Building this capability will require web-based forms that walk users through the user-options worksheet and translate their responses into an appropriately formatted text file (USER_OPTIONS.csv). Users would also need the ability to upload their project files that may include a boundary shapefile (required), a user-locations shapefile, a Landsat image, and a custom user raster. An FTP server (ftp://ftp.geodatacrawler.com) is already available with password-protected user accounts where geodatacrawler output can be accessed. Output often requires large amounts of disk space and the FTP site provides remote data storage where users can browse data and download only what is required for their specific needs.

Transferring Geodata Crawler to a cluster computing platform would drastically increase its processing efficiency by allowing it to process many projects simultaneously. This would broaden its capacity to support the potentially large number of jobs being submitted through a web-based portal. Geodata Crawler already includes multi-processing functionality that allows simultaneous processing of multiple jobs on separate processing cores. This capability should transfer to allow utilization of dozens of processing cores available with cluster computing. The University of Arkansas High Performance Computing Center (http://hpc.uark.edu/) provides ideal computing platforms for developing and testing this capability. ArcGIS for Server (Esri 2013b) provides an appropriate software package.

A shift from file geodatabases currently used by Geodata Crawler to ArcSDE databases would allow multi-user editing capability to streamline the processes of data archiving and project subdivision for faster processing. Archiving is required to share data among projects and this can drastically improve processing efficiency. This would be even more important with potentially increased traffic through a Geodata Crawler web portal because benefits of sharing data among projects, and costs of re-collecting previously collected data, would increase with more data requests. Multi-user editing capability also enables large projects to be subdivided and then processed as multiple simultaneous sub-units. Sub-dividing projects and archiving are currently supported in Geodata Crawler only by using an inefficient work-around that requires duplication of project geodatabases, because file geodatabases (ArcGIS for Desktop) do not allow multi-user editing. ArcSDE databases with multi-user editing would allow projects to be subdivided across many processing cores (*e.g.* on a computing cluster) without requiring additional disk space for duplicating project geodatabases.

Geodata Crawler has the capability to delineate linear paths and stream paths that connect all pairwise combinations of user-locations, but these features are still being developed. These features were designed for the purpose of studying animal movements, but there are many other applications of GIS path analysis. Analysis of animal movements has a long history using radio or GPS tracking devices on individual animals (Rodgers 2001), and landscape genetics is an exciting new field that measures movement of genes (and individuals that carry them) among populations (Manel *et al.* 2003, Storfer *et al.* 2010). For both approaches, it is necessary to delineate paths among animal locations (*e.g.* linear or stream paths) and to quantify landscape features along those paths. Different buffers are often applied to create two-dimensional sample areas, rather than linear transects. Developing new spatial scales for data collection, such as stream paths or linear paths, is made easier by Geodata Crawler's pre-existing infrastructure of national geodatabases and multi-scale functionality.

New national geodatabases are constantly being added to Geodata Crawler as they become available or become of-interest to new projects. Future climate data represent a significant body of available GIS data that have not yet been incorporated into Geodata Crawler. This would expand possible research topics to include climate change research such as modeling potential effects on hydrology, species distributions, and gene flow among populations. Future climate data (Hijmans *et al.* 2005) have already been acquired, but have not yet been processed for Geodata Crawler. Processing will include identifying preferred climate models or multi-model averages, clipping data to include only the continental United States, and re-projecting data to the Albers NAD 1983 datum. New climate-related variables will then be developed and included in the user-options worksheet and supporting documentation. Geodata Crawler was designed to simplify the process of incorporating new data.

Geodata Crawler was envisioned as a centralized national GIS database and automated data crawler that accepts data requests through a website, and serves data through an FTP server to support any research that could benefit from rapid access to custom multi-scale GIS data for locations in the continental United States. This powerful new tool provides a template for a system to better distribute custom multi-scale GIS data to non-GIS researchers in support of a broad range of sciences. Special emphasis has been given to my own research interests and the interests of my collaborators for applications in modeling species distributions, hydrology, and gene flow, but the basic concept and the tool itself could be applied in many disciplines. Many national GIS servers have been implemented, but Geodata Crawler differs in several important ways: the output data contain site-specific samples from larger GIS datasets, the spatial scales of sample areas can be customized to meet specific research goals, the national geodatabase contains data from many different sources, project-specific GIS databases are created in the process, and the unique data archiving system can drastically reduce processing time by sharing data among user's projects. Widespread-use of such a system could improve comparability of data among projects and disciplines, encourage new research topics by increasing data availability, support multi-scale analyses, and foster interdisciplinary collaboration.

Acknowledgments

Geodata Crawler has been developed with constant encouragement and feedback from Scott Longing and Pablo Bacon. The Center for Advanced Spatial Technology (CAST) at the University of Arkansas provided software licensing, technical expertise, computers, and facilities, particularly Dr. Jason Tullis who gave valuable advice as a member of my dissertation committee, introduced me to Python programming, and provided access to his Windows server for Geodata Crawler development. Gary Huxel, Dan Magoulick, and Steve Stephenson also served on my dissertation committee throughout Geodata Crawler's development and regularly offered constructive criticism and encouragement. James Falcone generously shared his national GIS datasets developed for hydrological modeling. Many collaborators have participated by utilizing Geodata Crawler's data, providing feedback, and recommending new datasets. Collaborators have included Scott Longing, Dan Magoulick, Pablo Bacon, Marlis Douglas, Brad Austin, J. D. Wilson, Jackie Guzy, Bill Wolfe, Wyatt Hoback, and Thea Christianson.

Work Cited

[AOGC] Arkansas Oil and Gas Commission. 2013. *Arkansas Oil and Gas Well Map*. Arkansas Oil and Gas Commission: Little Rock, AR. Available at http://www.aogc.state.ar.us/Maps.htm

Biewick LRH. 2008. *Areas of Historical Oil and Gas Exploration and Production in the United States.* U.S. Geological Survey Digital Data Series DDS-69-Q. U.S. Geological Survey: Reston, VA. Available at http://pubs.usgs.gov/dds/dds-069/dds-069-q/

Birth GS, McVey G. 1968. Measuring the color of growing turf with a reflectance spectroradiometer. *Agronomy Journal* **60**:640-643.

Carlisle DM, Falcone J, Wolock DM, Meador MR, Norris RH. 2010. Predicting the natural flow regime: Models for assessing hydrological alteration in streams. *River Research and Applications* **26**:118-136.

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology* **28**:2031-2064. Data available at http://www.prism.oregonstate.edu/normals/

Di Luzio M, Johnson GL, Daly C, Eischeid JK, Arnold JG. 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *Journal of Applied Meteorology and Climatology* **47**:475-497.

Esri. 2013a. ArcGIS 10.2 for Desktop. Version 10.2.0.3348. Esri Inc.: Redlands, CA.

Esri. 2013b. ArcGIS for Server. Esri Inc.: Redlands, CA. Available at http://www.esri.com/software/arcgis/arcgisserver

Fry J, Xian G, Jin S, Dewitz J, Homer C, Yang L, Barnes C, Heold N, Wickham J. 2011. Completion of the 2006 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* **77**:858-864. Data available at http://www.mrlc.gov/

Gao, BC. 1996. NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of Environment* **58**:257-266.

Hamon WR. 1961. Estimating Potential Evapotranspiration. *Journal of the Hydraulics Division, Proceedings of American Society Civil Engineers* **87**:107–120.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**:1965-1978. Data available at http://www.worldclim.org/current

Homer C, Dewitz J, Fry J, Coan M, Hossain N, Larson C, Herold N, McKerrow A, VanDriel JN, Wickham J. 2007. Completion of the 2001 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* **73**:337-341. Available at http://www.mrlc.gov/

Huete AR, Justice C. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of the Environment* **25**:295-309.

Hunt CB. 1979. *Surficial geology of the conterminous United States*. U.S. Geological Survey, National Atlas of the United States of America. Available at http://www.nationalatlas.gov/

Jensen, JR. 2005. *Introductory digital image processing: A remote sensing perspective*. Pearson Education Inc.: Upper Saddle River, NJ.

Kauth RJ, Thomas GS. 1976. The tasseled cap—A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. *Proceedings, Symposium on Machine Processing of Remotely Sensed Data*, West Lafayette, IN: Laboratory for Applications of Remote Sensing. p. 41-51.

Key CH, Benson NC. 2005. Landscape assessment: remote sensing of severity, the normalized burn ratio and ground measure of severity, the composite burn index. In 'FIREMON: Fire Effects Monitoring and Inventory System'. Eds. Lutes DC, Keane RE, Caratti JF, Key CH, Benson NC, Gani LJ. General Technical Report RMRS-GTR-164-CD. U.S. Department of Agriculture, Rocky Mountains Research Station: Ogden, UT.

King PB, Beikman HM. 1974. *Explanatory Text to Accompany the Geologic Map of the United States.* United States Geological Survey professional paper 901. Available at http://pubs.usgs.gov/dds/dds11/pp901_html/1_PP901.HTML

Krug WR, Gebert WA, Graczyk DJ. 1989. *Preparation of average annual runoff map of the United States, 1951-80.* U.S. Geological Survey Open-File Report 87-535.

Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* **18**:189-197.

Musick HB, Pelletier RE. 1988. Response to soil moisture of spectral indexes derived from bidirectional reflectance in thematic mapper wavebands. *Remote Sensing of Environment* **25**:167-184.

Nakagaki N. 2007a. *Grids of Agricultural Pesticide Use in the Conterminous United States, 1992*. Version 2.0. United States Geological Survey: Reston, VA. Metadata available at http://water.usgs.gov/GIS/metadata/usgswrd/XML/agpest92grd.xml

Nakagaki N. 2007b. *Grids of Agricultural Pesticide Use in the Conterminous United States, 1997*. Version 2.0. United States Geological Survey: Reston, VA. Metadata available at http://water.usgs.gov/GIS/metadata/usgswrd/XML/agpest97grd.xml

[NASA] National Aeronautics and Space Administration. 2011. *Landsat 7 Science Data Users Handbook.* NASA Goddard Space Flight Center, Landsat Project Science Office: Greenbelt, MD. Available at http://landsathandbook.gsfc.nasa.gov/

[NOAA] National Oceanic and Atmospheric Administration. 2012. *Global DMSP-OLS Nighttime Lights Time Series 1992-2012 (Version 4).* NOAA National Geophysical Data Center: Boulder, CO. Available at http://ngdc.noaa.gov/eog/download.html

Price CV, Nakagaki N, Hitt KJ, Clawges RM. 2007. *Enhanced Historical Land-use and Land-cover Data Sets of the U.S. Geological Survey: Raster Format Files.* United States Geological Survey: Reston, VA. Metadata available at http://water.usgs.gov/GIS/metadata/usgswrd/XML/ds240_landuse_raster.xml

Python. 2012. Python v2.7.3. Python Software Foundation. Available at: http://www.python.org

Qi J, Cabot F, Moran MS, Dedieu G. 1995. Biophysical parameter estimations using multidirectional spectral measurements. *Remote Sensing of Environment* **54**:71-83.

Reed JC, Bush CA. 2005. *Generalized Geologic Map of the Conterminous United States.* U.S. Geological Survey: Denver, Colorado. Metadata available at http://nationalatlas.gov/metadata/geology48.xml

Rock BN, Vogelmann JE, Williams DL, Vogelmann AF, Hoshizaki T. 1986. Remote detection of forest damage. *BioScience* **36**:439.

Rodgers AR. 2001. Tracking animals with GPS: the first 10 years. *Proceedings of the Conference on Tracking Animals with GPS.* The Macaulay Land Use Research Institute: Aberdeen, Scotland.

Rouse JW, Haas RH, Schell JA, Deering DW. 1974. Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium* **1**:48-62.

Schreve RL. 1966. Statistical law of stream numbers. *The Journal of Geology* **74**: 17-37.

Schruben PG, Arndt RE, Bawiec WJ, King PB, Beikman HM. 1994. *Geology of the Conterminous United States at 1:2,500,000 Scale – A Digital Representation of the 1974 P. B. King and H. M. Beikman Map.* United States Geological Survey: Reston, VA. Metadata available at http://pubs.usgs.gov/dds/dds11/

Seirup L, Yetman G. 2006. *U.S. Census Grids, 2000.* NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, NY. Available at http://dx.doi.org/10.7927/H4B85623

Smiley JD, Carswell WJ Jr. 2009. *The National Map—Hydrography*. U.S. Geological Survey Fact Sheet 2009-3054, 4 pp. Available at http://nhd.usgs.gov/

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. 2010. Landscape genetics: where are we now? *Molecular Ecology* **19**:3496-3514.

Strahler AN. 1952. Hypsometric (area-altitude) analysis of erosional topology. *Geological Society of America Bulletin* **63**:1117-1142.

[USDA] United States Department of Agriculture. 2006. *Digital General Soil Map of U.S.* United States Department of Agriculture, Natural Resources Conservation Service: Fort Worth, TX. Metadata available at http://catalog.data.gov/harvest/object/69d5a4d1-a4ef-4208-bf4b-898c0b67a425/html

[USDA] United States Department of Agriculture. 2011. *Active Fire Mapping Program.* U.S. Department of Agriculture, Forest Service, Remote Sensing Applications Center: Salt Lake City, Utah. Available at http://activefiremaps.fs.fed.us/

[USDC] United States Department of Commerce. 2011. *TIGER/Line Shapefile, 2011, State, Arkansas, Primary and Secondary Roads State-based Shapefile.* U.S. Department of Commerce, U.S. Census Bureau, Geography Division. Available at http://www2.census.gov/geo/tiger/TIGER2011/PRISECROADS/tl_2011_05_prisecroads.zip

[USEPA] United States Environmental Protection Agency, [USGS] United States Geological Survey. 2012. *National Hydrography Dataset Plus – NHDPlus.* Edition 2.10. Available at http://www.horizon-systems.com/nhdplus/

[USGS] United States Geological Survey. 2003. *Principal Aquifers of the 48 Conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands.* Version 1.0. U.S. Geological Survey: Reston, VA. Metadata available at http://water.usgs.gov/GIS/metadata/usgswrd/XML/aquifers_us.xml

[USGS] United States Geological Survey. 2010. *National Land Cover Version 1.* US Geological Survey, National Biological Information Infrastructure, Gap Analysis Program (GAP). Available at http://www.gap.uidaho.edu/Portal/DataDownload.html

Vogelmann JE, Howard SM, Yang L, Larson CR, Wylie BK, Van Driel JN. 2001. Completion of the 1990's national land cover data set for the conterminous United States. *Photogrammetric Engineering and Remote Sensing* **67**:650-662. Available at http://www.mrlc.gov/

Wolock DM. 1997. *STATSGO Soil Characteristics for the Conterminous United States*. U.S. Geological Survey Open-File Report 97-656.

Wolock DM. 2003a. *Base-flow Index Grid for the Conterminous United States.* U.S. Geological Survey: Reston, VA. U.S. Geological Survey Open-File Report 03-263. Metadata available at http://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml

Wolock DM. 2003b. *Estimated Mean Annual Natural Ground-water Recharge in the Conterminous United States.* U.S. Geological Survey: Reston, VA. Open-file report 03-311. Metadata available at http://water.usgs.gov/GIS/metadata/usgswrd/XML/rech48grd.xml

Wolock DM. 2003c. *Infiltration-excess overland flow estimated by TOPMODEL for the conterminous United States*. U.S. Geological Survey Open-File Report 03-310.

Wolock DM. 2003d. *Saturation overland flow estimated by TOPMODEL for the conterminous United States.* U.S. Geological Survey Open-File Report 03-264.

Wolock DM, Ayers MA, Hay LE, McCabe GJ Jr. 1989. Climate change effects on watershed runoff. *American Society of Civil Engineers, Hydraulic Engineering '89 Proceedings*, p. 673-678.

Wolock DM, McCabe GJ. 1995. Comparison of single and multiple flow-direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Research* **31**: 1315-1324.

Wolock DM, McCabe GJ. 1999. Estimates of runoff using water-balance and atmospheric general circulation models. *Journal of the American Water Resources Association* **35**:1341-1350.

Research with Geodata Crawler

Papers:

Leasure DR, Magoulick DD, Longing SD. 2014. Natural flow regimes of the Ozark and Ouachita Mountain region. *In review* for *River Research and Applications*.

Leasure DR. 2014. Landsat-based monitoring of an endangered beetle: Addressing issues of high mobility, annual life history, and imperfect detection. *In review* for *Landscape Ecology*.

Funded Proposals:

Magoulick DD, Leasure DR. 2014. Quantification of hydrologic alteration and relationships to biota in Arkansas streams: Development of tools and approaches for un-gaged streams. State Wildlife Grant, Arkansas Game & Fish Commission. $53,001.00

Professional Reports:

Willson JD, Guzy JC. 2014. Occupancy and habitat relationships of stream-associated salamanders in intensively managed forests of the Ouachita Mountains Ecoregion. Progress Report for Weyerhaeuser NR Company.

Willson JD, Guzy JC. 2013. Occupancy and habitat relationships of stream-associated salamanders in intensively managed forests of the Ouachita Mountains Ecoregion. Progress Report for Weyerhaeuser NR Company.

Professional Presentations:

Leasure DR. 2014. *Invited Symposium Speaker*—A foundation for Arkansas eFlows: Hydrologic classification and flow alteration modeling. Symposium: Environmental Flows: What, Why, and How? Arkansas Water Resources Center, Annual Conference: Fayetteville, AR.

Leasure DR. 2014. *Invited Symposium Speaker*—"Big data" squared: GIS, hydrological, and remote sensing data coupled with gene flow and species distributions. Symposium: Big Data Science and Its Impacts on Fish Conservation and Management. American Fisheries Society: Quebec City, Quebec, Canada. Abstract available online at https://afs.confex.com/afs/2014/webprogram/Paper16999.html

Leasure DR. 2014. *Invited Seminar Speaker*—Habitat of endangered American burying beetle in western Arkansas: Addressing issues of spatial scale and detection for satellite-based monitoring. Department of Biology, University of Nebraska, Kearney: Kearney, NE.

Leasure DR. 2014. *Contributed Oral Presentation*—Natural flow regimes of the Ozark-Ouachita Interior Highlands region. USGS Cooperators Meeting: Mayflower, AR.

Leasure DR. 2014. *Contributed Oral Presentation*—Assessment of hydrologic alteration at un-gaged streams. USGS Cooperators Meeting: Mayflower, AR.

Leasure DR. 2013. *Invited Symposium Speaker*—The modern era of "big data" in GIS: Multi-scale modeling of species distributions, hydrology, and gene flow. Symposium: Finding

Simplicity In Complexity: Matching Models To Data. American Fisheries Society: Little Rock, AR. Abstract available online at http://afs.confex.com/afs/2013/webprogram/Paper12193.html

Leasure DR. 2013. *Invited Seminar Speaker*—From NASA satellites to DNA microsatellites: 21st century conservation tools. Conservation Biology, Department of Biological Sciences, University of Arkansas - Fort Smith: Fort Smith, AR.

Leasure DR. 2013. *Contributed Oral Presentation*—Geodata Crawler: A centralized national geodatabase and automated multi-scale data crawler to overcome GIS bottlenecks in data analysis workflows. Ecological Society of America: Minneapolis, MN. Abstract available online at http://eco.confex.com/eco/2013/webprogram/Paper43858.html

Leasure DR. 2013. *Contributed Oral Presentation*—Landsat to monitor an endangered American burying beetle population. Arkansas Entomological Society meeting, Southern Arkansas University: Magnolia, AR.

Leasure DR. 2013. *Invited Speaker*—Finding endangered beetles from space: Landsat to monitor American burying beetle habitat. U.S. Fish and Wildlife Service American burying beetle science meeting, Oklahoma State University: Stillwater, OK.

Longing DD. 2013. *Invited Seminar Speaker*—Biogeography and conservation of diving beetles in the southeastern U.S. Department of Biological and Environmental Sciences, Georgia College and State University: Milledgeville, GA.

Longing SD, Bacon PA, Harp GL. 2013. *Contributed Oral Presentation*—Distribution, conservation and current status of three endemic *Heterosternuta* (Coleoptera: Dytiscidae: Hydroporinae) in Arkansas. Arkansas Entomological Society, Southern Arkansas University: Magnolia, AR.

Longing DD, Wolfe GW, Leasure DR. 2013. *Invited Symposium Speaker*—Distributions, ecology and conservation of diving beetles across endemic hotspots in Arkansas and Tennessee, USA. Symposium: When a Blind Beetle Crawls Over the Surface of the Globe, or Under the Water: Biodiversity and Systematics of Aquatic Beetles. Entomological Society of America Annual Meeting: Austin, TX.

Leasure DR, S Longing, PA Bacon, GR Huxel. 2012. *Contributed Oral Presentation*—A new automated tool for multi-scale sampling of spatial environmental data to predict the distribution of the Sulphur Springs diving beetle in northwest Arkansas. Entomological Society of America: Knoxville, TN. Abstract available online at http://esa.confex.com/esa/2012/webprogram/Paper68051.html

Leasure DR. 2012. *Contributed Oral Presentation*—Streamscapes: An automated GIS tool for sampling landscapes associated with streams. Ecomunch Seminar, University of Arkansas: Fayetteville, AR.

Leasure DR. 2011. *Poster Presentation*—Remote sensing and GIS to model endangered American burying beetle abundance across a landscape and to determine the optimal spatial scale for habitat samples. Ecological Society of America: Austin, Texas. Abstract available online at http://eco.confex.com/eco/2011/webprogram/Paper31938.html

Student Reports:

Beacher J, Magoulick DD. 2013. Effect of land use and hydrologic disturbance on crayfish assemblages in the Ozark Highlands. Research Experience for Undergraduates (REU) Program, University of Arkansas: Fayetteville, AR.

Coffman M, Longing SD. 2013. Monahans Sandhills: Land cover change from oil and gas production, 1900 – 2005. Agricultural Compounds Course, Department of Plant and Soil Science, Texas Tech University: Lubbock, TX.

Davis L, Longing SD. 2013. Pesticide use on cultivated crops in five Texas regions. Agricultural Compounds Course, Department of Plant and Soil Science, Texas Tech University: Lubbock, TX.

Gust S, Magoulick DD. 2013. Influence of natural and anthropogenic factors on stream fish assemblages in Arkansas. Research Experience for Undergraduates (REU) Program, University of Arkansas: Fayetteville, AR.

McKelvey D, Longing SD. 2013. Analysis of factors affecting the distribution of the red imported fire any, *Selnopsis invicta*. Agricultural Compounds Course, Department of Plant and Soil Science, Texas Tech University: Lubbock, TX.

Parks S, Longing SD. 2013. A landscape comparison of two Texas olive orchards. Agricultural Compounds Course, Department of Plant and Soil Science, Texas Tech University: Lubbock, TX.

Poole J, Longing SD. 2013. Pesticide use in four Southern High Plains watersheds in Swisher County, Texas. Agricultural Compounds Course, Department of Plant and Soil Science, Texas Tech University: Lubbock, TX.

Tables

Table 1.

These spatial scale designations are found at the beginning of column headings in Geodata Crawler's output. # represents the user option *LOCAL_RADIUS*, and * represents the user option *RIPARIAN_BUFFER_WIDTH* (see section *User Options*).

Beginning of Column Heading	Spatial Scale of Data
P	Point
L#	Local
W	Watershed
L#W	Local-watershed
R*	Riparian
L#R*	Local-riparian

Figures

Figure 1.
A sample area at the *local* spatial scale is centered on a user-defined location, and its size is based on a user-defined sample radius *x*.

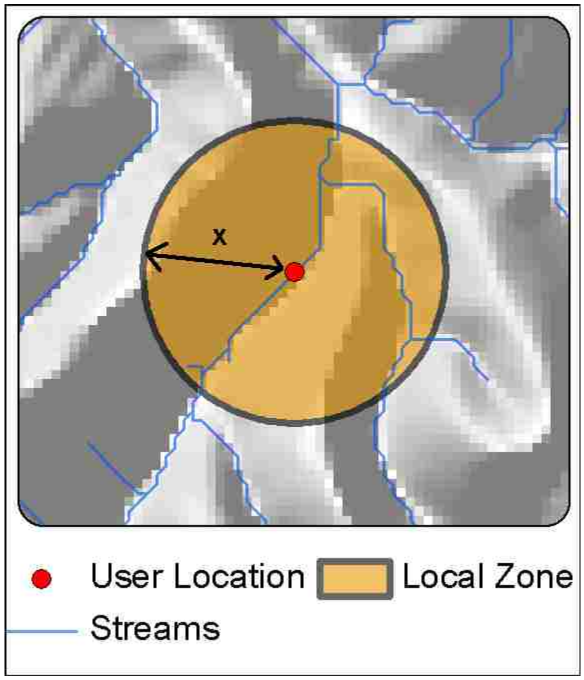


Figure 2.
The watershed spatial scale contains all land areas that drain into a user-defined location.

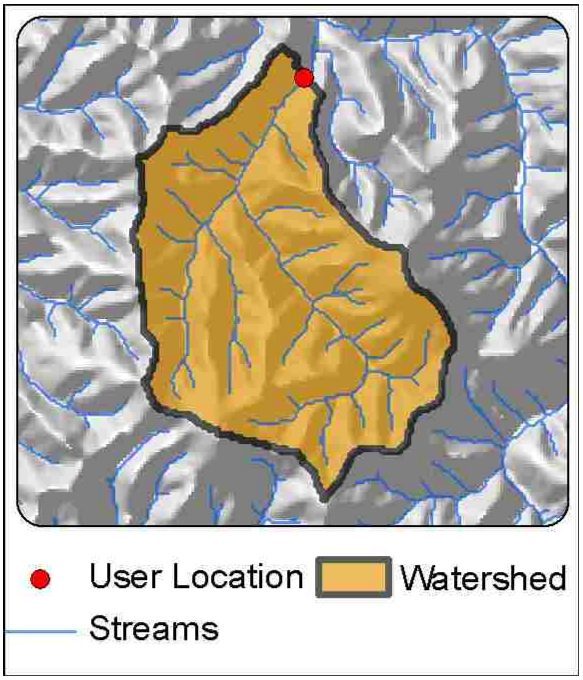


Figure 3.
The local-watershed spatial scale is the intersection of the local and watershed scales.

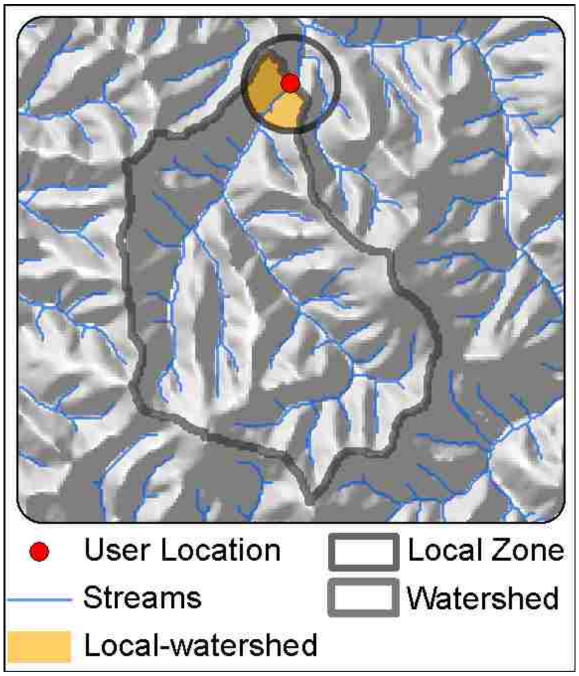


Figure 4.
The local-watershed spatial scale includes all land areas that drain into a user-defined location, but that are also within some user-defined radius *x* of sites.

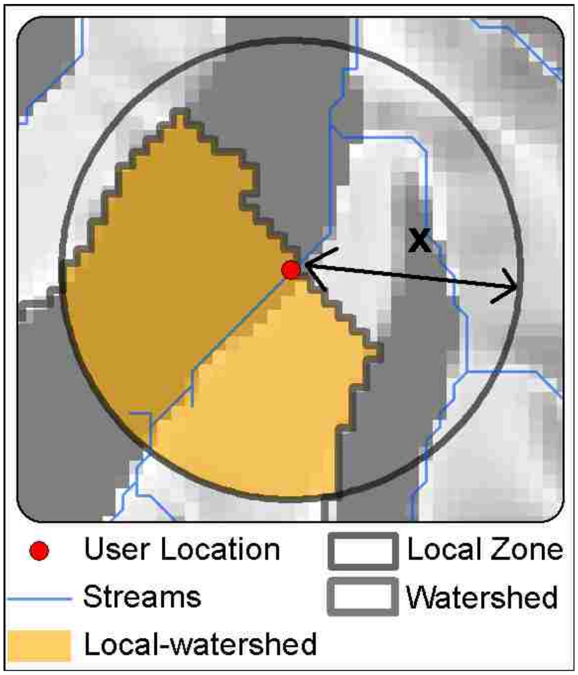


Figure 5.
The riparian spatial scale includes all that areas within a site's watershed, but that are also within some user-defined distance from streams (see Fig. 6).

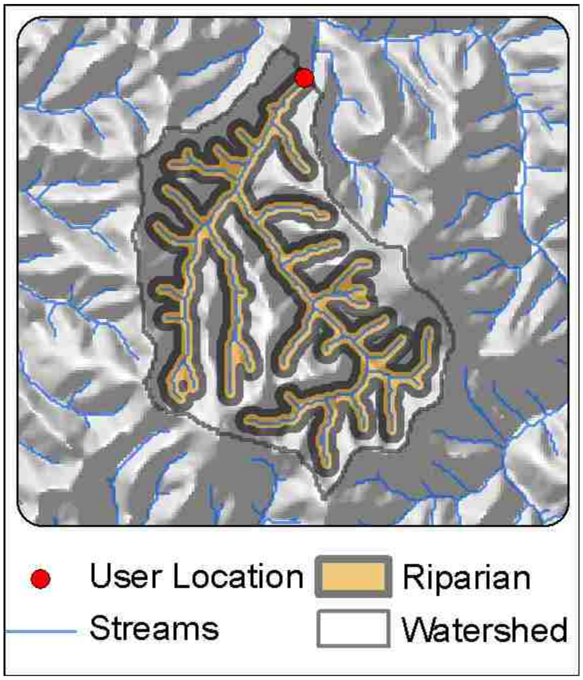


Figure 6.

The riparian spatial scale is delineated based on a user-defined stream buffer distance *x*.

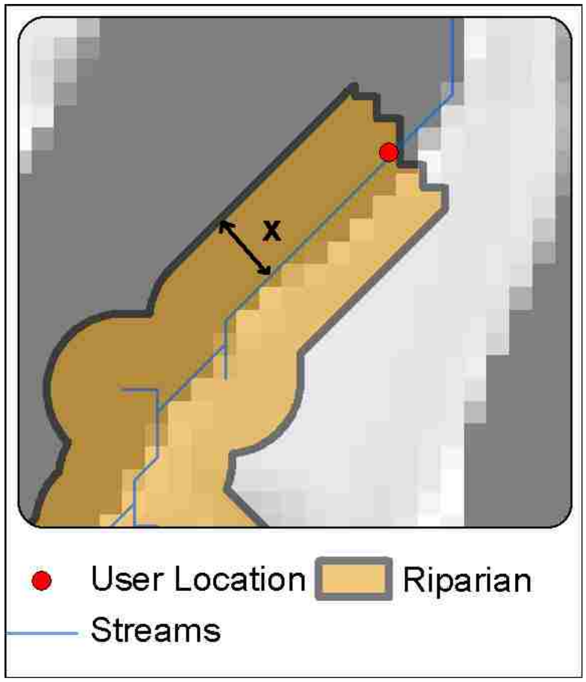


Figure 7.
The local-riparian spatial scale is the intersection of the local and riparian scales.

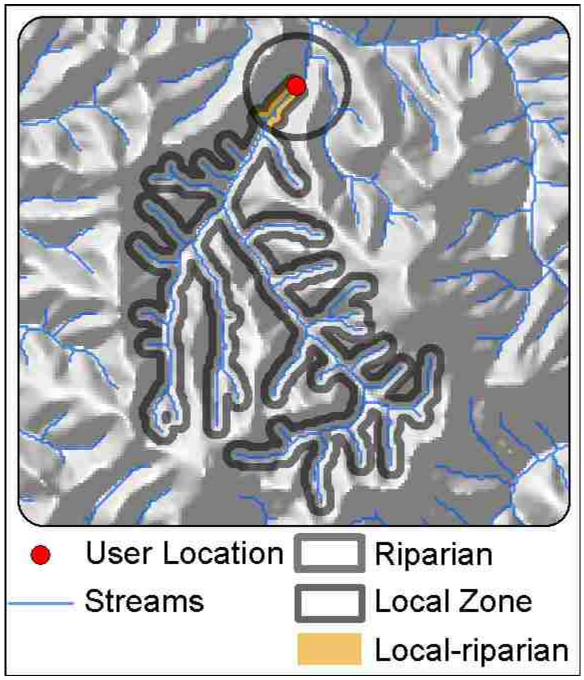
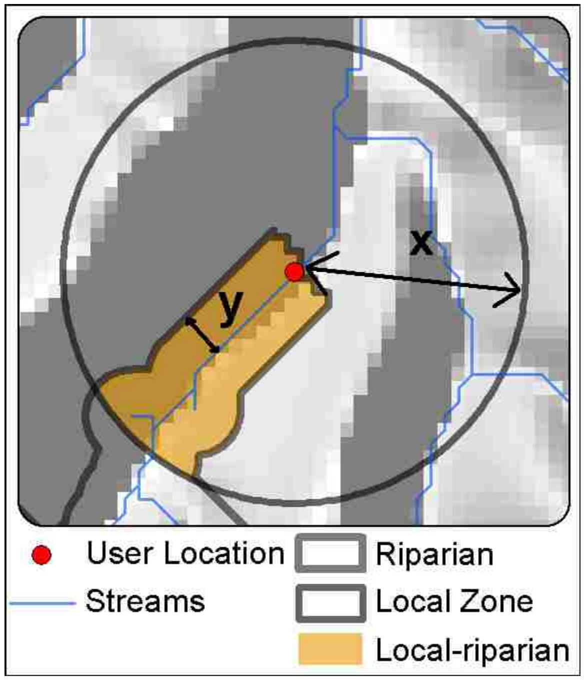


Figure 8.
The local-riparian spatial scale includes all areas within a site's watershed that are within a user-defined radius from sites *x* and also within a user-defined distance from streams *y*.



Appendixes

Appendix 1.
User-options worksheet that allows customization of Geodata Crawler projects and data collection runs. See sections *Variables* and *User Options* for descriptions of all options and variables.

Appendix 2.
Descriptions are provided for all Geodata Crawler variables including a list of spatial scales available for each variable, descriptions of user-input and user-options, and attributes of source data including citations and spatial resolutions. Some variables require additional selection of specific map units or years for data collection. These variables are identified in this appendix and lists of relevant map unit codes are provided. Map unit codes may be entered in the user-options worksheet as a bracketed list (*e.g.* [12, 42, 44]). Spatial resolutions that include an * describe datasets that were converted from polygon features to a raster with the indicated spatial resolution. Citations that include an ** indicate data that were processed by James Falcone and provided for use in this project via personal communication.

Appendix 3.
Descriptions are provided for Geodata Crawler output data found in the output geodatabase of the project folder (*i.e.* /*PROJECT_NAME*/OUTPUT.gdb/*OUTPUT_FILE_NAME*). Attributes of each output column are given including which variable the column is associated with, spatial scale availability, measurement units and data descriptions, the national geodatabase it originated from, and the original data's spatial resolution.

Conclusion

American Burying Beetle

Nicrophorus americanus abundance at Fort Chaffee, Arkansas was associated with native grasslands and open-canopy oak woodlands with rolling topography and sandy loam soils. Results suggested an association with vegetation communities recovering from moderate disturbances, like wildfires the previous year. The optimal spatial scale for measuring *N. americanus* habitat was an 800 m site radius, matching the estimated effective sample range of baited pitfall traps (USFWS 2014). Our field-based measurement of sample radius that used beetles released at known distances from traps suggested that 800 m was the maximum sample range of traps and that beyond a 400 m radius detection rates dropped below 5 percent (Fig. 1). This field work was conducted during a drought year and we suspected that weather conditions may have reduced flight activity of beetles potentially biasing our estimate of sample radius towards smaller estimates. We showed that hot dry conditions were negatively associated with detection of *N. americanus* (Fig. 2), and this effect was likely related to decreased flight activity (Merrick & Smith 2004). We expect that detection probabilities and sample radii are both related to flight activity and they likely respond similarly to weather conditions. Our results suggest that flight activity may be maximized on humid nights with temperatures in the high 20s (^{o}C) and with moderate winds. The sample radius of baited pitfall traps was much greater than the 20 m trap spacing used for multi-trap transects (USFWS 2014; Figs. 1 & 3), and this suggested that multiple traps along a transect should not be treated as individual sample units. Count data from single above-ground bucket traps were not comparable to transects with eight pitfall traps unless transects were treated as single units of sample effort, not eight independent traps (Fig. 4).

These results contributed to *N. americanus* conservation by providing data-driven recommendations of suitable weather conditions for trapping, optimal spatial scale for habitat assessments, a description of *N. americanus* habitat in Arkansas, and recommendations for improving standard data handling procedures. A Landsat-based monitoring tool was developed for Fort Chaffee, Arkansas (Fig. 5), and this concept could be applied for habitat assessments and monitoring of other *N. americanus* populations. A new trap method using above-ground bucket traps was developed and evaluated in comparison to standard pitfall trap transects (Fig. 6). This new trap method was initially developed for use in military training areas where digging was prohibited, but compared to standard pitfall trap transects, the new method was safer for trapped beetles, more resistant to disturbances from vertebrate scavengers, and more time efficient for workers to install. Our above-ground bucket traps have now been adopted by the U.S. Fish and Wildlife Service for *N. americanus* surveys nationally (USFWS 2014).

Several features and datasets were added to Geodata Crawler to support investigation of *N. americanus* spatial ecology. The “local” scale sample area was developed to allow data collection using multiple site radii to assess *N. americanus* habitat associations at multiple scales and to identify an appropriate spatial scale for conservation. A suite of Landsat-derived variables were incorporated into Geodata Crawler to help assess dynamics of vegetation condition in relation to *N. americanus* abundance. These variables included the normalized difference vegetation index (NDVI; Rouse *et al.* 1974), normalized difference water index (NDWI; Gao 1996), tasseled cap index (Kauth & Thomas 1976), and other vegetation indices (see Chapter 4 for a complete list). Other variables like terrain slope, land cover, and soil texture were also initially developed for Geodata Crawler to support *N. americanus* research.

Eco-hydrology

Our hydrologic classification identified seven natural flow regimes of the Ozark-Ouachita Interior Highlands (Fig. 7). This provides a foundation for future work to describe flow alteration—ecological response relationships for each flow regime (Poff *et al.* 2010). Geodata Crawler was used to collect data necessary to predict the natural flow regimes of all stream segments in the study area and an interactive Google Earth map document was produced (Chapter 3, Appendix 2). This provided critical information to support development of a risk-based water management strategy in the region. We conducted the first uncertainty analysis of 170 commonly used flow metrics from the Hydrologic Index Tool (Kennard *et al.* 2010, Henrikson 2006). Measurement uncertainty can occur when flow metrics are calculated from stream gages with short periods of record (*e.g.* less than 15 years). Our uncertainty analysis suggested that a 15 year period of record was adequate to minimize measurement uncertainty for most metrics (Fig. 8), but some metrics were identified with high measurement uncertainty with 15 years of flow data (Fig. 9). These results will assist with metric selection for future projects using the Hydrologic Index Tool (Henrikson 2006). Our results were similar to those of an uncertainty analysis for flow metrics from the River Analysis Package calculated for Australian streams (Kennard *et al.* 2010, Marsh 2012). We also identified sets of flow metrics that best quantified variation among streams for several ecologically-relevant components of each natural flow regime (Olden & Poff 2003; Table 1). These results will assist with metric selection for ecological research in each natural flow regime of the Ozark-Ouachita Interior Highlands.

Geodata Crawler's role in the hydrologic classification project demonstrated its ability to improve data collection efficiency and flexibility in support of new methods in eco-hydrology. Other applications may include assessments of hydrologic alteration (Carlisle *et al.* 2010) and predicting potential effects of climate change on regional hydrology (*e.g.* Liermann *et al.* 2011). We are currently pursuing novel methods to assess hydrologic alteration at un-gaged stream sites by modifying the method of Carlisle *et al.* (2010) and using Geodata Crawler to collect necessary data at all streams segments in the region.

Geodata Crawler

Development of Geodata Crawler associated with this dissertation has demonstrated its applicability for hydrological modeling and species distribution modeling, for both terrestrial and aquatic species. It also provides a template for a GIS data serving system that can efficiently provide project-specific geodatabases and site-specific multi-scale GIS data in tabular form to users with limited GIS experience for applications across a broad range of research disciplines. Future development of Geodata Crawler will pursue a web-based user-interface to allow users to submit jobs to a cluster computing server and to retrieve results from a password-protected FTP server. This will expand Geodata Crawler's processing capacity and increase its accessibility to the public.

Future development of Geodata Crawler will also incorporate new capabilities to support research in climate change and landscape genetics, while continuing to expand capacity for species distribution modeling and hydrological modeling. Geodata Crawler already includes current climate data and incorporating future climate data is now a priority. Landscape genetics is a relatively new discipline that studies gene flow among populations in relation to characteristics of potential paths or dispersal corridors that connect them (Manel *et al.* 2003, Storfer *et al.* 2010). This field is based on relatively new genetic approaches that use allele frequencies at highly variable microsatellite loci to quantify contemporary gene flow among populations. This provides a powerful tool to assess dispersal limitation on today's highly fragmented landscapes. In support of landscape genetics research, preliminary functionality has been developed for Geodata Crawler to tabulate GIS data from linear- and stream-paths connecting all pair-wise combinations of user-provided locations.

Four of the most pressing ecological issues of recent decades were targeted during Geodata Crawler's development: species distribution modeling, hydrology, landscape genetics, and climate change. Species distribution modeling has experienced a major boom in recent years supported by increased availability of GIS data and machine learning methods, and motivated in large part by interest in climate change. Dispersal limitation is an important factor when assessing potential effects of climate change on species distributions, particularly in freshwater systems with significant hydrologic alterations that may prevent dispersal. Eco-hydrology and landscape genetics have also received a flurry of interest and rapid advances that have been supported by increased availability of GIS data, machine learning methods, and new molecular tools. These four research disciplines are fundamentally related to one another, they all benefit from increased availability of GIS data, and they all suffer from similar difficulties acquiring and processing spatial data. Although these fields are often focused on the same GIS datasets, they use very different spatial scales to delineate site-specific sample areas for data collection, and this may result in perceived incompatibilities among fields. Geodata Crawler development will continue with the goals of supporting research in each of these fields and encouraging interdisciplinary research among these interrelated fields that are too often studied independent of one another.

Works Cited

Carlisle DM, Falcone J, Wolock DM, Meador RM, Norris RH. 2010b. Predicting the natural flow regime: Models for assessing hydrological alteration in streams. *River Research and Applications* **26**:118-136.

Gao, BC. 1996. NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of Environment* **58**:257-266.

Henrikson JA, Heasley J, Kennen JG, Nieswand S. 2006. Hydrologic Index Tool version 1.4. U.S. Geological Survey. Available at https://www.fort.usgs.gov/products/22798

Kauth RJ, Thomas GS. 1976. The tasseled cap—A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. *Proceedings, Symposium on Machine Processing of Remotely Sensed Data*, West Lafayette, IN: Laboratory for Applications of Remote Sensing. p. 41-51.Kozol AJ, Scott MP, Traniello JFA. 1988. The American burying beetle, *Nicrophorus americanus*: Studies on the natural history of a declining species. *Psyche* 95:167-176.

Kennard MJ, Mackay SJ, Pusey BJ, Olden JD, Marsh N. 2010a. Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. *River Research Applications* **26**:137-156.

Liermann CAR, Olden JD, Beechie TJ, Kennard MJ, Skidmore PB, Konrad CP, Imaki H. 2011. Hydrogeomorphic classification of Washington State rivers to support emerging environmental flow strategies. *River Research and Applications* **28**:1340-1358.

Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* **18**:189-197.

Marsh N. 2012. River Analysis Package, eWater Toolkit. eWater Innovation Centre, University of Canberra: Canberra, Australia. Available at http://www.toolkit.net.au/Tools

Merrick JM, Smith RJ. 2004. Temperature regulation in burying beetles (*Nicrophorus spp.*: Coleoptera: Silphidae): effects of body size, morphology and environmental temperature. *The Journal of Experimental Biology* **207**:723-733.

Olden JD, Poff NL. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research Applications* **19**:101-121.

Olden JD, Kennard MJ, Pusey BJ. 2011. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Ecohydrology* **5**:503-518.

Poff NL, Richter B, Arthington AH, Bunn SE, Naiman RJ, Kendy E, Acreman M, Apse C, Beldsoe BP, Freeman M, Henriksen J, Jacobson RB, Kennen J, Merritt DM, O'Keefe J, Olden J, Rogers K, Tharme RE, Warner A. 2010. The Ecological Limits of Hydrologic

Alteration (ELOHA): A new framework for developing regional environmental flow standards. *Freshwater Biology* **55**:147-170.

Rouse JW, Haas RH, Schell JA, Deering DW. 1974. Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium* **1**:48-62.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. 2010. Landscape genetics: where are we now? *Molecular Ecology* **19**:3496-3514.

[USFWS] United States Fish and Wildlife Service. 2014. *American Burying Beetle Nicrophorus americanus Oklahoma Presence/Absence Live-trapping Survey Guidance*. U.S. Fish and Wildlife Service, Ecological Services Field Office: Tulsa, OK. Available at http://www.fws.gov/southwest/es/oklahoma/ABB_Add_Info.htm

Tables

Table 1.

Sets of non-redundant flow metrics selected to represent ecologically important components of each natural flow regime. Lists contain a flow metric to represent each of the top three principal components with eigenvalues greater than one. Underlined metrics were selected for cluster analysis. Metrics in bold were considered the best descriptors (*i.e.* highest component loading on the 1st principal component) for nine ecologically important aspects of each flow regime. Metrics in parentheses had equal component loadings.

	All Streams (n=64)	**Stable Groundwater** (n=5)	**Groundwater** (n=6)	**Groundwater Flashy** (n=12)	**Perennial Runoff** (n=13)	**Runoff Flashy** (n=17)	**Intermittent Runoff** (n=7)	**Intermittent Flashy** (n=4)
Magnitude:								
Average Flow	**MA4**, MA41, MA13	**MA19**, MA8, MA2	**MA37**, MA1,MA17	**MA44**, MA26, MA8	**MA29**, MA40, MA13	**MA5**, MA26, MA13	**MA44**, MA18, MA2	**MA34**, MA1, MA43
Low Flow	**ML17**	**ML8**, ML22	**ML7**, ML3	**ML10**, ML13, ML14	**ML19**, ML1, ML13	**ML19**, ML2, ML1	**ML8**, ML12, ML13	**ML10**, ML9, ML4
High Flow	**MH14**, MH13, MH18	**MH27**, MH18, MH23	**MH27**, MH17, MH18	**MH25**, MH20, MH6	**MH27**, MH17, MH18	**MH27**, MH20, MH13	**MH21**, MH4, MH5	**MH27**, MH1, MH14
Frequency:								
Low Flow	**FL3**, FL1	**FL1 (FL2)**, FL3	**FL1 (FL2)**, FL3	**FL2**	**FL1**	**FL3**, FL2	**FL1**	**FL1**
High Flow	**FH7**, FH9, FH11	**FH9**, FH11	**FH3**, FH2, FH10	**FH6**, FH4, FH10	**FH1**, FH3, FH2	**FH1**, FH4, FH2	**FH6**, FH10	**FH1**, FH11, FH9
Duration:								
Low Flow	**DL4,** DL16, DL18*	**DL14**, DL16	**DL11**, DL17	**DL3**, DL5, DL9	**DL3**, DL5, DL16	**DL12**, DL10, DL5	**DL3**, DL10, DL9	**DL2**, DL10, DL16
High Flow	**DH4,** DH8, DH7	**DH12**, DH14, DH19	**DH1**, DH15, DH24	**DH5**, DH1, DH7	**DH8**, DH11, DH23	**DH12**, DH15, DH6	**DH18**, DH2, DH14	**DH2**, DH18, DH21
Timing:								
Average, Low, & High Flow	**TA1**, TL2, TH1	**TA1**, TH1	**TA2**, TH2	**TA2**, TA3, TL1	**TA1**, TH1, TA3	**TA2**, TH2, TL1	**TH1,** TA1, TL2	**TA1**, TH2
Rate of Change:								
Average Flow	**RA3**, RA4	**RA9**, RA2	**RA1**, RA9	**RA3**, RA2	**RA1**, RA4	**RA1**, RA4, RA9	**RA1**, RA9	**RA3**, RA9

* DL18 (no-flow days) was selected based on ecological and hydrological relevance even though it was excluded from the metric selection process due to high measurement uncertainty

Figures

Figure 1.
Predicted recapture probabilities as a function of release distance. Sample ranges appeared to be similar between trap methods, but sample sizes were too small to support robust conclusions. Recapture rates were less than 25% for both methods, even when beetles were released nearby.

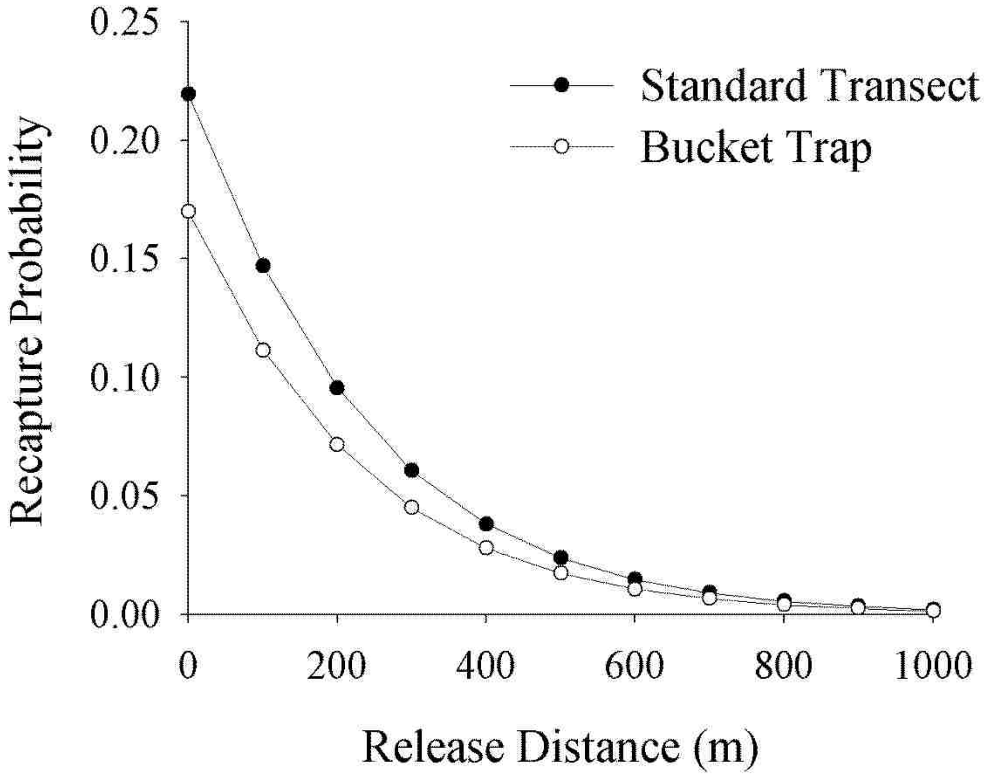


Figure 2.

Modeled relationships between observation covariates and detection probabilities. Thick lines are model predictions when other covariates are held to their means and thin lines are 95% confidence intervals. Scatter plots show model predictions using site specific values of all covariates each year. When site-specific model predictions deviate from the line, it is due to the influence of other covariates in the model, not model error. The gray regions indicate the range of optimal flight temperatures reported by Merrick and Smith (2004) for *Nicrophorus hybridus*.

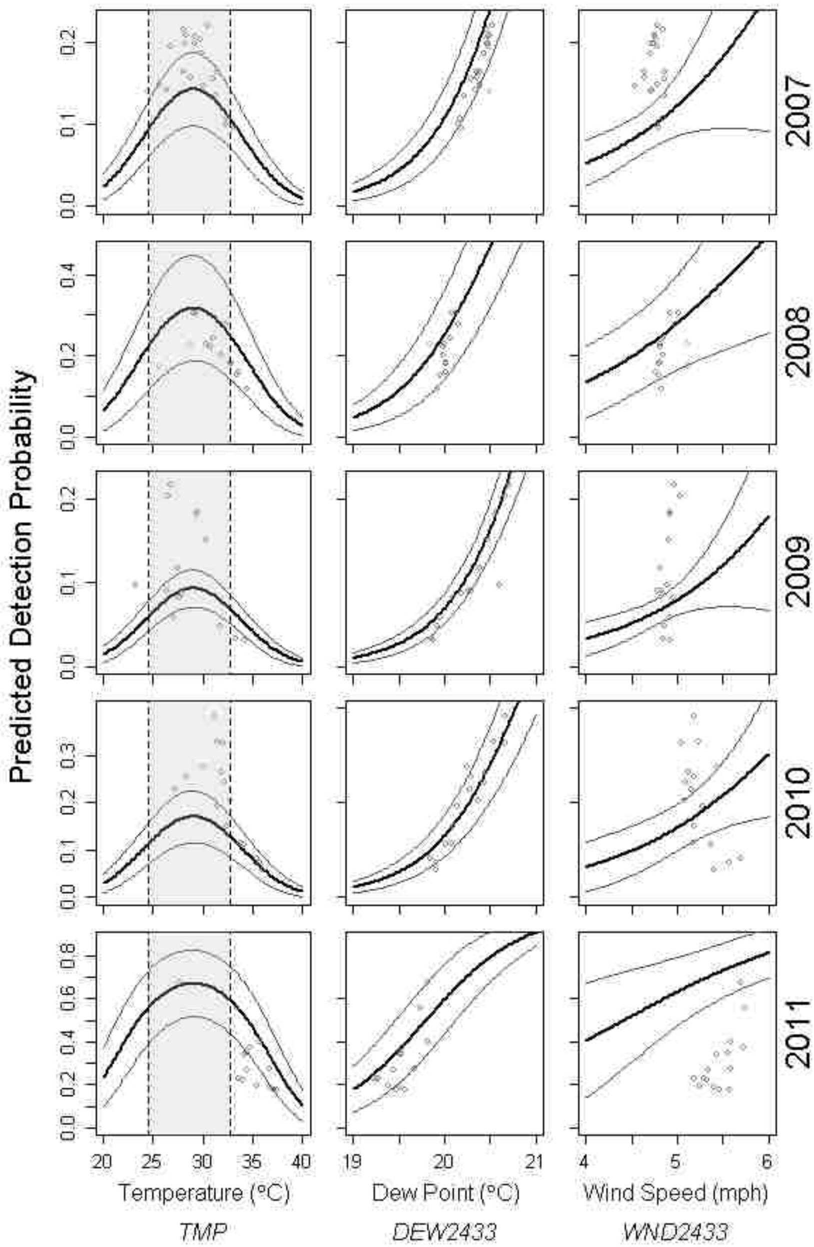


Figure 3.
Layout of a standard pitfall trap transect showing an 800 m trap sample range (USFWS estimate) and a 200 m trap sample range in comparison to 20 m trap spacing to illustrate the lack of independence among traps in a transect.

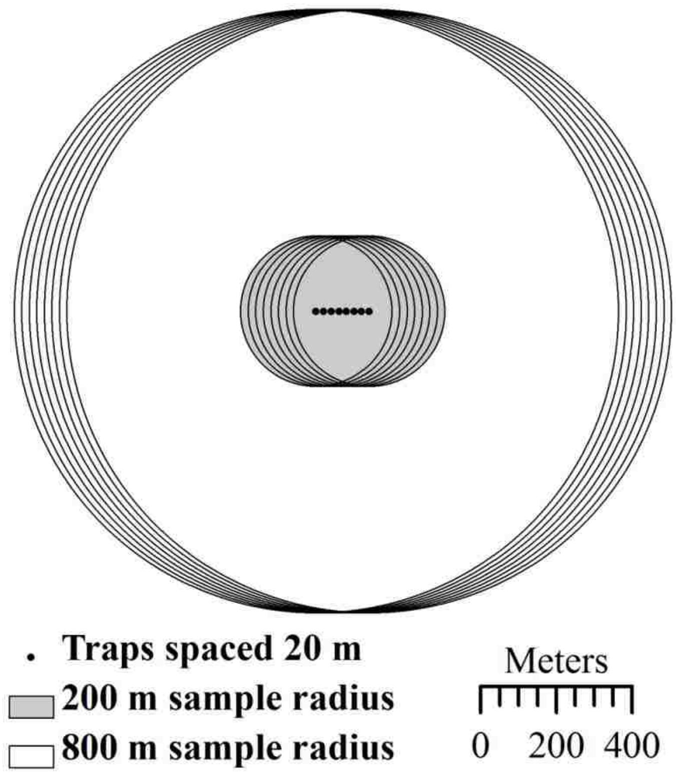


Figure 4.
Average differences in *N. americanus* abundance estimates between methods using various sample effort conversion rates (n = 20). Normalized average differences are mean differences divided by their standard deviations to provide a standardized scale for comparisons because abundance estimates—and therefore differences between them—are inherently smaller when trap-nights are artificially increased.

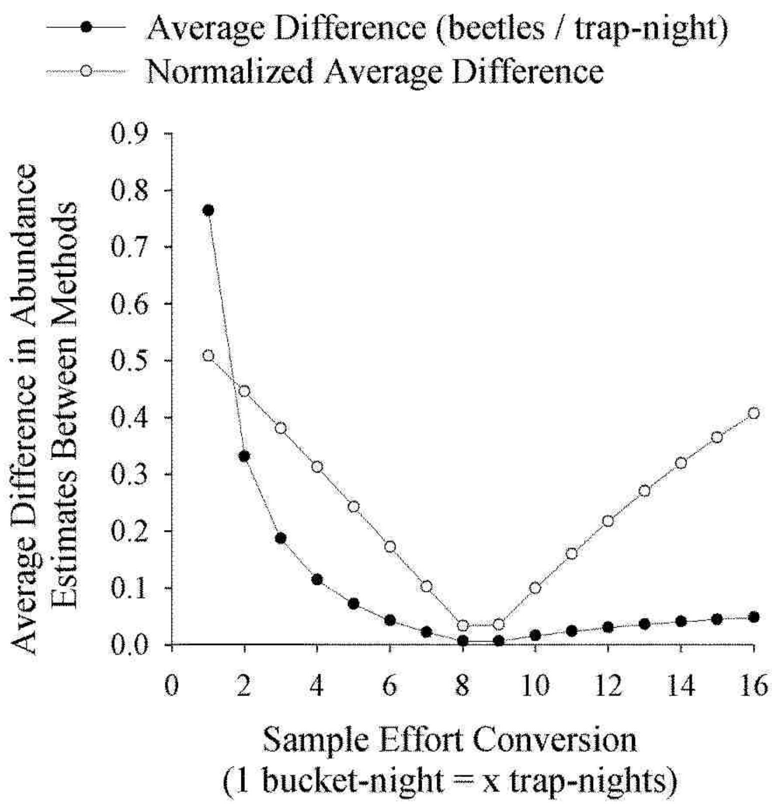


Figure 5.
Spatio-temporal dynamics of three site covariates and abundance model predictions at Fort Chaffee. The habitat model holds the *YEAR* factor constant at “2007” while the abundance model allows *YEAR* to vary. Red corresponds to low values and blue corresponds to high values in all maps.

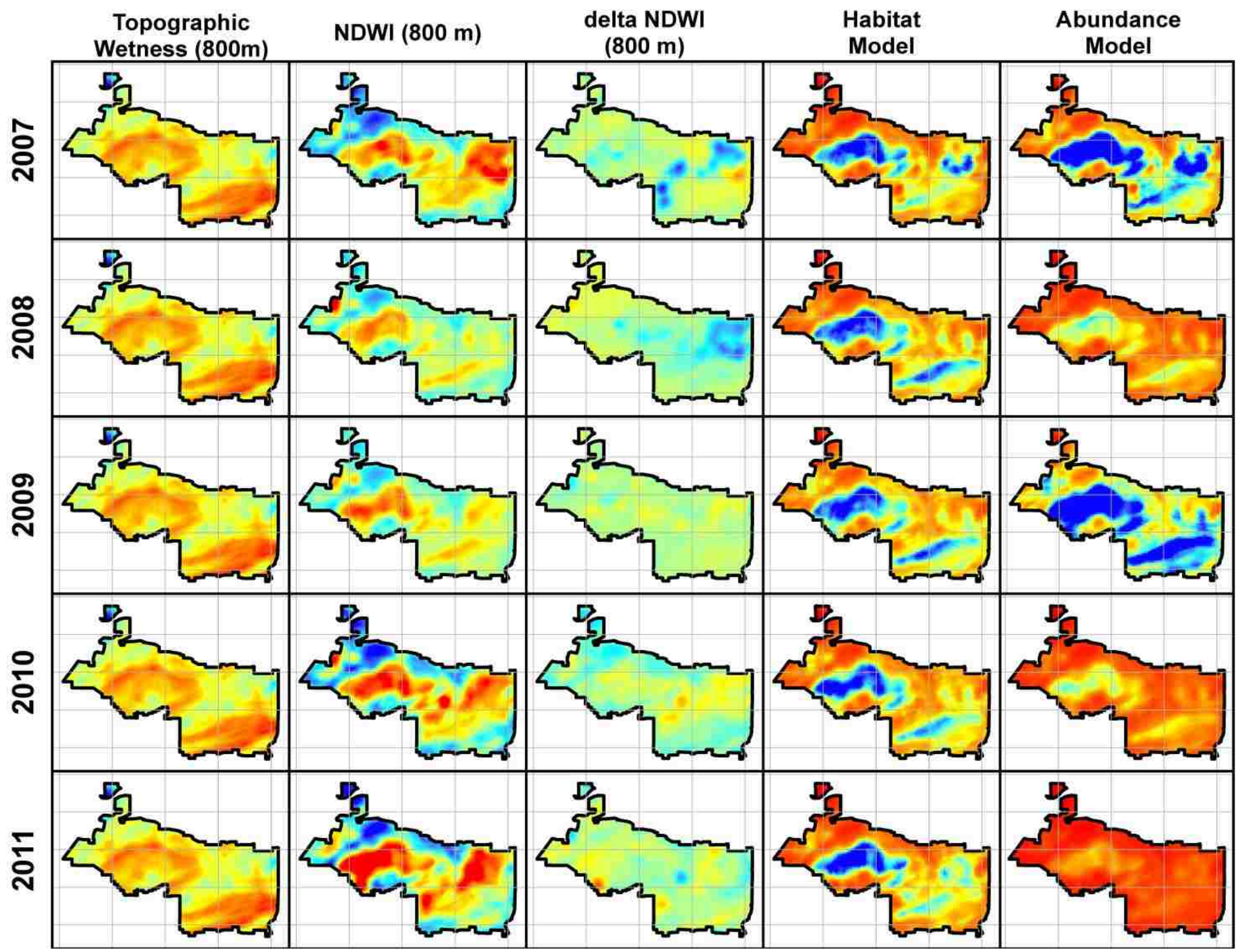


Figure 6.
Suggested improvements to above-ground bucket trap design using a wooden rain cover and landing pad to reduce disturbances from scavengers, reduce maintenance, and increase bait-life.

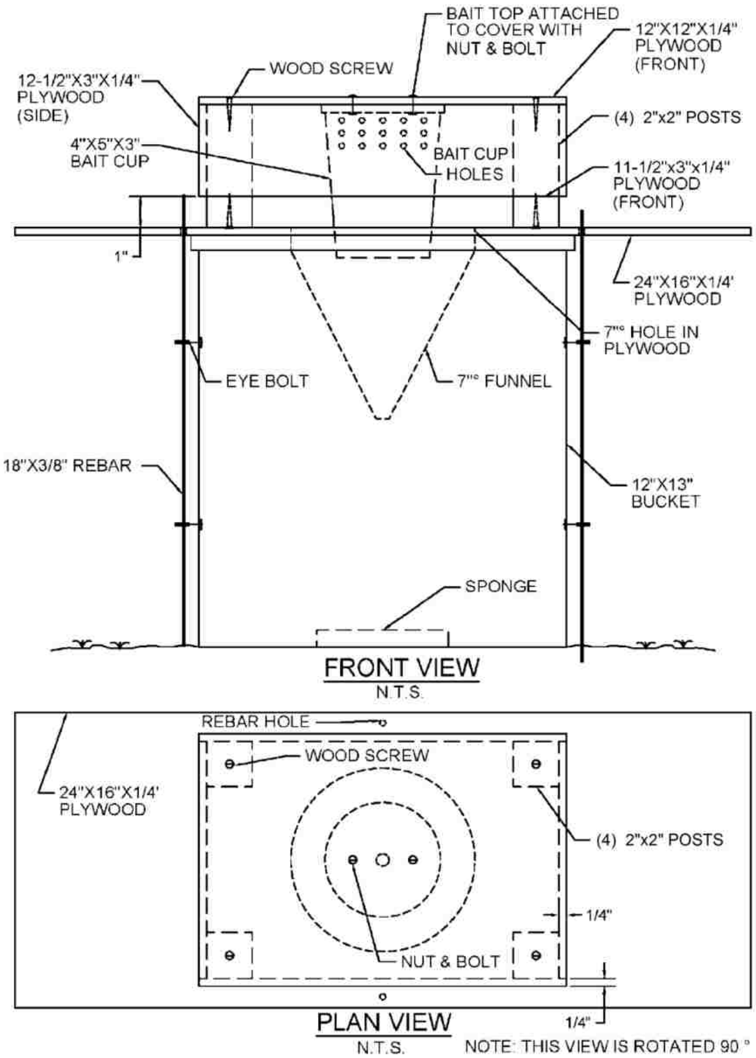


Figure 7.
Natural flow regimes of 64 reference gages were identified using mixture- model cluster analysis based on 10 flow metrics. Natural flow regimes of all stream segments were predicted based on climate and catchment characteristics using a random forest model.

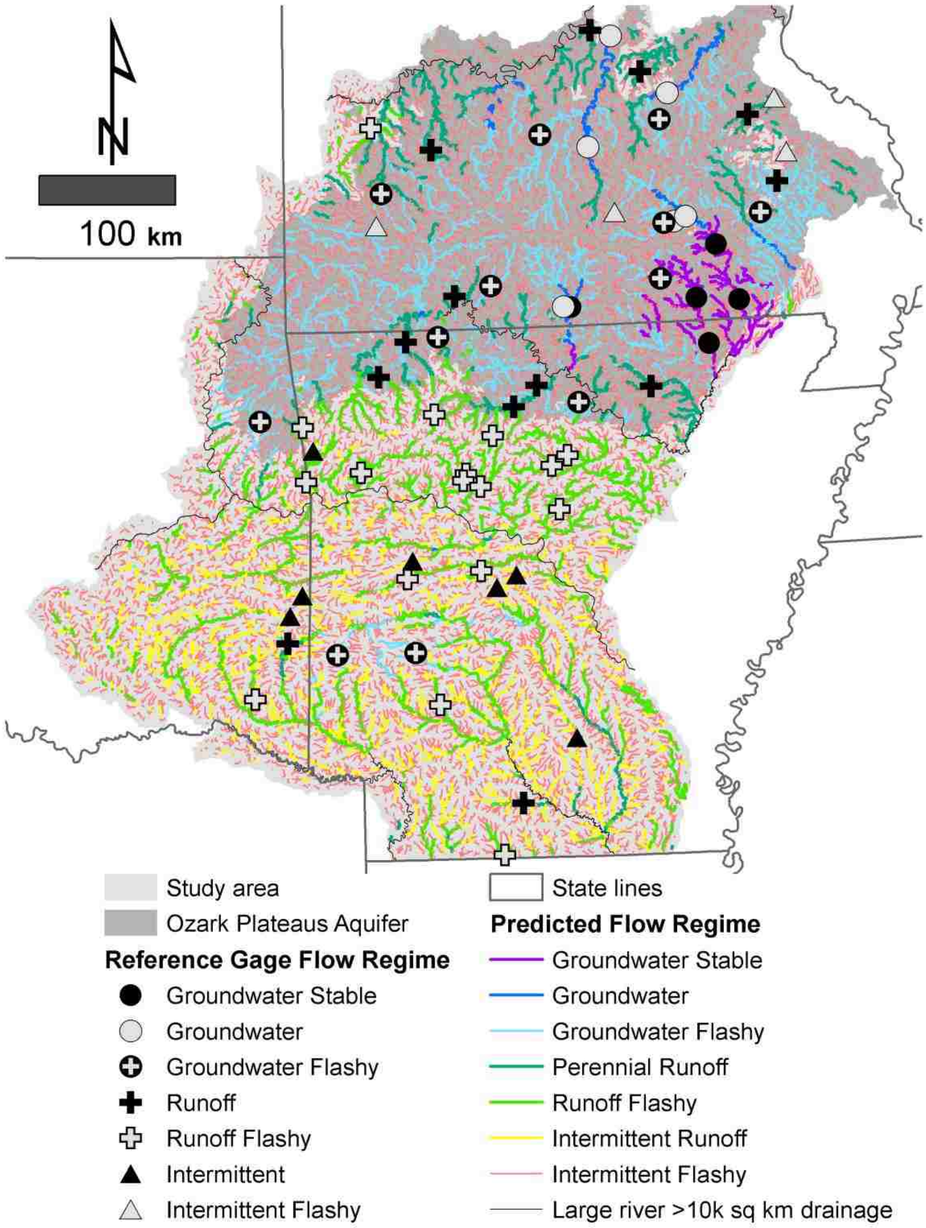


Figure 8.
Distribution of measurement uncertainties (bias, precision, and accuracy) among 170 flow metrics for increasing period of record lengths from 1 to 30 years. Each data point represents average uncertainty for a given flow metric among nine streams analyzed.

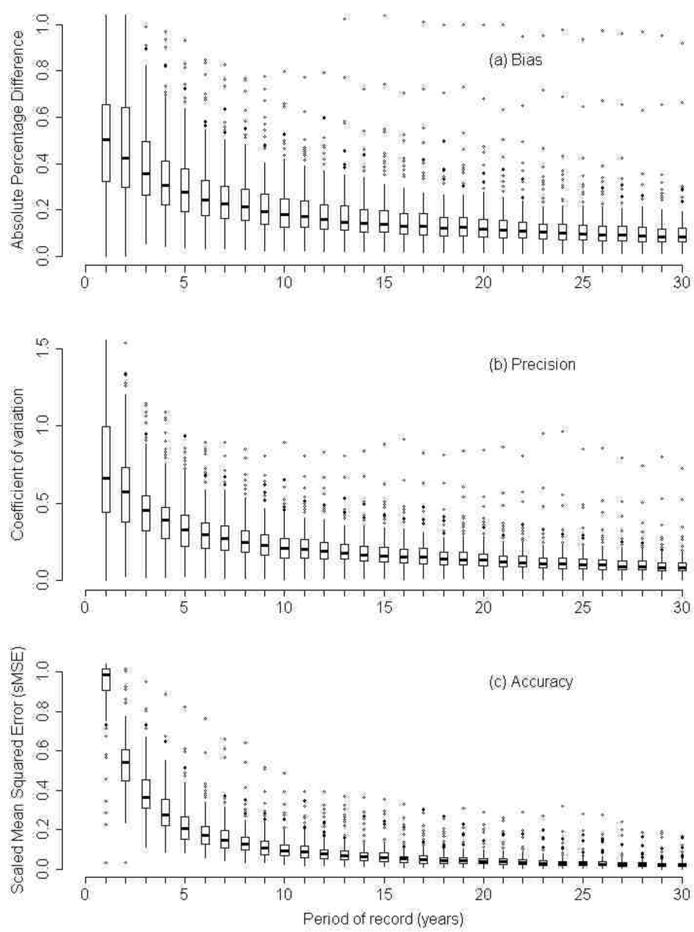


Figure 9.
Measurement uncertainties in terms of bias, precision, and accuracy for 170 flow metrics based on a 15 year period of record. Bars represent average values ± one standard deviation among the nine streams analyzed.