Habitat Distribution Modeling and Estimating Minimum Viable Area for Population Persistence for Three Arachnids of Conservation Interest in Gauteng Province

By

Ian Engelbrecht

Research report submitted in partial fulfillment of the
requirements for the degree of
MSc in Resource Conservation Biology
University of the Witwatersrand, Johannesburg, South Africa

2005

“I hereby declare that this thesis has been composed by myself and has not been accepted in any previous application for a degree. Information drawn from other sources and assistance received have been duly acknowledged.”

Ian Engelbrecht

30th May 2005

Table of Contents

Abstract

Three arachnid species, the rock scorpion *Hadogenes gunningi*, the burrowing scorpion *Opistophthalmus pugnax* and the baboon spider *Harpactira hamiltoni* have been identified as species of conservation interest for inclusion in a bioregional systematic conservation planning project by the provincial conservation authority in Gauteng province, South Africa. The systematic conservation planning procedure requires information on the spatial distribution and an estimate of the minimum viable area (MVA) required to support a population for species of conservation interest. The purpose of this report is to provide this information for these three arachnid species. 47 sites were sampled on a regular grid across Gauteng province where data were collected for habitat distribution modeling and density estimation for MVA calculation. Sites were sampled by two field workers. Distance sampling methodology was used for the estimation of density and the genetic algorithm for rule set production (GARP) was used for habitat distribution modeling. Analysis of distance data comprised fitting several alternative models to both continuous and interval data, and data for each field worker were analyzed both separately and pooled. To calculate MVA from estimates of population density a minimum viable population size of 2000 adult individuals was assumed. Based on composite models fitted to continuous data collected by both field workers mean MVA for *Opistophthalmus pugnax* was 431.57 ha (279.44 ha to 666 ha, 95% confidence interval), while that for *Harpactira hamiltoni* was 909.09 ha (518.00 ha to 1594.90 ha, 95% confidence interval). Insufficient data were collected for the estimation of population density for *Hadogenes gunningi*, but based on encounter rate relative to the other two species a subjective estimate of MVA between 380 ha and 570 ha is presented. Habitat distribution modeling was conducted at two grains of predictor variable data. As GARP produces highly variable results models were selected according to the criteria of having less than 5% omission and less than 10% non-prediction. Selected models were stacked and predictions of presence and absence summed for each map pixel across all models. The resultant maps of proportion of positive predictions per pixel were multiplied to obtain a final composite map of probability of occurrence. Accuracy of the coarse, fine and composite maps was assessed using receiver operating characteristic analysis. Mean AUC for models for *Hadogenes gunningi* were 0.893, 0.857 and 0.886. For *Opistophthalmus pugnax* and *Harpactira hamiltoni* mean AUC values were 0.831, 0.790, 0.856 and 0.783, 0.765, 0.805 respectively. Probabilities of occurrence were converted to presence absence at the threshold where false positive and false negative prediction rates were equivalent. *Hadogenes gunningi* is predicted to occur on most ridges within the province, while *Opistophthalmus pugnax* and *H. hamiltoni* are predicted to have a patchy distribution in the southern two thirds of the province. The results presented are a significant improvement on the data previously available for these species and it is recommended that their conservation status be revised in light of the results. Concerns regarding the utility of GARP in conservation planning and suggestions for further research are outlined.

1. Introduction:

1.1 Context

Gauteng Province, South Africa, is the most highly human impacted province in the country which poses unique challenges to the conservation of biodiversity. Population density for the province was 432 people per square kilometer at the most recent national census in 2001, having shown a 20% population increase in five years (Statistics South Africa 2001). High rates of urban expansion have led to the prediction that the cities of Pretoria, Johannesburg and Vereeniging will constitute a single, polycentric urban metropolis within the province by 2015 (SACN 2004). With approximately 25% of the province under agriculture in addition to this, potential future conflicts between land use transformation and biodiversity conservation have been identified for this region of the country (Wessels *et al.* 2003).

With these challenges in mind the Gauteng Department of Agriculture, Conservation and Environment (GDACE) is currently implementing a conservation planning project with the objective of maximizing the representation of a set of predetermined biodiversity elements, including rare and endangered species, biodiversity surrogates and ecotypes within a network of protected areas within the province. Three arachnid species, namely the scorpions *Hadogenes gunningi* and *Opistophthalmus pugnax*, and the baboon spider *Harpactira hamiltoni* have been included on the list of species to receive attention within this project.

Systematic conservation planning (Margules and Pressey 2000, Sarkar 2002 and references therein, Arthur *et al.* (2002) and references therein) uses the principle of complementarity to select areas that represent complementary sets of species or other entities in order to optimize the area/species representation relationship. While this method does have some weaknesses, such as the tendency to select marginal populations (Araujo and Williams 2001) it is generally accepted as a progressive approach to bioregional conservation (Younge and Fowkes 2003). The process is data driven with a features by planning units matrix as the fundamental data element (Williams *et al.* 2002, Margules *et al.* 2002,). Hence the outcome of the complementary sites identification process is strongly dependent on the approximation of the spatial distribution of the entities of conservation interest (Williams *et al.* 2002), such as species distributions, the distribution of certain ecotypes or vegetation communities, certain topographic features (such as ridges) or surrogates for important ecosystem process such as ecotones (Cowling and Pressey 2003).

Another important consideration for systematic conservation planning is that of population persistence (Williams and Araujo 2002, Araujo and Williams 2000, Cabeza and Moilanen 2001). A measure of the probability of population persistence at sites is seldom incorporated into the planning process when only presence and absence data are used. To illustrate the importance of this measure, Allen *et al.* (2001) showed that due to the highly fragmented nature of mammal habitat in Florida, only nine percent of total available suitable habitat consists of patches large enough to sustain a population robust to demographic stochasticity alone. The concept of minimum viable population allows one means for including population persistence in systematic conservation planning, as is being done by GDACE.

The aim of this report is to provide distribution information and a surrogate for population persistence, minimum viable area, for use by conservation practitioners in including these three arachnid species in the GDACE conservation planning process. The principle objectives are to provide potential distribution maps derived from habitat suitability modeling as well as an estimate of the minimum area required to support a population based on density estimates. Presented is an overview of biology of the three species of interest and a description of how they were selected for inclusion in the conservation plan by GDACE, followed by a review of the relevant minimum viable population and habitat distribution modeling literature.

1.2 Arachnid Conservation in Gauteng

South Africa has been praised for its progressive approach to the conservation of biological resources (Balmford 2003). While conservation attention is usually focused on larger, more charismatic and better known species, invertebrates have received special attention in this country (McGeoch 2002). However, this group of organisms is characterized by the lack of information upon which sound conservation assessments and conservation planning can be based (Redak 2000, Hammer and Slotow 2002, Armstrong 2002).

IUCN (2001) provides a formal system of assessing species' risk of extinction based on a set of standardized criteria. These criteria have been shown to be fairly robust and applicable to most taxa and take into consideration factors of population decline, habitat destruction, total distribution area and population viability. The criteria are intended for the assessment of the conservation status of the global population of a species, defined as all individuals of that species occurring worldwide, and species are classified as Critically Endangered, Endangered or Threatened.

An informal derivation of the IUCN red data categorization process is used by GDACE to rank species of conservation concern, with categories 1 - 3 being approximately correlated with the IUCN categories of Critically Endangered, Endangered and Vulnerable. Species identified as being at high risk are then included in the systematic conservation planning process as an entity requiring conservation attention. This may take the form of species specific conservation management plans or the addition of areas of suitable habitat for the species as high priority areas to the network developed by the systematic planning process described above. An important consideration is that rankings by

GDACE are not intended to apply to the global population of a species as is the case with IUCN (2001), but to those within the province only and hence assessment criteria are adjusted accordingly.

Due to the lack of hard data on which conservation assessments could be based by GDACE, preliminary conservation assessments for invertebrates were conducted through an expert consultation process where potential candidate species for conservation action were identified subjectively. Additionally, locality records from museum collections were collated in order to determine relative rarity of species under the assumption that the spatial distribution and number of records would adequately reflect the extent of occurrence or commonness of each species. Through this process the three arachnid species of interest were identified, and further details of their biology and the results of the assessment follow.

1.2.1 *Hadogenes gunningi* (Scorpiones, Liochelidae)

This medium-sized rock scorpion can be recognized from its black body colour, laterally compressed cauda I, well developed spines on the ventrolateral keels of cauda V and the relatively short tail in adult males. *Hadogenes* species are dorso-ventrally compressed which along with specialized tarsal claws is an adaptation to living in rock cracks (Lamoral 1979, Newlands and Prendini 1997, Prendini 2001a) They are fairly sedentary, with adult females occupying a single rock crack for a number of years. This can often be confirmed by the presence of a number of exuviae found in the rock crack with the scorpion. These sit and wait foragers will prey on anything they can overpower, including insects, arachnids and even small vertebrates such as geckos. Most scorpion species reproduce annually. Female *Hadogenes* give birth to between 15 and 30 offspring between January and April. These offspring remain with the mother for approximately two weeks before undergoing their first ecdysis and then dispersing (pers obs.). It is estimated that these scorpions may take eight years to reach adulthood (Newlands and Cantrell 1985), although this is probably an overestimate.

According to Prendini (1995) this species is limited to Gauteng Province and the immediately adjacent areas in North West Province. *Hadogenes* are usually allopatrically distributed, barring the *H. tityrus* group of species which co-occurs with larger *Hadogenes* in Namibia and the Northern Cape (pers obs, Prendini pers comm.). Prendini (1995) states that *H. gunningi* is limited to the Magaliesberg and associated ridges and is separated from *H. gracilis* which is distributed on the Pyramid Koppies ridge just north of the Magaliesberg. It has been suggested that this separation is maintained by a valley approximately 10km wide which acts as a dispersal barrier to these highly specialized species (Prendini 1995). However, the distribution of this species south of the Magaliesberg is not discussed, and the potential exists for it to occur on all ridges within Gauteng. Due their highly specialized microhabitat requirements geology is likely an important limiting variable on the spatial distribution of this species, where only rock types that weather to provide suitable rock cracks would be occupied.

1.2.2 *Opistophthalmus pugnax* (Scorpiones, Scorpionidae)

This species of burrowing scorpion is very stocky with large, powerful pincers. They usually burrow under rocks but may be found on open ground. They are usually associated with hard, clay soils, and are very common in the Magaliesberg area. They also reproduce once per year, with between 10 and 25 offspring being born. A study on a smaller *Opistophthalmus* species indicated that it takes four years to reach adulthood (Engelbrecht 2002). Due to its large size this species may take a little longer, with the total lifespan of an adult female potentially being in the region of 10 years. Adult males probably only live in the region of two years following their final ecdysis. This is due to increased mortality associated with searching for females on the surface during the summer months. This high mortality rate in males leads to a skewed sex ratio in most scorpion species, (Polis and Sissom 1990)

Most *Opistophthalmus* species are limited to occurring on soil substrates with specific hardness and texture (Lamoral 1979, Prendini 2001b, Prendini *et al.* 2003). Thus soil type or parent material (geology) might be a useful proxy for the limiting environmental gradients for distribution in this species.

1.2.3 *Harpactira hamiltoni* (Araneae, Theraphosidae)

H. hamiltoni is a medium-sized baboon spider (total length of females up to 55mm). They burrow to a depth of approximately 20-30cm. The only available life history, behavioral and population data for this species were collected by Martin Paulsen over a five year period from 1996 to 2001. These indicate that individuals spend most of their time in their burrows, adopting a sit and wait hunting strategy. They feed mostly on ground dwelling insects and other arthropods. Females reproduce annually laying between 20 and 40 eggs. The young remain within the maternal burrow for several months to a year, after which they disperse to find or dig their own burrows. This pattern of dispersal is found in some other Mygalomorph spiders (Cutler and Guarisco 1995). It is likely that the highest mortality rate occurs in the dispersal stage. Males take one to two years to reach adulthood while females take two to three years. The only longevity data for the species are based on records in captivity (M. Paulsen unpublished data). Adult males only live for approximately three months once they mature. In the wild they spend this time wandering in search of females and are subject to mortality through predation and exposure to adverse weather conditions. Females may live for a total of ten years. Observed predators include ants and the Fiscal Shrike. Potential predators are wasps, lizards, toads, shrews, mongooses, owls and other birds.

This species has been found to occur in a variety of habitats, from open plains to rocky ridges (Smith 1990, M. Paulsen pers comm., pers obs.). Due its burrowing habits it is possible that substrate might be an important limiting variable on the spatial distribution of this species.

1.3 Conservation Status

The preliminary risk assessments for these three species based on expert consultation ranked them as category 3 (equivalent to Vulnerable) in Gauteng. Conservation targets, based on the number or records and the number of other provinces in which the species are recorded to occur are 50%, 50% and 30% of all recorded localities to be protected for the three species respectively. MVA estimates were 60ha, 50ha and 50 ha respectively, based on the assumption of an MVP of 100 individuals (M. Forsyth pers comm.) Towards the end of this project *Opistophthalmus pugnax* was removed from the list of species requiring special conservation action due to the high number of localities at which it has been recorded (M. Forsyth pers. comm). At project conception, the distribution limits of these species within Gauteng, and within South Africa were not known.

1.4 Population Viability

Shaffer (1981) defines minimum viable population size (MVP) as the smallest population size required for a species to have a predetermined probability of persistence (or extinction) over a certain time horizon. Small populations have an inherent risk of extinction due to factors attributable to their size (Caughley 1994). These factors include demographic, environmental and genetic stochasticity and act together to decrease population fitness and hence the probability of persistence over long time periods.

MVP for a particular species or population is most usually estimated using population viability analysis (PVA). This modeling process aims to determine the viability or risk of extinction of a particular population over a certain time frame by modeling stochastic population processes and extrinsic threats (Gilpin and Soule 1986). Importantly, population viability should be defined in probabilistic terms in order to adequately represent uncertainty in the predictive nature of modeling required to estimate it (Beissinger and McCullough 2002 and various authors therein). Population viability analysis has received criticism, particularly for being data hungry and for the number of simplifying assumptions that need to be made in order for modeling to be tractable (eg Beissinger and Westphal 1998), which may have implications when applying model outputs in decision making processes. However this approach has a wide field of application despite these shortcomings, such as ranking management options according to impacts on a species and the development of conservation plans for endangered species (e.g.Brooke and Kikkawa 1998, Bustamante 1996, Carroll *et al.* 2003, Forys and Humphrey 1999, Kelly and Durant 2000, Marmontel *et al.* 1997 Lindenmayer and Possingham 1996).

PVA approaches offer an effective way for determining minimum viable population size. By accepting a certain risk of extinction over a specified time period, the initial population size in a model can be varied until these criteria are met. However, PVA approaches require population data for the species of interest or at the very least for species that are closely related. Very often, and in particular for endangered species, these data are not available and estimation of population parameters with any degree of certainty is difficult. Thus the importance of generating support for the use of general estimates of MVP for organisms with certain life history strategies or for particular taxa (Shaffer *et al.* 2002) has been stressed as an important field in conservation biology research.

Reed *et al.* (2003) investigated estimates of MVP for vertebrates using PVA and published data for 102 vertebrate species. They found a mean and median MVP of 7316 and 5816 individuals required for long term persistence. The authors include some support for these very high estimates of MVP, such as the extinction of 13 populations of lagomorphs within median population size of 3276 individuals. Other evidence based on genetic considerations suggests MVP of between 2000 (Schultz and Lynch 1997, Whitlock 2000) and 5500 (Thomas 1990). Reed and Hobbs (2004), when examining the relationship between population size and variability in population size, state that a population size greater than 2000 is required for long term persistence. This simple correlative exercise, using data from 2387 populations of 203 species, found that variation in population size relative to total population size is greater for small populations than for large ones, and that this high level of variability makes small populations more susceptible to extinction.

These estimates of MVP are based on data for vertebrates only. However, Reed *et al.* (2003) found that MVP size was generally greater for species with more variable populations, i.e. those with a higher per generation growth rate. Scorpions have been shown to have a *K*-selected life history strategy relative to other invertebrate species (Polis 1990), which would suggest they show low variability in population size. This is supported by findings of Polis and Farley (1980) where *Smerengurus mesaensis* showed very little variation in population size over a five year study period. Similarly, given its longevity and reproductive parameters it is likely that *Harpactira hamiltoni* is also more *K*-selected relative to most other arthropods. Thus, under the assumption that the three focal species of this report generally exhibit stable population dynamics MVP sizes could be inferred to be closer to the lower end of the range presented by Reed *et al.* (2003).

It is impossible to determine a ‘magic number’ for MVP as each population is exposed to a unique set of environmental and threat conditions (Reed *et al.* 2003). The values indicated above are only very general estimates that would indicate that MVP sizes are in the regions of several thousands of individuals. The importance of defining ‘population’ is also stressed. The above estimates of MVP define population as the number of potentially reproductive adults within the total population. Minimum viable area is defined as the area large enough to support a minimum viable population.

1.5 Habitat Suitability and Habitat Distribution Modeling

1.5.1 Overview of species habitat distribution modeling

Developing a comprehensive understanding of organism environment relationships is the central tenet of the science of ecology (Guisan and Zimmermann 2000). Fundamental to this is the concept of a species habitat, vaguely defined as the environmental conditions under which a species occurs (Morrison and Hall 2002). The concept of habitat has proved particularly challenging for ecologists to clarify due to the dynamic relationship between organisms and their environment, governed predominantly by the proximate processes of resource acquisition, population dynamics and species ecophysiology, and by the ultimate process of evolution.

One approach to better understanding this relationship is that of habitat distribution modeling, where the overarching axiom is that predictable relationships exist between the occurrence of an organism and certain aspects of its environment (Heglund 2002). This relationship is defined by the Grinnellian niche concept, quantified by Hutchinson (1957) as the subregion within the total n-dimensional hypervolume of environmental space where individuals of a species can survive and reproduce. This niche space is best described by the response of a species along a set of environmental gradients (Austin 2002a) Here an important distinction between the fundamental and realized niche of a species is required. The fundamental niche refers to the set of environmental conditions under which a species is physiologically capable of surviving and reproducing. Through competition, predation or other interactions with other species this fundamental niche space may be restricted. This smaller portion of the fundamental niche in which the species will be observed to survive and reproduce is referred to as the realized niche. Pearson and Dawson (2003), when discussing the applicability of climate envelope modeling methods to predicting impacts of climate change on species distributions, point out that the data upon which such correlative models are based are extracted from the species realized niche. Thus, while a species fundamental niche will remain constant, differing responses by different species to changing climate may alter biotic relationships and hence the realized niche may change through time. However, such biotic relationships may also vary spatially across a species distribution (Soberon and Peterson 2005), and the representation of the fundamental niche by the distribution data will depend on how the abiotic and biotic factors limiting distribution overlap within the species geographic distribution. Whereas Pearson and Dawson (2003) state that mechanistic models based on a species ecophysiology are required to model the fundamental niche, Soberon and Peterson (2005) state that correlative modeling methods may model the fundamental or realized niche, depending of the method used and the extent of this overlap.

1.5.2 Model applications and methods

Models predicting the habitat affinities and potential distributions of organisms have a wide range of applications from predicting the potential distributions of species of conservation interest and those of invasive species, as well as applications in biogeography and other disciplines and thus the modeling process has received much attention (see Guisan and Zimmermann 2000, Scott *et al.* 2002 and references therein for a thorough review of the topic). Issues relating to the modeling procedure, model evaluation, data quality and quantity and the scale of investigation have received much attention in the literature in order to obtain a better level of understanding of model reliability which is essential in both theoretical and practical applications (Guisan and Zimmermann 2000, Vaughan and Ormerod 2003)

A large number of habitat modeling methods are available for investigating relationships between organisms and their environment and for making predictions of where species may occur. These methods may be grouped into two broad categories, models based on mechanistic relationships between the organism and its environment and those based on correlative relationships. Vaughan and Ormerod (2003) have stressed the need to carefully state the objectives of the modeling exercise in order to clarify what model types might be most suitable, and Guisan and Zimmermann (2000) provide a framework for choosing methods based on such objectives. While mechanistic methods might be appealing for their ability to capture the dynamic nature of organism environment relationships and have been suggested as the superior methodology (O'Connor 2002), the data for such models is often not available or difficult to collect (Heglund 2002, Vaughan and Ormerod 2003). However, correlative modeling methods provide a useful means for developing postulates relating to such mechanistic relationships through the careful choice of predictor variables with low levels of autocorrelation or which might best approximate those stated to be important for a species in the literature or based on expert opinion (Vaughan and Ormerod 2003, Pearce *et al.* 2001)

Austin (2002a) emphasizes that any particular modeling exercise should be based on a sound conceptual ecological model of the system of interest and as such a number of the available statistical methods model the species response to environmental gradients. Gradients may be divided into those that represent resource, direct and indirect environmental variables (Austin 2002a). Resource gradients are those representing an entity that is consumed by the organism. Direct gradients have a direct measurable effect on the organism, for example the effect of temperature on growth rate, but are not consumed. Indirect gradients have no direct effect on the organism but may be highly correlated with environmental variables that do. For example, elevation is an indirect gradient which influences temperature. In developing the conceptual ecological model upon which statistical modeling will be based it is crucial to consider the form of response to these gradients shown by the organism. Evidence shows that these responses are very seldom linear (Austin 2002a), as is assumed by some statistical methods and may take a number of forms, some of which are fairly complex (Austin 2002, Guisan *et al.* 2002). Assumptions about the form of response are important in selecting a statistical modeling method (Guisan and Zimmermann 2000).

Guisan and Zimmermann (2000) give a comprehensive overview of available modeling methods. These range from simple GIS based implementations to complex machine learning methods. Habitat suitability index (HSI) methods use habitat information on a particular species provided by an expert or obtained from the literature to develop a habitat suitability map of the area. The index is determined by the combination of the number of environmental variables that qualify as suitable in each pixel over the area, as determined by the expert information. While this is a subjective process, it is easily understood by conservation practitioners and it is amenable to situations where data requirements for statistical methods are not met. Alternatively, statistical and machine learning methods investigate relationships between a set of environmental variables, the predictors, and some form of response variable, such as presence or absence or population density. These include regression methods such as generalized linear and additive models (see Guisan *et al.* 2002 for an overview) and nonparametric multiplicative regressions (McCune 2004), classification techniques such as classification and regression trees (CART 1984), environmental envelopes such as BIOCLIM (Busby 1991) and Bayesian modeling methods. More complex analyses include artificial neural networks (Manel *et al.* 1999), genetic algorithms (Stockwell and Noble 1992, Stockwell and Peters 1999) and machine learning methods (Phillips *et al.* 2004).

Generalized linear and generalized additive models (GLMs and GAMs) are examples of regression methods usually employed. The latter have been suggested as ideal to the purpose of modeling species responses to environmental gradients in that they do not assume an underlying distribution to the data and are said to be entirely data driven (Guisan *et al.* 2002). Artificial neural networks and genetic algorithms have the same advantage. A number of software packages have been developed to implement these methods, such as BIOMOD (Thuiller 2003) which implements GAMs, CART and neural networks, and GRASP (Lehmann *et al.* 2003), an implementation of GAMs specifically for this purpose.

Van Horne (2002) outlines some problems with the correlative modeling approach, such as lack of model generality across scales, the choice of appropriate predictor variables, assumptions on response shapes to environmental gradients and the difficulty in inferring causality. Parameterization of statistical models may prove particularly challenging with the very large number of predictor variables available, often with a high degree of collinearity between them. Rushton *et al.* (2004) argue that an information theoretic approach be adopted within the framework provided by Burnham and Anderson (2002). This method suggests that a number of candidate models should be developed based on a sound ecology theoretical background as suggested by Austin (2002b), and that the Akaike Information Criterion be used to identify the most parsimonious models from these. Utilizing expert opinion in choosing variables in stepwise parameter selection procedures may select biologically important, uncorrelated variables with some insight into causal relationships (Pearce *et al.*

2001). However, Seoane *et al.* (2005) have shown that the subjectivity of expert opinion does not necessarily result in models with superior predictive performance, and its utilization may be time consuming and difficult to standardize.

Model generality is an important consideration if the model is to be applied outside of the range within which it was developed, such as under scenarios of climate change. Whittingham *et al.* (2003) found their generalized linear mixed models for skylarks to give good predictions when tested outside the geographic range in which they were developed. On the other hand Osborne and Suarez-Seoane (2002) found that geographic data partitioning resulted in significantly better model performance for birds in Spain which was attributed to the possibility of geographically varying habitat use patterns. This was also observed by Fielding and Haworth (1995) where models showed very low levels of generality, leading to the suggestion that some species geographic distributions might not be predictable. Again the issue of modeling the fundamental versus the realized niche is critical in terms of model generality, where it is generally accepted that mechanistic models based on the fundamental niche will be more generally applicable than correlative ones based on the realized niche (O'Connor 2002). In addition to this model accuracy is highly variable across taxa with no consistent pattern associated with species ecological characteristics apparent (Karl *et al.* 2002, Stockwell and Peterson 2002a, McPherson *et al.* 2004).

1.5.3 Data and Habitat Distribution modeling

Data is another central issue in habitat distribution modeling and research relating the type of data available, its quantity and quality is commonplace. Most statistical correlative modeling approaches require presence/absence data as is usually collected in biodiversity surveys. Methods also exist for modeling using presence data only. Data most easily accessible for the purpose of habitat distribution modeling are primary biodiversity data, the data that accompany specimens in museum and herbarium collections. (Peterson *et al.* 2002) and these usually constitute presence data only. Due to information technology innovations such as online database integration protocols these data are fast becoming easily available. Remotely sensed data from which a range of fine scale environmental predictor variables can be extracted are also becoming more accessible. Together these allow for investigations into organism environment relations in a level of detail that was not possible or at least difficult before (Soberon and Peterson 2004). Thus presence only modeling methods have received significant attention (Stockwell and Peterson 2003, Soberon and Peterson 2004, Stockwell and Noble 1992, Stockwell and Peters 1999, Hirzel *et al.* 2002)

For a number of statistical modeling methods skewed data in the form of presences only are insufficient. Thus methods for generating absence data have also been suggested. These 'pseudo-absences' are usually generated by randomly selecting a number of sites from the study region and assuming that the species is absent in those locations. However, Engler *et al.* (2004) show how pseudo- absences may be generated by first determining the range over gradients where a species does occur based on presence data only and then assigning absences outside of that range.

Zaniewski *et al.* (2002), in a controlled experimental approach where true absences were recorded, showed how GAMs developed with pseudo absences generated using this method better approximated GAMs of true presence/absence for ferns in New Zealand than did GAM's with randomly generated absences.

The quality of survey recorded absence data has also received attention, with emphasis on the reliability of that data. Unless a species has a detection probability of one, it is likely that it may occur at certain sites were it is recorded absent. Inclusion of such non detections as absences in modeling may result in spurious environmental relationships being identified (Gu and Swihart 2004). MacKenzie *et al.* (2002) present a method for the estimation of site occupancy rates when detection is not certain, and Stauffer *et al.* (2002) show the binomial model may be adjusted to account for non detection. Anderson (2003) presents statistical methods for detecting false absences, and Edwards *et al.* (in press) suggest improved sampling methods for increasing detection rates of rare species.

Also important are issues of environmental data quality. Lim *et al.* (2002) indicate that distributions of species are difficult to model when good quality environmental data are lacking. Reese *et al.* (2005) also indicate that model performance is a function of environmental data quality. Peterson and Cohoon (1999) show that model accuracy reaches an asymptote with the inclusion of four to five environmental layers in their analysis. They do state though that these results are specific to their analysis and should not be generalized without further testing. Seoane *et al.* (2004a,b) investigate the use of competing topo-climatic and vegetation variables, finding that a combination of both provide the best models, and indicate that coarse scale vegetation layers are as good predictor variables as fine scale vegetation layers. Thomas *et al.* (2002) have shown that field collected environmental data may also include some errors, highlighting the need for proper training of field workers for survey work. A number of authors have identified the need for habitat models to account for spatial autocorrelation (Guisan and Zimmermann 2000, Keitt *et al.* 2002, Legendre *et al.* 2002)) and methods for investigating this property using semivariograms are discussed by Cablk *et al.* (2002).

Essentially, data quantity and quality are the principle driving factors determining the potential for good correlative models to be developed (Lim *et al.* 2002, Stockwell and Peterson 2002b, Reese *et al.* 2004) and with the current use of these models in land use planning decisions it is vital that good data be available for making good models (Vaughan and Ormerod 2003). The quantity and quality of data available for modeling should be assessed beforehand and if needs be they should be augmented with additional data collected for the purpose. Survey design for data collection is another important aspect relating to the quality of data available for modeling. Vaughan and Ormerod (2003) stress the requirement that surveys should aim to adequately sample important environmental gradients. While the number of gradients and potential combinations may be huge, a subset of these usually actually occur in the field. Hirzel and Guisan (2002), in using a virtual species to ascertain the efficacy of different sampling designs for providing data sets for habitat distribution show that a regular grid sampling design provides data for modeling as good that obtained from a random stratified sample design. The gradsect method has also been suggested as a cost effective manner of sampling major environmental gradients in a region (Austin and Heyligers 1989, 1991) and Chown and Freitag-Ronaldson (2002) provide an extension of the this method that includes considerations of spatial data completeness.

1.5.4 Model Accuracy and its Assessment

The issue of model accuracy is also an important consideration. Fielding (2002) defines accuracy as the closeness of model predictions to the real value. Assessment of habitat models is an important step in the modeling process as it allows for a certain level of confidence to be placed in that model and its predictions. While modeling methods have received wide attention, methods for assessing model accuracy have been largely ignored (Fielding 2002) with only default assessment statistics available with software packages usually reported.

The essential elements of a habitat model evaluation are the error rates in its predictions, drawn from a confusion matrix. Errors are of two principle types, omission (false negatives) where the model predicts the species not to occur where it has been recorded, and commission, where the model predicts it to occur where it hasn't been recorded. Commission has two components, true commission error, where the model predicts the species to occur where it does not (false positives), and apparent commission error where the model predicts the species to occur where it has not been recorded but does occur (Anderson *et al.* 2003, Schaefer and Krohn 2002). Identifying this latter component is often the objective of habitat distribution modeling. It is not however easily separated from true commission error and the interpretation overall commission error must take this into account.

Fielding and Bell (1997) provide a thorough overview of model assessment methods, their strengths and weaknesses. An important model assessment statistic has been Cohens Kappa (Cohen 1960). However, this measure requires pseudo absences be generated where absence data are not available and has been shown to be prevalence dependent, meaning that the outcome is not independent of the ratio of presence to absence records in the data (Manel *et al.* 2001). Fielding (2002) describes how a good accuracy measure should have several desirable properties. It should be prevalence independent, it should allow for the incorporation of relative costs of false positive and false negative predictions and it should be independent of thresholds for the conservation of predicted probabilities to presence and absence. The receiver operating characteristic (ROC) is a measure that meets these criteria (Fielding and Bell 1997, Manel *et al.* 2001, Fielding 2002). This measure calculates the false positive and false negative rates over a range of threshold values from zero to one, and plots these against each other. Evaluation of the model is by the area under the curve (AUC) and is compared against a null model of totally random prediction where AUC equals 0.5. Pearce and Ferrier (2000) present guidelines to interpreting AUC, where a value from 0.5 - 0.7 indicates a poor model, 0.7 - 0.9 a relatively good model and values above 0.9 indicating very good models. The significance of departure from random or differences between curves may also be tested using the methods of Delong *et al.* (1988) or Hanley and McNeil (1982). Insensitivity to prevalence and the ease of calculation and interpretation make ROC analysis amenable to model evaluation.

1.5.5 Spatial Scale and Habitat Distribution Modeling

An essential consideration in any aspect of ecology is spatial scale (Wiens 1989). As different patterns may be observed at different scales ecological theory should explicitly specify the scale domains over which its concepts are applicable. Levin (1992) states that the imposition of particular scales of investigation onto ecological systems by the investigator is essentially arbitrary in that organisms perceive their environments on a range of scales which may be very different to those perceived by humans. The suggestion is made that discovery of the scales over which important processes occur should be a key focus of ecological research. As such, scale is a crucially important consideration in modeling species distributions, and issues relating to model generality, model accuracy and the approximation of fundamental versus realized niches may be a result of failure to adequately address scale issues (Wiens 1989)

Spatial scale in ecology is a vague concept that refers to both the extent and the grain of data (Wiens 1989, Morrison and Hall 2002). For example, ‘large scale’ may refer to both data of large extent, or to data with large grain. Both grain and extent limit the ability of a researcher to ascertain underlying processes from patterns in the data (Wiens 1989, Huston 2002). Given that some ecological processes may give rise to patterns on different scales to those processes (Huston 2002) and that data may be collected on scales assumed to be important beforehand (Levin 1992) the detection of the underlying processes is often difficult. Blackburn and Gaston (2002) state that different scales of study may give rise to different answers to the same question and that these answers may be equally interesting within the context of the particular scale that they are dependent upon.

Pearson and Dawson (2003) have proposed a hierarchical framework of scale within which modeling can take place. This framework suggests that variables affecting distribution change across scales, with climatic variables being important at broad scales, and biotic interactions being important at fine scales. Guisan and Zimmermann (2000) also suggest a scale based approach to modeling methods be used, where correlative statistical models are appropriate at coarse scales and mechanistic models at fine scales. Despite the suggestions by Levin (1992) it is still common practice to determine the scale of study *a priori*, usually for reasons relating to the practical application of the exercise. For example, regional conservation planning requires species distribution information for a particular extent and grain determined by the planning region and the size of the planning units. However, some authors have taken the approach of modeling at multiple scales in order to determine which environmental variables might be important at different scales in explaining patterns of distribution. For example, Grand and Mello (2004) found that landscape patch metrics were important predictor variables of patch occupancy rates for moths and were more important than within patch environmental variables. A similar result was found by Johnson *et al.* (2002) following multiscale modeling. This multiscale modeling approach is still a relatively new practice, and further research is required on multiscale modeling methodologies (Vaughan and Ormerod 2003) and on spatial scaling in general (Wiens 1989).

Overall, habitat distribution modeling methods provide a powerful set of tools for investigating species environmental relationships and for providing useful information for a range of purposes. In light of the considerations above it is critical that outputs from such modeling exercises be viewed critically bearing in mind the shortcomings of the method or quality of the data used for the exercise.

1.6 Habitat Distribution Modeling in Conservation Science

Gaps in the geographical coverage of museum and herbarium data have obvious implications for selection of protected area networks in systematic conservation planning based on complementarity. Habitat distribution modeling may serve to fill these gaps to a certain degree and hence dampen the effect of spatial sampling bias in identification of areas of conservation priority (Williams *et al.* 2002, Polasky *et al.* 2000). Models of this type have a long history of use in conservation planning. For example they have been used for over ten years for more than 2300 species in New South Wales (Ferrier *et al.* 2002). Some habitat suitability modeling methods have been advocated as superior to traditional gap analysis methods such as HSI (Peterson and Kluza 2003, Stockwell and Peterson 2003). Model predictions have also been suggested as a proxy for population persistence probability and hence may be used to improve reserve site selection (Araujo and Williams 2000, Williams and Araujo 2002, Cabeza and Moilanen 2001, Cabeza *et al.* 2004, Fleishman *et al.* 2002). Habitat distribution models have been applied to a wide range of single species conservation problems, such as for mice in Ecuador (Anderson and Martinez-Meyer 2004), a multitude of birds (eg Gibson *et al.* 2004, Jeganathan *et al.* 2004) and for insects (eg Meggs *et al.* 2004, Maginni *et al.* 2002).

While their usefulness and wide range of application is apparent for conservation applications, attention has been drawn to certain shortcomings such as sensitivity of reserve selection outcomes to thresholds used in converting probabilities to presence absence (Wilson *et al.* 2005). Importantly again, Vaughan and Ormerod (2003) emphasize how models used for decision support in resolving land use conflicts should be sound. An integrated, iterative process of model development, postulate formulation, data collection, model evaluation and model improvement should accompany the use of habitat distribution models in conservation practice.

2. Methods

2.1 Site Selection and Sampling

Sampling points were placed across the province at a regular interval of 0.2 degrees latitude and longitude (approximately every 20 km), making a sampling grid starting at 27.0000S and 27.0000E, following the method of Hirzel and Guisan (2002), Sampling at these sites was carried out by Martin Paulsen, consulted by GDACE as an expert on baboon spiders, and Ian Engelbrecht as an expert on scorpions.

To find a site during the fieldwork its location on the ground was estimated from its relative position to notable landmarks, such as road intersections, on a map of the province and a GPS was used to locate the point precisely. Sampling took place as close as possible to these provisional sites, depending on whether the area constituted natural cover or not. Sites under agriculture, urban development or other major disturbance where not sampled, and the nearest accessible area with natural cover was sampled instead. Sampling took place from mid-October to mid-November, so as to take advantage of high ground visibility in fields that had been burnt the previous winter.

To determine species presence at a site line transects were walked for a minimum of two hours and the ground scanned for burrow entrances. Where rocks and rocky outcrops were present they were also investigated by rock turning and checking rock cracks. If a species was not located at a site after two hours, another, secondary site nearby was chosen based on whether it was felt any of the species may occur there, and searched. Thus, secondary site selection was subjective. If any of the focal species was located distance sampling methodology, described below, was carried out until a total of four hours had been spent at the site. Thus a total of 47 primary and 10 secondary sites were sampled.

2.2 Estimating Minimum Viable Area

In order to estimate the minimum viable area required to support a viable population of each species it is necessary to determine minimum viable population size. As the collection of the appropriate life history data for PVA modeling was not possible within the timeframe of this project, a generic MVP size of 2000 adult individuals is used based on the findings of Reed *et al.* (2003) and Reed and Hobbs (2004).

Once a target species was located at a site, distance sampling methodology using line transects was used to collect the relevant data for estimating population density (see Buckland *et al.* 2001 for a full treatment of this method). This method fits a curve of detection probability of objects of interest with distance from the line and uses this to estimate the proportion of objects that are missed up to a certain distance. A GPS (Garmin Geko 301 or Garmin GPSMAP60C) was used to record the co-ordinates of the start and end points of each line (with between three and six meters accuracy), and these were used to calculate line length using the formula in Buckland *et al.* (2001, p273). Objects of interest included *Opistophthalmus* and *Harpactira* burrow entrances on open ground or under rocks and rock cracks currently in use by *Hadogenes*. Once an object was spotted the perpendicular distance from the line to the object was measured using a 30m tape measure. Distances were measured to the nearest 0.05m. In estimating population density from the density of burrows an important assumption is that each burrow is occupied. Most movement from burrows occurs in the summer and autumn months where males search for females. Therefore it is reasonable to assume that a low proportion of burrows were unoccupied during the study period. If this was the case the population density estimate will be biased high, and the MVA estimate biased low

In order to estimate the proportion of the population that is adult a note was made of whether specimens of *Harpactira hamiltoni* and *Hadogenes gunningi* collected or observed in the field were adult or juvenile and the proportion calculated from this. For *Opistophthalmus pugnax* only adult voucher specimens were collected, which would bias such an estimate if it was used for this species too. As it would be logistically impossible and unethical to excavate all burrows observed for this species, burrows were categorized as large, medium or small, under the assumption that burrow size is correlated with the size of the occupying individual.

The software program DISTANCE (Thomas *et al.* 2003) was used to estimate population densities for *O. pugnax* and *Harpactira hamiltoni*. Sample size was insufficient to calculate density for *Hadogenes* gunningi. Population density was estimated for suitable habitat for these species in Gauteng, thus data were used only from sites where the species was recorded to occur and line transects from sites where the species were not found were excluded from the density estimation procedure in order to avoid underestimating density.

The analysis procedure started with visual inspection of histograms of the distance data per species to check for heaping (where distances are rounded by the fieldworker to make data collection easier resulting in non uniform decrease in detection probability with distance from the line) and the need for data truncation to eliminate outliers. Outliers are small numbers of observations (usually only one or two) that are unusually far from the line and other observations. These are essentially unusual observations that may arise for example when the observer spots an entity while away from the line for some other reason, and then measures the distance back to the line. These observations may skew the results of the model fitting procedure. Data were truncated by setting a maximum distance from the line up to which observations would be used after visual inspection of histograms so as to eliminate such outliers. A series of models were then fitted to the data using automatic selection of adjustment terms with a sequential selection method and the Akaike Information Criterion as the automatic model selection criterion. This latter criterion selects the models that best fit the data but includes a penalty for an increase in the number of model parameters. Thus models are chosen that fit the data well with the fewest parameters. This is an iterative process which generates a number of possible model fits, each with a set of parameters that can be used by the investigator to select which models will be used for estimation of population density. This procedure was carried out for both untransformed data as well as data transformed into intervals.

Those models resulting from this iterative model generation procedure that fit the data best were then selected out using AIC and χ^2 values for interval data, and AIC, χ^2 values and qq-plots for untransformed data (Appendix I). Once the best, or group of best model fits were identified these were used to estimate densities. The results of the density estimation were then used to estimate MVA assuming a MVP size of 2000 individuals. This was done for mean density values, as well as for the upper and lower confidence levels for the density estimate. Buckland *et al.* (2001) advocate reporting of results from several fitted models so as to give an indication of the quality of the data used for density estimation. If results are consistent across different models then a high degree of confidence can be placed in those results. Thomas *et al.* (2003) also suggests the use of a composite model where differences between model fit are small (ΔAIC<2) so as to get the most robust estimate of density. This method was applied here for comparison with other results.

An important consideration with this method is the assumption that detection probability of an object on the line be one, i.e. all objects on the line should be detected in the survey. There was no way to test this assumption with the data collected (although probability of detection is highest close to the line as determined from the histograms of observations per distance interval). Violation of this assumption may lead to an underestimation of population density. While provision can be made for such a violation in the analysis of distance data, an estimate of the proportion of individuals on the line missed is necessary, which can be very difficult to estimate. Thus it is assumed that probability of detection on the line is one in this instance.

2.3 Habitat Distribution Modeling

2.3.1 Genetic Algorithm for Rule Set Production (GARP)

Genetic algorithms have shown to be robust a method for finding optimal solutions to a range of complex problems (Goldberg 1989). They are based on the process of biological evolution, where selection criteria are set and elements of the solution are modified iteratively through a process that approximates mutation and natural selection. Optimality of the solution is not defined as the best solution, but rather that which is best relative to other possible solutions (Goldberg1989).

The GARP modeling method (Stockwell and Noble 1992, Stockwell and Peters 1999) is a genetic algorithm implementation where the elements of the solution, or model, are rules that predict the presence or absence of a species based on a set of environmental predictor variables (Stockwell and Noble 1992). Initially, a number of rules are generated randomly. These rules are then evaluated independently for their predictive performance and the best predictors are selected for inclusion in the rule set. This rule set is then modified ('mutated') through the inclusion of new, randomly-generated rules, values of parameters within rules are randomly modified and rules are randomly concatenated. This new rule set is evaluated again, and only the best rules maintained. The procedure is then repeated. The stopping condition for the algorithm is either where it reaches a certain maximum number of iterations, or where the predictive ability of the overall rule set reaches a preset convergence limit.

An important aspect of GARP is that rules are developed based on a range of rule types, some of which are based on statistical modeling methodology, such as envelope and regression models. As such GARP has been suggested as a superior modeling method as it is essentially a superset of these other modeling methods (Peterson *et al.* 2002). It has also been suggested that it may perform particularly well on ad hoc data assemblages where other multivariate methods may be inadequate due to imperfect sampling design (Stockwell and Peters 1999). In empirical studies it has been shown to outperform or at least provide similar results to other modeling methods, including logistic regression and expert based habitat classification methods (Peterson and Kluza 2003, Mandelberg 2004,). Additionally, each rule is specified by a set of preconditions for that rule to be applied, which provides the potential for developing rules that may apply in some parts of species range where these preconditions are met, and not in others(Stockwell and Noble 1992). This may be important where species occur under different environmental conditions in different places. GARP model predictions have been used in the comparison of the ecological niche spaces of closely related species where postulates about evolutionary or biogeographic histories have been inferred (see Peterson *et al.* 1999 and Anderson *et al.* 2002a, b for examples). To make a prediction the rule set is applied to each combination of environmental variables for each pixel of a map of the area under investigation. Values are returned as either species predicted as present or absent. Where rules conflict the value of no prediction is returned for that pixel.

The GARP algorithm is currently implemented in the freely available software package DesktopGARP (Scachetti-Pereirra 2005). The original implementation was at the Environmental Resources Information Network (ERIN, Boston and Stockwell 1994) followed by an implementation for the Biodiversity Species Workshop (BSW, Payne and Stockwell - no date) where the algorithm was based at the San Diego Super Computer Centre. The original two implementations differ from DesktopGARP in the rule types used and in the output. The ERIN and BSW implementations used atomic, BIOCLIM, range and logit rules. Atomic rules are simple rules specifying a single value for a predictor variable, eg temperature = 15C. The BIOCLIM rule type was a rule based approximation of bioclimatic envelope modeling method of Busby (1991) while the logit rule is based on logistic regression. DesktopGARP replaces the BIOCLIM rule type with a negated rule type (Scachetti-Pereirra 2005). Outputs differ between the ERIN and BSW implementations and DesktopGARP in how non-predictions are treated. In the ERIN and BSW implementations such cases were included in the outputs, whereas they are not in DesktopGARP. Here instead non-predictions are output as absences, based on the assumption that only areas where the species is predicted to occur are important. No mention is made in the documentation of how conflicts are dealt with. Finally, the ERIN and BSW implementation included a module that output the rule sets in a format that could be interpreted rule by rule, whereas this facility is not available in DesktopGARP. The ERIN and BSW implementations are no longer available for general use.

Running the GARP algorithm produces a large number of resultant models. Due to random rule production and mutation in the algorithm model results may vary quite markedly (Anderson *et al.* 2002b, Elith and Burgman 2002). Thus the evaluation of these models and the selection of a subset of models for the purpose is necessary. Anderson *et al.* (2003) provide a ‘best subsets’ selection procedure, based on subjective expert evaluation of model predictions. Here a series of predictions are ranked by an expert as good, average or poor according to how well they are thought to approximate the distribution of the species. The relevant omission and commission statistics for these models are then plotted and the region on the commission axis where most of the good models fall is identified. This region is defined as an interval and, in addition to an arbitrarily defined acceptable omission threshold, e.g. 5%, is used to select ‘good’ models. The predictions of these models are then stacked and the number of positive predictions per cell summed across the models. The final model is taken as this overlay of all model predictions.

A best subsets procedure is implemented in DesktopGARP. However, it does not allow for a commission interval to be set, but rather a commission threshold. Thus a larger number of models qualify as candidates from a single run of the algorithm than would be the case for the procedure of Anderson *et al.* (2003).

2.3.2 Distribution Modeling for the Species of Interest

DesktopGARP was used to model the distributions of the three species of interest for this project. Point locality records from this field survey were combined with those collected on GDACE general biodiversity surveys, in order to obtain datasets of 20 or more records required for GARP modeling (Stockwell and Peterson 2002a). Museum record data were not included in the modeling exercise due to low precision of the georeferenced data compared to that of the survey records collected here and by GDACE. A total of 22, 62 and 41 unique localities entered the analysis for *Hadogenes gunningi*, *Opistophthalmus pugnax* and *Harpactira hamiltoni* respectively.

Predictor variables for modeling were chosen to represent a range of climatic and environmental variables understood to be important in limiting species distributions in general (Appendix II). 16 climatic and nine environmental variables were selected for the modeling. Climatic variables were selected according to the presumption that temperature and water stress might be important limiting variables. Thus, chosen variables represented extremes, such as mean maximum temperature in summer or water stress in winter, as well as average conditions such as mean annual temperature. Other environmental variables included elevation, slope, aspect, soil properties, geology, vegetation type and landcover. Choice of predictor variables was heavily constrained by data availability. For example, while soil type may be an important limiting variable for burrowing scorpion species (Prendini 1995), there is no complete soil type spatial data set for Gauteng.

Due to constraints relating to resolution (raster layers) or minimum mapping unit (MMU for vector layers) two separate modeling exercises were conducted at different grains. Choice of variables for each modeling exercise was predominantly influenced by the original resolution or MMU of the data layers. This method was chosen as resampling all layers to the same resolution resulted in loss of important information in some layers. Thus, climatic variables were used as predictor variables for the first modeling exercise, with a pixel resolution of 1.8km and topographic, landcover, vegetation and substrate variables were used in the second with a pixel resolution of 0.05km.

For each analysis DesktopGARP was set to use 70% of the data for training and 30% for testing in order to obtain internal model performance statistics. All rule types were used and the option for all rule type combinations was set. The number of runs was set to 100, which yielded a total of 1500 models per species at both coarse and fine grains. The stopping criterion was 1000 iterations or a convergence of 0.01. In order to select a best subset of models, the method of Anderson *et al.* (2003) was attempted. The subjectivity of this procedure proved to be prohibitive in its implementation in that the potential distribution of these species is not known, making the selection of an appropriate commission interval difficult, and in that the ‘goodness’ of the fine scale predictions could not be judged. In addition to this, it appeared that what was considered to be a good model depended on which models were previously ranked as good, average or poor. The possibility exists that the method of Anderson *et al.* (2003) may be more appropriate for very large scale model evaluation, such as that were they developed it on rodents over the northern half of South America. Hence, an alternative method for selecting a best subset was used here, based on the premise that commission should not be included as a selection criterion if the distribution of the species is not known and that models that fail to make a prediction for large areas are poor models. The first criterion for a model to be included in the subset was that intrinsic omission, or the number of training presence records predicted as absence, was less than 5%. The second was that the non predicted area be less than 10% of the total area. This is based on the assumption that for a model to be useful it should make a prediction for at least 90% of the total area.

Once a subset of models was identified from the results of each DesktopGARP run, the predictions for these were overlain to obtain the number of positive predictions per map pixel across the subset. This value was then divided by the total number of models in the subset, in order to obtain the proportion of positive predictions per pixel. This proportion is assumed to approximate the probability of the species occurring in that pixel, and is referred to hereafter as the probability of occurrence. The corresponding pixel values of the coarse and fine grain predictions of probability of occurrence for each species were then multiplied by each other to produce a composite prediction at the same resolution as the fine scale prediction. This way only those areas with a high probability of occurrence for a species at both scales maintained a high probability of occurrence in the composite prediction. This was based on the assumption that conditions governed by both coarse and fine scale environmental variables have to be favorable if a species is to be able to survive and reproduce in an area.

The final overlay and composite models were evaluated using the area under the curve statistic of the ROC plot (Hanley and McNeil 1982). The MS Excel extension package Analyse-It was used for this analysis, and ROC plots and estimates of AUC with confidence limits are presented. From these results the cutoff point where the number of false positive and false negative predictions were equal was taken as the cutoff to convert the probabilities of occurrence to presence and absence. Presence data for the ROC analysis were the same as those used to develop the GARP models while absence data were obtained from the results of the field survey under the assumption that failure to locate a species at a site meant that it was absent. While this assumption is weak given that detection probabilities of the species are likely to be less than one and hence there are likely false absences in the dataset it does provide the best possible option for the ROC analysis without extensive statistical interrogation of the data. The method for ranking models qualitatively of Pearce and Ferrier (2000) was applied to the ROC analysis results.

3. Results

3.1 Survey Results

The sites surveyed for this study are presented in figure 3.1. Four new records were collected for *Hadogenes gunningi* (Figure 3.2), while 25 and 22 new records were collected for *Opistophthalmus pugnax* (Figure 3.3) and *Harpactira hamiltoni* (Figure 3.4) respectively. These additional records increased the extent of occurrence for all three species from that previously known, but most markedly so for the latter two species.

Table 3.1: Recorded localities within Gauteng province for all scorpion and theraphosid baboon spider species collected during the field survey. Site numbers refer to those on figure 3.1.

Order	Family	Species	Site numbers
Scorpiones	Buthidae	*Uroplectes carinatus*	GD07a
		Uroplectes triangulifer marshali	GD31, GD44
		Uroplectes triangulifer triangulifer	GD01, GD08, GD11, GD12, GD14, GD18, GD20, GD22, GD23a, GD26, GD27, GD28a, GD32, GD33, GD35, GD36a, GD37, GD29, GD40, GD43
		Uroplectes sp (*U. formosus* relative)	GD30
	Liochelidae	*Hadogenes gracilis*	GD16
		Hadogenes gunningi	GD06, GD12, GD18, GD23a, GD26
		Hadogenes longimanus	GD05
		Cheloctonus jonesii	GD01
	Scorpionidae	*Opistophthalmus glabrifrons*	GD01, GD02, GD03, GD04, GD05, GD06, GD07, GD08, GD09, GD09a, GD10, GD11, GD17, GD18, GD21b
		Opistophthalmus pugnax	GD13a, GD15, GD16, GD19, GD20, GD21a, GD22, GD23, GD26, GD27, GD28, GD29, GD30, GD31, GD32, GD33, GD34, GD35, GD36, GD36a, GD37, GD40, GD42, GD44, GD45,
Araneae	Theraphosidae	*Augacephalus junodi*	GD01, GD02, GD04, GD05, GD07a
		Harpactira hamiltoni	GD06, GD12, GD13a, GD14, GD17, GD18, GD19, GD20, GD23a, GD28, GD28a, GD29, GD30, GD32, GD33, GD34, GD36, GD37, GD38, GD39, GD40, GD44, GD47a
		Harpactirella sp	GD02, GD07a, GD14, GD30, GD33, GD34
		Idiothele nigrofulva	GD38

Table 3.1 presents all other scorpion and baboon spider species collected during the field survey and the sites at which they were found. An important discovery was that of the rock scorpion *Hadogenes gracilis* on the Magaliesberg as this species was thought to be limited to the Pyramid Koppies range just north of the Magaliesberg. Another was a trapdoor baboon spider, *Idiothele nigrofulva*, near Carltonville. This species is more widely distributed in savanna having been recorded at numerous localities around Barberton and the Kruger National Park. This represents a significant westerly extension of its distribution range, as well as the first record of it occurring in the grassland

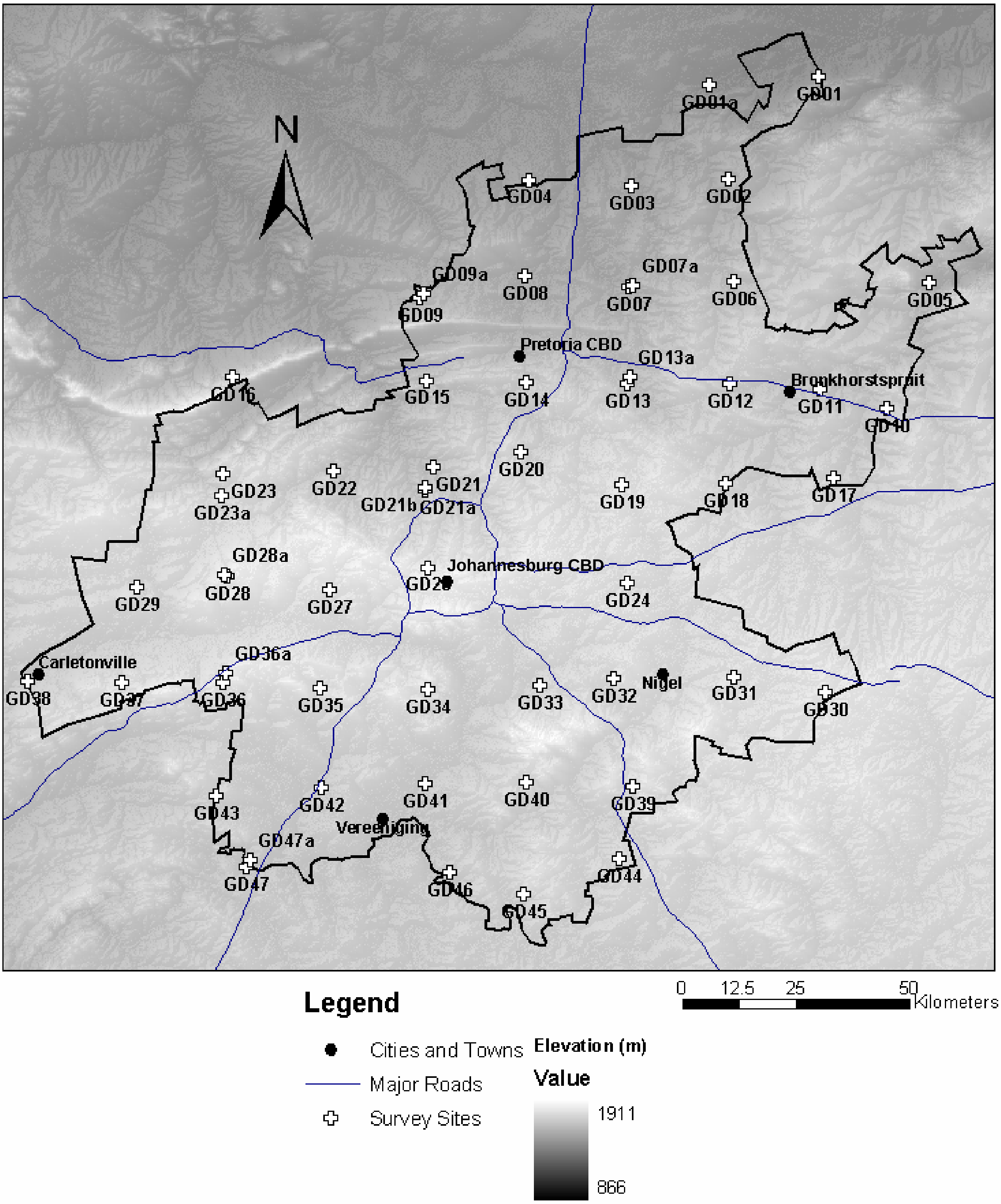

Figure 3.1: Gauteng Province with all field sites surveyed indicated. Refer to table 3.1 for arachnid species collected per site. Sites were arranged on a 0.2 degree grid. Major roads and urban centers are indicated.

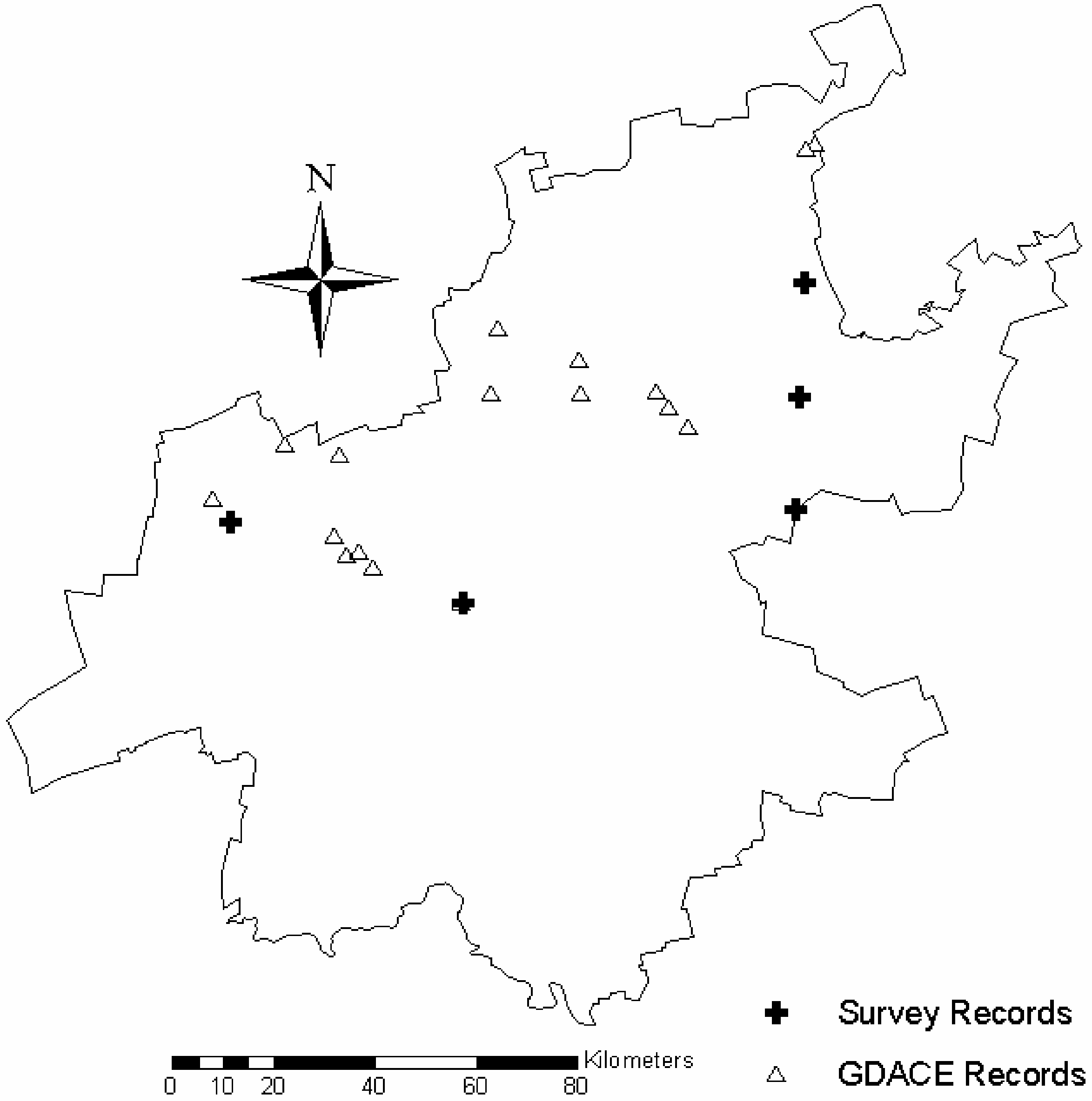

Figure 3.2 Locality records for *Hadogenes gunningi* indicating those collected during the field survey and those collected by GDACE. All locality records were used in developing GARP models.

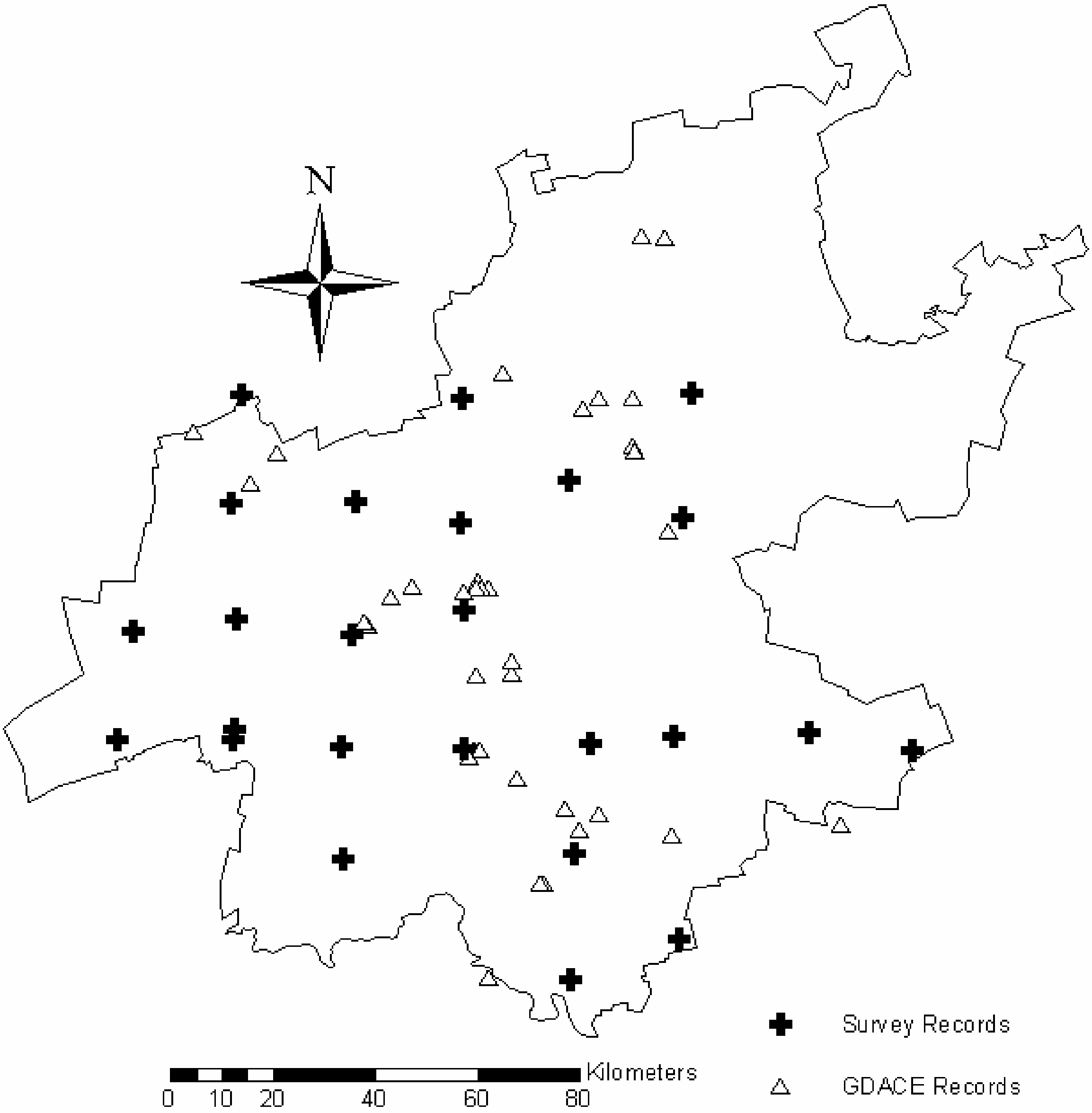

Figure 3.3 Locality records for *Opistophthalmus pugnax* indicating those collected during the field survey and those collected by GDACE. All locality records were used in developing GARP models.

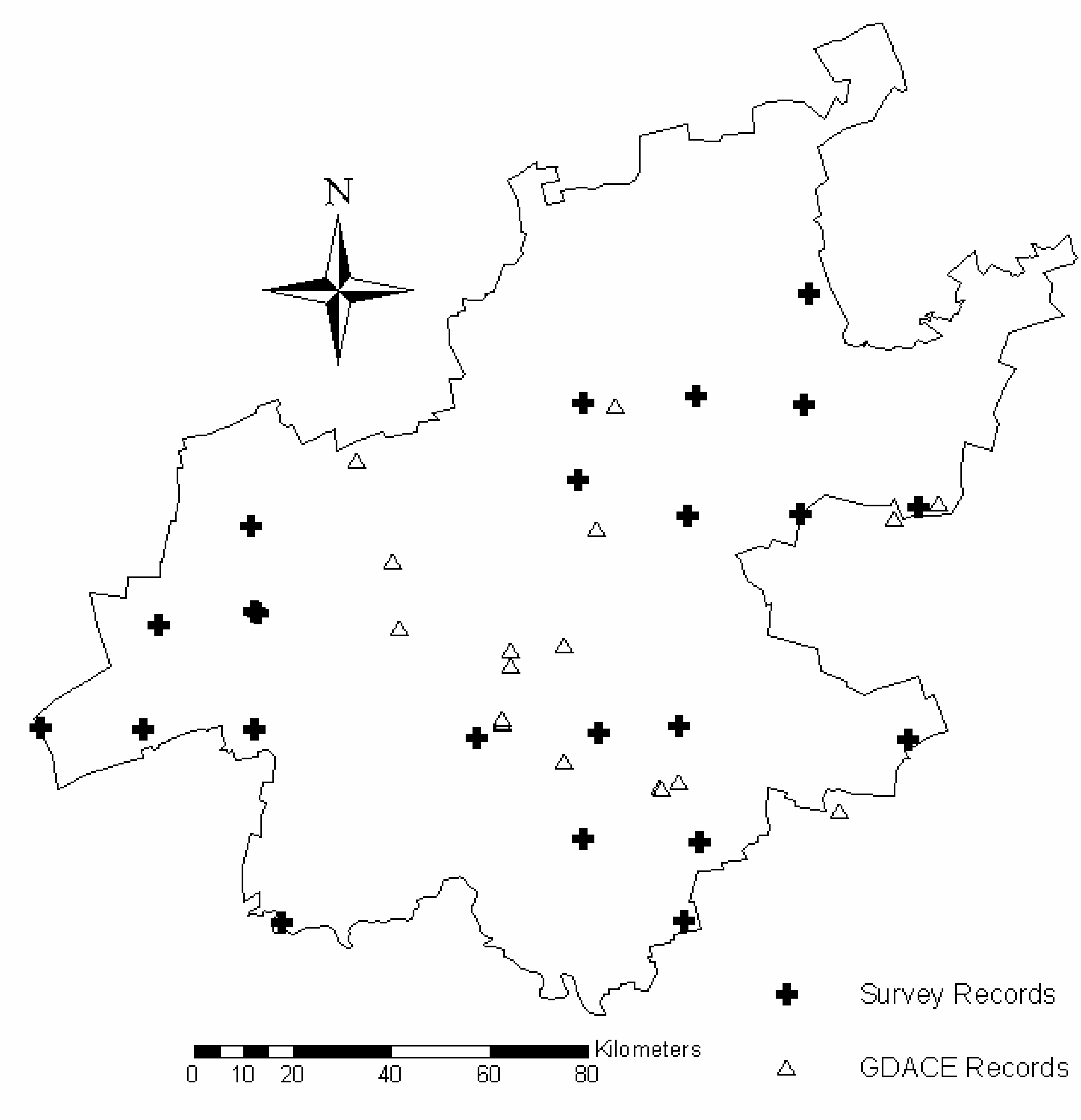

Figure 3.4 Locality records for *Harpactira hamiltoni* indicating those collected during the field survey and those collected by GDACE. All locality records were used in developing GARP models.

biome. Another interesting find was a potentially undescribed species of *Uroplectes* in the *U. formosus* species complex (Prendini, pers comm.). This species was also collected by M. Paulsen near Ermelo, so is possibly widely distributed on the Mpumalanga highveld. Other scorpion species recorded for the province but collected during this survey include *U. vittatus*, *Pseudolychas ochraceus* and *Parabuthus transvaalicus*. Another undescribed species of *Uroplectes* in the *U. triangulifer* complex has been collected from Bonaccord Dam in Pretoria and from near Cullinan. These are the only recorded localities for this species and it may be endemic to the province.

3.2 Density Estimation and MVA

Per hectare density estimates for *Opistophthalmus pugnax* and *Harpactira hamiltoni* using both interval and continuous data are presented in Tables 3.2 to 3.5, along with their corresponding MVA estimates. Results are reported per fitted detection probability curve, for each observer and an average for both observers. For *Opistophthalmus pugnax,* the data recorded at site GD13a (Figure 3.1) were analyzed separately from all other sites where the species was found, as the population density at this site was markedly higher than at other sites. While both field workers, IE and MP, collected data for all species of interest, data analysis showed large discrepancies between the two, both in the shape of detection curves and in estimated densities. Thus densities are reported per observer, and from a range of model fitting methods. .

Using burrow size categories as a rough approximation of population structure showed 23% of individuals to be large, 41% to be medium-sized and 36% to be small for *Opistophthalmus pugnax*. For *Harpactira hamiltoni* 55% of individuals collected or observed were adult, while 44% of *Hadogenes gunningi* collected or observed were adult.

The density estimates across all models for *Harpactira hamiltoni* based on the data collected by Martin Paulsen were between 2.03 and 7.84.ha^{-1}(95% confidence intervals) with a mean of approximately four adult individuals per hectare. For the same species the data collected by Ian Engelbrecht yielded density estimates of between 1.66 and 9.20.ha^{-1} across all models with a mean of between three and four adult individuals per hectare. For *Opistophthalmus pugnax* density estimates for sites other than GD13a were from 2.17 to 26.25.ha^{-1} with a mean of approximately seven adult individuals per hectare based on data collected by Martin Paulsen and 13.28 to 119.98.ha^{-1} with a mean of 31 adult individuals per hectare from that collected by Ian Engelbrecht. For site GD13a these estimates were between 4.55 and 17 264.03.ha^{-1} with a mean of about 260 adult individuals per hectare based on data by Martin Paulsen and 48.18 and 2812.21.ha^{-1} with a mean of about 440 adult individuals per hectare for that by Ian Engelbrecht. Estimates based on the data collected by both field workers gives an estimate from 2.28 to 7.87.ha^{-1} with a mean of about four adult individuals per hectare for *H. hamiltoni*. For *O. pugnax* density at sites other than GD13a based on all data were between 9.85 and 29.14.ha^{-1} with a mean of about 16 or 17 adult individuals per hectare. For site GD13a these estimates were between 175.01 and 862.81.ha^{-1} with a mean of about 303 adult individuals per hectare.

Table 3.2: Population Density and MVA estimates for *Opistophthalmus pugnax* using continuous data. Variance estimates are by bootstrap resampling of line transects. Where the Model is 'Composite', this refers to the AIC weighted average method of Burnham and Anderson (2002). See Medium-sized for the models included for this. Minimum viable area (MVA) estimates for the mean, lower and upper confidence intervals are in hectares. LCL is the lower confidence level, UCL is the upper confidence level and CV is the coefficient of variation.

Site	Collector	Model and adjustment term	Mean Density	Density LCL	Density UCL	Density CV	Mean MVA	MVA LCL	MVA UCL
Ex 13a	MP	Composite	8.21	3.78	17.81	0.40	840.02	1824.48	387.23
		Halfnormal/Cosine	5.17	2.51	10.64	0.37	1333.96	2747.63	648.17
		Hazardrate/polynomial	10.62	4.30	26.25	0.48	649.39	1603.85	262.73
		Uniform/polynomial	4.46	2.17	9.16	0.37	1546.31	3178.13	752.90
	IE	Composite	33.01	17.85	61.04	0.32	208.92	386.36	112.98
		Hazardrate/Cosine	33.01	17.85	61.04	0.32	208.92	386.36	112.98
		Hazardrate/Polynomial	33.01	17.85	61.04	0.32	208.92	386.36	112.98
		Uniform/Cosine	33.05	19.52	55.97	0.27	208.67	353.31	123.22
	MP and IE	Composite	15.98	10.35	24.68	0.22	431.57	666.33	279.44
		Halfnormal/Cosine	15.98	10.35	24.68	0.22	431.57	666.33	279.44
		Hazardrate/Polynomial	16.82	10.62	26.65	0.24	410.02	649.39	258.78
		Uniform/Cosine	16.40	10.58	25.44	0.22	420.52	651.85	271.09
13a	MP	Composite	268.81	5.19	13926.54	0.40	25.66	1328.82	0.50
		Halfnormal/Cosine	268.81	5.19	13926.54	0.40	25.66	1328.82	0.50
		Hazardrate/Polynomial	224.99	6.58	7687.82	0.41	30.65	1048.11	0.90
		Uniform/Cosine	268.81	5.19	13926.54	0.40	25.66	1328.82	0.50
	IE	Composite	399.24	117.24	1359.52	0.23	17.27	58.82	5.07
		Halfnormal/cosine	399.24	117.24	1359.52	0.23	17.27	58.82	5.07
		Halfnormal/hermite	399.24	117.24	1359.52	0.23	17.27	58.82	5.07
		Hazardrate/polynomial	560.00	253.48	1237.15	0.36	12.32	27.21	5.57
	MP and IE	Composite	328.05	175.01	614.90	0.23	21.02	39.41	11.22
		Halfnormal/cosine	328.05	175.01	614.90	0.23	21.02	39.41	11.22
		Hazardrate/polynomial	453.84	238.73	862.81	0.30	15.20	28.89	7.99
		Uniform/Cosine	333.94	176.14	633.10	0.22	20.65	39.15	10.89

Based on these density values and using the proportions of the population estimated to be adult MVA estimates are in the order of between 200 and 700 ha for *O. pugnax* (at sites other than GD13a) based on data from both field workers, and between 450 and 2000 ha for *H. hamiltoni*.

For *Hadogenes gunningi*, the sample size collected during the survey was too small to allow for effective model fitting during data analysis due to difficulty in walking transects over the terrain where they occur and the time taken to investigate rock cracks. An attempt was made to collect sufficient data by night at a particular site taken to represent 'typical' habitat for this species in Gauteng Province (Kloofendal Nature Reserve, Roodepoort, Johannesburg). Night searching was done using an Ultra Violet light (8W fluorescent tube (Sylvania) run off a 12V, 8W inverter ballast and portable 12V 7AH lead acid battery). Six hours of collecting using this method revealed six specimens, which is comparable to the rate at which they were observed by sampling during the day. At this rate of encounter the additional time required to obtain a useable sample size for this species was not available, and hence density could not be estimated empirically.

For the purposes of the report for GDACE only a subjective estimate of population density based on encounter rate relative to that of the other two species can be offered for *Hadogenes gunningi*. It is speculated that the population density of this species on quartzite ridges of the Witwatersrand and Pretoria formations, which make up the majority of suitable habitat for this species in the study region, is slightly higher than that obtained for *Harpactira hamiltoni*, being in the region of eight to 12 individuals per hectare. Taking the proportion of adults to juveniles into consideration this yields an MVA estimate of between 380 and 570 ha. This estimate applies only to quartzite ridges, as densities on other geologies might differ based on how well they weather to provide suitable home sites for this species. One particular site, GD18, had a much higher rate of encounter of individuals of this species in comparison to any other sites visited, implying a higher population density. While this ridge is also quartzite, the inclination of the strata is such that weathering produces many more suitable home sites than on other quartzite ridges, which is most likely the reason for the higher number of specimens observed.

Table 3.3: Population Density and MVA estimates for *Opistophthalmus pugnax* using interval data. Variance estimates are by bootstrap resampling of line transects. Where the Model is 'Composite', this refers to the AIC weighted average method of Burnham and Anderson (2002). See Appendix I for the models included for this. Minimum viable area (MVA) estimates are in hectares. LCL is the lower confidence level, UCL is the upper confidence level and CV is the coefficient of variation.

Site	Collector	Model/adjustment term	Mean Density	Density LCL	Density UCL	Density CV	Mean MVA	MVA LCL	MVA UCL
Ex 13a	MP	Composite	7.70	3.60	16.45	0.39	895.66	1915.71	419.24
		Halfnormal/cosine	7.70	3.60	16.45	0.39	895.66	1915.71	419.24
		Uniform/Cosine	7.70	3.60	16.45	0.39	895.66	1915.71	419.24
	IE	Composite	41.05	14.11	119.38	0.58	168.00	488.77	57.77
		Halfnormal/cosine	29.46	17.62	49.25	0.26	234.10	391.40	140.03
		Halfnormal/hermite	21.80	13.28	35.79	0.25	316.36	519.32	192.69
		Uniform/polynomial	25.72	15.40	42.96	0.26	268.14	447.83	160.53
	MP and IE	Composite	18.06	11.19	29.14	0.25	381.87	616.31	236.67
		Halfnormal/cosine	15.13	9.85	23.25	0.22	455.82	700.16	296.63
		Hazardrate/cosine	18.06	11.19	29.14	0.25	381.87	616.31	236.67
13a	MP	Composite	280.36	4.55	17264.03	0.40	24.60	1515.73	0.40
		Halfnormal/cosine	284.11	9.72	8306.84	0.41	24.27	709.52	0.83
		Uniform/Cosine	280.36	4.55	17264.03	0.40	24.60	1515.73	0.40
	IE	Composite	368.08	48.18	2812.21	0.21	18.74	143.14	2.45
		Hazardrate/polynomial	585.44	258.33	1326.75	0.39	11.78	26.70	5.20
		Uniform/polynomial	368.08	48.18	2812.21	0.21	18.74	143.14	2.45
	MP and IE	Composite	341.58	182.18	640.45	0.23	20.19	37.86	10.77
		Halfnormal/cosine	341.58	182.18	640.45	0.23	20.19	37.86	10.77
		Halfnormal/hermite	341.58	182.18	640.45	0.23	20.19	37.86	10.77

Table 3.4: Population Density and MVA estimates for *Harpactira hamiltoni* using continuous data. Variance estimates are by bootstrap resampling of line transects. Where the Model is 'Composite', this refers to the AIC weighted average method of Burnham and Anderson (2002). See Appendix I for the models included for this. Minimum viable area (MVA) estimates are in hectares. LCL is the lower confidence level, UCL is the upper confidence level and CV is the coefficient of variation.

Collector	Model and Adjustment term	Mean Density	Density LCL	Density UCL	Density CV	Mean MVA	MVA LCL	MVA UCL
MP	Composite	3.99	2.03	7.84	0.35	911.37	1791.31	463.82
	Halfnormal/cosine	4.26	2.37	7.66	0.30	853.61	1534.33	474.72
	Uniform/cosine	3.59	2.07	6.21	0.28	1012.91	1756.70	585.57
IE	Composite	3.16	1.66	6.03	0.33	1150.75	2190.58	603.05
	Halfnormal/cosine	3.97	1.72	9.20	0.44	915.96	2114.16	395.26
	Uniform/cosine	3.16	1.66	6.03	0.33	1150.75	2190.58	603.05
MP and IE	Composite	4.00	2.28	7.02	0.29	909.09	1594.90	518.00
	Halfnormal/cosine	4.73	2.85	7.87	0.26	768.79	1275.92	462.05
	Hazardrate/cosine	4.00	2.28	7.02	0.29	909.09	1594.90	518.00
	Uniform/cosine	3.81	2.30	6.29	0.26	954.43	1581.03	578.12

Table 3.5: Population Density and MVA estimates for *Harpactira hamiltoni* using interval data. Variance estimates are by bootstrap resampling of line transects. Where the Model is 'Composite', this refers to the AIC weighted average method of Burnham and Anderson (2002). See Appendix I for the models included for this. Minimum viable area (MVA) estimates are in hectares. LCL is the lower confidence level, UCL is the upper confidence level and CV is the coefficient of variation.

Collector	Model and Adjustment term	Mean Density	Density LCL	Density UCL	Density CV	Mean MVA	MVA LCL	MVA UCL
MP	Composite	3.60	2.08	6.24	0.28	1010.10	1748.25	582.75
	Halfnormal/cosine	3.77	2.09	6.81	0.30	964.55	1739.89	533.97
	Uniform/cosine	3.60	2.08	6.24	0.28	1010.10	1748.25	582.75
IE	Composite	3.89	2.30	6.56	0.27	934.80	1581.03	554.32
	Halfnormal/cosine	3.89	1.86	8.12	0.38	934.80	1955.03	447.83
	Hazardrate/cosine	3.89	2.20	6.86	0.29	934.80	1652.89	530.08
	Uniform/cosine	3.89	2.30	6.56	0.27	934.80	1581.03	554.32
MP and IE	Composite	4.00	2.28	7.02	0.29	909.09	1594.90	518.00
	Halfnormal/cosine	4.59	2.75	7.68	0.26	792.24	1322.31	473.48
	Hazardrate/cosine	4.00	2.28	7.02	0.29	909.09	1594.90	518.00
	Uniform/cosine	3.88	2.41	6.26	0.24	937.21	1508.86	580.89

3.3 Habitat Distribution Modeling

The method used here for identifying best subsets from each GARP run for each species resulted in 286, 262 and 105 coarse grain and 377, 124 and 291 fine grain models for *Hadogenes gunningi, Opistophthalmus pugnax and Harpactira hamiltoni* respectively. Overlays of these coarse and fine grain models, as well as the composite models are presented for each species in figures 3.5 to 3.13. Presence data used in developing the GARP models are also presented in these figures The corresponding ROC plot for each prediction is presented with the figures and the AUC's presented in the figure legends. Figures 3.14 to 3.16 present presence absence maps for each of the species given after conversion using the chosen cutoff probability. Absence records which were used in the ROC analysis are also presented in these figures, while the presence records were the same

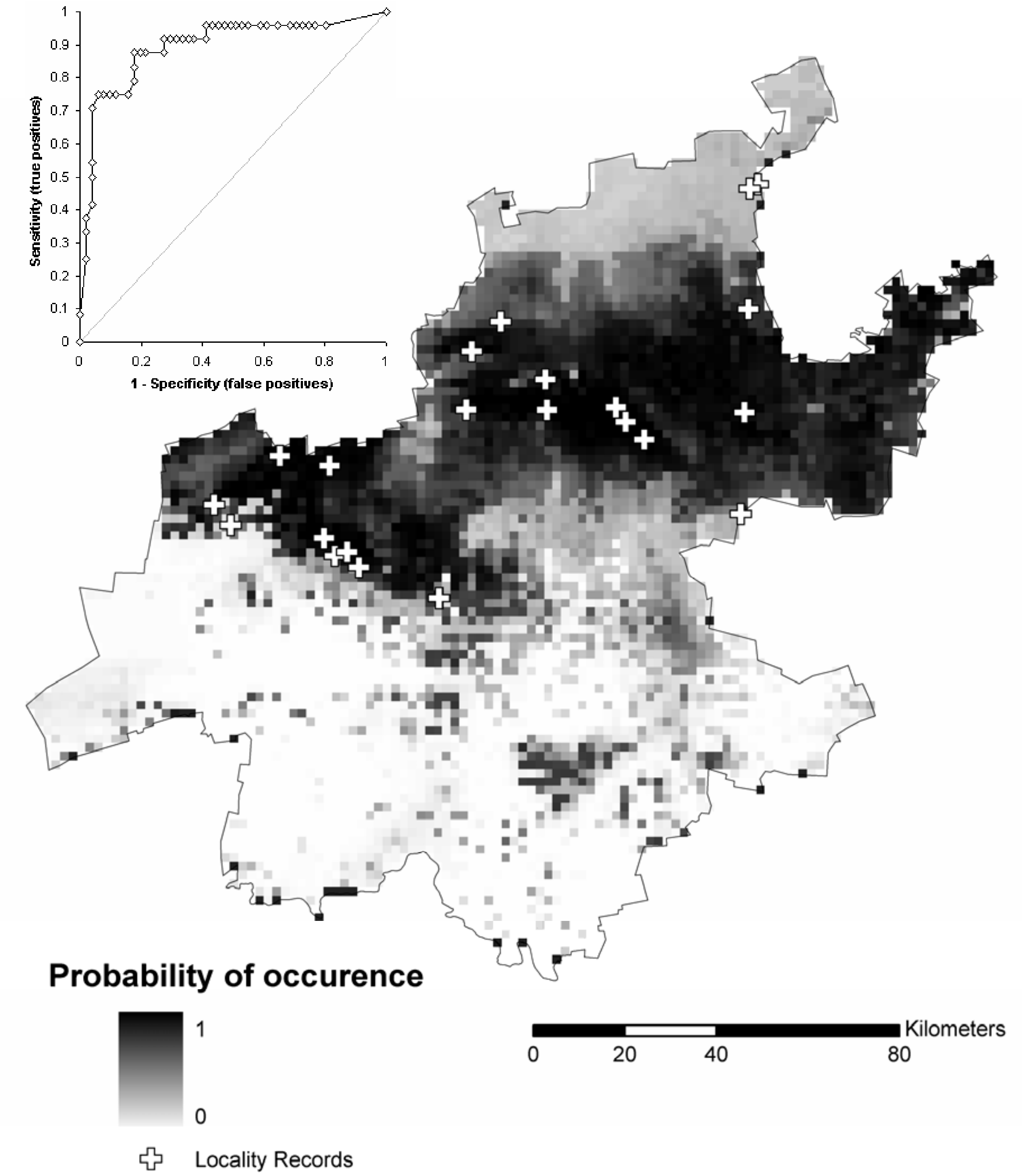

Figure 3.5: Coarse grain prediction of probability of occurrence of *Hadogenes gunningi* from overlays of GARP models based on climatic predictor variables only. Locality records used to develop the GARP models are indicated. The receiver operating characteristic (ROC) plot for this prediction is inset top left. The area under the curve (AUC) for the ROC plot is 0.893 ± 0.0447 (mean ± SE).

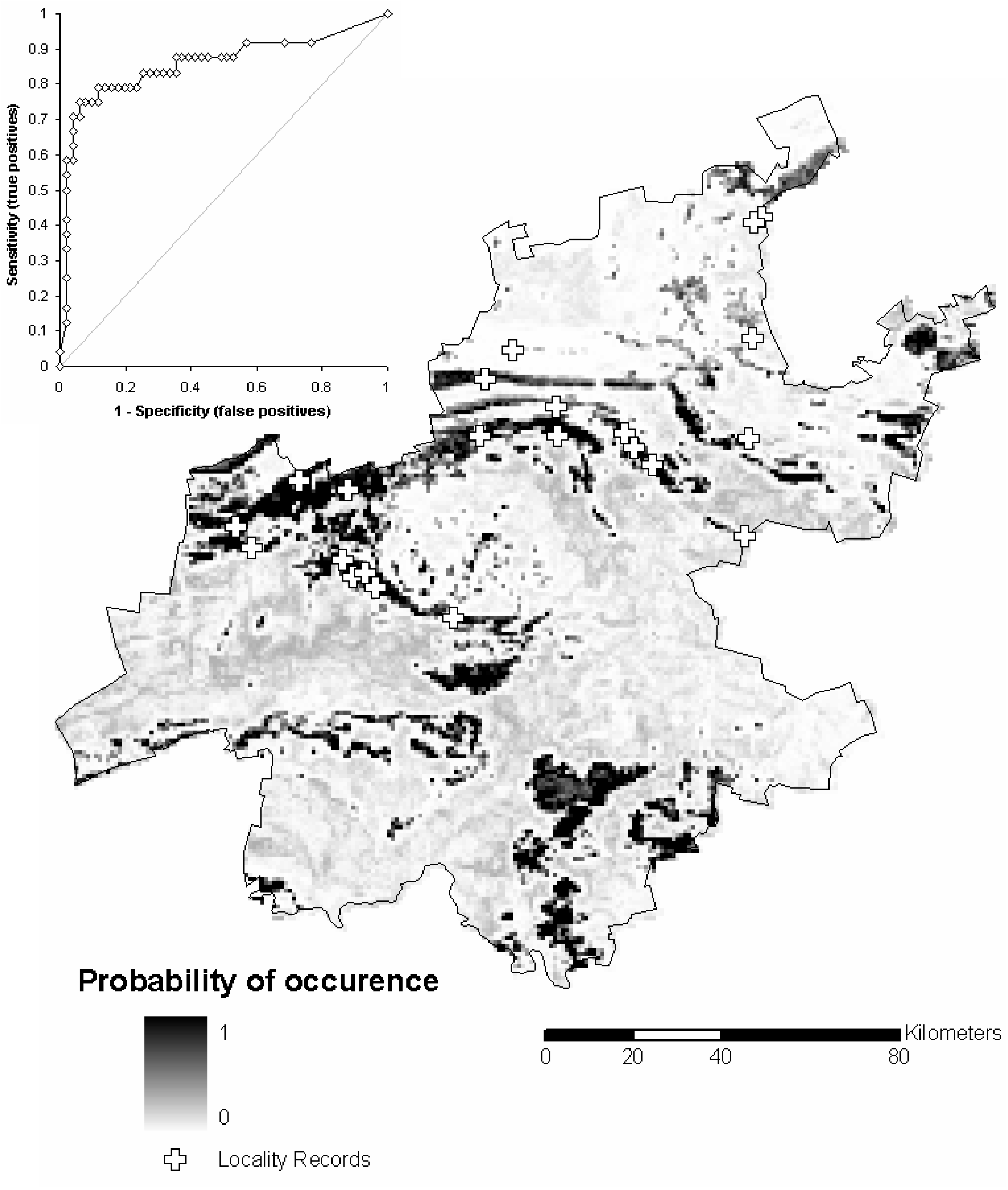

Figure 3.6: Fine grain prediction of probability of occurrence of *Hadogenes gunningi* from overlays of GARP models based on topographic, vegetation, landcover and substrate predictor variables. Locality records used to develop the GARP models are indicated. The receiver operating characteristic (ROC) plot for this prediction is inset top left. The area under the curve (AUC) for the ROC plot is 0.857± 0.0561 (mean ± SE).

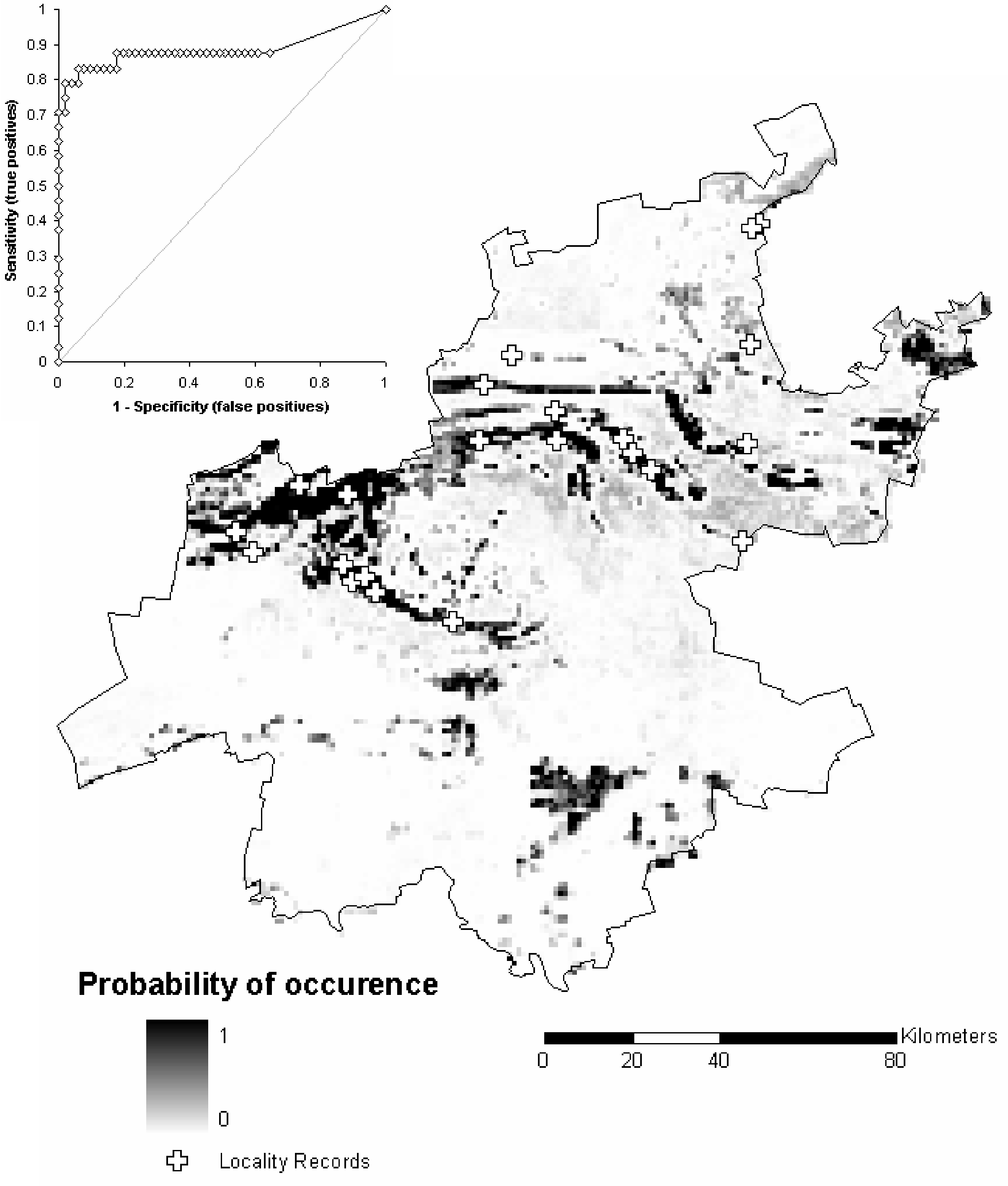

Figure 3.7: Composite prediction of probability of occurrence of *Hadogenes gunningi* from the multiplicative overlay of coarse and fine grain predictions. Locality records used to develop the GARP models are indicated. The receiver operating characteristic (ROC) plot for this prediction is inset top left. The area under the curve (AUC) for the ROC plot is 0.886 ± 0.0564 (mean ± SE).

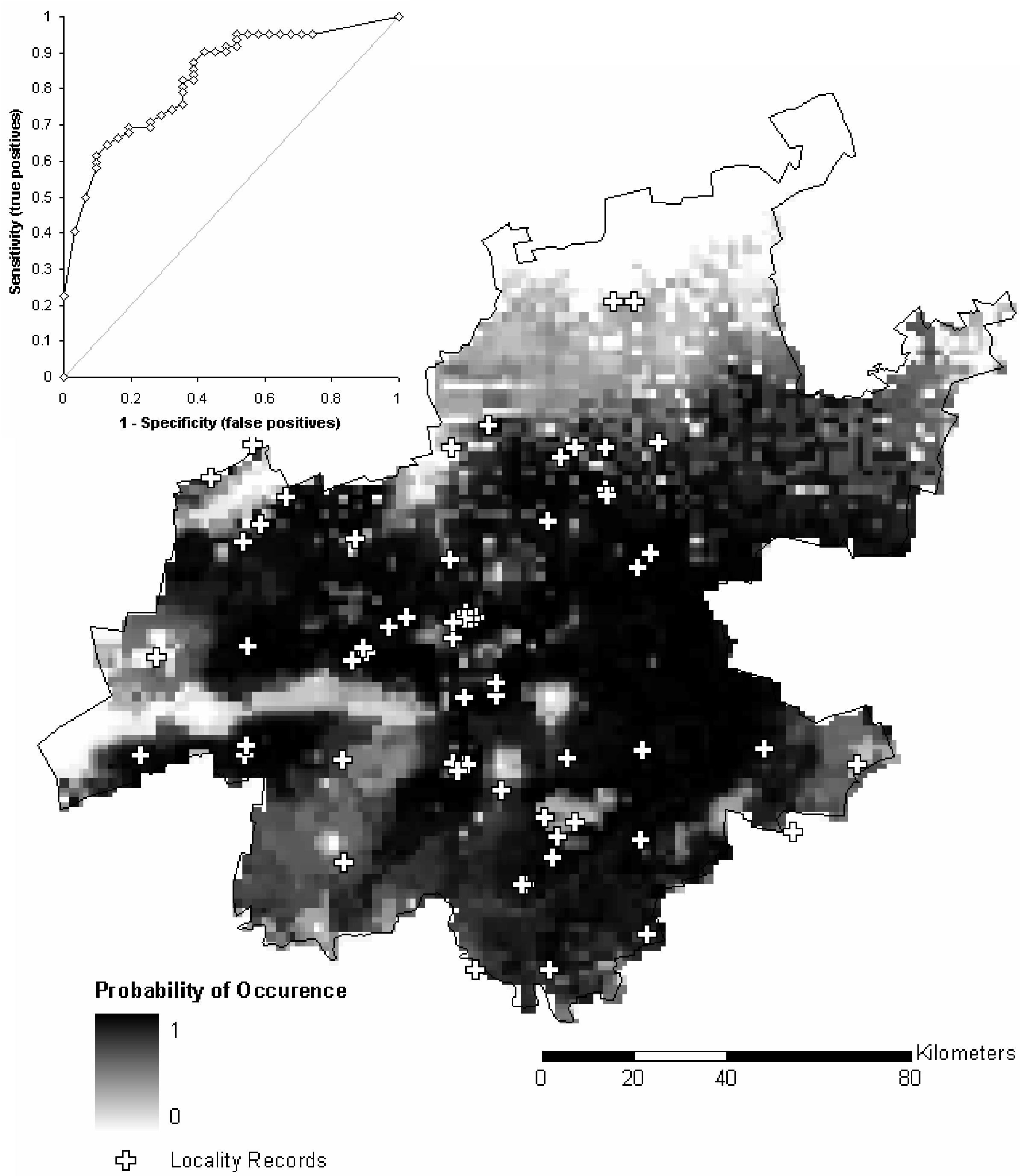

Figure 3.8: Coarse grain prediction of probability of occurrence of *Opistophthalmus pugnax* from overlays of GARP models based on climatic predictor variables only. Locality records used to develop the GARP models are indicated. The receiver operating characteristic (ROC) plot for this prediction is inset top left. The area under the curve (AUC) for the ROC plot is 0.831 ± 0.0428 (mean ± SE).

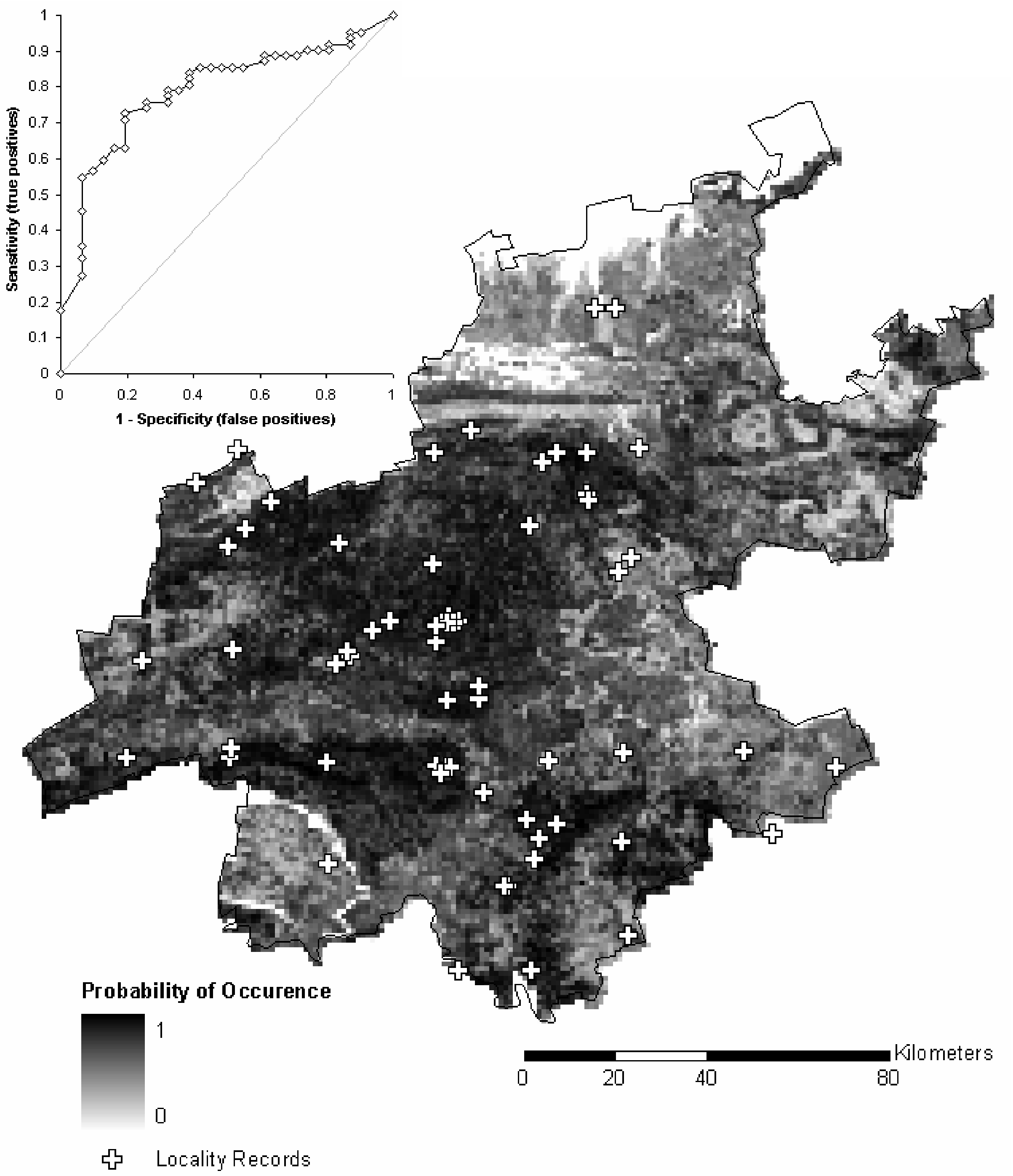

Figure 3.9: Fine grain prediction of probability of occurrence of *Opistophthalmus pugnax* from overlays of GARP models based on topographic, vegetation, landcover and substrate predictor variables. Locality records used to develop the GARP models are indicated. The receiver operating characteristic (ROC) plot for this prediction is inset top left. The area under the curve (AUC) for the ROC plot is 0.790 ± 0.0485 (mean ± SE).

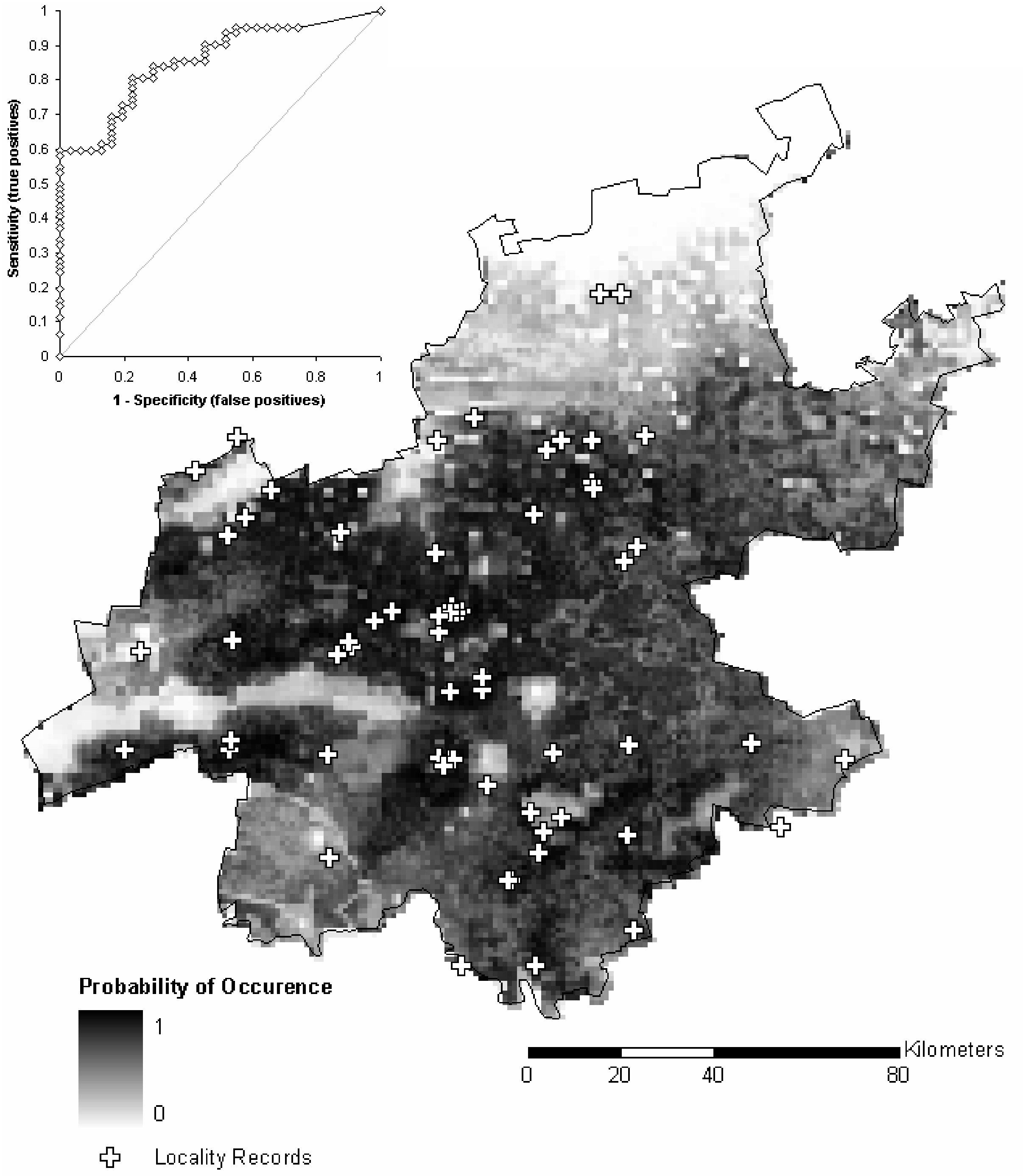

Figure 3.10: Composite prediction of probability of occurrence of *Opistophthalmus pugnax* from the multiplicative overlay of coarse and fine grain predictions. Locality records used to develop the GARP models are indicated. The receiver operating characteristic (ROC) plot for this prediction is inset top left. The area under the curve (AUC) for the ROC plot is 0.856 ± 0.0380 (mean ± SE).

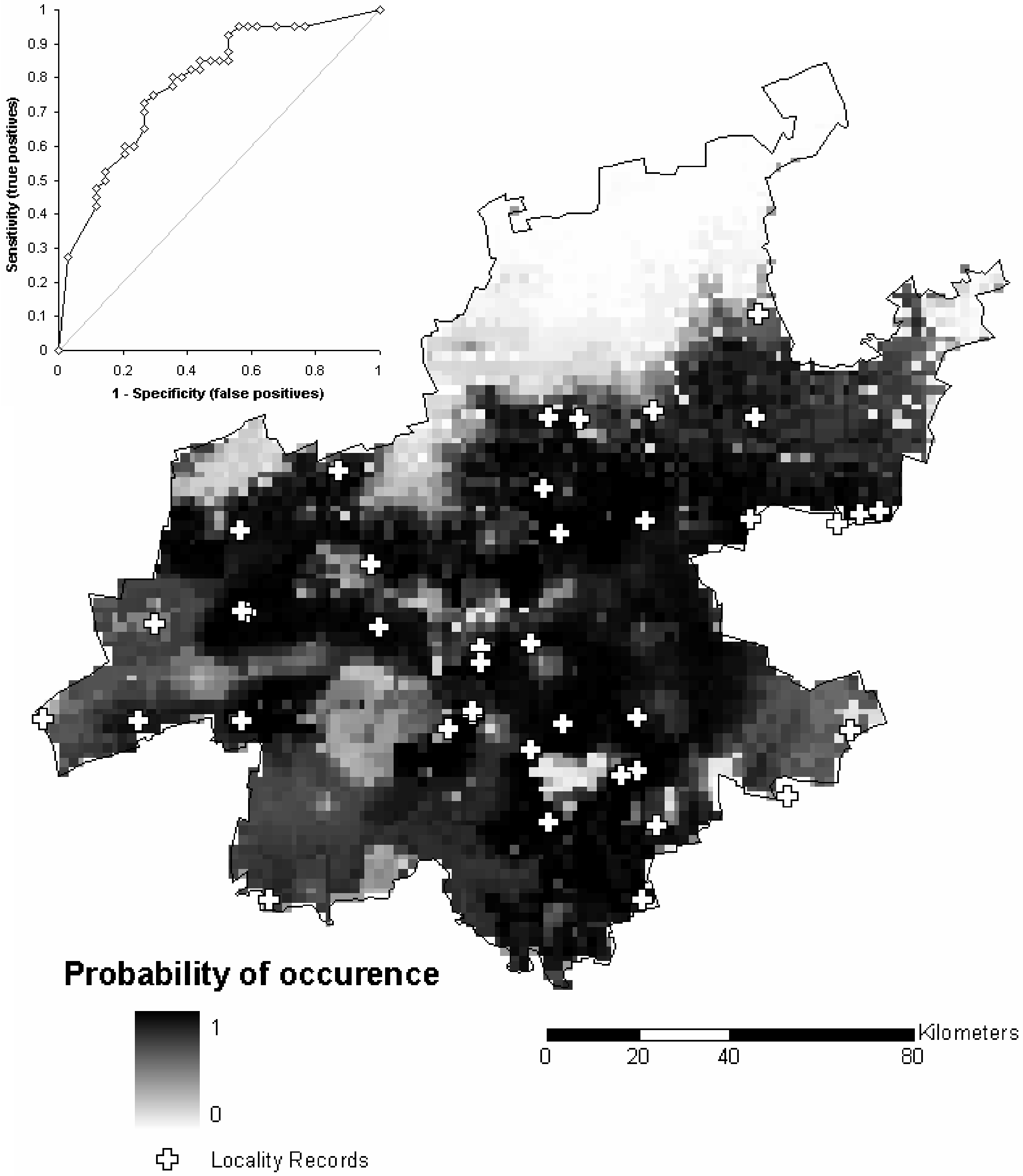

Figure 3.11: Coarse grain prediction of probability of occurrence of *Harpactira hamiltoni* from overlays of GARP models based on climatic predictor variables only. Locality records used to develop the GARP models are indicated. The receiver operating characteristic (ROC) plot for this prediction is inset top left. The area under the curve (AUC) for the ROC plot is 0.783 ± 0.0535 (mean ± SE).

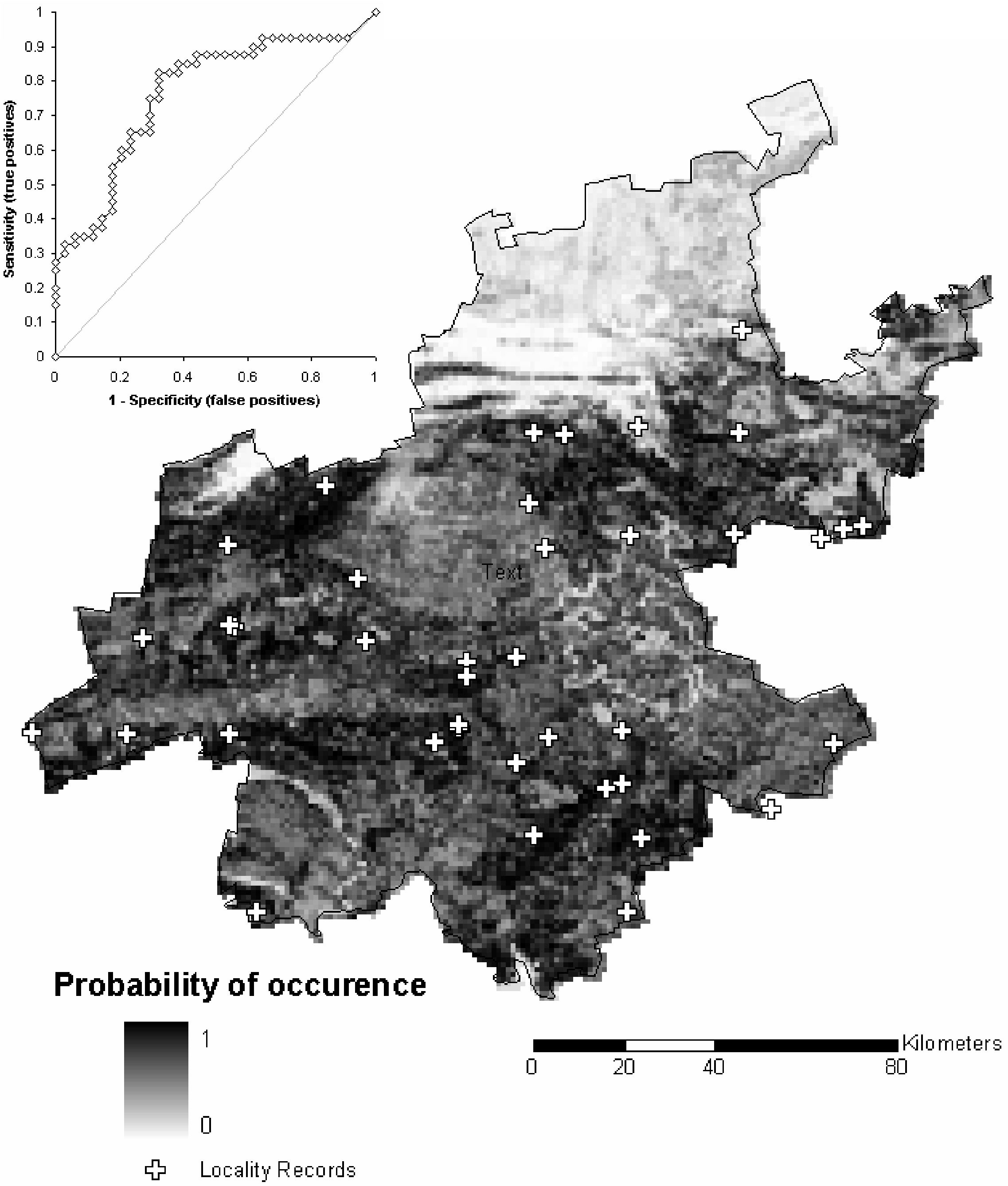

Figure 3.12: Fine grain prediction of probability of occurrence of *Harpactira hamiltoni* from overlays of GARP models based on topographic, vegetation, landcover and substrate predictor variables. Locality records used to develop the GARP models are indicated. The receiver operating characteristic (ROC) plot for this prediction is inset top left. The area under the curve (AUC) for the ROC plot is 0.765 ± 0.0561 (mean ± SE).

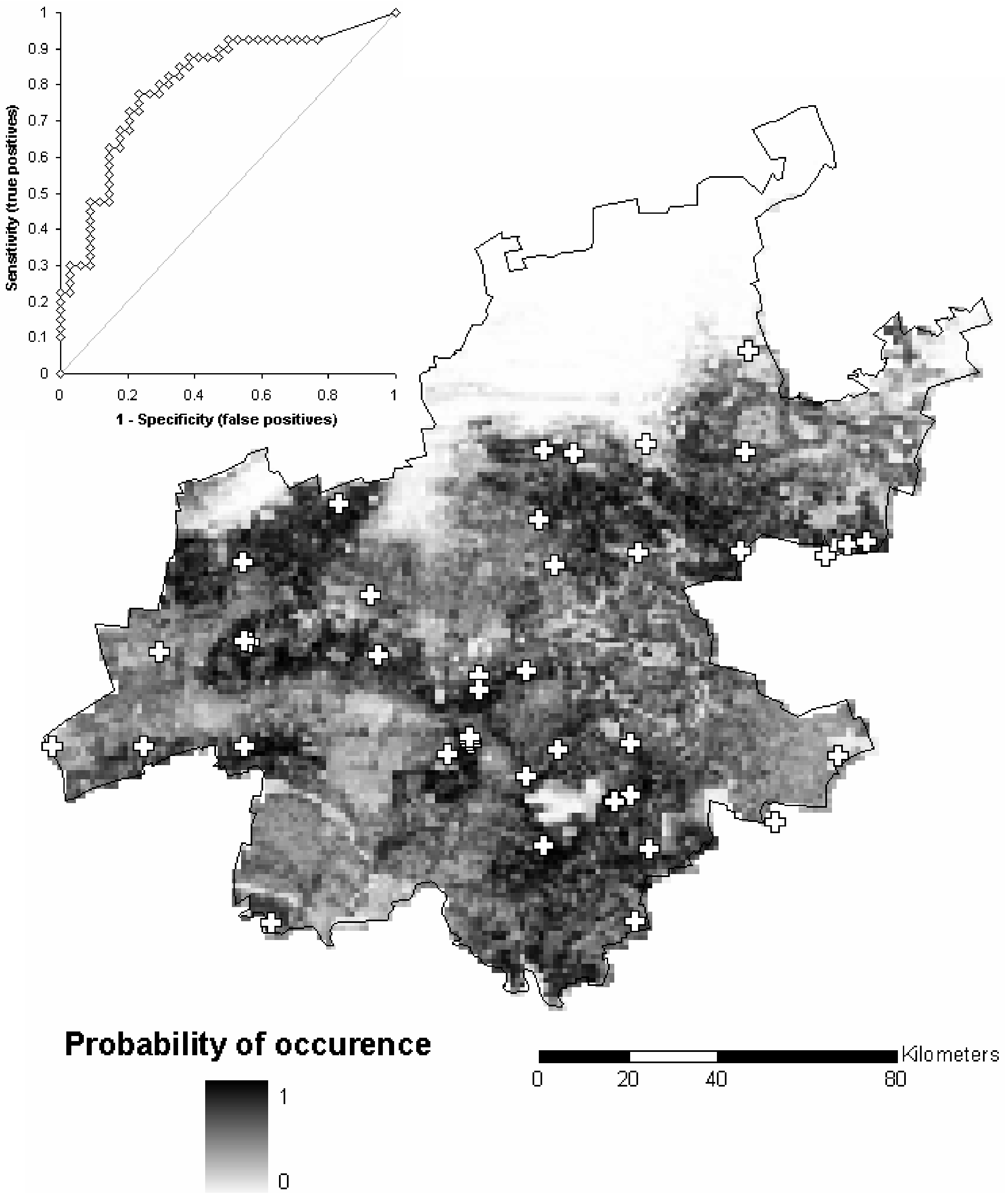

Figure 3.13: Composite prediction of probability of occurrence of *Harpactira hamiltoni* from the multiplicative overlay of coarse and fine grain predictions. Locality records used to develop the GARP models are indicated. The receiver operating characteristic (ROC) plot for this prediction is inset top left. The area under the curve (AUC) for the ROC plot is 0.805 ± 0.0520 (mean ± SE).

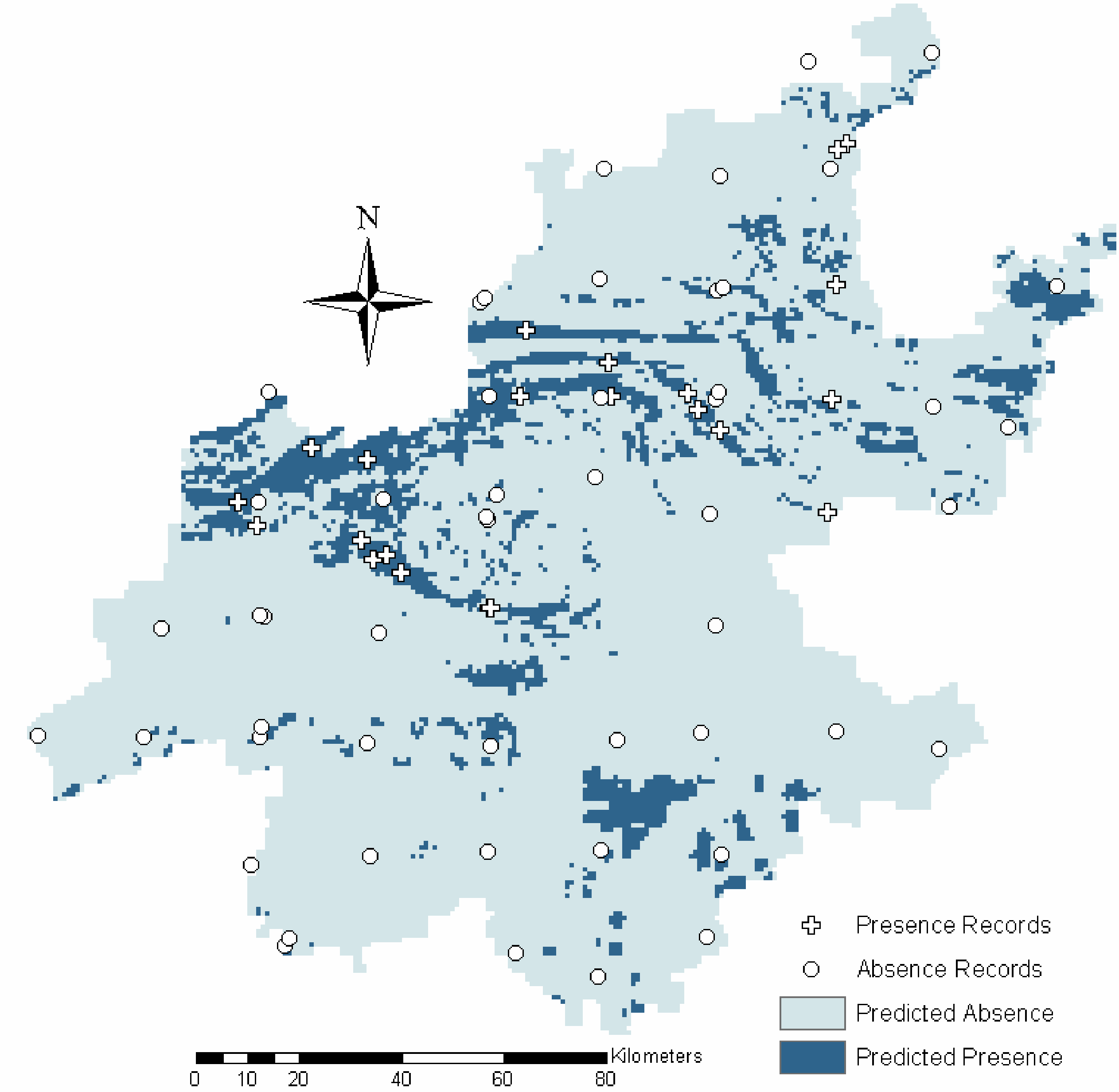

Figure 3.14: Predicted presence and absence for *Hadogenes gunningi* in Gauteng province. The cutoff level used to convert probabilities of occurrence to presence/absence was that where the false positive and false negative prediction rates were equal (~0.1496).

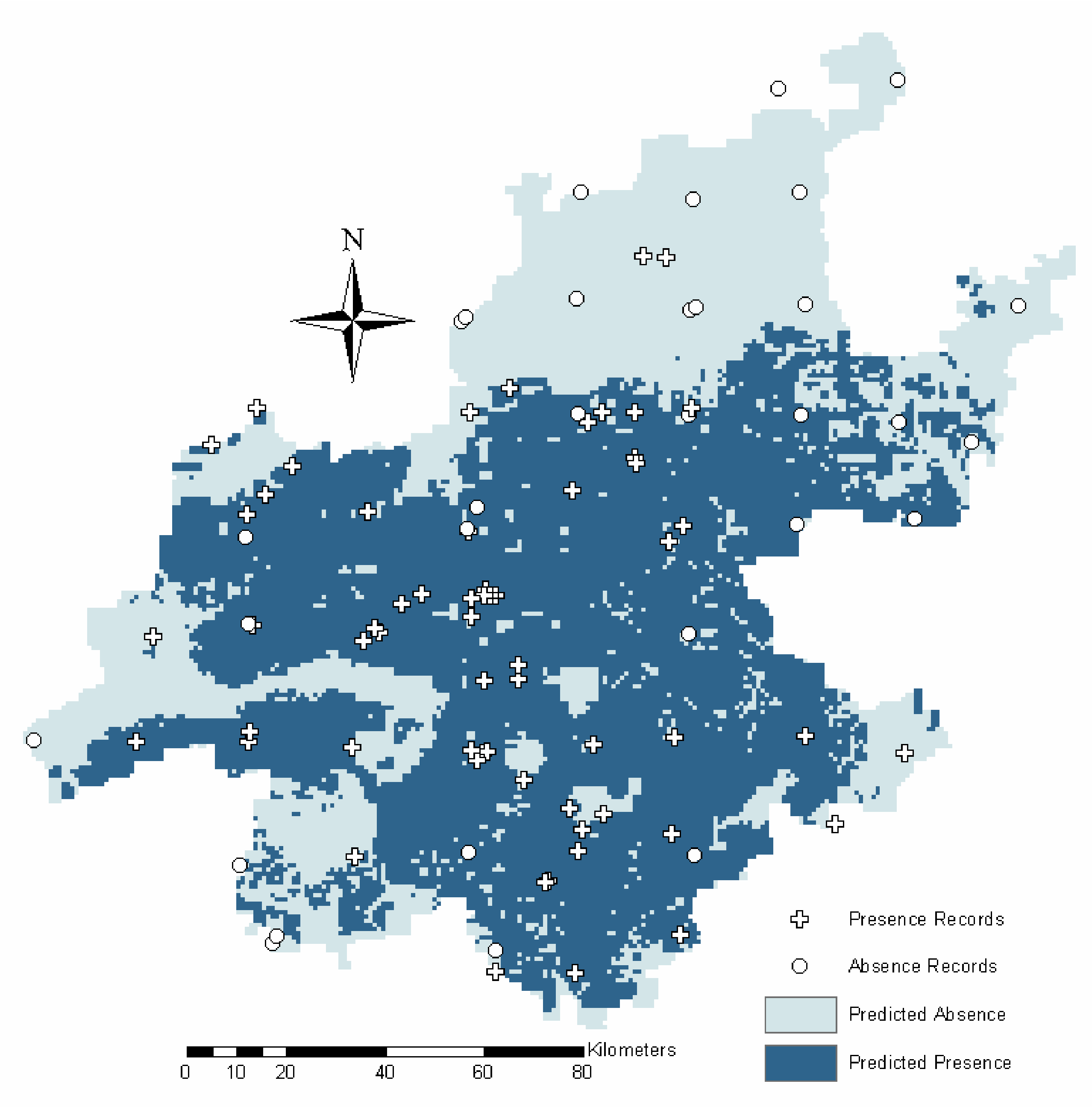

Figure 3.15: Predicted presence and absence for *Opistophthalmus pugnax* in Gauteng province. The cutoff level used to convert probabilities of occurrence to presence/absence was that where the false positive and false negative prediction rates were equal (~0.5607).

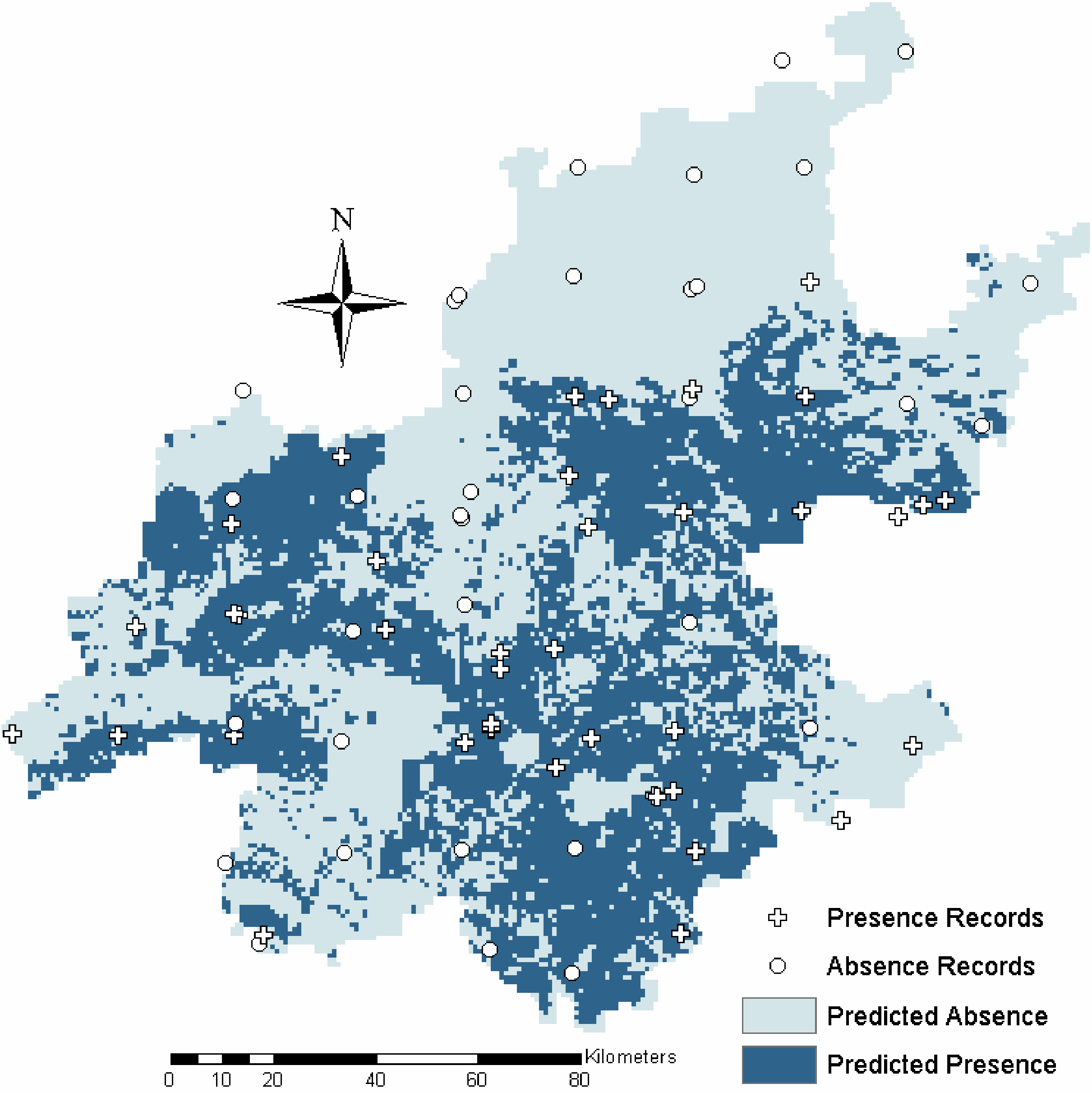

Figure 3.16: Predicted presence and absence for *Harpactira hamiltoni* in Gauteng province. The cutoff level used to convert probabilities of occurrence to presence/absence was that where the false positive and false negative prediction rates were equal (~0.6249).

as those used to develop the GARP models. The 95% confidence intervals for the estimate of AUC for each species are presented in Table 3.6.

Table 3.6: Means and 95% confidence intervals (in parenthesis) for estimates of area under the curve (AUC) of the receiver operating characteristic (ROC) plots based on coarse and fine grain and composite probability of occurrence models for *Hadogenes gunningi*, *Opistophthalmus pugnax* and *Harpactira hamiltoni*.

	Coarse Grain prediction	Fine Grain Prediction	Composite Prediction
Hadogenes gunningi	0.893 (0.806 - 0.981)	0.857 (0.747 - 0.966)	0.886 (0.775 - 0.996)
Opistophthalmus pugnax	0.831 (0.747 - 0.915)	0.790 (0.694 - 0.885)	0.856 (0.781 - 0.930)
Harpactira hamiltoni	0.783 (0.679 - 0.888)	0.765 (0.655 - 0.875)	0.805 (0.703 - 0.907)

The AUC values for all models for all species show that the models are better than a random prediction. These values indicate that the models for *Hadogenes gunningi* are consistently better than for the other two species, and that those for *Opistophthalmus pugnax* are better than those for *Harpactira hamiltoni*. The composite models are also better than the coarse or fine grain models for *Opistophthalmus pugnax* and *Harpactira hamiltoni*, while the coarse grain prediction is the best for *Hadogenes gunningi.* However, based on the confidence intervals presented for each of these AUC values it would appear that none of these differences are significantly different. Based on the ranking criteria of Pearce and Ferrier (2000) examination of the 95% confidence intervals the models for *Hadogenes gunningi* are ranked as good to very good, and those for *Opistophthalmus pugnax* and *Harpactira hamiltoni* as average to good.

The coarse and fine grain predictions for *O. pugnax* are fairly congruent with lower probabilities of occurrence in the north of the province. A similar pattern is evident for *Harpactira hamiltoni*. The pattern for *Hadogenes gunningi* differs in that the coarse grain predictions indicate an east west band of suitable climatic conditions across the northern half of the province, while the fine grain predictions identify all ridges as having a higher probability of occurrence than non ridge areas.

The presence absence map for *Hadogenes gunningi* (Figure 3.14) shows this species to be limited to ridges across Gauteng province. That for *Opistophthalmus pugnax* (Figure 3.15) would indicate that this species is widely distributed across the southern two thirds of the province but is notably excluded from the eastern margins of this area as well as from a broad region in the west. The map for *Harpactira hamiltoni* (Figure 3.16) shows it to be excluded from the northern third of the province, and having a very patchy distribution in the southern two thirds. Interestingly the latter two species are both predicted not to occur on the top of the Suikerbosrand ridge near Heidelberg. This the highest elevation area in Gauteng province, and based on the overlay models for climate variables only it would appear that the former two species are excluded due to unfavorable climate conditions.

4. Discussion

The results of this study provide a significant improvement on the information that was previously available for the purpose of conservation assessments by providing empirical estimates of minimum viable areas and developing a better understanding of the distributions of the three arachnid species of interest in Gauteng province. Primarily, it is suggested that the threat status of each species be reassessed in light of these results, and careful consideration be given to the means of their further inclusion in the GDACE Biodiversity Gap Analysis project. The results pertaining to each of the areas of sampling design, MVA and habitat distribution are discussed below, and the report is concluded with suggestions for future research directions.

4.1 Survey Design

The site selection method used for this study proved effective for collecting additional locality records for the species of interest. It simultaneously allowed for additional records to be collected for other species as well, which may provide additional useful information to GDACE. While this sampling strategy provided useful biodiversity survey data its primary function was to provide data that could be used in developing accurate and general habitat distribution models. Thus it is required that such a sampling strategy adequately sample the space along environmental gradients that are known or thought to be important in limiting a species distribution (Vaughan and Ormerod 2003). Hirzel and Guisan (2002) proposed that a regular grid of sample points, as the one used here, provided data that resulted in habitat models with comparable accuracy to those developed using data from a sample design using environmental stratification to select sites. However, from this study it was obvious that the grid sampling method was inadequate in that certain important environmental gradients were not adequately sampled. For example several categories within the geology layer, assumed to be important in limiting the distribution of *Hadogenes gunningi*, were not sampled at all. It is likely that the sampling completeness resulting from the Hirzel and Guisan (2002) gird based method is dependent on sample size i.e increasing sample size is required to adequately sample gradients as environmental heterogeneity increases. The sampling adequacy of this field survey could potentially have been improved if sampling sites had been specifically chosen to maximize environmental coverage given the constraints of time and funding available for the purpose.

On the other hand the grid-based sampling did provide additional records for the species of interest well beyond those previously collected by GDACE (Figures 3.2 - 3.4), which has important implications for the calculation of extents of occurrence for threat assessments. This is particularly evident for *Opistophthalmus pugnax* and *Harpactira hamiltoni* where new localities collected during this survey extended the known distributions of these species westwards of all previous records. Additionally, the number of records for *H. hamiltoni* was greater than those previously collected by

GDACE. This species is possibly the most elusive of the three investigated here due to its low population densities and its tendency to camouflage its burrow entrance with debris making it hard to find. The use of experts in surveying such elusive species is thus justified as evidenced by these new additional records.

4.2 Population Density and MVA

Population densities for scorpions have been shown to be dependent on environmental variables such as substrate (Brown *et al.* 2002, Bradley 1986, Bradley and Brody 1984) with densities generally higher in sandy deserts. Densities have been measured at eight to 12 individuals per square meter in the littoral zone dwelling *Vaejovis littoralis* in Baja California (Due and Polis 1985) to 0.0002 individuals per square meter for *Tityus fasciolatus*, living in termite mounds in Brazil (Lourenço 1978). Polis and McCormick (1986) showed densities of *Smeringurus mesaensis*, a sandy desert living species, to be between 0.15 to 0.40 individuals per square meter. In southern Africa, Harington (1978) has reported the density of *Cheloctonus jonesii* at 0.67 individuals per square meter and Lamoral (1978) reported densities of *Opistophthalmus carinatus* and *O. wahlbergii* at 0.07 individuals per square meter in the Kalahari. Engelbrecht (2002) reported a density of 0.16 to 0.26 individuals per square meter for *O. pictus* near Kimberley in South Africa.

When converted to individuals per square meter for comparison, the density estimates presented here for *O. pugnax* based on data by both field workers ranges from 0.0011 to 0.0025 individuals per square meter. The density at site GD13a ranges from 0.0191 and 0.0681 individuals per square meter. While these densities are lower than those recorded for other *Opistophthalmus* species in southern Africa this is likely due to the estimation of density at the landscape scale using distance sampling, compared to the results of the quadrat based methods usually used in scorpion density estimation (Lamoral 1978, Polis and Farley 1980). Many species of *Opistophthalmus*, especially those occurring on the drier western side of southern Africa, show an aggregated dispersion pattern, where individuals form ‘clumps’ within large areas of apparently suitable habitat (L. Prendini pers comm, pers. obs). No dispersal data exists for any *Opistophthalmus*, but such an aggregated spatial dispersion pattern may be explained by low dispersal rates (for a mathematical explanation of how this pattern may emerge see Hanski, 1999, p 77 - 79). From observations in the field it would appear that *O. pugnax* shows some degree of aggregated dispersion, but the densities within these aggregations are averaged out over the landscape by the distance sampling method. Another explanation for the lower population densities in *O. pugnax* relative to other *Opistophthalmus* species may be higher dispersal rates, higher predation rates or a combination of the two. It has been suggested that prey availability may not be limiting on scorpion population size due to their unusually low metabolic rate enabling high food resource use efficiency (Lighton *et al.* 2001), but this remains to be tested in Southern African examples. No explanation for the significantly higher population density estimated for site GD13a relative to other sites in Gauteng can be offered at this stage.

Population densities for other theraphosid spiders have not been measured explicitly. Yanez and Floater (2000) examined population size of the tarantula *Brachypelma klaasi* in Mexico, but don't present any estimate of density, other than stating that burrows are between 3 - 257m apart, with a mean of 103m. Field observations indicate that population density of *Harpactira* species is always relatively low (M. Paulsen pers comm., R. Gallon, pers comm., pers obs.)

Minimum viable area estimates presented here for the three study species are higher than those previously used by GDACE and are generally in the order of several hundred hectares. In this instance it is assumed that 2000 adult individuals constitute a viable population with a high probability of persistence. This number is lower than numbers in the 5000 to 7000 region suggested by Reed *et al.* (2003). However, as the purpose of this project is supply information to GDACE that can be used in their conservation planning process MVA may be recalculated from the density estimates using larger, more conservative MVP sizes if needs be. Importantly it must be remembered that a single magic number for MVP (and hence MVA) cannot be substantiated (Reed *et al.* 2003) and that these estimates must be interpreted as general. Another consideration is that the MVP sizes suggested by this author are based on studies of a range of vertebrates. How these estimates apply to invertebrate populations is not known, but given the more K-selected life history strategies of these three species it is likely that such estimates for vertebrates are applicable in this instance.

The domain of application of the estimated MVA for each species needs to be carefully considered by the relevant conservation practitioners. The original MVA estimates used in the conservation planning process were not based on any hard empirical data. The estimates presented here only offer an improvement on those used previously and represent the best available information that could be provided given time and budget constraints. The generic MVP estimates presented by Reed *et al.* (2003) do not consider threats to individual populations or processes occurring within those populations which are central to a comprehensive population viability analysis, and these should be considered explicitly should it be determined necessary to develop species specific management plans for any of these arachnid species.

When Allen *et al.* (2001) investigated implications of including estimates of MVA in the selection of a network of areas for the conservation of mammals in Florida it was found that only a small percentage of areas containing suitable habitat were large enough to sustain a viable population as a result of habitat fragmentation. High levels of habitat fragmentation in Gauteng will lead to the same problem for identifying suitable areas for conservation for these arachnid species. The possibility that such habitat fragmentation may lead to a metapopulation structure for these species needs to be considered when selecting these areas. How such fragmentation will affect population viability, and MVA, depends on the degree of fragmentation (inter-patch distance), on the dispersal capacity of the species and on autocorrelation amongst processes that influence rates of extinction in subpopulations. For example, important population processes such as source - sink dynamics should be considered to allow for the selection of areas that will most likely ensure persistence. This is an area which would benefit greatly from further research.

4.3 Habitat Distribution Modeling

GARP has been applied and shown to perform well in a number of applications in the literature, from investigating ecological divergence in closely related species (Anderson *et al.* 2002a) to offering an alternative method for modeling species distributions and identifying priority hotspots in conservation programs (Peterson and Kluza 2003, Stockwell and Peterson 2003). Soberon and Peterson (2005) discuss how correlative ecological niche based modeling is best suited to modeling the broad scale distribution (termed the geographic distribution) of organisms, validating the use of predictor variables in the modeling exercise thought to limit distribution at this scale. Hence GARP has most usually been applied with heavy emphasis on climatic variables as predictors. The resulting broad scale outputs have been refined to smaller scales by ‘clipping’ of the outputs with landcover types deemed to be favorable to the species (Peterson and Kluza 2003, Anderson and Martinez-Meyer 2004).

Pearson and Dawson (2003) provide a schematic representation of how different environmental variables may affect the distribution of species at different scales, where climate is the most important broad scale predictor, and other factors, such as topography and substrate properties, limit distributions at finer scales. They specify the imperfection and oversimplification of this schema, but state that it may allow for postulate formulation on limiting factors on species distribution at different scales. Wiens (1989) describes how factors that are limiting at different scales may function in a hierarchical manner to determine species distributions, and that the scales at which they function may differ for different organisms. Blackburn and Gaston (2002), in a review of the issue of scale in macroecology cite several examples where interactions between processes occurring at different scales give rise to observed ecological patterns, such as species distribution or species richness, lending support for this method for developing composite habitat distribution models.

The method of using climatic variables for coarse grain modeling and topographic, vegetation, landcover and substrate variables for fine grain modeling corresponds to the scale dependency of these variables in limiting distribution as proposed by Pearson and Dawson (2003). Given that climatic conditions only vary significantly over much wider extents than do other variables it can be expected that these would limit distribution on a scale in the same order of magnitude as that extent over which they vary, and thus should limit species distribution at a broader scale than other environmental variables (Pearson and Dawson 2003). Within this broad scale region of suitable climate for a species, finer scale environmental variables, such as those relating to vegetation and substrate variables, will limit the distribution of a particular species at a finer grain than climatic variables, again as represented in the schema of Pearson and Dawson (2003). While the choice of variables used in coarse and fine grain modeling here was determined predominantly by constraints relating to the properties of the data layers for those variables they correspond well to the scales at which they may limit distributions of species in general. How such scale processes might affect the distributions of the three species modeled here can only be speculated, where climate probably sets the regional scale limits on their distributions, while the availability of suitable substrate conditions affects where they will occur within that suitable climatic region. The method presented here of combining coarse and fine grain predictions as the product of their probabilities of occurrence presents a novel approach to downscaling GARP outputs.

In evaluating the results of the GARP modeling process Holland (1975) describes how outputs from genetic algorithms should be expected to be highly variable, as was observed here as well as by other authors (Anderson *et al.* 2002a, Elith and Burgman 2002) and hence the method used in selecting best subsets from the models produced is important. The procedure used here offers an alternative to that suggested by Anderson *et al.* (2003) where models are selected on the criteria of keeping omission error and non-prediction low. This latter aspect is important as non-prediction was as high as 99% for some models. It also doesn't make the assumption that area of occupancy might be known with some degree of confidence by an expert beforehand, which is required to rank models using the method of Anderson *et al.* (2003).

The method used here for selecting a cutoff level for use in transforming probabilities of occurrence to presence absence simply assumes the costs of false presences and false absences to the be equal. Zweig and Campbell (1993) show how an appropriate cutoff level may be chosen when these costs are not equal. For example, should the cost of identifying an area as suitable habitat when it is unsuitable be high, then minimization of the false positive rate is important. This does however carry the consequence where some areas of suitable habitat will be excluded concurrently. The method requires that the relative costs of false positive and false negative predictions be estimated, and in conjunction with prevalence these can be use to generate a slope which is tangential to the ROC curve at the appropriate cutoff.

An interesting outcome of the cutoff selection method used was the very low cutoff value (0.1496) that yielded equal false positive and false negative rates for *Hadogenes gunningi*, in comparison to the values obtained for the other two species (0.5607 and 0.6249). This indicates a low level of over-prediction by the model with increasing cutoff levels, which may be attributed to this species stenotopic, or specialist, habitat requirements.

The presence absence models for *Opistophthalmus pugnax* (Figure 3.15) and *Harpactira hamiltoni* (Figure 3.16) indicate that these species are limited to the grassland biome areas of the southern two thirds of Gauteng. The areas where these species are predicted to be present appear quite patchy, which can be expected as a result of using fine scale predictor variables in modeling. The model for *O. pugnax* predicts its occurrence over large areas in the north eastern parts of the province around Bronkhorstspruit. A number of sites were surveyed in this area during fieldwork and it would appear that it is not present in that region. A closely related species, *O. glabrifrons* is widespread in those areas. It is unlikely that these two species exclude each other competitively as they have been found to co-occur at a number of localities (Lang pers comm., Hawkes, pers comm., pers obs.). It would appear that they are ecologically separated in that the former prefers harder, more clay soils and the latter softer, sandier soils. Substrate specialization is common in this genus (Prendini 2001, Lamoral 1979). The model also predicts its absence on the western side of the province in an east west band near Carltonville. It is also predicted as absent on the eastern side of the province near Devon. *H. hamiltoni* is predicted as absent from the same area near Devon, as well as from most of the highveld granite areas between Johannesburg and Pretoria. These species were both recorded in these areas during the survey and are likely more widely distributed within those than would be indicated by the models.

The presence/absence model for *Hadogenes gunningi* (Figure 3.14) shows this species to occur on most of the ridges within Gauteng. It has only been recorded to occur on the ridges of the Pretoria and Witwatersrand geological groups north of Johannesburg. The model predicts this species to occur on the ridges in far eastern part of the province where another species, *H. longimanus* is recorded. It is also predicted to occur on the Klipriviersberg ridge south of Johannesburg, as well as the Suikerbosrand ridge near Heidelberg. No locality records exist from these two ridges for the species despite previous survey work in this study and by GDACE. Importantly, these latter three ridges differ from the ridges where this species has been found to occur in geological composition. *H. gunningi* appears to be restricted to sedimentary rock types, while *H. longimanus* is recorded to occur predominantly on granite (Prendini 2001). The Klipriviersberg and Suikerbosrand ridges, where no *Hadogenes* are recorded, are composed of fine grained igneous rocks such as basalt and andesite. These issues of prediction of presence of *H. gunningi* where it has recorded as absent and absence for *O. pugnax* and *Harpactira hamiltoni* in areas where they have been recorded as present highlight the principle shortcoming of the GARP modeling method - interpretation of model predictions.

In the context of developing models for conservation planning a map of the potential distribution of the organism is the primary objective. However it is important that conservation planners have faith in the model used to develop that map. The choice predictor variables used in modeling will have a significant effect on the results obtained and requires careful consideration (Vaughan and Ormerod 2003, Pearce *et al.* 2001). Proper interpretation of model outputs requires that the investigator be able to determine how the model made a certain prediction and it is important to understand how different predictor variables may affect the model results. Using more conventional statistical methods it is possible to determine what contribution a certain predictor variable may make to a model output. In the case of the current GARP implementation this is not possible. Taking *Hadogenes gunningi* as an example, an explanation is required as to why the species was predicted on the Klipriviersberg and Suikerbosrand. From looking at the maps of the predictor variables used in the GARP analysis it appears that the majority of models that produced fine scale overlay predicted map for this species were based on a map of the ridges in the province. The ridges map was originally derived from slope. It is not possible to say whether the GARP models identify slope, or ridges as the important predictor variable. Secondly, based on the fact that this species is highly adapted to living in rock cracks it can be inferred that it would only be likely to occur on particular rock types that provide this refuge, making geology a potentially important limiting variable. The geological difference between the Klipriviersberg and Suikerbosrand ridges and the ridges where this species occurs have already been mentioned, but it cannot be ascertained from the models as to whether geology was an important predictor variable in the GARP analysis.

The same problem occurs in interpreting the model for *O. pugnax*. The substrate affinities for this species have already been discussed, but it is not possible to determine whether the substrate variables that entered the analysis are important in driving the resulting models. The original implementations of GARP at ERIN and BSW included a function to translate the rules developed by the genetic algorithm into a form that could be interpreted. As interpreting overlays in such a manner might be more difficult from sets of rules for individual models, simple measures such as the number of times a predictor variable contributed to a model could be used to give some level of interpretability. This rule translation function is not available in the current implementation of GARP. Peterson and Cohoon (1999) however, outline a statistical method for investigating the relative importance of different environmental predictors where individual predictors are successively left out of the analysis and the resulting impact on model accuracy measured. This leads to the next shortcoming of GARP - practicality.

For this study it took a period of nearly four weeks of data manipulation, trial and error before the first successful run of DesktopGARP was achieved. This can be attributed principally to poor documentation that accompanies the software. The description of how data should be prepared, with particular reference to the importance of geographic projections of the data is only briefly mentioned. Secondly there is no reference for troubleshooting when problems are encountered. Additionally, important aspects of the implementation are not explained. For example, there is no documentation for the best subsets procedure, so the impact of the use of a commission threshold as opposed to a commission interval is not described. The method for inclusion of a mask layer in the analysis is not adequately described either, and the omission of BIOCLIM and inclusion of negated rules and the possible impact on the performance of the algorithm is not mentioned. It is also not clear as to how combinations of rule types and environmental layers are effected in the algorithm. Obviously if a certain rule type or environmental layer is not included in a particular combination it will not be included in the final model, but it is not apparent as to whether the process whereby rules are mutated in the algorithm may exclude certain environmental variables or rule types if their predictive ability is low. Lastly, the stopping criterion is problematic. The algorithm stops when a mutation to the rules results in no significant increase in the predictive ability of the rule set. However, these mutations are random suggesting that stopping the algorithm based on the predictive ability between two iterations is itself a random stopping point. No significant increase in the predictive ability of the model over a number of successive iterations might be a better stopping condition.

Genetic algorithms are computationally intensive procedures (Goldberg 1989). One run of GARP using all possible combinations of rule types and 100 iterations per run resulted in 1500 iterations. This took longer than 48 hours for one run when fine scale data were used. Obviously this is impractical where large numbers of modeled species distributions are required for conservation purposes.

Lastly there is the problem of non-prediction. When prediction results layers are exported by DesktopGARP non-predictions are converted to absences under the assumption that only predicted presences are important. The conservation literature shows that there are costs associated with both false positive and false negative prediction, depending on conservation objectives. For example, the cost of false negative prediction would be high if the purpose of the modeling exercise is to locate unknown populations of a particular species. This problem would be easily dealt with if two maps were produced per model, with the second one indicating predicted and non predicted cells. Overlays of these, in conjunction with prediction overlays, would indicate spatial patterns in non-predictions and hence give a better indication of uncertainty in the presence absence overlays. Such depictions of uncertainty are important for the interpretation of model outputs (Elith *et al.* 2002).

Goldberg (1989) stresses that robustness, defined as the ability to perform well for a range of different conditions, is the key desirable feature of search and optimization methods such as genetic algorithms. Genetic algorithms have received a lot of interest and have been shown to comply with this requirement (Goldberg 1989, Mitchell 1998, Haupt 1998). They are applied in a wide range of fields including engineering and business. Thus genetic algorithms likely hold interesting potential as a method for exploring species environment relations. The problems associated with GARP mentioned above are predominantly problems of implementation, which can be attended to. However, given that the GARP algorithm was developed over 20 years ago and that both ecological modeling and genetic algorithms theory have advanced in that period the potential exists for the application of superior genetic algorithm based methods to the problem of modeling species distributions. For example they may be used in conjunction with other methods like artificial neural networks for the generation of rule sets that may be more biologically interpretable (Markowska-Kaczmar and Wnuk-Lipinski 2004) or for the generation of fuzzy rules from data mining methods (Ishibuchi and Yamamoto,2002).

Levins (1966), in an overview of ecological modeling, states that models cannot be judged against the criteria of perfect representation of the systems they explore. Instead, models are intended as tools for exploration of these systems where assumptions and postulates are made explicit in an attempt to better understand those systems. A models relevance to the task for which it was developed is the principle criterion against which it should be judged. This requires that the purpose of any habitat distribution modeling exercise should be clarified before it commences. Vaughan and Ormerod (2003) show the importance of this in determining what data will be collected in the field. A decision should be made as to whether modeling is intended solely for the purpose of generating distribution maps or where it is intended to explore important relationships between the organism and its environment. The latter requires that data on direct and resource gradients be collected, with the result that models are likely to be more generally applicable as the species response to these gradients is less likely to vary.

Given the importance of a map of predictions of presence and absence that can be given to a site selection algorithm used in systematic conservation planning GARP is useful in providing just that. The presence absence maps for each of the species presented here are a significant improvement on the use of presence records only, where all areas where the species has not been recorded are taken by the algorithm as absence areas. Thus it is recommended here that these modeled distributions replace the use of locality records for these species in the GDACE Biodiversity Gap Analysis project.

The shortcoming of the use of these maps in the conservation planning process is that they were essentially derived from a black box modeling method, with problems relating to understanding of how those maps were obtained described above. Should it be required that the models should provide for a better understanding of habitat selection by the species of interest, which is often the case with many species of conservation interest, then a modeling method other than GARP should be employed. ENFA (Hirzel *et al.* 2002) provides a statistical method that uses presence data only that provides results that can be interpreted biologically. The issue of practicality in the use of GARP by conservation practitioners could be markedly improved with the provision of proper documentation for DesktopGARP to decrease the learning curve required.

4.4 Future Research

Arthropods have only recently been incorporated into large scale conservation planning strategies (Chown and Freitag-Ronaldson 2002) and are often characterized by locality data sets where species are poorly sampled spatially (Koch *et al.* 2000). It is important to consider the uncertainty inherent in these data when they are used for conservation planning programs and extinction threat assessments. Issues relating to detection probability are particularly important for invertebrates, as many are only detectable during certain periods, or exhibit habits that make them very hard to find. As such detectability should be a key factor in evaluation of available data for these organisms.

Reassessment of the threat status and conservation targets for each of these three arachnid species should explicitly consider the results presented in this report. Criterion B of IUCN (2001), which takes into account parameters relating to the geographic distribution of species, would be appropriate for such a reassessment. Another important consideration in reassessment is that the distributional extents of these species beyond the borders of Gauteng province should be established. For example, collection or anecdotal records exist for all three species well outside of the boundaries of Gauteng province. There is a record of *Hadogenes gunningi* in the South African Museum from near Mafikeng, *Opistophthalmus pugnax* has been recorded at Bloemfontein and near Gaberone in Botswana, and *Harpactira hamiltoni* has been collected near Lydenburg.

While the advantages of the both the MVA and habitat distribution modeling results presented in this report over the previous available information have been highlighted, there are several areas where improvements are still possible.

Another, more information-hungry approach to conservation assessment that is probably more appropriate for the context of rapid land-use change in Gauteng is to consider threats to populations explicitly. While the MVA results presented here are useful in themselves, improvement could start with investigation of population density as a function of habitat quality. Threats to populations occurring in high quality habitat through proximity to unsuitable land use classes for areas with high likelihood of transformation could then be assessed. Population level processes important for population viability can also be investigated, such as natality, mortality and dispersal and population models parameterized for each species may be developed.

Improvements may also be made to the habitat distribution models. First and foremost would be to collected new, independent locality data for further evaluation of the models presented. The data used in model evaluation utilized some data not used in model development (absence data) but an evaluation based on entirely independent data would be preferable. Multiscale modeling might also provide some useful insight into the relationships between these species and their environments, and models with a higher degree of accuracy and generality may be developed with the use of new locality data from carefully selected survey sites and using modeling methods that utilize both presence and absence data.

Should such population dynamics data and accurate fine scale habitat suitability data become available concurrently these may be used to develop spatially realistic metapopulation models (Hanski 1999, Akçakaya 2000) for these species. Such models would be useful in assessing the impacts of species specific and general conservation strategies, land cover transformation and possibly even climate change on the viability of populations of these species within Gauteng.

5. Acknowledgements

Funding for this project was provided by a research grant from GDACE (Grant number: D 295220), a grant from the South African Norwegian Tertiary Education Development Initiative (SANTED) and a

Postgraduate Merit Award from the University of the Witwatersrand. I would like to thank Marianne Forsyth of GDACE for the opportunity to undertake this project, as well as Barend Erasmus for supervision, advice and for providing the opportunity for a greater insight into the field of habitat distribution modeling and conservation biology in general. I would like to acknowledge all the landowners who granted us access to their property to search for specimens, as well management at Roodeplaat Dam, Suikerbosrand and Abe Bailey Nature Reserves for hosting us during the fieldwork. I would like to thank the Council of Geosciences, as well as GDACE for access to the spatial environmental data layers. Lastly to Martin Paulsen with whom the fieldwork was undertaken - it was an experience.

6. References:

Akçakaya, R.H. (2000) Viability analysis with habitat based-metapopulation models. *Population Ecology* **42**: 45 - 53

Allen, C.R., Pearlstine, L.G. and Kitchens, W.M. (2001). Modeling viable mammal populations in gap analysis. *Biological Conservation* **99**: 135 - 144

Anderson, R.P. (2003). Real vs. artefactual absences in species distributions: test for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. *Journal of Biogeography* **30**: 591 - 605

Anderson, R.P., Gomez-Laverde, M. and Peterson, A.T. (2002a). Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography* **11**: 131 - 141

Anderson, R.P., Lew, D. and Peterson, A.T. (2003) Evaluating predictive models of species distributions: criteria for selecting optimal models. *Ecological Modelling* **162**: 211 - 232

Anderson, R.P. and Martinez-Meyer, E. (2004). Modeling species geographic distributions for preliminary conservation assessments: and implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biological Conservation* **116**: 167 - 179

Anderson, R.P., Peterson, A.T. and Gomez-Laverde, M. (2002b). Using niche based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos* **98**: 3 - 16

Araujo, M.B. and Williams, P.H. (2000). Selecting areas for species persistence using occurrence data. *Biological Conservation* **96**: 331 - 345

Araujo, M.B. and Williams, P.H. (2001). The bias of complementarity hotspots towards marginal populations. *Conservation Biology* **15(6)**: 1710 - 1720

Armstrong, A.J. (2002). Insects and determination of biodiversity conservation areas in Kwazulu-Natal province, South Africa. *African Entomology* **10(1):** 11 - 27

Arthur, J.L., Haight, R.G., Montgomery, C.A. and Polasky, S. (2002). Analysis of the threshold and expected coverage approaches to the probabilistic reserve site selection problem. *Environmental Modeling and Assessment* **7(2):** 81 - 138

Austin, M.P. (2002a) Case studies of the use of environmental gradients in vegetation and fauna modeling: theory and practice in New Zealand and Australia. In Scott *et al.* (Eds) *Predicting species occurences: Issues of scale and accuracy*. Island Press, Washington

Austin, M.P. (2002b). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* **157**: 101 - 118

Austin, M.P. and Heyligers, P.C. (1989). Vegetation survey design for conservation: gradsect sampling of forests in north eastern New South Wales. *Biological Conservation* **50**: 13 - 32

Austin, M.P. and Heyligers, P.C. (1991). New approach to vegetation survey design: gradsect sampling In Margules and Austin (Eds) *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. CSIRO, Australia.

Balmford, A. (2003). Conservation planning in the real world: South Africa shows the way. *Trends in Ecology and Evolution* **18(9):** 435 - 438

Beissinger, S.R. and McCullough, D.R. Eds (2002). *Population Viability Analysis*. University of Chicago Press, Chicago.

Beissinger, S.R. and Westphal, M.I. (1998). On the use of demographic models of population viability in endangered species management. *Journal of Wildlife Management* **62**: 821 - 841

Blackburn, T.M. and Gaston, K.J. (2002). Scale in macroecology. *Global Ecology and Biogeography* **11:** 185 - 189

Boston, T. and Stockwell, D.R.B. (1994). Interactive species distribution reporting, mapping and modeling using the World Wide Web. *Computer Networks and ISDN Systems* **28:** 231 - 238.

Bradley, R.A. (1986). The relationship between population density of *Paruroctonus utahensis* (Scorpionida: Vaejovidae) and characteristics of its habitat. *Journal of Arid Environments* **11:** 165-171

Bradley, R.A and Brody, A. J. (1984). Relative abundance of three vaejovid scorpions across a habitat gradient. *Journal of Arachnology* **11**: 437 - 440

Brook, B.W. and Kikkawa, J. (1998). Examining threats faced by island birds: a population viability analysis of the Capricorn silvereye using long term data. *Journal of Applied Ecology* **35**: 491 - 503

Brown, C.A., Davis, J.M., O'Connell, .D.J. and Formanowitz, D.R. Jr. (2002) Surface density and nocturnal activity in a west texas assemblage of scorpions. *The Southwestern Naturalist* **47(3):** 409 - 419

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. (2001) *Introduction to Distance Sampling: Estimating abundance of biological populations*. Oxford University Press, Oxford.

Burnham, K.P. and D.R. Anderson (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd Edition Springer-Verlag, New York.

Busby, J.F. (1991) BIOCLIM - a bioclimatic analysis and prediction system. In Margules and Austin (Eds) *Nature Conservation: Cost effective biological surveys and data analysis*. CSIRO, Melbourne.

Bustamante, J. (1996) Population viability analysis of captive and released bearded vulture populations. *Conservation Biology* **10(3):** 822 - 831

Cabeza, M. and Moilanen, A. (2001) Design of reserve networks and the persistence of biodiversity. *Trends in Ecology and Evolution* **16(5):** 242 - 248

Cabeza, M., Araujo, M.B., Wilson, R.J., Thomas, C.D., Cowley, M.J.R. and Moilanen, A.(2004). Combining probabilities of occurrence with spatial reserve design. *Journal of Applied Ecology* **41:** 252 - 262.

Cablk, M., White, D. and Kiester, A.R. (2002). Assessment of spatial autocorrelation in empirical models in ecology. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Carroll, C., Noss, R.F., Paquet, P.C. and Schumaker, N.H. (2003), Use of population viability analysis and reserve selection algorithms in regional conservation plans. *Ecological Applications* **13(6):** 1773 - 1789

CART (1984) Classification and regression tree software. California Statistical Software, Lafayette.

Caughley, G. (1994) Directions in conservation biology. *Journal of Animal Ecology* **63**: 215 - 244

Chown, S.L. and Freitag-Ronaldson, S. (2002). Survey site selection for data deficient arthropod taxa: from RAGS to riches. *African Entomology* **10(2):** 345 - 349

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Education and Psychological Maesurement* **20:** 37 - 46

Cowling, R.M. and Pressey, R.L. (2003) Introduction to systematic conservation planning in the Cape Floristic Region. *Biological Conservation* **122(1-2):** 1- 13

Cutler, B. and Guarisco, H. (1995) Dispersal Aggregation of *Sphodros fitchi* (Araneae: Atypidae). *Journal of Arachnology* **23:** 205 – 206

DeLong, E.R., DeLong, D.M. and Clarke-Pearson, D.L. (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**: 837-845.

Due, A.D. and Polis, G.A. (1985). Biology of the intertidal scorpion, *Vaejovis littoralis*. *Journal of Zoology* **207**: 563 - 580

Edwards, T.C. Jr, Cutler, R., Zimmermann, N., Geiser, L. and Alegria, J. (in press) Enhancing survey detection rates of rare species using model based stratification: lichens as a case study. *Ecology.*

Elith, J. and Burgman, M.A. (2002) Predictions and their validation: rare plants in the central highlands, Victoria, Australia. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Elith, J., Burgman, M.A. and Regan, H.M. (2002) Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling* **157**: 313 - 329

Engelbrecht, I.A. (2002) Morphometrics and population structure in the burrowing scorpion *Opistophthalmus pictus* (Scorpionidae, Scorpiones). Honours Thesis, University of the Witwatersrand, South Africa.

Engler, R., Guisan, A. and Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species using occurrence and pseudo absence data. *Journal of Applied Ecology* **41**: 263 - 274

Ferrier, S., Watson, G., Pearce, J. and Drielsma, M. (2002) Extended statistical approaches to modeling spatial pattern in biodiversity in northeast New South Wales. I Species level modeling. *Biodiversity and Conservation* **11:** 2275 - 2307

Fielding, A.H. (2002) What are the appropriate characteristics of an accuracy measure? In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Fielding, A.H. and Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24(1):** 38 - 49

Fielding, A.H. and Haworth, P.F. (1995) Testing the generality of bird habitat models. *Conservation Biology* **9(6):** 1466 - 1481

Fleishman, E., Murphy, D.D. and Sjögren-Gulve, P. (2002). Modeling species richness and habitat suitability for taxa of conservation interest. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Forys, E.A. and Humphrey, S.R. (1999). Use of population viability analysis to evaluate management options for the endangered Lower Keys Marsh Rabbit. *Journal of Wildlife Management* **63(1):** 251 - 260

Gibson, L.A., Wilson, B.A., Cahill, D.M. and Hill, J. (2004) Spatial prediction of rufous bristlebird habitat in the coastal heathland: a GIS-based approach. *Journal of Applied Ecology* **41**: 213 - 223

Gilpin, M.E. and Soule, M.E. (1986) Minimum viable populations: Processes of species extinction. In Soule, ME (ed). *Conservation Biology: The Science of Scarcity and Diversity*. Sinauer, Sunderland, MA. Pp 19 - 34

Goldberg, D.E. (1989) *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley Publishing Company, Inc. New York

Grand, J. and Mello, M.J. (2004) A multi scale analysis of species environment relationships: rare moths in a pitch pine- scrub oak (*Pinus rigida - Quercus ilicifolia*) community. *Biological Conservation* **119**: 495 - 506

Guisan, A., Edwards, T.C. and Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* **157**: 89 - 100

Guisan, A. and Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* **135**: 147 - 186

Gu, W. and Swihart, R.K. (2004) Absent of undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation* **116**: 195 - 203

Hammer, M.L. and Slotow, R.H. (2002) Conservation application of existing data for South African millipedes (Diplopoda). *African Entomology* **10(1):** 29 - 42

Hanley, J.A. and McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**: 29-36.

Hanski, I. (1999) *Metapopulation ecology*. Oxford University Press, New York

Harington, A. (1978) Burrowing biology of the scorpion *Cheloctonus jonesii* (Arachnida: Scorpionida: Scorpionidae). *Journal of Arachnology* **5:** 243 - 249

Haupt, R.L. (1998) *Practical Genetic Algorithms*. John Wiley and Sons, New York

Heglund, P.L. (2002) Foundations of species environment relations. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Hirzel, A.H. and Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modeling. *Ecological Modelling* **157**: 331 - 341

Hirzel, A.H., Hausser, J., Chessel, D. and Perrin, N. (2002) Ecological niche factor analysis: how to compute habitat suitability maps without absence data? *Ecology* **83(7):** 2027 - 2036

Holland, J.H. (1975) *Adaptation in natural and artificial systems*. Ann Arbor: The University of Michigan Press.

Huston, M.A. (2002) Critical issues for improving predictions. In Scott *et al*. (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington.

Hutchinson, G.E. (1957) Concluding Remarks. *Cold Spring Harbour Symposium on Quantitative Biology* **22**: 415 - 427

Ishibuchi, H. and Yamamoto, T. (2002) Fuzzy rule selection by data mining criteria and genetic algorithms. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO):* 399-406. Available from www.ie.osakafu-u.ac.jp/~hisaoi/ci_lab_e/index.html

IUCN (2001) *IUCN Red List Categories: Version 3.1* Prepared by the IUCN Species Survival Commission. IUCN, Gland, Switzerland and Cambridge, UK.

Jeganathan, P., Green, R.E., Norris, K., Vogiatzakis, I.N., Bartsch, A., Wotton, S.R., Bowden, C.G.R., Griffiths, G.H., Pain, D. and Rahmani, A.R. (2004) Modeling habitat selection and distribution of the critically endangered Jerdon's courser *Rhinoptilus bitorquatus* in scrub jungle: an application of a new tracking method. *Journal of Applied Ecology* **41**: 224 - 237

Johnson, C.M., Johnson, L.B., Richards, C. and Beasley, V. (2002) Predicting the occurrence of amphibians: an assessment of multiple scale models. In Scott *et al.* (Eds) *Predicting species occurrences: Issues of scale and accuracy*. Island Press, Washington.

Karl, J.W., Svancara, L.K., Heglund, P.J., Wright, N.M. and Scott, M.J. (2002) Species commonness and the accuracy of habitat relationship models. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington.

Keitt, T.H., Bjornstad, O.N., Dixon, P.M. and Citron-Pousty, S. (2002) Accounting for spatial pattern when modeling organism environmental interactions. *Ecography* **25:** 616 - 625

Kelly, M.J. and Durant, S.M. (2000) Viability of the Serengeti cheetah population. *Conservation Biology* **14(3):** 786 - 797

Koch, S.O., Chown, S.L., Davis, A.L.V., Endrody-Younga, S. and van Jaarsveld, A.S. (2000) Conservation strategies for poorly surveyed taxa: a dung beetle (Coleoptera, Scarabaeidae) case study from southern Africa. *Journal of Insect Conservation* **4**: 45 - 56

Lamoral, B.H. (1978) Soil hardness, an important limiting factor in burrowing scorpions of the genus *Opistophthalmus* CL Kock 1837 (Scorpionidae, Scorpionida) *Symposium of the Zoological Society of London* **42**, 171 - 181

Lamoral, B.H. (1979) The scorpions of Namibia (Arachnida: Scorpionida). *Annuls of the Natal Museum* **23**: 498 - 783

Legendre, P., Dale, M.R.T., Fortin, M., Gurevitch, J., Hohn, M. and Myers, D. (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* **25**: 601 - 615

Lehmann, A., McC. Overton, J. and Leathwick, J.R. (2002) GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling* **157**: 189 - 207

Levin, S.A. (1992) The problem of pattern and scale in ecology. *Ecology* **73(6):** 1943 - 1967

Levins, R. (1966) The strategy of model building in population ecology. *American Scientist* **54**: 421 - 431

Lighton, J.R.B., Brownell, P.H., Joos, B. and Turner, R.J. (2001) Low metabolic rate in scorpions: Implications for population biomass and cannibalism. *Journal of Experimental Biology* **204**: 607 - 613

Lim, B.K., Peterson, A.T. and Engstrom, M.D. (2002) Robustness of ecological niche modeling algorithms for mammals in Guyana. *Biodiversity and Conservation* **11**: 1237 - 1246

Lindenmayer, D.B. and Possingham, H.P. (1996) Ranking conservation and timber management options for Leadbeater's Possum in Southeastern Australia using Population viability analysis. *Conservation Biology* **10(1):** 235 – 251

Lourenço, W.R. (1978) Etude sur les scorpions appartenants au "complex" *Tityus trivittatus* Kraepelin, 1898, et, en particulier, de la sous-espece *Tityus trivittatus fasciolatus* Pessoa, 1935 (Buthidae). PhD Thesis, L'Universite Pierre et Marie Curie, Paris.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. and Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83(8):** 2248 - 2255

Maggini, R., Guisan, A. and Cherix, D. (2002) A stratified approach for modeling the distribution of a threatened and species in the Swiss National Park. *Biodiversity and Conservation* **11**: 2117 - 2141

Mandelberg, L. (2004) A comparison of the predictive abilities of four approaches for modeling the distribution of cetaceans. MSc Thesis, University of Aberdeen, UK.

Manel, S, Dias, J and Ormerod, SJ (1999) Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling* 120: 337 - 347

Manel, S., Williams, H.C. and Ormerod, S.J. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* **38:** 921 - 931

Margules, C.R., Pressey, R.L. and Williams, P.H. (2002) Representing biodiversity: data and procedures for identifying priority areas for conservation. *Journal of Biosciences* **27(4)** Suppl 2: 309 - 326

Margules, C.R. and Pressey, R.L. (2000) Systematic conservation planning. *Nature* **405**: 243 - 253

Markowska-Kaczmar, U. and Wnuk-Lipinski, P. (2004) Rule extraction from neural network by genetic algorithm with pareto optimization. In Rutkowski L *et al.* (Eds) *ICAISC 2004, LNAI* **3070**: 450 - 455. Available at http://sci2s.ugr.es/keel/

Marmontel, M., Humphrey, S.R. and O'Shea, T. (1997) Population viability analysis of the Florida Manatee (*Trichechus manatus latirostris*), 1976 - 1991. *Conservation Biology* **11(2):** 467 - 481

McCune, B. (2004) Nonparametric multiplicative regression for habitat modeling. Available from www.pcord.com/NPMRintro.pdf

McGoech, M.A. (2002) Insect conservation in South Africa: an overview. *African Entomology* **10(1):** 1 - 10

McPherson, J.M., Jetz, W. and Rogers, D.J. (2004) The effects of species range sizes on the accuracy of distribution models: ecological phenomenon or statistical artifact? *Journal of Applied Ecology* **41**: 811 - 823

Meggs, J.M., Munks, S.A., Corkrey, R. and Richards, K. (2004) Development and evaluation of predictive habitat models to assist the conservation planning of a threatened lucanid beetle, *Hoplogonus simsoni*, in north east Tasmania. *Biological Conservation* **118**: 501 - 511

Mitchell, M. (1998) *An Introduction to Genetic Algorithms (Complex Adaptive Systems).* The MIT Press.

Morrison, M.L. and Hall, L.S. (2002) Standard Terminology: toward a common language to advance ecological understanding and application. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington.

Newlands, G. and Cantrell, A.C. (1985) A re-appraisal of the rock scorpions (Scorpionidae: *Hadogenes*) *Koedoe* **28**: 35 - 45

Newlands, G. and Prendini, L. (1997) Redescription of *Hadogenes zumpti* Newlands and Cantrell 1985: an unusual rock scorpion (Scorpiones: Ischnuridae) from the Richtersveld, South Africa. *South African Journal of Zoology* **32**: 76 - 81

O'Connor, R.J. (2002) The conceptual basis for species distribution modeling: time for a paradigm shift? In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Osborne, P.E. and Suarez-Seoane, S. (2002) Should data be partitioned spatially before building large scale distribution models? *Ecological Modelling* **157**: 249 - 259

Payne, K. and Stockwell, D.R.B. (no date) *GARP Modeling System Users Guide and Technical Reference*. Available at http://biodi.sdsc.edu/

Pearce, J.L., Cherry, K., Drielsma, M., Ferrier, S. and Whish, G. (2001) Incorporating expert opinion and fine scale vegetation mapping into statistical models of faunal distribution. *Journal of Applied Ecology* **38**: 412 - 424

Pearce, J. and Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* **133**: 225 - 245

Pearson, R.G. and Dawson, T.P. (2003) Predicting the impacts of climate change on the distributions of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* **12**: 361 - 371

Peterson, A.T. and Cohoon, K.P. (1999) Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecological Modelling* **117**: 159 - 164

Peterson, A.T. and Kluza, D.A. (2003) New distributional modeling approaches for gap analysis. *Animal Conservation* **6**: 47 - 54

Peterson, A.T., Soberon, J. and Sanchez-Cordero, V. (1999) Conservatism of ecological niches in evolutionary time. *Science* **285**: 1265 - 1267

Peterson, A.T., Stockwell, D.R.B. and Kluza, D.A. (2002) Distributional prediction based on niche modeling of primary occurrence data. In Scott *et al*. (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Phillips, S.J., Dudik, M. and Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling. *Proceedings of the 21st International Conference on Machine Learning*, Banff, Canada.

Polasky, S., Camm, J.D., Solow, A.R., Csuti, B., White, D. and Ding, R. (2000) Choosing reserve networks with incomplete species information. *Biological Conservation* **94**: 1 - 10

Polis, G.A. (1990) Ecology. In: Polis, GA (ed) *The Biology of Scorpions*. Stanford University Press, Stanford, California

Polis, G.A. and Farley, R.D. (1980) Population biology of a desert scorpion: survivorship, microhabitat and the evolution of life history strategy. *Ecology* **61(3):** 620 - 629

Polis, G.A. and McCormick, S.J. (1986) Patterns of resource use and age structure among species of desert scorpion. *Journal of Animal Ecology* **55**: 59 – 73

Polis, G.A. and Sissom, W.D. (1990) Life History. In: Polis, GA (ed) *The Biology of Scorpions*. Stanford University Press, Stanford, California

Prendini, L. (1995) Patterns of scorpion distribution in Southern Africa: a GIS approach. Unpubl BSc(Hons) Thesis. University of Cape Town.

Prendini, L. (2001a) Two new species of *Hadogenes* (Scorpiones: Ischnuridae) from South Africa, with a redescription of *Hadogenes bicolor* and a discussion of the phylogenetic position of *Hadogenes*. *Journal of Arachnology* **29**: 146 - 172

Prendini, L. (2001b) Substratum specialization and speciation in Southern African scorpions: the Effect Hypothesis revisited. In '*Scorpions 2001:* In Memorium *Gary A. Polis'*. Eds Fet V and Seldon PA. British Arachnological Society, Buckinghamshire, UK

Prendini, L., Crowe, T.M. and Wheeler, W.C. (2003) Systematics and biogeography of the family Scorpionidae (Chelicerata: Scorpiones) with a discussion on phylogenetic methods. *Invertebrate Systematics* **17**: 185 – 259

Redak, R.A. (2000) Arthropods and multispecies habitat conservation plans: are we missing something? *Environmental Management* **26** Suppl 1: S97 - S107

Reed, D.H., O'Grady, J.J., Brook, B.W., Ballou, J.D. and Frankham, R. (2003) Estimates of minimum viable population sizes for vertebrates and factors affecting those estimates. *Biological Conservation* **113**: 23 - 34

Reed, D.H. and Hobbs, G.R. (2004) The relationship between population size and temporal variability in population size. *Animal Conservation* **7**: 1 - 8

Reese, G.C., Wilson, K.R., Hoeting, J.A. and Flather, C.H. (2005) Factors affecting species distribution predictions: a simulation modeling experiment. *Ecological Applications* **15(2):** 554 - 564

Rushton, S.P., Ormerod, S.J. and Kerby. G. (2004) New paradigms for modeling species distributions? *Journal of Applied Ecology* **41**: 193 - 200

SACN (2004) *State of the Cities Report*. South African Cities Network, available at www.sacities.net

Sarkar, S. (2002) Preface to special edition on conservation biology. *Journal of Biosciences* **27(4)** Suppl 2: i - iv

Scachetti-Pereirra, R. (2005) Desktop GARP version 1.1.6. University of Kansas Biodiversity Research Centre. Available from http://beta.lifemapper.org/desktopgarp/

Schaefer, S.M. and Krohn, W.B. (2002) Predicting vertebrate occurrences from species habitat associations: improving the interpretation of commission error rates. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Schultz, S.T. and Lynch, M. (1997) Mutation and extinction: the role of variable mutation effects, synergistic epistasis beneficial mutations and degree of outcrossing. *Evolution* **51**: 1361 – 1371

Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B. (2002) *Predicting species occurrences: issues of accuracy and scale*. Island Press, Washington.

Seoane, J., Bustamante, J. and Diaz-Delgado, R. (2004a) Are existing vegetation maps adequate to predict bird distributions? *Ecological Modelling* **175**: 137 - 149

Seoane, J., Bustamante, J. and Diaz-Delgado, R. (2004b) Competing roles for landscape, vegetation, topography and climate in predictive models for bird distribution. *Ecological Modelling* **171**: 209 - 222

Seoane, J., Bustamante, J. and Diaz-Delgado, R. (2005) Effect of expert opinion on the predictive ability of environmental models of bird distribution. *Conservation Biology* **19(2):** 512 - 522

Shaffer, M.L. (1981) Minimum population size for species conservation. *Bioscience* **31**: 131 - 134

Shaffer, M.L., Hood, L., Snape, W.J.I. and Latchis, I. (2002) Population viability analysis and conservation policy. In: Beissinger, SR and McCullough, DR (eds), *Population Viability Analysis*. University of Chicago Press, Chicago, USA.

Smith, A.M. (1990) *Baboon Spiders and Tarantulas of Africa and the Middle East*. Fitzgerald Publishing, London.

Soberon, J. and Peterson, A.T. (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London Series B* **359**: 689 - 698

Soberon, J. and Peterson, T.A. (2005) Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiversity Informatics* **2**: 1 - 10

Statistics South Africa (2001) *Census 2001: Key Results*. Statistics South Africa, Pretoria. Available at www.statssa.gov.za

Stauffer, H.B., Ralph, C.J. and Miller, S.L. (2002) Incorporating detection uncertainty into presence absence surveys for Marbled Murrelet. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Stockwell, D.R.B. and Noble, I.R. (1992) Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation* **32**: 249 - 259

Stockwell, D.R.B. and Peters, D. (1999) The GARP modeling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Science* **13**: 143 - 158

Stockwell, D.R.B. and Peterson, A.T. (2002a) Effects of sample size on accuracy of species distribution models. *Ecological Modelling* **148**: 1 - 13

Stockwell, D.R.B. and Peterson, A.T. (2002b) Controlling bias in biodiversity data. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Stockwell, D.R.B. and Peterson, A.T. (2003) Comparison of methods used in mapping biodiversity patterns from point occurrence data. *Ecological Indicators* **3**: 213 - 221

Thomas, C.D. (1990) What do real population dynamics tell us about minimum viable population size? *Conservation Biology* **4**: 324 - 327

Thomas, K., Keeler-Wolf, T. and Franklin, J. (2002) A comparison of fine and coarse resolution environmental variables toward predicting vegetation distribution in the Mojave Desert. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington

Thomas, L., Laake, J.L., Strindberg, S., Marques, F.F.C., Buckland, S.T., Borchers, D.L., Anderson, D.R., Burnham, K.P., Hedley, S.L., Pollard, J.H. and Bishop, J.R.B. (2003). Distance 4.1. Release 2. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. http://www.ruwpa.stand.ac.uk/distance/

Thuiller, W. (2003) BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under climate change. *Global Change Biology* **9**: 1353 - 1362

Van Horne, B. (2002) Approaches to habitat modeling: the tensions between pattern and process and between specificity and generality. In Scott *et al.* (Eds) *Predicting Species Occurrences: Issues of Scale and Accuracy*. Island Press, Washington.

Vaughan, I.P. and Ormerod, S.J. (2003) Improving the quality of distribution models for conservation by addressing shortcoming in the field collection of training data. *Conservation Biology* **17(6):** 1601 - 1611

Wessels, K.R., Reyers, B., van Jaarsveld, A.S. and Rutherford, M.C. (2003) Identification of potential conflict areas between land transformation and biodiversity conservation in north eastern South Africa. *Agriculture, Ecosystems and Environment* **95:** 157 - 178

Wiens, J.A. (1989) Spatial scaling in ecology. *Functional Ecology* **3**: 385 - 397

Whitlock, M.C. (2000) Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles and sexual selection. *Evolution* **54**: 1855 – 1861

Whittingham, M.J., Wilson, J.D. and Donald, P.F. (2003) Do habitat association models have any generality? Predicting skylark *Alauda arvensis* in different regions of southern England. *Ecography* **26**: 521 - 531

Williams, P.H. and Araujo, M.B. (2002) Apples, oranges and probabilities: integrating multiple factors into biodiversity conservation with consistency. *Environmental Modeling and Assessment* **7(2):** 139 - 151

Williams, P.H., Margules, C.R. and Hilbert, D.W. (2002) Data requirements and data sources for biodiversity priority area selection. *Journal of Biosciences* **27(4)** Suppl 2: 327 - 338

Wilson, K.A., Westphal, M.I., Possingham, H.P. and Elith, J. (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation* **122**: 99 - 112

Yanez, M. and Floater, G. (2000) Spatial distribution and habitat preference of the endangered tarantula, *Brachypelma klaasi* (Araneae: Theraphosidae) in Mexico. *Biodiversity and Conservation* **9**: 795 - 810

Younge, A. and Fowkes, S. (2003) The Cape Action Plan for the Environment: overview of an ecoregional planning process. *Biological Conservation* **112**: 15-28

Zaniewski, A.E., Lehmann, A. and McC. Overton, J. (2002) Predicting species spatial distributions using presence only data: a case study of native New Zealand ferns. *Ecological Modelling* **157**: 261 - 280

Zweig, M.H. and Campbell, G. (1993) Receiver operating characteristic (ROC) plots: a fundamental tool in clinical medicine. *Clinical Chemistry* **39**: 561 - 577

7. Appendices

Appendix I: Model fitting results for Distance Analysis

A: Model fitting and selection criteria using untransformed distance data. Under ‘Sites’, ex 13a refers to all sites where *Opistophthalmus pugnax* was collected and distance data recorded. ‘All Sites’ for *Harpactira hamiltoni* is all sites where this species was collected and distance data recorded. Under ‘Collector’, MP is Martin Paulsen and IE is Ian Engelbrecht. ‘# params’ is the number of adjustment parameters included in a model. ‘AIC’ is Akaike Information Criterion, ‘AICc’ is the small sample size adjusted AIC, ‘K-S GOF p-value’ is the p-value for the Kolmogorov-Smirnoff Goodness of Fit Statistic and ‘χ^2 p-value’ is the p value of the Chi Squared Goodness of Fit Statistic for 10 equal data intervals.

Species	Site	Collector	Model	Adjustment	# params	AIC	AICc	K-S GOF p-value	χ^2 p-value
Opistophthalmus pugnax	13a	MP	Uniform	Cosine	1	50.82	50.91	0.12	0.79
			Uniform	Polynomial	1	51.60	51.70	0.12	0.55
			Halfnormal	Cosine	1	50.84	50.93	0.12	0.77
			Halfnormal	Hermite	1	50.84	50.93	0.12	0.77
			Hazardrate	Cosine	2	52.20	52.49	0.12	0.64
			Hazardrate	Polynomial	2	52.20	52.49	0.12	0.64
		IE	Uniform	Cosine	1	75.33	75.39	0.39	0.19
			Uniform	Polynomial	1	75.78	75.85	0.13	0.24
			Halfnormal	Cosine	1	74.90	74.97	0.36	0.23
			Halfnormal	Hermite	1	74.90	74.97	0.36	0.23
			Hazardrate	Cosine	2	75.44	75.64	0.62	0.18
			Hazardrate	Polynomial	2	75.44	75.64	0.62	0.18
		IE and MP	Uniform	Cosine	1	124.33	124.37	0.13	0.38
			Uniform	Polynomial	1	125.76	125.80	0.05	0.19
			Halfnormal	Cosine	1	123.90	123.94	0.11	0.40
			Halfnormal	Hermite	1	123.90	123.94	0.11	0.40
			Hazardrate	Cosine	No Data	No Data	No Data	No Data	No Data
			Hazardrate	Polynomial	2	126.00	126.11	0.14	0.28
	ex13a	MP	Uniform	Cosine	1	137.50	137.57	0.05	0.12
			Uniform	Polynomial	1	137.63	137.70	0.00	0.07
			Halfnormal	Cosine	3	130.87	131.31	0.13	0.78
			Halfnormal	Hermite	1	136.63	136.70	0.09	0.15
			Hazardrate	Cosine	2	134.75	134.96	0.13	0.32
			Hazardrate	Polynomial	4	133.18	133.92	0.13	0.51
		IE	Uniform	Cosine	4	114.85	115.22	0.99	0.31
			Uniform	Polynomial	3	120.2	120.45	0.15	0.06

		IE				3		
			Halfnormal	Cosine	4	114.64	115.01	0.99	0.29
			Halfnormal	Hermite	1	120.72	120.76	0.02	0.03
			Hazardrate	Cosine	2	112.95	113.05	0.94	0.37
			Hazardrate	Polynomial	2	112.95	113.05	0.94	0.37
		IE and MP	Uniform	Cosine	1	124.33	124.37	0.13	0.38
			Uniform	Polynomial	1	125.76	125.80	0.05	0.19
			Halfnormal	Cosine	1	123.90	123.94	0.11	0.41
			Halfnormal	Hermite	1	123.90	123.94	0.11	0.41
			Hazardrate	Cosine	No Data	No Data	No Data	No Data	No Data
			Hazardrate	Polynomial	2	126.00	126.11	0.14	0.28
Harpactira hamiltoni	All sites	MP	Uniform	Cosine	1	34.39	34.56	0.64	0.92
			Uniform	Polynomial	3	36.40	37.49	0.98	0.86
			Halfnormal	Cosine	1	34.55	34.72	0.78	0.91
			Halfnormal	Hermite	1	34.55	34.72	0.78	0.91
			Hazardrate	Cosine	2	35.02	35.54	0.99	0.90
			Hazardrate	Polynomial	2	35.02	35.54	0.99	0.90
		IE	Uniform	Cosine	0	-5.36	-5.36	0.70	0.71
			Uniform	Polynomial	0	-5.36	-5.36	0.70	0.71
			Halfnormal	Cosine	1	-3.90	-3.50	0.97	0.68
			Halfnormal	Hermite	1	-3.90	-3.50	0.97	0.68
			Hazardrate	Cosine	2	-2.15	-0.81	0.97	0.62
			Hazardrate	Polynomial	2	-2.15	-0.81	0.97	0.62
		IE and MP	Uniform	Cosine	2	38.34	38.68	0.85	0.86
			Uniform	Polynomial	3	39.95	40.66	0.86	0.78
			Halfnormal	Cosine	2	38.00	38.35	0.81	0.90
			Halfnormal	Hermite	1	39.76	39.87	0.55	0.68
			Hazardrate	Cosine	2	37.76	38.10	0.95	0.89
			Hazardrate	Polynomial	2	37.76	38.10	0.95	0.89

B: Model fitting and selection criteria using distance data transformed into equal intervals. Under 'Sites', ex 13a refers to all sites where *Opistophthalmus pugnax* was collected and distance data recorded. 'All Sites' for *Harpactira hamiltoni* is all sites where this species was collected and distance data recorded. Under 'Collector', MP is Martin Paulsen and IE is Ian Engelbrecht. The number of intervals for each dataset was selected by visually inspecting the shape and assessing shape criterion as well as to minimize heaping in the data. '# params' is the number of adjustment parameters included in a model. 'AIC' is Akaike Information Criterion, 'AICc' is the small sample size adjusted AIC and 'χ^2 p-value' is the p value of the Chi Squared Goodness of Fit Statistic.

Species	Site	Collector	No. of Intervals	Model	Adjustment	# params	AIC	AICc	χ^2 p-value
		MP	5	Uniform	Cosine	1	130.27	130.37	0.73
				Uniform	Polynomial	1	131.16	131.26	0.65
				Halfnormal	Cosine	1	130.39	130.48	0.85
				Halfnormal	Hermite	1	130.39	130.48	0.85

Opistophthalmus pugnax	13a			Hazardrate	Cosine	2	132.27	132.56	0.44
				Hazardrate	Polynomial	2	132.27	132.56	0.44
		IE	10	Uniform	Cosine	1	276.74	276.80	0.14
				Uniform	Polynomial	1	275.80	275.87	0.17
				Halfnormal	Cosine	1	276.13	276.19	0.16
				Halfnormal	Hermite	1	276.13	276.19	0.16
				Hazardrate	Cosine	2	276.78	276.98	0.18
				Hazardrate	Polynomial	2	276.78	276.98	0.18
		IE and MP	5	Uniform	Cosine	1	318.66	318.69	0.09
				Uniform	Polynomial	1	318.50	318.54	0.09
				Halfnormal	Cosine	1	318.26	318.30	0.11
				Halfnormal	Hermite	1	318.26	318.30	0.11
				Hazardrate	Cosine	4	318.71	319.10	No Data
				Hazardrate	Polynomial	4	320.11	320.50	No Data
	ex13a	MP	10	Uniform	Cosine	3	253.81	254.25	0.81
				Uniform	Polynomial	5	257.41	258.54	0.62
				Halfnormal	Cosine	2	252.83	253.05	0.79
				Halfnormal	Hermite	1	258.33	258.40	0.20
				Hazardrate	Cosine	2	253.98	254.20	0.65
				Hazardrate	Polynomial	2	253.98	254.20	0.65
		IE	10	Uniform	Cosine	4	323.22	323.59	No Data
				Uniform	Polynomial	3	326.90	327.12	0.01
				Halfnormal	Cosine	2	321.78	321.89	0.21
				Halfnormal	Hermite	1	330.06	330.09	0.00
				Hazardrate	Cosine	2	319.33	319.44	0.69
				Hazardrate	Polynomial	2	319.33	319.44	0.69
		IE and MP	15	Uniform	Cosine	3	795.09	795.23	0.10
				Uniform	Polynomial	1	851.90	851.93	0.00
				Halfnormal	Cosine	2	792.58	792.65	0.16
				Halfnormal	Hermite	1	806.98	807.00	0.00
				Hazardrate	Cosine	2	789.37	789.44	0.44
				Hazardrate	Polynomial	2	789.37	789.44	0.44
Harpactira hamiltoni	All sites	MP	5	Uniform	Cosine	1	76.40	76.57	0.79
				Uniform	Polynomial	1	78.21	78.38	0.42
				Halfnormal	Cosine	1	76.42	76.58	0.86
				Halfnormal	Hermite	1	76.42	76.58	0.86
				Hazardrate	Cosine	2	77.68	78.20	0.86
				Hazardrate	Polynomial	2	77.68	78.20	0.86
		IE	5	Uniform	Cosine	0	57.94	57.94	0.99
				Uniform	Polynomial	0	57.94	57.94	0.99
				Halfnormal	Cosine	1	59.94	60.19	0.95
				Halfnormal	Hermite	1	59.94	60.19	0.95
				Hazardrate	Cosine	2	61.94	62.74	0.85
				Hazardrate	Polynomial	2	61.94	62.74	0.85
		IE and MP	5	Uniform	Cosine	2	100.84	101.18	0.93
				Uniform	Polynomial	3	102.85	103.56	0.95
				Halfnormal	Cosine	2	100.88	101.22	0.79
				Halfnormal	Hermite	1	100.94	101.05	0.99
				Hazardrate	Cosine	2	100.61	100.96	0.90
				Hazardrate	Polynomial	2	100.61	100.96	0.90

Appendix II: Environmental predictor variables used in GARP habitat distribution modeling for *Harpactira hamiltoni*, *Hadogenes gunningi* and *Opistophthalmus pugnax* in Gauteng Province, South Africa.

A: Coarse scale environmental variables

Environmental Variable	**Data source**	**Description**
Mean annual temperature	AGIS	
Daily Maximum temperature in January	AGIS	Possible limiting variable in organism environmental temperature relations
Daily mean temperature in January	AGIS	Possible limiting variable in organism environmental temperature relations
Daily minimum temperature in July	AGIS	Possible limiting variable in organism-environmental temperature relations
Daily mean temperature in July	AGIS	Possible limiting variable in organism-environmental temperature relations
Average annual evaporative water stress	AGIS	Possible limiting variable in organism-environmental water relations
Evaporative water stress in January	AGIS	Possible limiting variable in organism-environmental water relations
Evaporative water stress in July	AGIS	Possible limiting variable in organism-environmental water relations
Mean Annual precipitation	AGIS	Possible limiting variable in organism-environmental water relations
Coefficient of variation of mean annual precipitation	AGIS	Possible limiting variable in organism-environmental water relations
Mean number of frost days	AGIS	Possible limiting variable in organism environmental temperature relations
Standard deviation number of frost days	AGIS	Possible limiting variable in organism environmental temperature relations
Water stress in January	AGIS	Possible limiting variable in organism-environmental water relations
Water stress in July	AGIS	Possible limiting variable in organism-environmental water relations
Heat stress from April to September	AGIS	Possible limiting variable in organism environmental temperature relations
Heat stress from October to March	AGIS	Possible limiting variable in organism environmental temperature relations

B: Fine Scale Environmental variables

Environmental Variable	**Data source**	**Description**
Geology	1:250 000 geology vector data, Council for Geosciences	Possible important variable for those species with substrate dependencies
Lancover	1994 National Landcover dataset, sourced from GDACE	Possible important variable in determine which landcover types a species may occur in

Presence/absence of ridges and a ridge buffer zone of 1.5km	GDACE ridges vector dataset	Important for those species that may be associated with ridges
Vegetation type	GDACE VegHab vector dataset	May serve as an important limiting variable, or correlate well with other important limiting variables
Soil clay percentage	Dataset sourced from GDACE	Possible important variable for those species with substrate dependencies. Soil clay percentage may approximate soil texture.
Slope	Derived from SRTM 90m DEM	
Aspect	Derived from SRTM 90m DEM	
Elevation	SRTM 90m DEM	