Mapping the spring and winter distribution of Kalmykia's Saiga population

Graham Elliott

Supervisors: E.J. Milner-Gulland & Navinder Singh

1 December, 2011

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science and the Diploma of Imperial College London

Declaration of own work

I declare that this thesis:

“Mapping the spring and winter distribution of Kalmykia's Saiga population”

is entirely my own work and that where material could be construed as the work of others, it is fully cited and referenced, and/or with appropriate acknowledgement given.

Signature:

Graham Elliott

Supervisor 1: E.J. Milner-Gulland

Supervisor 2: Navinder Singh

Table of Contents

“By the example of the dynamics of the saiga population on the territory of our country, it is shown that the protection of any fauna species implies not only its direct conservation but also a system of interrelated measures of preserving its habitat and ordering human economic activities.” B. D. Abaturov, 2007.

Glossary

CZBR - Chernye Zemli Biosphere Reserve

EVI - Enhanced vegetation index

GAM – Generalised additive model

GIS – Geographic Information System

GLM – Generalised linear models

HSM - Habitat suitability model (also known as species distribution model)

NDVI – Normalised difference vegetation index

NIR - Near-infrared radiation

PA - Protected area

PC - Pre-Caspian

RS - Remote sensing

SDM - Species distribution model

Within the text, unless otherwise stated, “winter” refers to the November-February monitoring period, “spring” refers to the March-June period.

Abstract

The saiga is a critically endangered, migratory antelope of the Eurasian steppe, which declined by more than 95% in a decade, from a population of a million in the early 1990s to an estimated 30,000 by 2003. The species has received little attention, though in recent years, levels of research have increased. The saiga is a keystone species; the decline of the saiga has lead to severe ecological changes in the steppe ecosystem.

Recent research suggests the migratory behaviour of the saiga in Kalmykia has been disrupted and their seasonal distribution is unclear. The main aim of this study was to identify saiga distribution in Kalmykia during winter and spring, based on data from a participatory monitoring programme. Models were produced to identify the drivers of distribution and predict likely areas of saiga occupancy.

An extensive area of likely saiga habitat was identified across central and southern Kalmykia, centred around the Stepnoi and Chernye Zemli Biosphere reserves. Key predictors of saiga presence were distance to the protected areas and distance to water; the probability of saiga presence increased in areas closer to water and the reserves. Based on this study, participatory monitoring has clear potential to contribute to our understanding of saiga distribution and migratory behaviour.

Word count: 14486

Acknowledgements

I would like to thank my supervisors, E.J. Milner-Gulland & Navinder Singh for their advice, insight and guidance. I appreciate their efforts in restricting my various determined attempts to head off at tangents. I also thank Anna Lushchekina for the many translations between English and Russian (and sorry for all the grief over my flight). Thank you to Yuri Arylov and Anatoly Khludnev for the information on monitors and waterholes. A number of respondents in the Imperial R group offered help and advice, which was much appreciated. Thanks is also owed to the following organisations for funding this project:- the Saiga Conservation Alliance and the Rufford foundation. Without the much appreciated efforts of the monitoring team, this project would not have been possible.

Above all else, I owe a huge debt of gratitude to my mother, for all her support and guidance over the years. Thank you.

Introduction

1.1 The saiga antelope

The enigmatic saiga (*Saiga tatarica*) is a keystone species of the steppes of central Asia and is the only extant representative of the genus *Saiga* (EDGE, no date). Yet the saiga is one of the fastest declining and most threatened species (EDGE, no date). As a keystone species, any reduction in the saiga population may also have a negative impact on the ecology of the region.

The Red List status for the saiga is Critically Endangered (Mallon, 2008) and it is listed on Appendix II of CITES and Appendix II of the Convention on Migratory Species. As a migratory species, the saiga poses a number of challenges in terms of research, management and conservation. In part, this stems from the financial cost and logistical challenge of studying a species, whose range may extend over hundreds or thousands of kilometres. The situation may be compounded where migrations cross political borders and so require international cooperation between states, which may have differing views on conservation and resources. That said, the migrations of wildebeest within the Serengetti-Mara ecosystem and caribou in north America are well known and relatively well studied, but less is known about the saiga migration (Singh et al, 2010a).

Compared with non-migratory species, the increased range requirement of migratory species in order to encompass resource variation, incurs greater risks (Singh et al, 2010a). These risks include threats such as poaching, habitat loss or fragmentation, increased vulnerability to climate change, and challenges posed by obstacles, be they natural, such as rivers or mountain ranges, or anthropogenic barriers, such as roads or railways (Berger, 2004).

The factors which determine saiga distribution, may change over time and are likely to become more dynamic as climate change impacts increase. Poaching may be replaced by habitat loss or invasive species as the major threat. Migratory routes can vary on a seasonal basis (Bolger et al, 2008). For the saiga this variation is driven by forage quality, distribution of water sources and disturbance (Singh et al, 2010a); in the case of the Serengeti migration, this is due to fine-scale variation in rainfall and vegetative production (Boone et al, 2006). The saigas' habitat has reduced and become fragmented; poaching and general disturbance means that they are probably occupying sub-optimal habitat (Bekenov et al, 1998; Milner-Gulland et al, 2001). All these aspects make planning of appropriate conservation action (such as static protected areas (PAs) and corridors) difficult.

The saigas' range used to stretch across the northern hemisphere from Canada, east to Siberia (Arylov et al, 2004; Harington, 1998). Today, this migratory antelope is confined to 5 populations in Russia, Kazakhstan (which supports the bulk of the species), Uzbekistan and W. Mongolia (subspecies: *Saiga tatarica mongolica*). The Mongolian population has, like the nominate species, experienced a severe reduction in recent years, although this population may now be increasing (Young et al, 2009).

1.2 The importance of saiga conservation

Conservation of the saiga is important for a number of reasons. Not least is the important, but often overlooked view, that all species have a right to exist. In addition, if the saiga goes extinct, we will have lost a species and the last representative of its genus. A brief explanation of the saigas' role as a keystone species provides additional points. Steppes are globally important ecosystems, supporting a wide variety of flora and fauna, many of which are endemic (GEF, 2008). The Kazakhstan steppe supports 2000 floral species (including 8 endemic vegetation communities) and 73 mammal species (including 9 endangered species) (GEF, 2008). The decline of the saiga has lead to severe and extensive ecological changes to the steppe, which were previously maintained in a more stable state by grazing by saiga and other ungulates (GEF, 2008). Clearly, the loss of the saiga would have serious implications for this key ecosystem.

1.3 The saiga migration

The saiga migration is one of the few remaining working ungulate migrations (Berger, 2004). However, their severe decline in Kalmykia has disrupted their normal migratory patterns and currently the distribution of this population is unclear (Bekenov et al, 1998; Singh et al 2010a). Although efforts have been made to identify the distribution, based on participatory monitoring (PM)(Whitebread, 2008; Leon, 2009) and monitoring by rangers (O'Neil, 2008). There is evidence to suggest many saigas may no longer migrate, but remain within the Chernye Zemli Biosphere Reserve (CZBR) and Stepnoi Reserve throughout the year (O'Neil, 2008).

Herein lies the problem which this project seeks to address: to conserve the saiga, it is crucial that we have a clear understanding of the species' current distribution. Animals migrate seasonally to avoid climatic extremes, predators, diseases, or in search of water, food and other resources (Bolger et al, 2008). We need to fully understand the saigas' spatial requirements; if some saiga do still migrate, we need to know where they go, in order to assess if and how they can be protected during this migration. We need to identify the drivers of saiga distribution and how these drivers influence their distribution. This data can then inform and increase the cost-effectiveness of conservation interventions.

1.4 Aims and objectives

This study builds on previous research in Kalmykia:- a 6 month pilot PM scheme established in 2008 (Whitebread, 2008) and an evaluation of the effectiveness of local knowledge in mapping saiga distribution (Leon, 2009). The aim of the project is to use PM data, to determine the distribution of the Kalmykian saiga population during the winter and spring seasons, and to determine those factors which influence this distribution. The results will then inform conservation and public engagement activities and represent a baseline with which further data can be compared. Furthermore, the results will be used to create a habitat suitability model to identify distribution drivers for conservation planning.

The following objectives are required to achieve this aim:-

- Use local monitoring to map (in ArcGIS) saiga distribution during the winter and spring periods
- Model a range of possible explanatory variables to determine key drivers of distribution
- Create a map of potential saiga habitat based on the drivers of distribution
- Assess if explanatory variables of distribution vary between spring and winter
- Evaluate the usefulness of PM data with regard to conservation of the saiga
- Evaluate the usefulness of Maxent with regard to identifying key drivers of distribution and predicting saiga the probability of saiga distribution

2 Background

2.1 Overview

To position this thesis within the context of existing studies, this section will review research which is germane to this study's objectives. The section begins with a brief description of the challenges involved in conserving ungulates, followed by an overview of the role of monitoring in conservation. The next 2 sections focus on saiga ecology and potential drivers of saiga distribution. Participatory monitoring will then be discussed, with a view to the advantages and disadvantages of this approach; the advantages and limitations of species distribution modelling will be the final area for appraisal. The section will conclude with a brief description of the study area.

2.2 Challenges of conserving ungulates

Ungulates are a diverse, successful group, distributed across all continents except Australasia and Antarctica and established in a variety of habitats from the tropics to the arctic circle (Geist, 1990). Yet this group includes a number of highly threatened species, in addition to the saiga, such as the Hirola (*Damaliscus hunteri*; Critically Endangered), Aders' duiker (*Cephalophus adersi*; Critically Endangered) and the Scimitar-horned Oryx (*Oryx dammah*), which is extinct in the wild.

Another challenge to the conservation of ungulates comes from their diversity of habitat and behaviour. Some ungulate species are migratory (i.e. saiga, wildebeest *Connochaetes taurinus*, zebra *Equus burchelli* and caribou *Rangifer tarandus*) and have extensive ranges. Others are found at altitudes up to 5000m (i.e. argali *Ovis ammon* and chiru *Pantholops hodgsonii*) or on precipitous slopes > 45^{o}(ibex *Capra sibirica* and markhor *Capra falconer*)(Singh & Milner-Gulland, 2011). Monitoring species at high altitude, on steep slopes or across extensive ranges is difficult. Furthermore, many of the range states lack the capacity and financial resources to implement rigorous monitoring schemes. Even when monitoring does occur, it may be sporadic and/or the methodology is inconsistent.

2.3 Monitoring

Monitoring is a fundamental part of conservation. Monitoring indicates trends in species' populations and provides insight into a species distribution, which are vital in identifying species threatened with extinction. Monitoring also informs how populations respond to management interventions or environmental stochasticity. It is vital that areas of occupancy are correctly identified to target efficient conservation action. Failure to do so can lead to scant resources allocated in unoccupied areas, while a target species declines in occupied areas.

There are a variety of monitoring techniques, including mark-release-recapture, direct counts, distance sampling and photographic monitoring (Sutherland, 2000). For this project, direct counts of saiga were effected on an opportunistic basis by the monitors. Relating monitoring data to environmental variables, provides an overview of the habitat where the saiga are present, and where they are absent. Identifying differences between the 2 habitats, such as vegetation or elevation, suggests potential predictors of saiga presence, which can be extrapolated to a broader scale. Therefore, from monitoring on a localised scale, predictions of saiga distribution can be mapped across a much larger scale. This means that conservation measures are more likely to be allocated to the appropriate areas.

2.4 Saiga ecology

The saiga is a medium sized antelope of the steppe and semi-desert of Russia and central Asia, (Bekenov et al, 1998). The species is sexually dimorphic, with males almost 50% heavier than females (male: 40.6 kg; female: 28.1 kg) (Fadeev & Sludskii, 1982). The coat is buff coloured with a white abdomen in summer; a thicker, all white coat is grown in winter (Bekenov et al, 1998). This stocky antelope is capable of speeds of 80 kph (Arylov et al, 2004). Generally, only males bear a pair of translucent amber horns (Bekenov et al, 1998); although within the pre-Caspian (PC) population the ratio of horned females is about 2-3 per 10,000 (Minoransky, 2009). The saigas' most noticeable feature is the proboscis-like nose. A number of theories exist regarding the purpose of this adaptation:- Clifford & Witmer (2004) suggest the nose serves as a filter; Arylov et al (2004) propose it acts as a counter-current heat exchanger for thermoregulation, while Frey et al (2007) suggest it has a prominent role in the rutting roar of the males.

The saiga is adapted to a harsh, extreme environment, but is vulnerable to a specific set of severe climatic conditions known as a *dzhut* (Bekenov et al, 1998). A dzhut is an extreme event comprising dense, deep (at least 30 cm) or ice covered snow which inhibits foraging (Bekenov et al, 1998); typically this situation is exacerbated by a combination of extremely low temperature and high winds. Bekenov et al (1998) reported that dzhuts normally occur every decade in Kazakhstan, but less severe dzhuts are experienced 3 or 4 times per decade. Four dzhuts occurred in Kazakhstan in the 1970s, resulting in the deaths of 400,000 saiga in 1971/2 and 100,000 saiga in 1975/6; thereafter the Kazakh saiga population reduced by 50% (Bekenov et al, 1998). The PC population experienced a dzhut in 2009/10, which along with the effect of poaching, has impacted this small population, with unofficial estimates suggesting numbers have reduced to 8,000 (Kuhl, 2010).

The species is extremely fecund; females are sexually mature at seven or eight months and can breed throughout their life, which spans about ten years (Milner-Gulland, 1994). Kuhl et al (2009) found 12% of yearlings produced twins and 77% were fecund – this rose to 95% in older females, while the twinning rate rose to 73%. Saiga have among the highest levels of maternal investment within ungulates, with an average birth litter mass of 16.9% of maternal bodymass (Kuhl et al, 2007). Saiga calves spend their first days hiding, as opposed to following their mothers, as seen among wildebeest (Kuhl, 2008). Given the saiga's high *in utero* investment and high birth litter mass (Kuhl et al, 2007), it is reasonable to expect that saiga would produce precocial young, as evident in wildebeest, i.e. followers rather than hiders. Furthermore, the hider strategy could make calves more vulnerable to poaching, if poachers know the locations of these static aggregations.

In addition to the calving periods, saiga form large aggregations during the rut and also at other times. In addition to representing an anti-predator strategy, in winter large aggregations help penetrate snow to provide access to forage, whilst in summer, aggregations may afford individuals some relief from biting insects, (Arylov et al, 2004). However, other threats to the saiga are the result of aggregations of domestic animals.

Before CZBR was established, overgrazing by sheep degraded the vegetation, resulting in desertification; although much of the reserve has since been returned to its former state, 25000 ha remains as bare soil or sand (Badmaev & Ubushaev, 2004). Attempts in the 1960s to restore forage, by ploughing 150,000 ha of pasture to create fields, failed as erosion transformed the land to desert (Arylov et al, 2004). This action further fragmented saiga distribution and increased competition with livestock for fodder (Arylov et al, 2004).

The PC saiga population typically inhabits the steppe and semi desert (Singh et al, 2010a). The saigas' preferred habitat is flat terrain with low vegetation, numerous water bodies, with only a moderate snowfall (Bekenov et al, 1998). Unlike the Kazakhstan populations, the PC population does not exhibit a clear north-south migratory pattern. Throughout the year the saiga distribution in the pre-Caspian occupies the majority of Kalmykia and extends into neighbouring Astrakhan province (Singh et al, 2010a).

2.5 Hunting

Saiga have long been hunted for hide and meat, with their horn being a key ingredient of many Chinese medicines (Bekenov et al, 1998). Official records show almost four million horns were exported to China, in the nineteenth century (Arylov et al, 2004). From the middle of the nineteenth century, the saiga population suffered a drastic decline until, by about 1930, they numbered only hundreds or a few thousand (Bekenov et al, 1998). After 1930, the population began to recover following the introduction of a hunting ban in 1919, until again being driven towards extinction by overhunting, following the collapse of the Soviet Union (Bekenov et al, 1998). Poaching lead to a 97% reduction in the nominate saiga population, from around a million animals in the early 1990s to an estimated 30,000 by 2003 (Milner-Gulland et al, 2001; Mallon, 2008). In the pre-Caspian the saiga has fared worse; numbers dropped by almost 99%, from around 800,000 in the 1970s (Mallon, 2008) to around 10,000 currently (Saiga News, 2011). The severe winter of 1998/99 saw 80,000 saiga migrate from Kalmykia to neighbouring Daghestan – the vast majority were killed by poachers over the course of a few weeks (Arylov et al, 2004).

In addition to the direct impact of reducing the population, hunting was biased towards males, since they bear horns which are valued in traditional Chinese medicine. This so skewed the sex ratio, that reproductive rates for this fecund, polygynous species crashed (Milner-Gulland, 2003), although it now appears reproductive rates have recovered (Milner-Gulland, pers. comm.). Nevertheless, this recovery is not a truly positive result for saiga conservation, since it probably occurred because the emphasis on males for horn has been replaced by general hunting of both sexes for meat.

2.6 Likely distribution drivers

Previous research highlighted the dynamic effect of drivers of saiga distribution (Singh et al, 2010a). The key drivers vary between seasons and over longer periods of many years (Singh et al, 2010a, 2010b). This variation is influenced by both biological and anthropogenic processes. For example, Bekenov et al (1998) reported that winter distribution was determined by snow depth and density which dictates access to food; whereas spring distribution was linked to quality of vegetation, access to water and, for calving, the level of disturbance (Singh et al, 2010b). It is likely that during the winter the priority is simply to find food, with more emphasis on quantity and less emphasis on the quality of food. However, this changes during spring as the females prepare to give birth and quality is more of an issue along with the need for calves to be born away from possible threats.

In a study of 40 years of data on the Kazakhstan & PC saiga populations, Singh et al (2010a) concluded productivity and precipitation were the key determinants of saiga distribution. However, Singh et al (2010a) noted that avoidance of human settlements had increased markedly in the last decade. It is worth noting the Kazakhstan data were based on aerial surveys, which were biased, did not include error measurement and were likely underestimates of true numbers (McConvile at al, 2009); although this surveying is probably more accurate than the surveying carried out in Kalmykia (Milner-Gulland, pers. comm).

The need to avoid human disturbance (or more specifically poaching) during calving and other times was also found to be a major driver of saiga distribution by Singh et al (2010b), along with man-made barriers and availability of water (Singh et al, unpub.). Saiga have a high dependence on water (Bekenov et al, 1998) and prefer to avoid barriers – natural and man-made (Bekenov et al, 1998).

2.7 Participatory monitoring

When faced with limited resources, participatory monitoring may offer a cost-effective means of identifying species' distributions. Given that conservation is invariably constrained by financial resources, the reduced cost of PM is arguably its greatest benefit. In addition, locals may have better knowledge of a species' ecology and distribution than rangers and enforcement officers (Kuhl, 2008). Whitebread (2008) reported that local monitors were as accurate as rangers in their monitoring of saiga, although generally, the accuracy of local knowledge is highly variable (Danielsen et al, 2005a; Gilchrist et al, 2005). Participatory monitoring can lead to rapid decision-making/responses (Danielsen at al, 2009) and local knowledge may be beneficial in identifying species to researchers in an unfamiliar habitat (Oba et al, 2008).

Migratory species are unlikely to be protected throughout their range, i.e. wildebeest (Thirgood et al, 2004), therefore conservation efforts within PAs may be negated by poaching in unprotected areas. By including local stakeholders, PM creates a connection between the monitored species and communities. Exposing communities to conservation programmes and informing them about conservation can generate local interest in, and value of, the species, which will be beneficial for, and complementary to, conservation. Evidence of this in Kalmykia comes from a study by Howe et al (2011), who concluded that locals were more willing to contribute to saiga conservation when exposed to a media campaign, than if exposed to traditional "fortress" conservation. Perhaps this is unsurprising, given the Kalmyk people have a long tradition of respect for nature and sustainable management of their environment (Basangova, 2004). Yet, other studies (i.e. Aipanjiguly et al, 2002; Dolisca et al, 2009) also concur that increased ecological knowledge promotes greater interest in conservation.

In addition, valuable intelligence on poaching may be gleaned through closer cooperation with locals (FFI, 2009). Before discussing the possible pitfalls of PM, it is worth noting the

CBD advocate the involvement of local people in conservation and encourage (CBD) Parties to ensure the participation of locals (COP 9 Decision IX/18 Bonn, 2008).

In contrast, Gilchrist et al (2005) cautioned against management interventions based solely on local knowledge, due to questions over the accuracy of monitoring. Likewise, Danielsen et al (2005) raised some criticisms, whilst remaining cautiously optimistic about the potential for participatory monitoring (a view shared by Gilchrist et al (2005)). Criticisms included a greater likelihood of bias and lack of precision, although they suggested that potential problems resulting from inexperience, lack of clearly documented protocols (leading to inconsistent monitoring) and unrepresentative sampling, could be overcome through proper planning.

A further drawback of PM is that locals perceive biodiversity as a utility (Oba et al, 2008) and their assessment will reflect these different values. For example, herders require grassland for their cattle, therefore an increase in shrubs and other woody species is considered to be a negative trend (Oba et al, 2008). Although, with awareness of this (understandably) biased view, it should be able to design monitoring schemes which accommodate these biases.

Nonetheless, there is a groundswell of opinion that combining local knowledge with a rigorous scientific approach offers great potential for monitoring species (i.e. Danielsen et al, 2005a; Gilchrist et al, 2005). It can be argued that conservation has been a fickle discipline which has followed different, often diametrically opposed, trends such as "fortress" conservation and community based conservation; time will tell if PM is simply another fad. As with the afore mentioned trends, there are aspects of PM which appear promising and advantageous.

2.8 Species distribution models

Detailed information on the distribution of species is a key component of conservation, but this data is often lacking. To understand the growth in the use of species distribution models (SDMs), it is worth considering the reasons why this species distribution data is lacking. Ground surveys of species and environmental variables can be costly in time and money. Therefore they tend to be fairly limited in their spatial extent, since cost is correlated with the extent of the survey area and cost is normally a constraint on research. Funding of conservation is most limited in developing countries, which often have the highest levels of biodiversity and threats to biodiversity (Rao & Ginsberg, 2010). Multiple researchers/ volunteers may reduce the time and financial cost but will introduce additional individual errors, which may bias results.

Remote sensing (RS) enables monitoring of habitat at broad spatial scales, involves far less time than ground surveys, is often freely available via the internet (i.e. MODIS, SPOT) and therefore does not require the researcher to travel to the area of interest. However, RS data is normally less accurate, and to achieve greater accuracy requires additional input such as expert opinion or ground sampling; though this would likely be less extensive than the afore mentioned ground survey (Strand et al, 2007).

Therefore a combination of RS and ground sampling is often incorporated into an SDM to determine species distribution. SDMs enable assessment of species distribution on a landscape or regional scale, without incurring prohibitive costs. SDMs have many applications within conservation, resource management and academic research, including protected area design, species and biodiversity action plans, and prediction/mitigation with respect to environmental change (Franklin, 2010). Guisan & Zimmermann (2000) comment that static predictive models are generally considered to be empirical, as they are based on field data. Therefore, they are a more realistic and precise representation of a specific species distribution, as opposed to mechanistic or analytical models, which are based on knowledge of the ecological mechanisms or processes and are more general (Guisan & Zimmermann, 2000).

SDMs typically relate the response variable (i.e. presence/absence or pseudo absence data), with possible explanatory environmental variables within a GIS. Explanatory variables will generally fall into one of three categories:- disturbances, resources and regulatory factors (Guisan & Thuiller, 2005). Species may avoid disturbances, both natural or anthropogenic, such as fire or hunting. By contrast, species distribution is limited by the need to access resources, i.e. water or forage. Regulatory factors involve biological conditions beyond the species tolerance, i.e. temperature, fresh or salt water. These data are held as layers within the GIS. The model determines the key explanatory variables of the species distribution, by statistical methods such as logistic regression (generalised linear model - GLM), discriminant analysis, or artificial neural networks (Manel et al, 1999). From this, the species potential distribution may be mapped, informing conservation planning and enabling future monitoring to be more efficiently directed.

In contrast to their advantages, there are disadvantages associated with these models. SDMs require technical knowledge (i.e. processing RS data and understanding of GIS) and are complex. Since they involve a number of elements, such as remotely sensed data, field data, a GIS and statistical methods, they may accumulate individual errors from each of the elements – leading to bias within the model. It is suggested they are data hungry (i.e. Hirzel & Guisan, 2002), although accurate models have been developed with small sample sizes (i.e. Elith et al, 2006). Strand et al (2007) suggest the models may encourage over-confidence in their ability. Since the data used in an SDM is temporally and spatially limited, any predictions are, in theory, only valid within those same limits (Guisan & Thuiller, 2005). In practice, predictions from the model are projected beyond these limits; however, it is worth remembering that the species and its habitat may not be in equilibrium, especially given the possible impact of climate change (Guisan & Thuiller, 2005). In addition, a species may exhibit different responses in different areas to the same explanatory variables. For example, rural species such as foxes are likely to be more wary of human presence than urban foxes. Therefore model predictions may not apply beyond the sample area.

Appropriate spatial and temporal scales are important for the accuracy of an SDM. The spatial scale involves two components:- resolution and extent (Franklin, 2010). Ideally all remotely sensed data will be at the same resolution (the pixel size), though often this is not possible. Data gathered in a field survey at a fine scale may be incompatible with environmental predictors which are measured on a coarse scale (Guisan & Thuiller, 2005).

At a coarse scale, RS may not accurately describe the habitat, i.e. a pixel may be deemed to be bare soil and therefore unsuitable habitat, yet 40% of the pixel may be suitable forage. This would lead to omission errors in the SDM. If the extent of the study area, and therefore the size of the assessed population, is constrained (i.e. by political borders or physical barriers) the response curve may be truncated (Guisan & Thuiller, 2005). This may suggest a different relationship between the explanatory and response variables (Guisan & Thuiller, 2005). For example, in Figures 2.1 and 2.2 the entire sample population exhibits a unimodal response, but the shaded subset in Figure 2.1, from a reduced study area suggests a linear response, whereas the response curve of the subset in Figure 2.2 (shaded) has a negative skew.

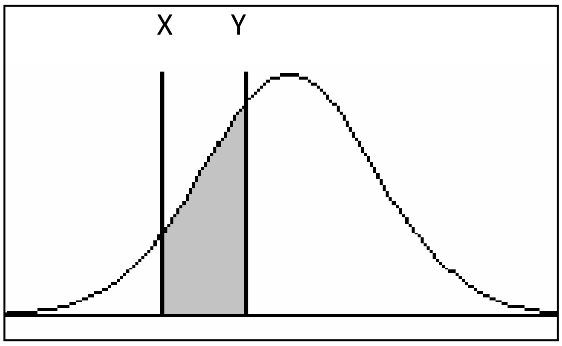

Figure 2.1. Apparent linear response from a population subset

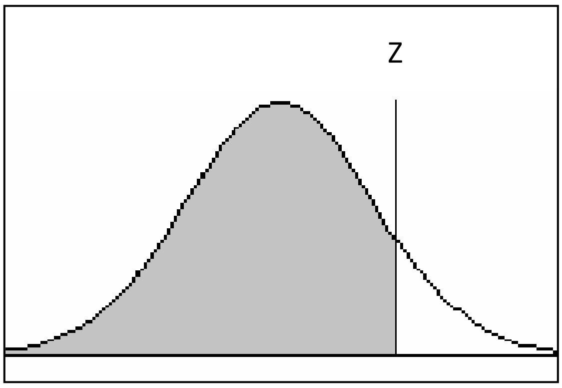

Figure 2.2 Apparent negative skew from a population subset

2.9 Study Area

Located in southwest Russia, Kalmykia forms part of the northern coast of the Caspian Sea. The area is characterized by steppe & semi-desert (Singh et al, 2010a). Vegetation includes *Stipa* and *Festuca* species and shrubs such as *Artemesi* and *salsola* (Singh et al, 2010a). There is an increasing precipitation gradient from the semi-desert of the southeast (170 - 200mm), to the northwest steppe (350 - 400 mm) (Singh et al, 2010a). Kalmykia is generally flat with a harsh climate comprising hot summers and cold winters; temperatures in Kalmykia range from – 35°C in winter to +44 °C in summer (Kühl, 2008). In winter, snow in the south east of Kalmykia is blown away, leaving the land looking black and desolate, hence the name Chernye Zemli, which means "Black Lands" (Badmaev & Ubushaev, 2004).

Saiga are predominantly found in the semi-desert, they tend to inhabit the steppe only during drought or in periods of deep snow. Overgrazing, predominantly by sheep, lead to 80% of Kalmykia being subject to desertification (Arylov et al, 2004; Namrueva, 2004). In the post-Soviet era, levels of livestock have declined dramatically and some recovery of the steppe is evident (Hölzel et al, 2002).

With a predominantly rural population (63%) (Lepretre, 2001), it is unsurprising that the Kalmyk economy is based on agriculture (Ochirova, 2004), with an emphasis on livestock (cattle and sheep). Widespread privatisation of state controlled farming enterprises occurred in the 1990s (Govt. Kalmykia, 2002). Despite the move to the private sector, various indicators confirm Kalmykia's position as one of the poorest regions of Russia. Kalmykia has one of the highest infant mortality rates in Russia, it has one of the lowest percentages of homes with running water and sewerage facilities and is 73rd out of 80 regions, based in descending order, on the HDI index (the Human Development Index incorporates income, level of education and life expectancy – higher is better)(UNDP, 2010).

In contrast, Kalmykia is estimated to have 8.2 billion tonnes of oil and gas condensate (Govt. Kalmykia, 2002; Zaabinvest, no date). To put this into context, if the reserves are proven, then if oil constituted 50% of this figure, that would place Kalmykia twelfth in world rankings of oil reserves (BP, 2011). Oil and gas production could have a notable positive impact on the Kalmykian economy.

3 Methods

3.1 Section outline

The methods section begins with a brief description of the project research strategy. Thereafter, an account of the method of data collection is followed by a description of data processing and preparation within ArcGIS. An explanation of habitat suitability modelling and an assessment of alternative modelling approaches precedes a description of Maxent software, which was used to determine saiga presence probability. The section concludes with a description of Maxent limitations and a sensitivity analysis.

3.2 Research strategy

The research strategy covers the broad process of moving from an initial question, to obtaining a set of results, upon which conclusions can be drawn. Figure 3.1 outlines the strategy involved in this project. Background research on the subject matter was based around the objectives outlined in the Introduction. The potential explanatory variables, along with saiga presence data, were incorporated into layers for analysis in ArcGIS (version 10.0, SP2). The data output from ArcGIS were then analysed by Maxent (version 3.3.3e) and a number of models were produced based on different parameters.

The initial phase of model selection involved no assumptions, in other words, all variables were incorporated and default settings were used. Thereafter, a heuristic process was used to refine the choice of variables and parameters used in subsequent models, in order to select the models used to define the predictive maps of saiga presence. This process took account of the AUC_{Test} scores, AUC Jackknife results, the contribution of variables to the model and a visual assessment of the models' maps. Through a combination of these indicators and knowledge based on the background research, 2 models (November-February and March-June) were selected which were considered to be the most accurate predictors of saiga presence during these seasons. The models were used by Maxent to produce maps showing the presence probabilities for saiga. The rationale behind the choice of software used is included in the software descriptions below.

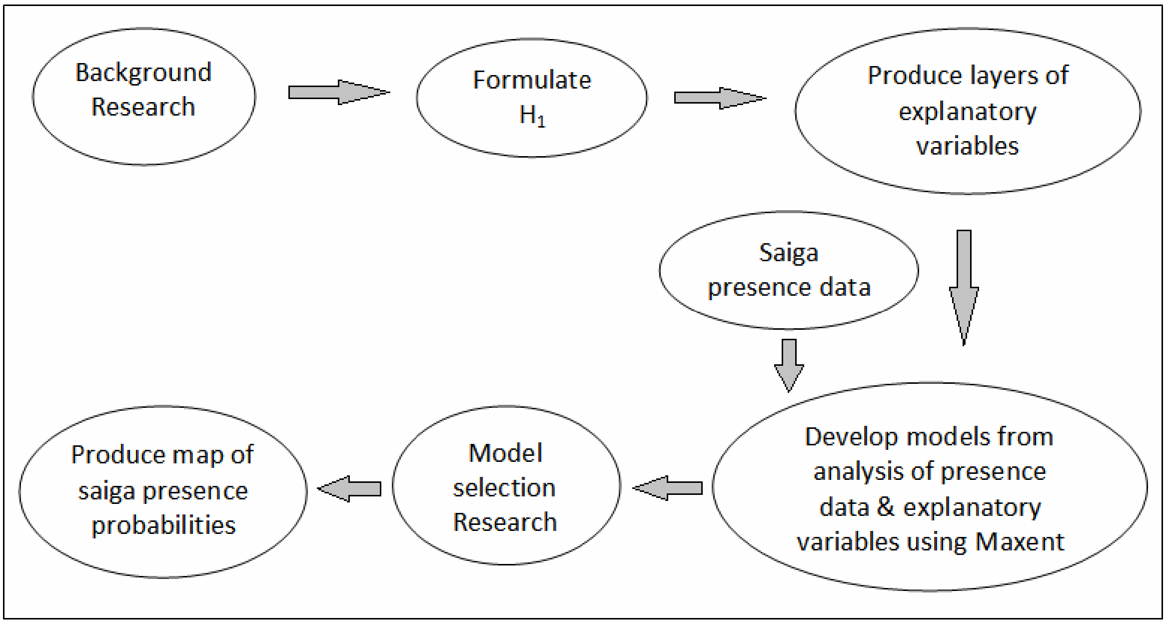

Figure 3.1. Flow diagram of model selection strategy

3.3 Data collection

3.3.1 Field Data

Opportunistic saiga monitoring was carried out by 25 monitors from the end of October 2010 to the end of June 2011. The monitors selected were likely to have the most contact with saigas, i.e. those living in isolated rural communities, especially farmers. Basic training and an equipment pack were provided, along with a small monthly payment. The monitors were clustered around 2 areas:- north of the PA - across central Kalmykia, and south of the PA - in southern Kalmykia (see Fig 3.2). One monitor was located in the unprotected buffer zone between the 2 reserves; one monitor did not report any saiga sightings and was excluded from data processing.

Twenty of the monitors were managed by Professor Yuri Arylov of the Kalmykia State University (hereafter the “CZ monitors”). The remaining 5 monitors were overseen by Anatoly Khludnev, Director of the Stepnoi reserve (hereafter the “Stepnoi monitors” – though this group includes the monitor located in the buffer zone between the 2 reserves). The monitors were located within, or close to, the extent of saiga distribution as identified by Lushchekina & Struchkov (2001), covering the period 1990 to 2000 (Figure 3.3).

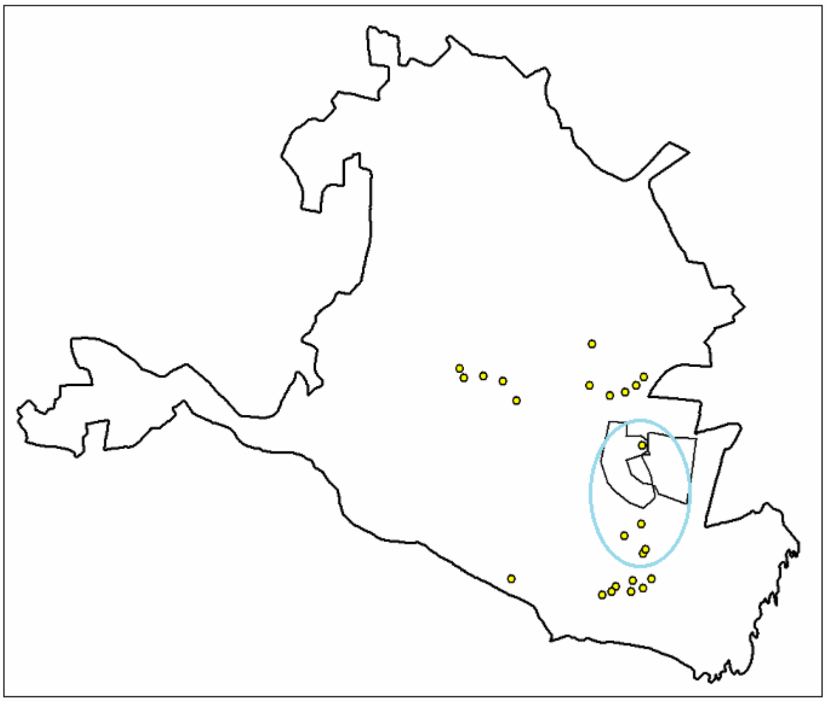

Fig 3.2. Distribution of monitors. The “Stepnoi monitors” are located within the blue circle, the eastern polygon is Stepnoi, the western polygon is CZBR.

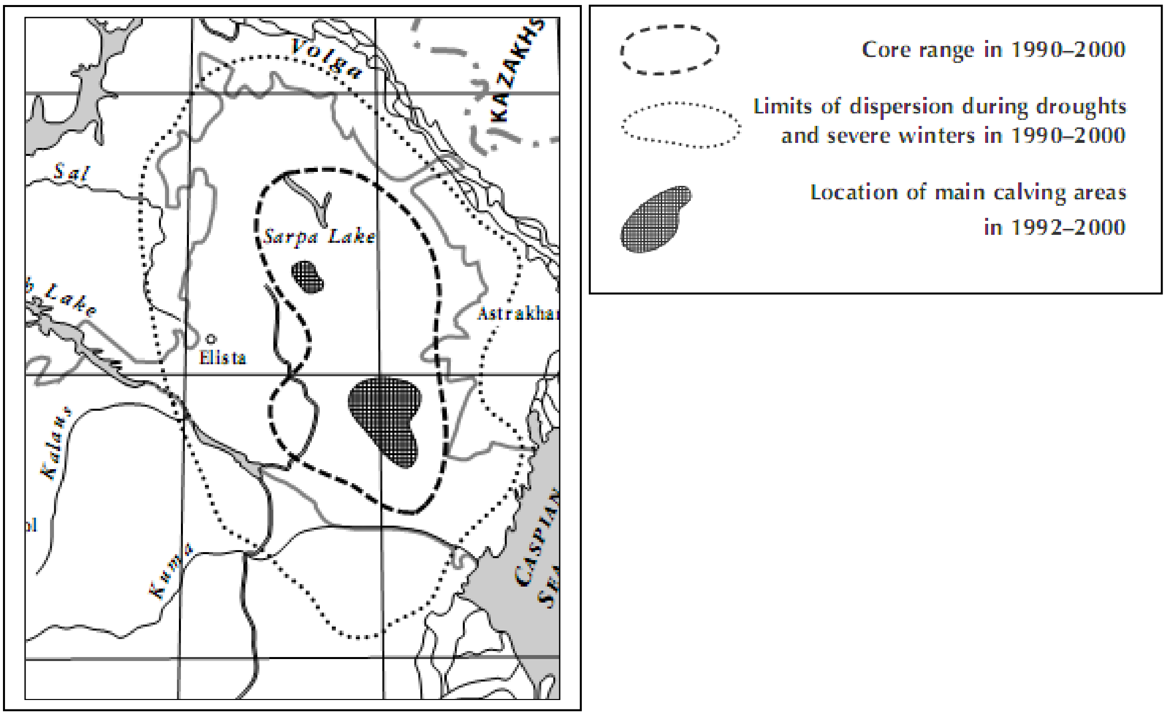

Figure 3.3. Map showing range of saiga in Kalmykia 1990-2000. Source: Lushchekina & Struchkov, 2001.

Although some monitors recorded absences, the majority did not. Some or all of the following variables were recorded by monitors:- actual, minimum and maximum number of saigas; date; time; sex; distance from observer; angle from path and comments. The distribution of sightings was positively skewed (Figure 4.1), therefore it was decided to simply use an indicator of abundance in the data analysis. This indicator specified whether a sighting had been made by a monitor during each month. The latitude and longitude for each monitor's location were provided, along with the distance travelled by each monitor (north, south, east and west). These distances indicated the extent of individual monitoring zones. The total area of the monitor zones (excluding overlapping areas) was 1120.6 km^2, which is 1.5% of the total area of Kalmykia. The mean size of the zones was 49.3 km^2, but individual monitoring zones varied considerably, from 506 km^2 down to zero - this monitor apparently only recorded from their home; while some monitors observed along line transects.

3.3.2 Potential explanatory variables

Distance to settlement (also referred to as "distance to town") was used as a proxy for poaching. This was considered a likely predictor of distribution based on the fact that saiga numbers have been greatly reduced due to hunting. In addition, Singh et al (2010b) assessed 40 years of data on saiga distribution in Kazakhstan and found that in the last 2 decades, the location of saiga calving aggregations is increasingly influenced by human disturbance. It was considered that saiga behaviour may have been affected by poaching, leading to avoidance of human settlements, but also increased occupancy of protected areas. Therefore distance to PA was included as a potential explanatory variable. In support of the theory of increased saiga presence in the PA, O'Neil (2008) reported that saiga were seen by rangers throughout the year. However, traditionally saiga migrated during spring to spend summer in the northern and western parts of the pre-Caspian (Singh et al, 2010a).

Distance to water was also considered a likely predictor as saiga need to drink regularly; Bekenov et al (1998) stated that waterholes were key predictors of saiga distribution. Singh et al (2010a) recommended further research on the influence of distance to water, while Singh et al (2010b) found calving aggregations were located at an intermediate distance from waterholes. A detailed map of Kalmykia featuring most of these variables appears in the Appendix (Figure A.1).

Enhanced Vegetation Index (EVI) was chosen as a proxy for vegetation productivity. This was in preference to the Normalised Difference Vegetation Index (NDVI), as EVI is less sensitive to atmospheric pollution and performs better in habitats with sparse vegetation and/or a high proportion of bare soil (Strand et al, 2007). Vegetation indices are based on measurements of the difference between red and near-infrared radiation (NIR) reflected by vegetation (Solano et al, 2010). As a fundamental part of the process of photosynthesis, radiation in the red wavelength is absorbed by leaf pigments, whereas NIR is reflected (Solano et al, 2010). There is a positive correlation between this difference and the quantity of vegetation (Solano et al, 2010). EVI uses the positive correlation between atmospheric aerosols and the difference in blue and red reflectances to account for atmospheric pollution (Solano et al, 2010). The EVI equation is

$$EVI = G \times (\rho NIR - \rho red)/(\rho NIR + (C1 \times \rho red - C2 \times \rho blue) + L)$$

where G = 2.5, C1 =6, C2 = 7.5, L = 1, ρNIR, ρred, ρblue is reflectance of NIR, red and blue bands (Solano et al, 2010).

Snow depth would have been a useful variable, but was not available for Kalmykia, instead percentage snow coverage (per pixel) was used. Distance to road and distance to rail were included, as these lines of communication represent physical obstacles to saiga. Investigation of the influence of barriers on saiga distribution was suggested by Singh et al (2010a). A rail line runs parallel with the canal, so in effect they represent 1 barrier. Fire may well be a driver of saiga distribution, but the peak period for fires was June – August and therefore outside the temporal scope of this study (Dubinin et al, 2010).

EVI and snow data were obtained from the MODIS Terra satellite (www.mrtweb.cr.usgs. gov). Shapefiles for settlements, water bodies and the Kalmykia border outline (extracted from a world map) were obtained from GIS-lab (http://gis-lab.info/qa/ vmap0-eng.html).

The roads and canal layers were obtained from gdata (http://www.biogeomancer.org/software.html).

3.4 ArcGIS

3.4.1 Data processing and preparation

All the potential explanatory variables were projected within ArcGIS to the UTM (Universal Transverse Mercator) coordinate system, zone 38N (Geographic Coordinates System: WGS 1984). In addition to preparing the data for Maxent processing, ArcGIS provided a useful overview of the data. In the same way that plotting results is useful prior to statistical analysis, mapping the various layers helps gain a spatial insight into the area of interest. For example, from the maps it was clear that monitors were clustered and therefore monitoring was spatially biased. The time series of EVI and snow tiles provided a view of temporal changes to the landscape. The maps also suggested areas which, based on background research, appeared to be good potential saiga habitat. These assumptions could then be contrasted with the predictive maps, promoting analysis of why assumptions do or do not match the predictive maps.

Due to inconsistencies in the level of absence reporting by monitors, it was decided to use only presence data. Since the latitude and longitude for the sighting were not recorded, the co-ordinates of the monitors' locations were used as a proxy for the presence points. Potential explanatory variables were created as raster files, having their boundaries set to the Kalmykian border and then converted to ASCII (American Standard Code for Information Interchange) grids. All grids had the same extent, cell size and resolution, as required for processing by Maxent. Many of the data preparation tasks were incorporated into the ArcGIS ModelBuilder, a workflow tool which allows different functions to be joined to produce a single data processing task.

Although there was high spatial variation in the monitoring zones, a minimum monitor buffer size of 200m x 200m was used, as smaller spatial elements would not be processed by the GIS raster statistics analysis functions. The water layer was edited to include a section of a canal bordering the western CZBR and man-made waterholes in the Stepnoi PA. A map of the waterholes was produced by Anatoly Khludnev; the canal edit was based on a Google Earth map. The canal extends from the south-east of Kalmykia to the north-west. Saiga in CZBR drink from the canal (E.J. Milner-Gulland, pers. comm.), however, only the section adjoining the PA was added to the layer, since most of the canal appeared dry on the Google Earth map. Much of the territory bordering the remaining section of the canal was determined to have 50% or higher probability of saiga presence by many of the models. Therefore exclusion of this section of the canal from the water layer is unlikely to have unduly influenced the model.

For most rasters, each pixel held the distance (m) to the nearest occurrence of a variable (i.e. distance to water or distance to PA), but the EVI raster comprised the EVI value (after multiplying by the scale factor of 0.0001 (Solano et al, 2010)) within each pixel and the snow raster held the percentage of snow cover within each pixel. Snow data were at 0.05^{o} resolution, the EVI resolution was nominally 250m (231.66m actual). Individual EVI rasters spanned 16 day periods, in total covering November 1, 2010 to June 24, 2011; monthly snow rasters covered November 2010 to June 2011. The data were aggregated to produce EVI and snow rasters for the 2 seasons of interest:- November-February (winter) and March-June (spring), using the Workspace To Raster Dataset function. The integrity of some of the EVI and snow tiles were affected by cloud and in the case of EVI, snow and ice. Details are described in the Data Limitations section.

3.4.2 Data limitations

The western and north-western extremities of Kalmykia were excluded from remote sensing - MODIS reported these areas as “no-data”, however this area is beyond the saiga core range suggested by Lushchekina and Struchkov (2001). The quality of the EVI data was poor for the period November 17 to February 1, although EVI is more likely to influence the spring models. The quality of the snow data was poor during December 2010. Again, this variable contributed little to the models.

There is a question over the accuracy of the water layer: 2 of the monitors locations were located within water bodies and 3 of the settlements overlapped water bodies. However, no other sources of hydrology in Kalmykia were available.

3.5 Habitat suitability modelling

Modelling methods fall within a number of broad categories such as generalised regression (including GLMs and generalised additive models - GAMs), classification techniques (classification tree and maximum-likelihood classification), machine learning (such as Maxent) and environmental envelopes (boxcar and convex hull) (Guissan and Zimmerman, 2006; Franklin, 2010). Regardless of which of these categories they derive from, habitat suitability modelling involves 2 key datasets:- the response variable (i.e. present or absent) and the potential explanatory variables (such as precipitation, elevation, habitat type). The methodology of the generalised regression approach provides a relatively simple insight into one method of habitat suitability modelling, including how these variables are incorporated (Figure 3.4)

Figure 3.4 illustrates the response (absent or present on islands) of a species of bird to 2 explanatory variables:- area of island (km^2) and isolation - measured as distance (km) from the mainland (Crawley, 2007). The data points on the isolation graph form 2 more or less distinct groups in relation to distance, producing a curve with a steeper gradient. Generally the bird was more likely to occupy islands close to the mainland, although there was some overlap on the graph, as absences were recorded down to 5km from the mainland. By comparison, there is a greater degree of overlap in the data points indicating response to area. Larger islands were likely to be occupied, but both presence and absence were recorded on smaller islands, which is reflected in the shallow gradient of the curve. This result suggests distance to mainland has more influence on occupancy than the size of the island.

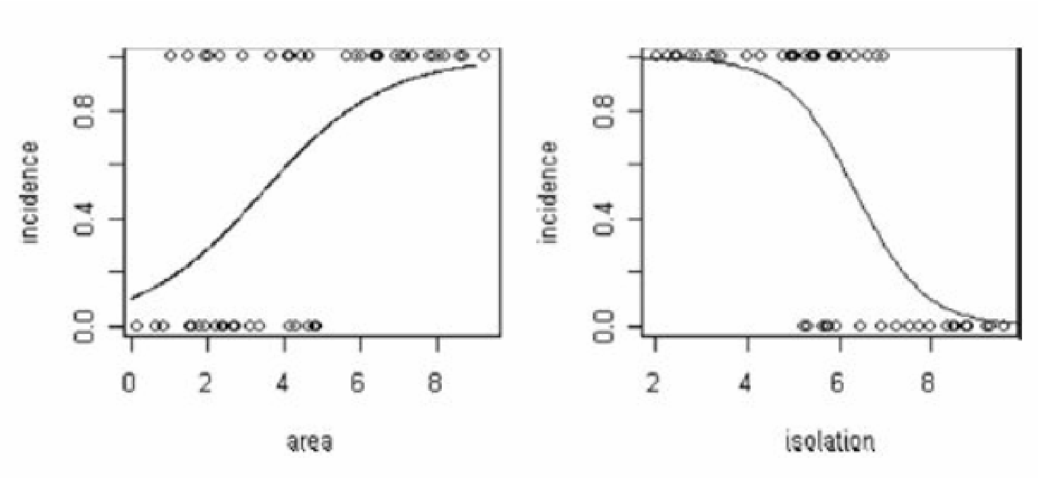

Fig 3.4 Example of response to 2 explanatory variables: area and distance from mainland (isolation). Source: Crawley, M.J. (2007) The R book..

Habitat suitability models are often equilibrium models (such as the models developed during this research), that is to say they do not reflect changes to the system over time but are static (Milner-Gulland and Rowcliffe, 2007). The severe decline in saiga numbers indicates the system is not in equilibrium, however one of the key project aims was to identify the current saiga distribution. Furthermore, repeated monitoring and modelling can produce a time series of predicted habitat, which will reflect temporal changes in saiga distribution.

3.6 Alternative modelling options

Two other approaches were assessed for their suitability to produce predictive distribution models. The first method was a GLM, however the R interface used to produce the model is a scripting programming language, which requires a higher level of technical competence than the Maxent application. The use of GLMs would have been more informative and transparent, but a key constraint when considering modelling options was the limited time frame. Maxent was chosen as it was relatively easy to learn and could readily produce results. Predictive maps were developed in R, but required a lot of computer memory and took a long time to produce. The high number of models produced in the available time would not have been possible using R.

The second method tested was the Geostatistical Analyst facility within ArcGIS. This software allows statistical and spatial analysis within the same application; examples of its use include predicting suitable habitat for the Tasmanian Devil (*Sarcophilus harrisii*) (McGovern, 2005) and predicting the distribution of fish larvae in the Dover Strait (Koubbi et al, 2006). The Geostatistical Analyst application requires a high technical capacity and involved unfamiliar statistical techniques such as radial basis functions, kriging and cokriging, which would have required a longer time to become proficient than the Maxent application.

3.7 Maxent

Maxent uses the maximum entropy (= maximum unpredictability) method for determining a species geographical distribution. Maximum entropy involves selecting the distribution which reflects only what is known, but is otherwise the most uninformative distribution from the data (Shipley, 2010; Warren and Seifert, 2011), which may appear counter-intuitive. However, any other approach would either incorporate unsupported assumptions or require ignoring available information. Maxent derives a distribution for which the values for each of the features (environmental variables), equal their mean value across the sample points (Phillips et al, 2006).

Maxent provides a user-friendly interface. In this respect, it provides a quick and relatively easy means of producing results. Maxent was also chosen because this application does not require absence data. Instead, presence data is compared with a number of background points (the default is 10,000) which the program selects. The software is in common use and was found to produce accurate predictions of species distributions, even when using small, undesigned, presence only samples (Elith et al, 2006; Franklin, 2010). Maxent also appears to be insensitive to location errors, with errors of up to 5 km reported to have little or no effect on the performance of Maxent, although other techniques such as GLM, GAM and multivariate adaptive regression splines (MARS) performed equally well (Graham et al, 2008).

Part of Maxent's robust performance is due to the fact that the user can control how accurately the model fits the data (Elith et al, 2006; Baldwin, 2009). This process - regularisation - penalises models with heavily weighted features or many parameters

(Anderson and Gonzalez, 2011). However, regularisation can also be a weakness of Maxent.

Since the sampling was spatially and temporally biased (Figure 3.2), a bias grid was produced which reflects the relative sampling effort across the target region. For example, values of 3 and 1 within the grid confirm the ratio of probability of sampling 2 areas, i.e. the former was probably sampled 3 times as often as the latter (Phillips, no date). Most of the target area was not sampled, but since 0 is not allowed in the grid, a value of 0.1 was used to indicate zero sampling (other lower values such as 0.001 and 0.00001 had produced erroneous results).

There were only 24 sampling values (i.e. > 0.1) within the grid, representing the 24 monitor locations. Each value was the number of days that monitor had monitored. An alternative grid was produced, in order to see if the model would be improved if these values were extended within the grid, to correspond with the monitors' buffer zones. In other words, this was a comparison between a grid which reflected only the specific single locations which were provided for the monitors and a grid which reflected the monitors' buffer zones. However, when run against the full period of data (November-June), the "buffer zone" grid only achieved an AUC value of 0.6. Consequently, the "presence points" grid was used to produce the various models. Maxent uses this grid in its selection of background points; areas with higher sampling levels are attributed higher numbers of background points, in order to balance out the sampling bias (Maxent help document, no date).

When considering the contribution made by variables, the "Percent contribution" (PC) in the Maxent output was used rather than "Permutation importance" (PI). The PC reflects the contribution of a variable to the increased gain of the model (Phillips, no date). The PI reflects how the AUC changes as a variable's value is permutated among the training and background points (Phillips, no date). The PC can vary with each iteration as a model converges (is fitted to the data), so repeated runs of the same model can produce different PC values, whereas the PI is calculated once the model has converged and so is fixed (Phillips, no date). On reflection, the PI may have been a better indicator of variable contribution. That said, the PC formed only part of the reasoning behind model choice. In addition, the same presence points were used for testing throughout. An iterative approach which varied the choice of training and testing points for each choice of variables and parameters would have been more informative.

3.8 Maxent Limitations

Overfitting is a problem with the maximum entropy approach (Hernandez et al, 2006; Rodda et al, 2011). A model which is a good fit to the data may not be transferable beyond the study area. On the other hand, a more generalised model is likely to be a less accurate representation of distribution drivers. To mitigate overfitting, Maxent provides the user with a regularisation option (the Beta setting), which can be applied in 2 ways. The model can be constrained by imposition of user determined ß settings on individual parameters (as opposed to the default ß settings), or a single ß multiplier can be set, which simply multiplies all the default ß values by the value selected by the user. However, Warren and Seifert (2011) comment on the tendency for users to rely on the default settings, without investigating the impact of changing these settings, or at best simply changing the multiplier. This may partly result from the lack of advice on appropriate alternative settings (Warren and Seifert, 2011).

Nevertheless, using different regularisation settings (either individual ß values or the ß multiplier) produces widely differing results (Anderson and Gonzalez, 2011; Warren and Seifert, 2011; Rodda et al, 2011). The sensitivity of Maxent to these ß settings, leading to high variation among models, is acknowledged by Phillips and Dudik (2008). This is an important point, as the aim of developing a species distribution model is to achieve the maximum possible accuracy (indicated by metrics, of which AUC is commonly used), whilst producing a model which accurately reflects the underlying biology and abiota. It is possible to achieve high AUC scores, without actually knowing how these parameters are affecting the model (beyond knowing that the complexity is increasing or decreasing) and also, therefore, how well the model reflects the underlying biology. This lack of transparency also applies to the fact that with the same data, different parameters can produce predictions that the same specific area has a high or low probability of saiga presence. Based on the Maxent results it is unclear whether this variation reflects a limitation of Maxent or an inconsistent response to the explanatory variables used in the models. In this sense, a GLM-based approach would have been more informative.

The Maxent default settings are based on the best average values obtained in studies on a number of species from all regions of the earth (Anderson and Gonzalez, 2011; Warren and Seifert, 2011), but these “average” settings may not be appropriate for the species under consideration. The Maxent default values were selected partly based on AUC (Phillips and Dudik, 2008) but Lobo et al (2008) advise against the use of AUC, citing several reasons. These include the fact that AUC applies equal weights to omission and commission errors and is unduly influenced by the extent of the target area (Lobo et al, 2008). If the area extends beyond the distribution of the species, predicted absences outside the normal distribution will be correct, which increases the AUC score (Lobo et al, 2008). Furthermore, since these default values were evaluated in the studies using set-aside test data, both training and test data are subject to the same sampling bias and are not truly independent (Anderson and Gonzalez, 2011; Warren and Seifert, 2011).

In addition to the question of whether test data are independent of training data, a number of papers (i.e. Fielding and Bell, 1997; Manel et al, 2001; Anderson and Gonzalez, 2011; Warren and Seifert, 2011) question the value of evaluating models on their predictive performance. Indeed, Manel et al (2001) state that many model evaluations are “inherently misleading” having found that predictive accuracy is affected by the species prevalence.

3.9 Sensitivity Analysis

Having identified these limitations from exploratory modelling and background research, a rigorous sensitivity analysis was performed, to test Maxent performance with respect to various parameters. Comparisons could be made between results obtained using the default settings and a variety of different parameters, which would then inform the model selection process.

Anderson and Gonzalez (2011) found that higher AUC values were obtained when the regularisation setting for linear, quadratic and product features (ß LQP) was set between

1.5 and 3.0. Therefore I used the mid-point of this range (2.3) when testing non-default LQP settings. The hinge parameter adds flexibility by allowing the gradient of the response to alter (Elith et al, 2011). The default setting for the hinge was 0.5, but extreme values of 0.9 and 0.1 were tried to test their effect. The Maxent help file recommends the Product parameter is used on a sample size > 79. This parameter fits simple pairwise interactions to the model (Elith et al, 2011). However, the GLMs (run within R) suggested interactions between variables were influential, therefore the Product threshold was set to 18 (= the number of training points in the models) for some runs, to test its effect.

4. Results

4.1 Results overview

This section begins with some descriptive statistics of the monitoring data. Thereafter the results obtained from the modelling process for November-February are detailed, followed by the results from March-June. Comparisons between the 2 seasons are also described, in addition to comparisons resulting from the sensitivity analysis.

4.2 Data description

Figure 4.1 highlights the skewed distribution of saiga numbers seen by individual monitors from November to June. Throughout the monitoring period, mean numbers per sighting by 21 of the 24 monitors ranged from 1 to 74 saigas; 14 of the monitors saw between 1 and 10 saigas. Mean numbers for the remaining 3 monitors were 127.7, 149.2 and 349.6. Although, as Figure 4.2 illustrates, there was a wide variation in actual numbers seen. Figure 4.3 combines the mean numbers of saiga per sighting with the monitors' positions and also the number of sightings. Several of the monitoring zones overlapped, in one case 3 zones overlapped and in another, one zone was entirely located within another. This may have lead to double counting - 2 monitors report the same group of saiga. All the monitors with the lowest mean numbers per sighting were located at greater distances from the PA, the three monitors with the highest mean numbers per sighting were located closest to the PA (excluding the monitor who was within the PA buffer zone).

A useful metric would have been the ratio of actual sightings to total observations, as an indication of presence probability. This was not possible as only 6 monitors recorded absences. The mean number of saigas seen by the Stepnoi Group was 130.5, for the CZ group the mean was 16.8. The mean number of days on which sightings were recorded was 31.8 for the Stepnoi group and 7.9 for the CZ group. There was a correlation between number of monitoring days when sightings were made and mean saiga per sighting (r_s = 0.59, P = 0.02).

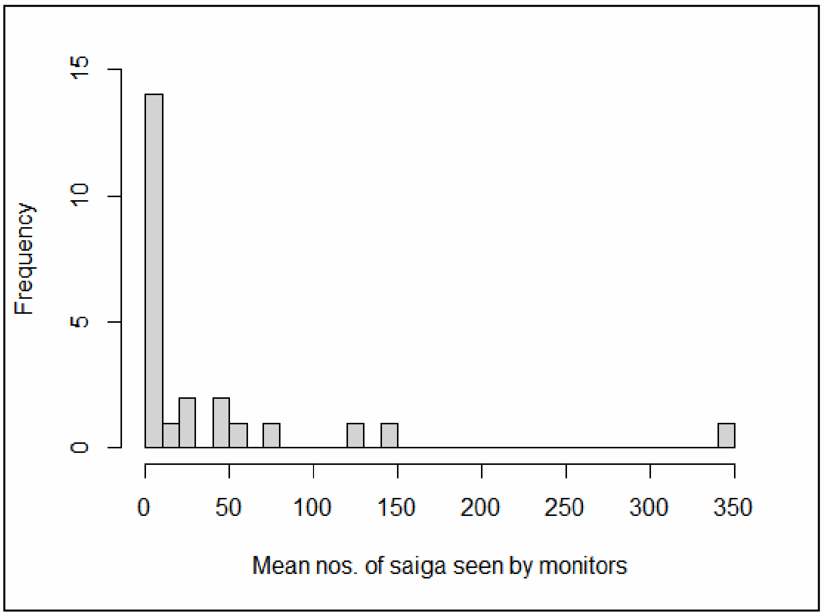

Fig 4.1. Histogram of mean saiga numbers per sighting by monitors Nov 2010 – Jun 2011.

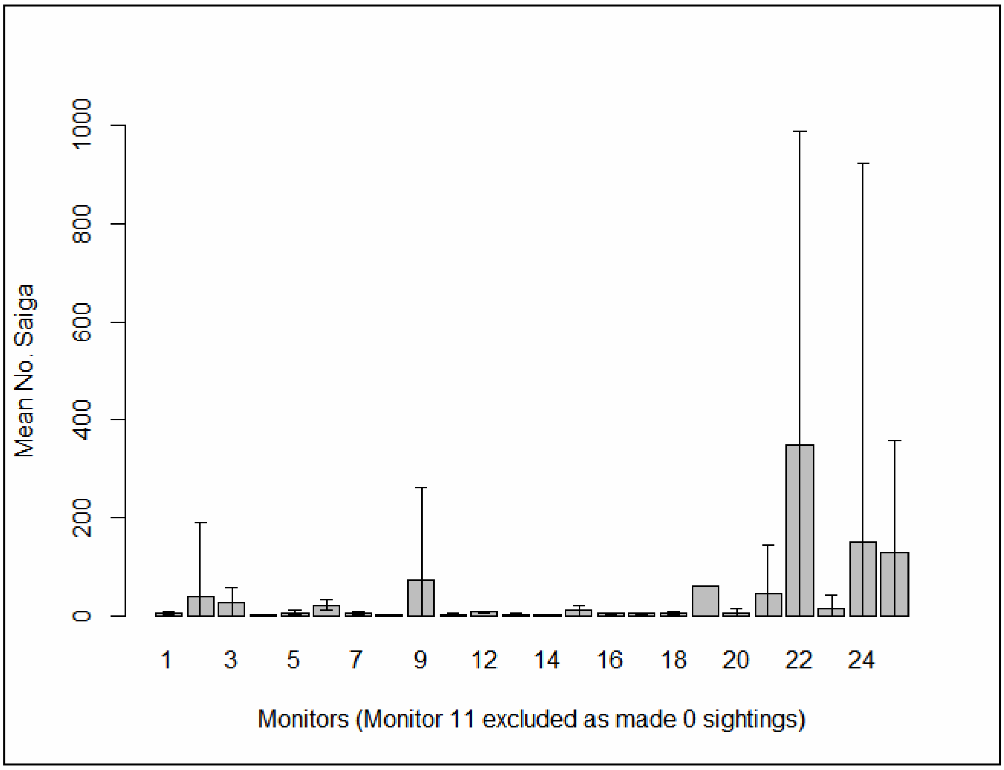

Fig 4.2. Mean saiga numbers per sighting by individual monitors (whiskers = 1 SD) Nov 2010 – Jun 2011.

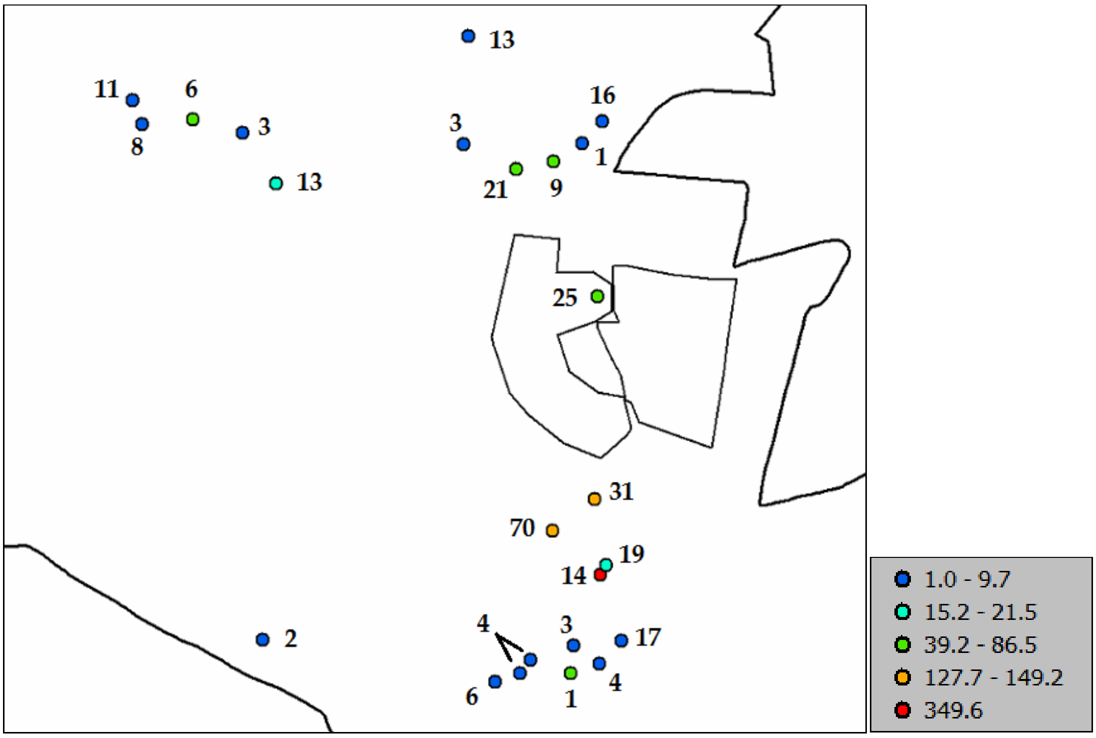

Figure 4.3 Map of monitor locations with mean number of saiga per sighting (colour coded) and total number of sightings (labels) for Nov 2010 – Jun 2011.

When comparing the seasons, Figure 4.4 illustrates that more saiga were seen during spring than winter (median: Mar-Jun = 9.5, Nov-Feb = 6.0; means: Mar-Jun 107.7, 47.8 Nov-Feb). The difference could not be tested as data were not normally distributed and the group sizes were unequal (in spring there were fewer observations and 5 monitors made no observations)

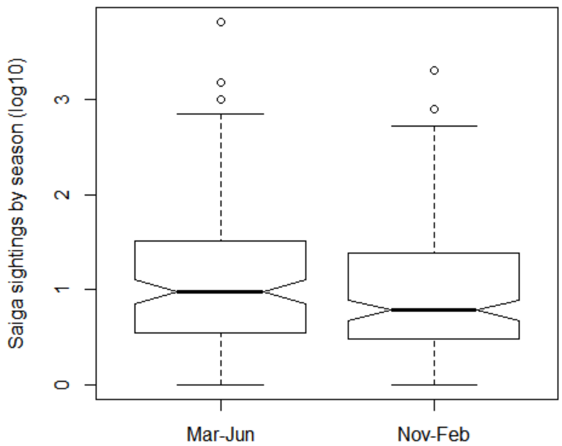

Fig 4.4. Boxplot of saiga numbers (log10) for Nov-Feb and Mar-Jun 2010/2011, bold line = median

For a spatial comparison of saiga numbers, monitors were divided into north (of the PA) and south (of the PA) groups (Figure 4.5); the monitor located within the PA buffer zone (mean sightings = 46.4) was excluded from this analysis. A higher number of saigas were seen in the south than in the north (Wilcoxon 2 sample test: $W = 6204.5$, $p = 7.0e\text{-}07$).

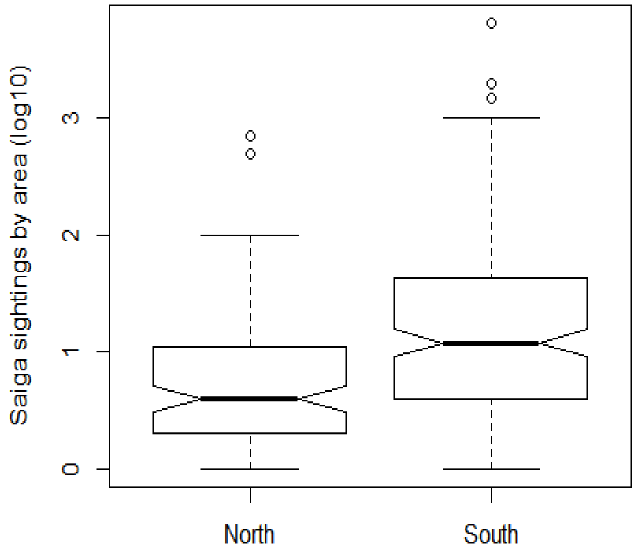

Figure 4.5. Boxplot of saiga numbers (log10) seen in areas north and south of the PA, bold line = median

4.3 Model assessment and selection

4.3.1 November – February

The starting point of the modelling process was a comprehensive model using all variables and default settings (model 16 - Figure 4.6). The map predicts high presence probability across a large part of southern Kalmykia, with additional areas across central Kalmykia. Most of the north and west of Kalmykia were predicted to have low or zero probability of saiga presence. Low or zero probability of occupancy was predicted for most of Stepnoi and CZBR (for reference, Figure 4.15 includes the PA outline), although an area of high presence probability is evident around the canal which adjoins CZBR. The border area to the north-east of Stepnoi has a high prediction of occupancy. The model achieved an AUC_{Test} score of 0.894, indicating a high level of transferability (Figure 4.7). The contribution of the variables are listed in descending order in Table 4.1. This order was generally seen for all models that included all variables; snow, EVI and distance to road contributed the least to the models, distance to PA and distance to water clearly contributed the most, with distance to PA making the highest contribution. Distance to town and distance to rail contributed relatively little to the models. Model descriptions and AUC values are listed in Appendix 1, Table A.1.

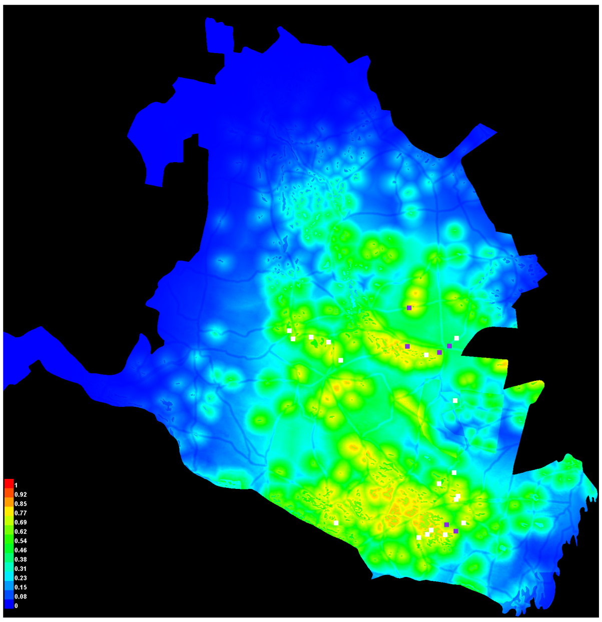

Figure 4.6. Predicted saiga presence during winter using all variables and default settings

Table 4.1 Contribution of variables to model shown in Figure 4.6

Variable	Percent contribution
Distance to PA	49.5
Distance to water	34.7
Distance to rail	8.9
Distance to town	5.9
Distance to road	0.9
EVI	0.1
Snow	0

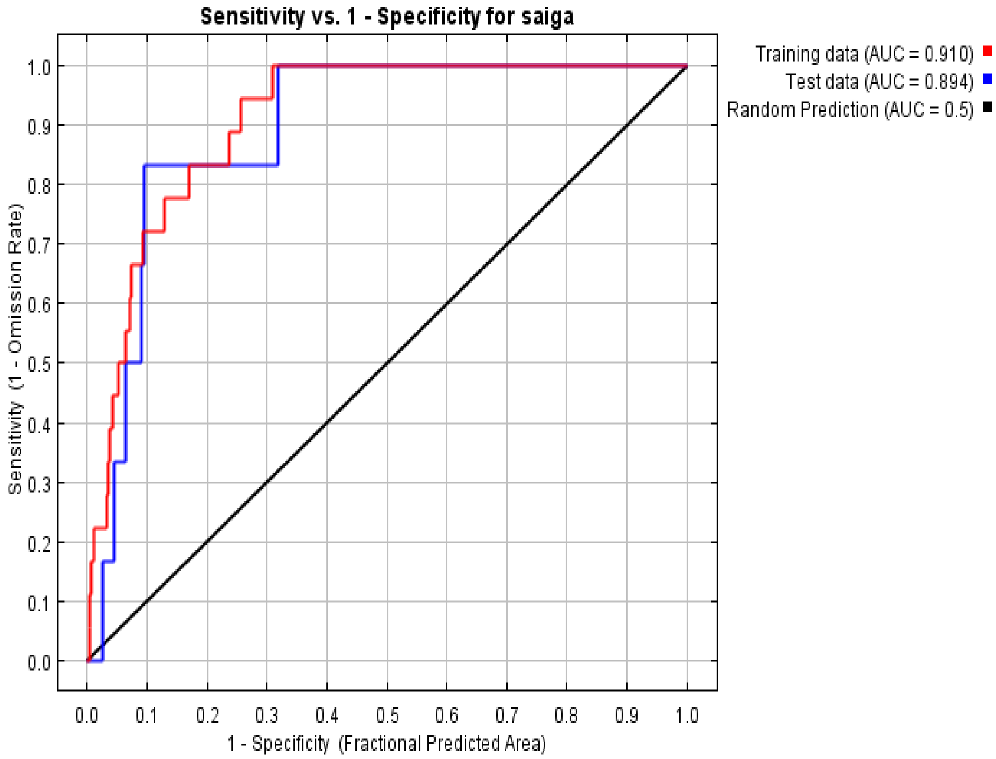

Figure 4.7 ROC, AUC_{Train} and AUC_{Test} for model in Figure 4.6

Figure 4.8 illustrates the extreme differences that were obtained from changing one parameter. The only difference between models is that model A had a hinge setting of 0.9, model B had a hinge setting of 0.1; all variables were included. Model A achieved the third highest AUC_{Test} score among the winter models, although the scores for both models were fairly similar:- A: 0.906; B: 0.828. Model A shows a large contiguous area of central and southern Kalmykia having medium/high presence probability. In contrast, the same region in Model B shows a more fragmented presence probability, with some areas having 0 probability while other localised areas have a probability around 90%. Notably, most of the Stepnoi PA has a medium/high probability of saiga presence in B, while in A most of Stepnoi has low or 0 presence probability. White dots indicate training points, violet dots are testing points.

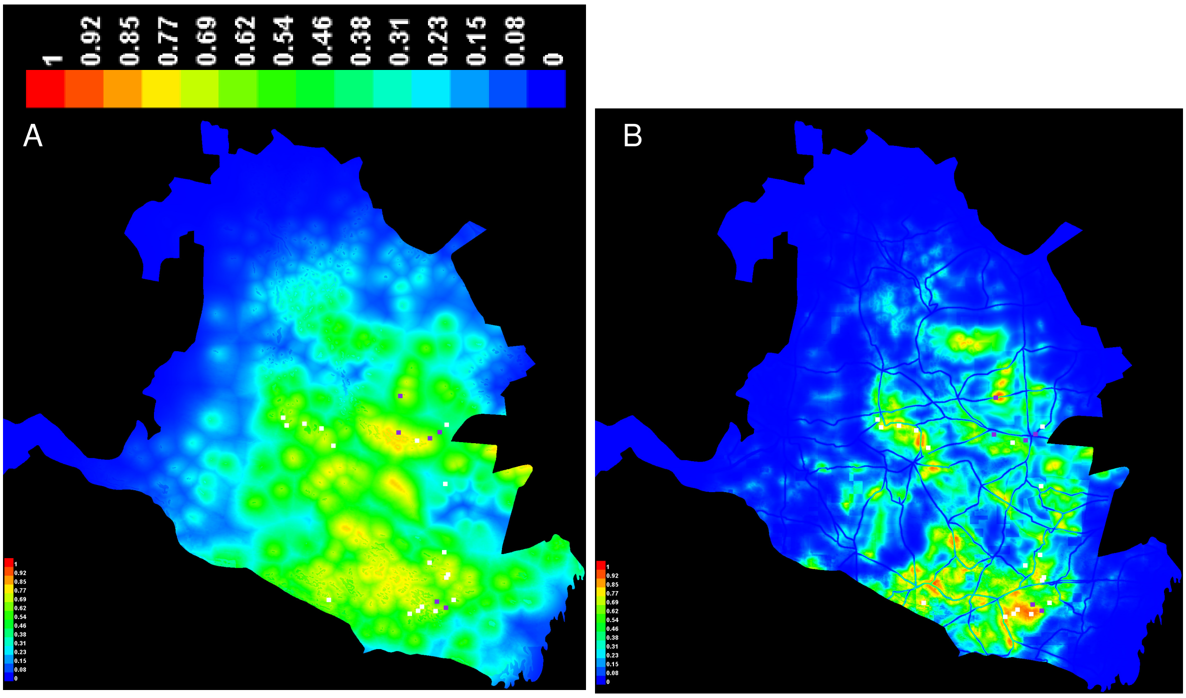

Figure 4.8. Maps of saiga presence probability based on Nov-Feb data. Both models used all variables and default settings, but with hinge set to 0.9 (A) and 0.1 (B). The Maxent colour scale indicates presence probability.

Figure 4.9 depicts 2 models which achieved the same AUC score (0.837) using distance to water and distance to PA. The only difference was that model A incorporated 2-way interactions. On a local scale, model A shows the area in and around the PA to have medium/high presence probability, whereas model B shows that much of CZBR and the area surrounding Stepnoi has low probability of saiga presence. This contrasts with the close agreement on areas of high presence probability at a broader scale.

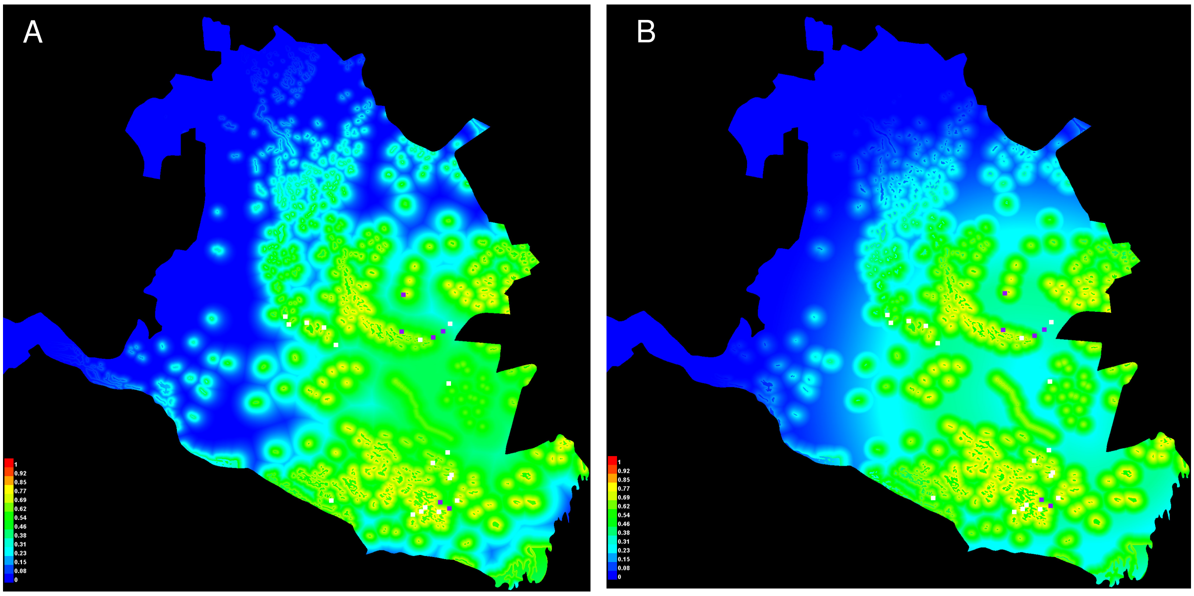

Figure 4.9. Maps of saiga presence probability based on Nov-Feb data. Both models used only distance to water & distance to PA and had the same AUC score (0.837). Model A incorporated 2-way interactions, otherwise default settings were used for both models.

Of the 6 presence points used for testing, 5 were between 2 and 4 km from water, the other point was 11.7 km from water. Distance to water was consistently the second highest contributor to the models. However, Jackknife results from models 10, 11, 15, 16 and 21 showed that AUC_{Test} would be improved by omitting this variable (Figure 4.10). In addition, Jackknife results from all models that used only distance to water and distance to PA, showed that AUC_{Test} would be improved by omitting distance to water, leaving distance to PA as the only explanatory variable. This seems counter-intuitive, given the contribution of distance to water. However, model 12 (Figure 4.11) excluded distance to water, snow, EVI and distance to road, and achieved the second highest AUC_{Test} score (winter) of 0.908. Generally there is similarity between predicted areas of occupancy in this model and most of the other models.

With regard to the issues of veracity of AUC and Maxent sensitivity, the analysis is based on a small dataset, which is sensitive to noise and sampling error. Unexpected results may derive from limitations of Maxent and/or the use of AUC as a metric, or due to noise and/or sampling error, or all of these aspects.

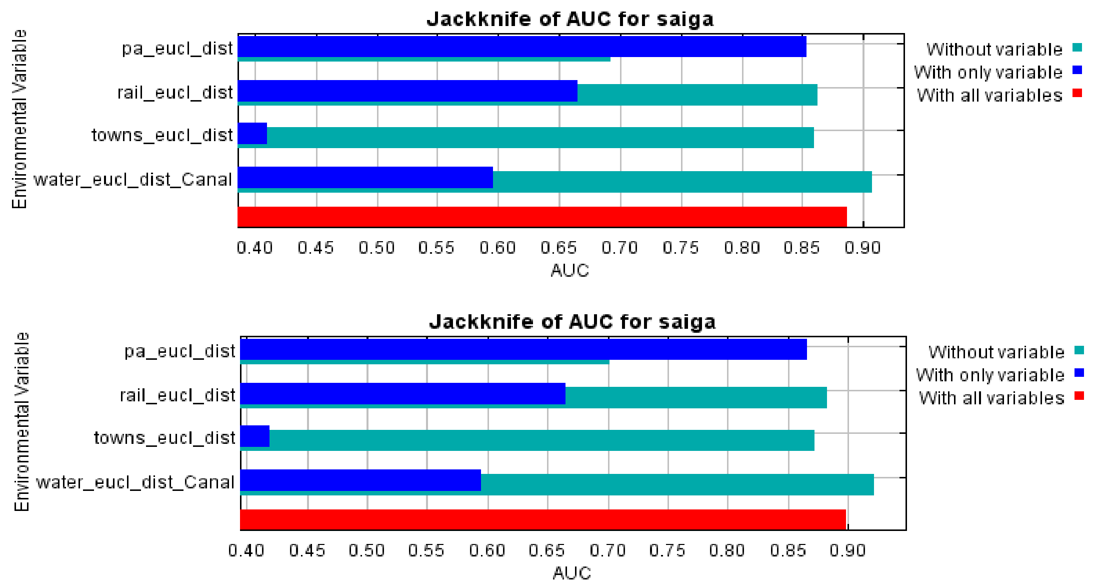

Figure 4.10. AUC Jackknife results for models 10 (top) and 11 show predictive power would be improved by excluding distance to water (water_eucl_dist_Canal).

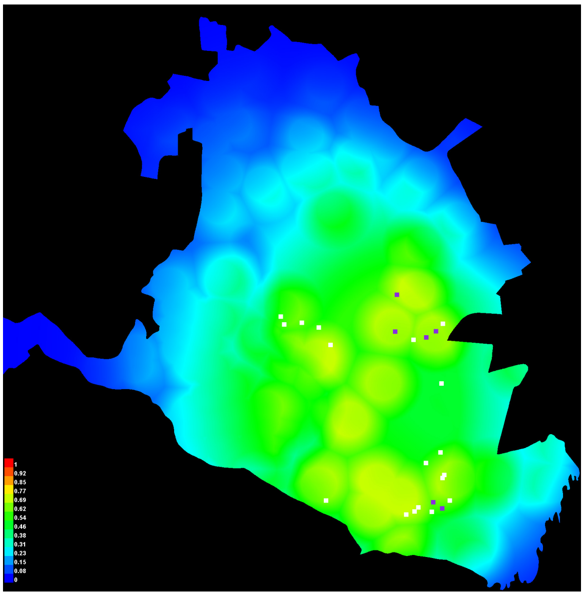

Figure 4.11. Map showing saiga presence probability based on 3 variables:- distance to town, distance to PA and distance to rail.

Phillips (no date) suggests it is normal for the AUC_{Train} to exceed AUC_{Test}, but many of the models had a higher AUC_{Test} score (Appendix 1). One of these models was re-run using half sample size (n=13), the AUC_{Test} score was again higher and the difference between AUC_{Train} and AUC_{Test} doubled, suggesting the model was correct (since fitting the model to half the sample size meant it was more generalised). This suggests that with a small dataset there is less likelihood of overfitting, since the models achieved good AUC_{Train} and AUC_{Test} scores.

In contrast to the marked variation achieved by changing parameters, a high degree of consistency is evident among many models with different parameters, in relation to areas of high presence probability (yellow and orange areas). This consistency is most noticeable among models using only distance to water and distance to PA (i.e. Figure 4.9(A & B)), but also to a lesser degree among other models (i.e. Figure 4.8(A)). Probability of saiga presence increases as distance to PA and distance to water decrease. Therefore, the choice of a definitive model was limited to models which used only these variables.

Given this general agreement between models, selection was based on assessment of the predictive maps (see Discussion). Figure 4.12 (model 22) highlights 6 areas of high presence probability outside the PA, during winter. Based on distance to water and distance to PA, the model achieved an AUC score of 0.859, which was not the highest score. It is likely the 3 highlighted border areas east of the PA are part of the same area, but appear fragmented due to the border between Kalmykia and Astrakhan, as this study only considered Kalmykia. Much of Stepnoi is predicted to be occupied, as is the western border of CZBR adjoining the canal, although the rest of CZBR has a relatively low likelihood of occupancy by saiga.

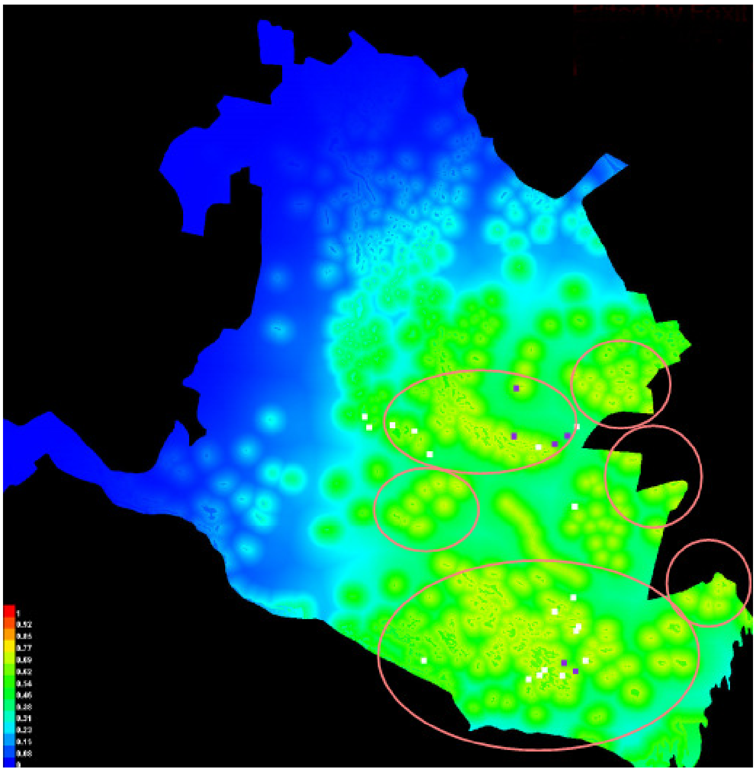

Figure 4.12. Areas with high probability of saiga presence during Nov-Feb. Those areas outside the PA are highlighted by red ovals. Only distance to water & distance to PA were used in this model. Hinge was set to 0.9, other parameters used default settings

4.3.2 March – June

Generally, results resembled the winter period: snow, EVI and distance to road contributed the least to the models, distance to PA and distance to water contributed the most, with distance to PA always making the highest contribution. The models predicted saiga presence increased as distance to PA and water decreased. Since snow and EVI were the only variables which could vary between seasons, AUC scores were similar between seasons for models sharing the same parameters and variables.

The predicted areas of occupancy for the comprehensive models (using all variables with default settings) for spring (Figure 4.13) and winter (Figure 4.6) were very similar with only subtle differences between the predictions. The contribution of variables to these 2 models was also similar, the main difference was that EVI made a slightly higher contribution to the spring model (Table 4.2). The spring model achieved an AUC_{Test} score of 0.898 (Figure 4.14). Many models produced Jackknife results showing predictive performance would be increased by omitting distance to water, which reflected the results for winter. Given these similar results, fewer models were run for the spring season, as choices were informed by the model selection process used for the winter period.

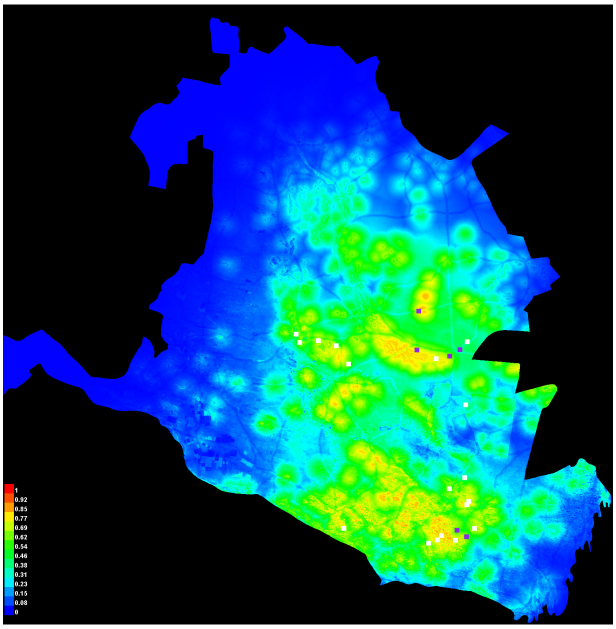

Figure 4.13. Predicted saiga presence during spring using all variables and default settings

Table 4.2 Contribution of variables to the model shown in Figure 4.13

Variable	Percent contribution
Distance to PA	49.1
Distance to water	32.8
Distance to rail	8.8
Distance to town	5.7
EVI	2.5
Distance to road	1.0
Snow	0

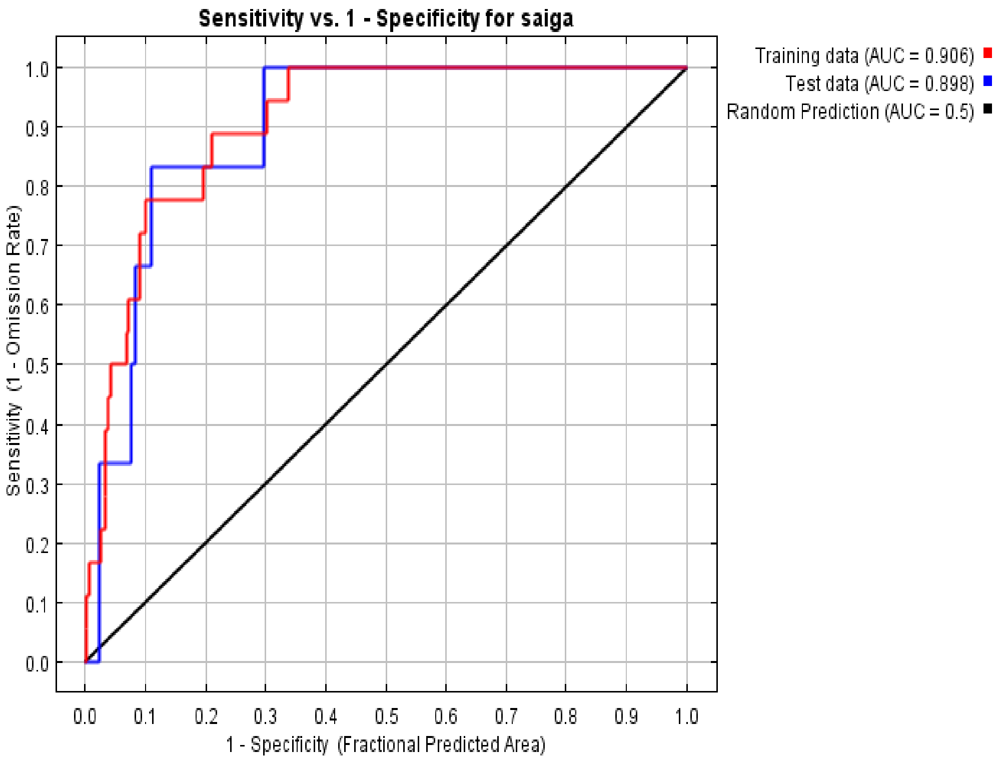

Figure 4.14 ROC, AUC_{Train} and AUC_{Test} for model shown in Figure 4.13

Changing parameters produced similar contrasts to those seen during the winter period. For example the most marked contrast was again evident when comparing a hinge setting of 0.9 with 0.1, using all variables. The mapped predictions were very similar to those produced with the same criteria for the November-February period (Figure 4.15). Nonetheless, there were subtle differences relating to the level of probability for some predictions. Models A and B in Figure 4.15 used the same criteria, but a higher probability was evident across central and southern areas in spring (A), with presence probability around 90% (orange area).

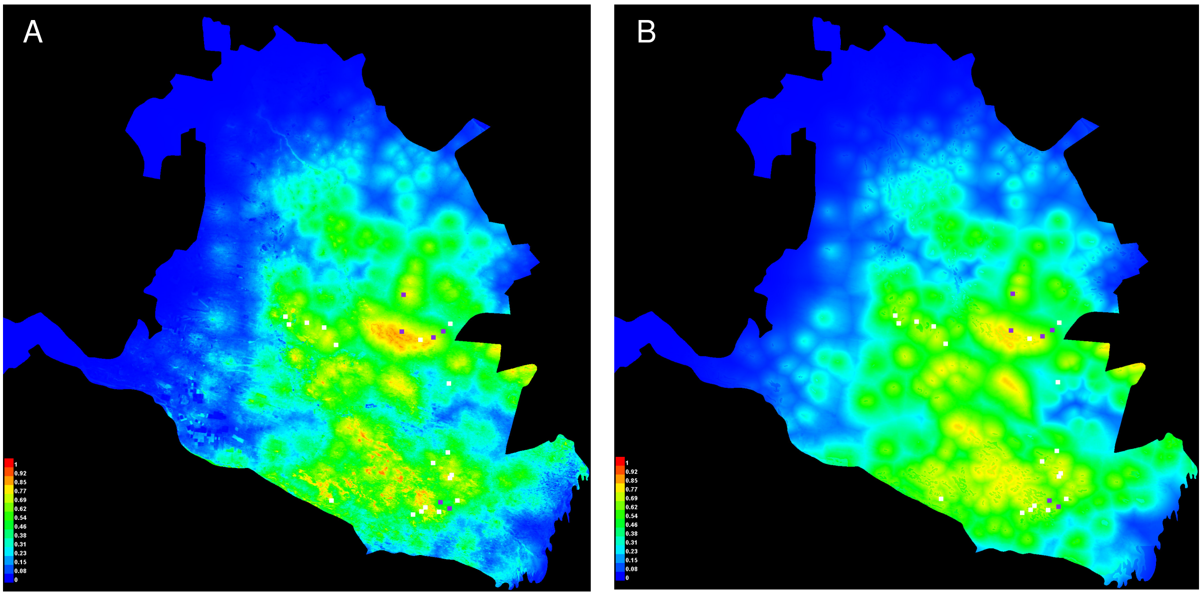

Figure 4.15. Maps of saiga presence probability for Mar-Jun (A) and Nov-Feb (B). Both models used all variables with hinge set to 0.9, all other parameters used default settings. Note the higher presence probability (orange/red) across central and southern areas in A.

The model in Figure 4.15 (A) achieved an AUC score of 0.908 and was unique in that EVI made the third highest contribution to the model (8.5%), slightly higher than distance to town (8.2%). Based on these results a model was run which included only distance to water, distance to PA, EVI and distance to town. This model (Figure 4.16) achieved the highest AUC score of 0.917 (although the contribution of EVI was only 5.4%).

As with the winter period, a consensus was evident among several models in relation to areas of high presence probability (though with less spatial agreement than for the winter period). Those areas outside the PA are shown in Figure 4.16 and generally correspond to the areas of high presence probability during winter (Figure 4.12). The main difference when compared with Figure 4.12, is that much of Stepnoi and the southern part of CZBR are predicted to have low or zero probability of saiga presence during Spring, although the north-western sector of CZBR has a high probability of occupation. These findings, in respect of saiga distribution in and around the PA, agree with Rangers' sightings in late June and early July 2008 reported by Whitebread (2008).

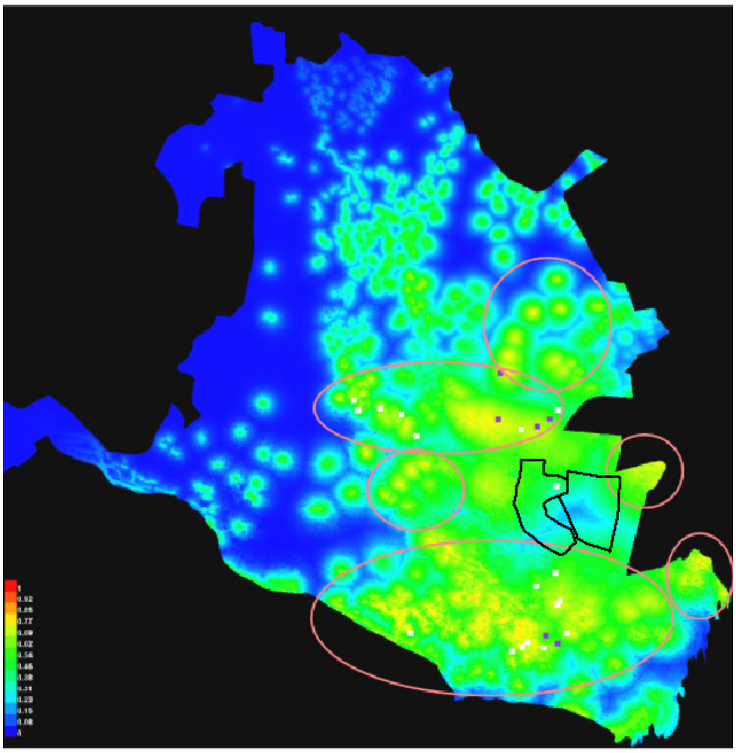

Figure 4.16. Areas with high probability of saiga presence during Mar-June. Those areas outside the PA are highlighted by red ovals. Model used distance to water, distance to PA, EVI and distance to town. Hinge was set to 0.9, otherwise default settings were used.

5 Discussion

5.1 Overview

This section discusses the results obtained, in light of the research objectives identified in the Introduction. It will summarise the findings and draw conclusions on the three main areas of research:- the distribution of saiga during the winter and spring seasons, including an assessment of the explanatory variables which drive this distribution and any seasonal differences within these variables; the usefulness of PM data for saiga conservation and the efficacy of the Maxent modelling application. Limitations of the research are discussed and recommendations made on further conservation action. Concluding remarks form the closing section.

5.2 Participatory Monitoring

The provision of PM data enabled this research project, so in that respect participatory monitoring has proven useful. In addition, PM data provided new information on saiga from the south of Kalmykia, effectively extending the study area of previous research, increasing our knowledge of saiga distribution and identifying new areas occupied by saiga. Although it was not possible to measure the accuracy of the data provided, Whitebread (2008) found that monitors in Kalmykia were as accurate as rangers. Even if improvements were needed in aspects of the PM data, it provides a starting point on which to build. Any weaknesses in monitoring can be identified and rectified; in this respect, some recommendations are made in this paper.

Whitebread (2008) determined that monitors were more cost effective than rangers in terms of absolute cost and time spent monitoring, but were less cost effective in terms of costs per sighting. Their cost-effectiveness in terms of absolute cost is a key issue, since this is the more important cost, in terms of the budgets of CZBR and Stepnoi. Whitebread (2008) compared salary with number of sightings to conclude that monitors were less cost effective per saiga sighting. However, sightings only represent part of the underlying biology; the importance of recording absences became apparent during this research (see Maxent discussion below).

Consistent reporting of absences by all monitors would improve the cost effectiveness of monitors, in addition to improving the quality of data. The project found a positive correlation between number of monitoring days which resulted in a sighting and number of saiga seen. However, the results do not support any conclusions on whether monitoring effort, saiga abundance or their interaction has the most influence on sightings.

A high level of sightings were reported in 3 localised areas, but a more even spatial distribution of monitors would produce a more uniform distribution of sightings across a greater area, which would be more informative. An additional advantage of increasing the monitoring area is that this would raise awareness among locals, of saiga conservation across more of Kalmykia. This may then result in a higher number of Kalmykians becoming actively involved in saiga conservation and research, and an increased profile for saiga among decision makers.

Local knowledge, from which participatory monitoring derives, already exists and is widespread among communities which have a close relationship with the natural environment (i.e. farmers, shepherds and other inhabitants of rural communities) (Oba et al, 2008). Considering the financial constraints on conservation (which are likely to become more severe, in the current economic climate), perhaps the question is not, whether we should make use of participatory monitoring, but whether we can afford not to. Training and rigorous scientific assessment of data quality, should increase the likelihood that PM data that informs saiga conservation, is both precise and accurate. Given that Whitebread (2008) reported a favourable level of accuracy already exists, participatory monitoring is a resource which has the potential to make a positive contribution to saiga conservation.

5.3 Model choice

The choice of model for the March-June period was relatively straight forward. There was agreement between a number of models and the model selected had achieved the highest AUC_{Test} score. Model selection for the November-February period was less precise. Again consensus was evident among several models and selection was limited to those models which used only distance to water and distance to PA. Although model 24 (see Appendix 1) had the highest AUC_{Test} score among models which only included these 2 variables, the model predicted high presence probability around waterholes, but excluded the Stepnoi waterholes and the canal which adjoins CZBR. With no evidence to support this exclusion, the model was discounted. In addition, the definitive maps selected for both periods exhibited a gradient from high to low probability areas. This was considered more realistic than the sharply defined border between areas with different probabilities which were evident in other maps, given the relatively homogeneous habitat in Kalmykia.

5.4 Saiga distribution

An extensive region predicted to have medium or high probability of saiga presence was clearly defined across much of central, eastern and southern Kalmykia. Leon (2009) found that the extent of saiga in the south had reduced markedly, based on survey data, although the surveys were spatially biased, as a much higher number of villages were surveyed across central Kalmykia, than in the south. The results of this research suggest the opposite; for both seasons, the largest areas with high presence probability were located in the south of Kalmykia. It appears that much of the border region between Kalmykia and Astrakhan has high presence probability, though this cannot be confirmed as no data were available from Astrakhan. This area of medium or high presence probability forms a more or less contiguous region surrounding most of the PA, evident for both seasons (Figures 4.12 and 4.16).

Compared with the saigas' core range for 1990-2000, identified by Luschekina & Struchkov (2001) and adapted by Singh et al (2010), the results describe a range which has contracted latitudinally and increased longitudinally. In other words, the predicted saiga presence has extended eastward but the northern part of the range has reduced (but see Limitations). The predicted saiga presence in Kalmykia has become more compact and centred about the PA; if this is an accurate reflection of the actual range, this would emphasize the importance of Stepnoi and CZBR. However, since saiga occupancy has now been recorded from areas which were not previously monitored, the results do not support a conclusion that the actual saiga range has altered, but this project has identified new areas for future monitoring.

5.5 Distribution Drivers

Given the sensitivity of Maxent to different parameters, the consistency of predictions of areas with high probability of saiga presence, were in contrast to the general variation in results. This consistency adds weight to these predictions. Distances to PA and water consistently contributed most to the models for both seasons. Based on the background research, these variables were expected to influence saiga distribution. The results suggest these 2 variables are the main drivers of distribution during the winter and spring seasons, explaining most of the variation in saiga presence; though given the issues over Maxent, caution needs to be exercised.

The fact that distance to PA was the key determinant of saiga distribution has both positive and negative implications. It suggests that Stepnoi and CZBR are fulfilling their functions as areas where saiga are protected. The importance of the PAs confirms the prudence of investing resources in protected areas and could provide leverage with regard to further investment. It should also provide motivation for PA personnel and promote an appreciation for the PA among locals. However, the fact that saiga are more likely to be seen in and around the PA, corroborates the existing body of evidence that poaching continues to be a problem in Kalmykia.

Saiga distribution was expected to be based close to water sources. These results echo the findings of Singh et al (2010b) and Bekenov et al (1998), who concluded that distance to water was a key predictor of saiga distribution in Kazakhstan. The importance of water for saiga is a double-edged sword. Construction of artificial waterholes in areas lacking water sources may increase the likelihood of saiga presence. Furthermore, waterholes can be located in areas which are optimal from a conservation perspective. For example, in areas which are easily managed and protected, or in areas which are between isolated populations, to encourage genetic flow between the populations. In contrast, saiga could be vulnerable to predators or poachers at waterholes, though with the reduced saiga population this may be less of a problem. If both saiga and livestock use the same waterholes, there is the potential for disease transmission between species.

The results contrast those of Singh et al (2010a), who found saiga migration was driven by precipitation and vegetation productivity. However, direct comparisons cannot be drawn between the 2 studies, due to spatial and temporal differences. Singh et al (2010a) assessed 40 years of data on saiga distribution across the pre-Caspian and Kazakhstan, compared with 8 months of monitoring data from central and southern Kalmykia, which were analysed in this research.

Nevertheless, the impact of winter precipitation is an interesting point. Deep and dense snow can affect saiga distribution, if it inhibits or prevents access to food. If winters are less severe and saiga can still forage, then distribution is likely to be dictated by other variables. Kalmykia experienced a dzhut in 2009/10, which forced saiga to move out of the PA (Khludnev, 2010). However, there were no reports of a dzhut for 2010/11. Therefore, it is not surprising that snow was not considered a contributing factor for saiga distribution during 2010-2011. Although dzhuts are sporadic, they have a strong negative impact on saiga, in terms of mortality and distribution. Furthermore, if saiga are forced to migrate away from protected areas, as in 1998/99 (Arylov et al, 2004) and 2009/10 (Khludnev, 2010), they may be exposed to poaching.

Given the impact of poaching on saiga, it was considered that distance to town (a proxy for hunting) would be influential in determining saiga distribution. The theory being that saiga would be more likely to be found away from settlements. Therefore, the fact that generally this variable made little contribution to the models was unexpected, especially as saiga presence is correlated with proximity to the PA. On reflection, this could be because poaching near settlements involves a higher risk of detection. Poachers are highly mobile; Kuhl (2008) reported a positive correlation between saiga poaching and motorbike ownership in Kalmykia. Therefore It would be prudent for poachers to operate away from settlements.

It was expected that EVI would also be influential in saiga distribution during spring, as previous research found that vegetation productivity was a key driver of saiga distribution during this season (Singh et al, 2010a; Bekenov et al, 1998). Results obtained in this research do not support that view, as generally EVI contributed little to the predictive models. This contrast may be due to the differences in time frames of the research. Singh et al (2010a) assessed data spanning 40 years (they used NDVI as a proxy for vegetation productivity); the time frame for Bekenov et al (1998) is unknown. Interannual variation of NDVI was evident and Singh et al (2010a) concluded that saiga opt for areas with low interannual variation. The time frame for this research project (8 months) was not long enough to reveal any interannual variation.

Despite the *prima facie* suggestion that EVI is not a key determinant of saiga distribution, it is worth noting that EVI made a higher contribution to model 16 (spring). The initial models for spring (which included all variables) predicted low or zero saiga occupancy in Stepnoi and CZBR, suggesting a change in the ecosystem, since saiga were predicted in the PA during winter. However, generally the contribution of the other variables did not indicate they were having a higher influence on saiga distribution during spring. Therefore, the influence of EVI is worth considering in future research, incorporating monitoring within the PA.

Having a model which comprises just 2 variables should make conservation planning more straightforward; although the contribution of other covariates should not be overlooked. Resources can be allocated to conservation with these 2 variables in mind. Cost-effectiveness of conservation interventions should be high, certainly compared with investing the same resources but with many variables in mind. For example, further monitoring can be focussed in clearly defined areas. Monitoring of areas relatively close to the PA and water, should result in higher levels of saiga presence, while lower numbers of saiga should be present in areas which are distant from water and the PA. Intelligence gathering and anti-poaching interventions can likewise be targetted in areas which are close to water and the PA. From a logistics viewpoint, all interventions should be more efficient when based on a simple model. On the negative side, arguably the advantage of a simple explanatory model also applies to poachers – they are more likely to find saiga in areas close to water and the PA, although this covers an area of several thousand km^2.

5.6 Efficacy of the Maxent application

On the positive side, the Maxent application is relatively easy to understand and use, and requires less technical expertise than a scripting programming language, such as R. Models can be quickly developed and a comprehensive output provides a map of the prediction, measurements of model accuracy, details on the contribution of individual variables to the model and also their impact on the model's predictive power. The choice of recommended default settings or user defined settings provides flexibility. Based on the AUC scores the predictive models were both accurate and transferable (but see Model Metrics below), since they exceeded the 0.75 threshold suggested by Elith et al (2006).

However, Maxent was rigorously tested using a variety of parameters and was found to be highly sensitive to different parameter values. Although the sensitivity analysis identified potential limitations with Maxent, it is acknowledged that this analysis was based on a small dataset, which incorporated spatial bias (although use of the bias grid should have mitigated this effect) and was sensitive to noise.

It is possible to achieve high variety in predictive outputs simply by changing values for one parameter. Depending on the entered values, models can then predict the same area as having high, low or zero probability of species presence. At best, the user is confounded by a wide choice of models, with little or no information on how to choose the best model, beyond the AUC score - the veracity of AUC is discussed below. At worst, the user may choose a high scoring model, which could be the worst indicator of the underlying biology.

Part of the problem stems from the lack of documented advice on the choice of value, also there is only limited information on how this choice affects the model and no advice regarding how different models relate to the underlying biology. Put simply, it would be useful if there was some guidance linking parameter settings to different biological scenarios, for example, taking account of complexity of community structure and habitat classification. In total, 56 models were produced during this research; one or more of these models may be very accurate predictions of saiga distribution, or they could all be moderate or poor approximations of saiga distribution. Ultimately, the choice of model should rely on a metric of model accuracy and background research or background/expert knowledge.

On the other hand, whereas some models may require a high degree of precision, with regard to presence probabilities (such as models used to direct monitoring programmes), that requirement did not apply to this project. The objectives were based around relative questions and identifying the most important variables. In that respect, Maxent was able to produce adequate predictive models.

There are various sources advocating the use of Maxent (i.e. Elith at al, 2006; Franklin, 2010) but recently papers have been published (Anderson and Gonzalez, 2011; Warren and Seifert, 2011; Rodda et al, 2011) highlighting limitations of Maxent which were evident in this research. The contrast between Maxent advocates and those who identify its limitations may confirm the view of Warren and Seifert (2011), that many users just rely on the default settings and have not tested the effect of changing parameters.

With contrasting views on the utility of Maxent, it would be prudent to effect further, similar research using different approaches and applications, i.e. GLMs. Comparison of results and validation through monitoring can then inform an assessment of the best application to use. For conservation generally, the issues over modelling could have serious implications, especially since funding is normally a constraint on any action. These implications relate to conservation action taken, allocation of resources for monitoring, allocation of hunting quotas and gazetting of protected areas.

5.7 Model metrics

Maxent uses AUC, which is accepted as a standard of model accuracy. However, there may be a tendency to rely overly (or solely) on this measure, for example in the case of students or in relation to a novel invasive species, if there is no existing knowledge relating the species to its new habitat. This research has identified issues relating to the use of AUC as a measure of model accuracy. The criticism of AUC is less about its veracity as a measurement, and more about its performance in indicating the best models which accurately reflect the underlying biology. Many models were produced which achieved the same or similar AUC scores, yet exhibited marked differences in their predictions. Also, many AUC scores for Jackknife tests suggested omitting distance to water would improve predictive power. This seems counter-intuitive, contradicts the contribution made by this variable to the model and also the conclusions of Singh et al (2010b) and Bekenov et al (1998). In these cases, AUC is not an informative measure. Similar concerns were raised by Warren and Seifert (2011), who found that various models with high predictive scores (including AUC_{Test} and AUC_{Train}), were poor representations of the underlying biology and abiota. Rodda et al (2011) suggest AUC is a poor metric for choosing between models.

At a broader level, the use of predictive power as a measure of model accuracy is also questionable. Use of additional metrics (such as AIC) would be advisable to confirm the accuracy of predictive models produced by Maxent.

5.8 Project limitations

The main limitations were sample bias, inconsistent reporting of absences and dataset size. There was wide variation among monitors in the number of days when sightings were reported. These days were used in the Maxent bias grid, but they may or may not be representative of sampling effort. The dataset was small, comprising 24 presence points. For model training and testing purposes these numbers were further reduced; Maxent used 18 presence points for training and 6 points for testing.

A small dataset will be more sensitive to noise and bias. Consistent reporting of absences would have allowed testing for a correlation between number of monitoring days and number of saiga and sightings. Absence points could also have been incorporated into the Maxent algorithm. Nevertheless, the context of these limitations is that this research adds to the existing knowledge on saiga distribution in Kalmykia. Limitations and errors are highlighted and can inform and improve subsequent research.

5.9 Recommendations

5.9.1 Future monitoring proposals

Consideration needs to be given to the monitoring protocol and this process should involve all stakeholders. In addition to a wider, more even distribution of monitors, a method for more unified monitoring is also needed, with regard to recorded data and monitoring periods. It is recommended that the role of absences in data processing is clearly explained to monitors; in addition to clarifying the importance of this information, it may provide motivation for monitoring during periods when saigas are absent. Latitude and longitude are key elements which need to be recorded, even if the details are approximations based on maps. Consideration should also be given to integrating ranger sightings and PM data and loading the data into a database. Increasing the number of monitors and the level of supervision should reap benefits.

The Government of Kalmykia approved a plan to carry out aerial surveys and thermal imaging of saigas (Saiga News, 2010). With this in mind, a novel and inexpensive form of aerial surveying which is worth considering is carried out in Maryland (USA), using radio-controlled model aircraft (see www.fws.gov/northeast/pdf/obrecht.pdf).

5.9.2 Political issues

Given the likelihood of a transborder saiga population, cooperation between Kalmykia and Astrakhan is likely to make a positive contribution to saiga conservation. Although an ongoing territorial dispute (Milner-Gulland, 2009) presents a challenge to transborder conservation. The majority of Kalmykia's oil reserves are located in the south (Kalmykia Govt., 2002; ZAAB, no date), which is where a large area of potential saiga habitat is predicted. Oil exploration and production could have serious implications for saiga conservation (i.e. through disturbance and pollution). This is a development which needs to be monitored.

5.9.3 Future research

Monitor 11 reported no saiga sightings but commented on the presence of wolves. Similarly, Leon (2009) found that 17% of locals believed wolves were responsible for controlling saiga distribution. Further research could analyse the impact of wolves on saigas and establish baseline data for future analysis of trends in wolf predation. Research on the impact of climate change on the migratory behaviour of the saiga is also needed. Due to the likelihood of a transborder saiga population, any future research should consider extending the study area into Astrakhan.

5.10 Concluding remarks

An extensive area surrounding the PA and possibly extending into Astrakhan were identified as potential suitable habitat for the saiga. The key predictors of saiga presence were distance to PA and distance to water. This confirms the central role of Stepnoi and CZBR in saiga conservation. It is important that decision makers and conservation planners fully appreciate this fact. A key result of this research was the identification of an extensive area of predicted saiga habitat across southern Kalmykia. This finding has added to our understanding of saiga distribution in Kalmykia and resulted from the participatory monitoring scheme. This result underlines the potential of PM data for saiga conservation.

Possible limitations of the Maxent application were identified, although the small size of the dataset precluded firm conclusions on these limitations. When using Maxent, the use of additional model metrics should be considered.

References

Aipanjiguly, S., Jacobson, S. & Flamm, R.(2002). Conserving manatees: knowledge, Attitudes and intentions of boaters inTampa Bay, Florida. Conservation Biology, 17, 1098–1105.

Anderson, R.P. & Gonzalez, I. (2011) Species-specific tuning increases robustness to sampling bias in models of species distributions. Eco. Modelling, 222

Arylov, Y., Lushchekina, A., Neronov, V. (2004). The saiga antelope in the drylands of Russia and how to ensure its sustainable future. In *Traditional knowledge and modern technology for the sustainable management of dryland ecosystems*. Proceedings of the International Workshop, Elista, Republic of Kalmykia, Russian Federation, 23–27 June. UNESCO–MAB Drylands Series, 4, 163-166.

Badmaev V., Ubushaev B. (2004). Chernyje Zemli Biosphere Reserve and its role in the conservation of biodiversity in dryland ecosystems. In *Traditional knowledge and modern technology for the sustainable management of dryland ecosystem*. Proceedings of the International Workshop, Elista, Republic of Kalmykia, Russian Federation, 23–27 June. UNESCO–MAB Drylands Series, 4, 155-157.

Basangova, T.G. (2004) The ecological traditions of the Kalmyks. In *Traditional knowledge and modern technology for the sustainable management of dryland ecosystem*. Proceedings of the International Workshop, Elista, Republic of Kalmykia, Russian Federation, 23–27 June. UNESCO–MAB Drylands Series, 4, 108-110.

Bekenov, A.B, IU. A. Grachev & Milner-Gulland, E.J. (1998) The ecology and management of the Saiga antelope in Kazakhstan, *Mammal Review*, 28, 1–52

Berger, J. (2004). The last mile: how to sustain long-distance migration in mammals. *Conserv.Biol*., 18, 320–331.

Bolger, D.T., Newmark, W.D., Morrison, T.A. & Doak, D.F. (2008) The need for integrative approaches to understand and conserve migratory ungulates. *Ecology Letters*, 11, 63–77.

BP (2011) BP Statistical Review of World Energy June 2011. URL: www.bp.com. Accessed on Dec 2, 2011.

Boone, R. B., Thirgood, S. J. & Hopcraft, J. G. C. (2006) Serengeti wildebeest migratory patterns modelled from rainfall and new vegetation growth. *Ecology.* 87 (8), 1987-1994.

Clifford, A.B. & Witmer, L.M. (2004) Case studies in novel narial anatomy: 3. Structure and function of the nasal cavity of saiga (Artiodactyla : Bovidae : Saiga tatarica) JOURNAL OF ZOOLOGY. Vol 264; 3; 217-230.

Crawley, M.J. (2007). *The R Book*. Wiley.

Danielsen, F., Burgess, N. D. & Balmford, A. (2005a) Monitoring matters: examining the potential of locally-based approaches. *Biodiversity and Conservation.* 14 (11), 2507-2542.

Danielsen, F., Jensen, A. E., Alviola, P. A., Balete, D. S., Mendoza, M., Tagtag, A., Custodio, C. & Enghoff, M. (2005b) Does monitoring matter? A quantitative assessment of management decisions from locally-based monitoring of protected areas. *Biodiversity and Conservation.* 14 (11), 2633-2652.

Danielsen, F., Burgess, N. D., Balmford, A., Donald, P. F., Funder, M., Jones, J. P. G., Alviola, P., Balete, D. S., Blomley, T., Brashares, J., Child, B., Enghoff, M., Fjeldsa, J., Holt, S., Hubertz, H., Jensen, A. E., Jensen, P. M., Massao, J., Mendoza, M. M., Ngaga, Y., Poulsen, M. K., Rueda, R., Sam, M., Skielboe, T., Stuart-Hill, G., Topp-Jorgensen, E. & Yonten, D. (2009) Local Participation in Natural Resource Monitoring: a Characterization of Approaches. *Conservation Biology.* 23 (1), 31-42.

Dolisca, F., McDaniel, J. M., Shannon, D. A., and Jolly, C. M. (2009). A multilevel analysis of the determinants of forest conservation behavior among farmers in Haiti. Society and Natural Resources. 22(5): 433–447.

Dubinin, M., Potapov, P., Lushchekina, A. & Radeloff, V.V. (2010) Reconstructing long time series of burned areas in grasslands of southern Russia by satellite remote sensing. *Remote Sensing of Environment,* 114.

EDGE (no date) ZSL URL: http://www.edgeofexistence.org/mammals/species_info.php?id=62 Accessed on 6 Apr, 2011.

Elith, J., Graham, C., Anderson, R., Dudik, M., Ferrier, S. & Guisan, A. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography.* 29 (2), 129-151.

Elith, J., Phillips, S. J., Hastie, T., Dudik, M., Chee, Y. E. & Yates, C. J. (2011) A statistical explanation of MaxEnt for ecologists RID B-5972-2011. *Diversity and Distributions.* 17 (1), 43-57.

Fadeev, V. A. & Sludskii, A. A. 1982 Saigak v Kazakhstane. Alma-Ata: Academy of Sciences, USSR.

FFI - Fauna & Flora International (2009) *Bridging the gap between conservation and community development in post-conflict Cambodia.* In: FFI 2009. URL: www.fauna-flora.com/ docs/Livelihoods_&_Conservation_Cambodia.pdf>. Accessed on October 28, 2009 (previous study).

Franklin, J. (2010) Mapping species distributions. Cambridge University Press. Cambridge, UK.

Frey, R., Volodin, I. & Volodina, E. (2007) A nose that roars: anatomical specializations and behavioural features of rutting male saiga. *Journal of anatomy* [0021-8782] vol:211; 6: 717 -736.

GEF (2008). Steppe Conservation and Management. Request for CEO endorsement/approval. Global Environment Facility, Washington, USA.

Geist, V. (1990) Even-toed ungulates in Gould, E. & McKay, G. (Eds) Encyclopedia of Animals - Mammals. Merehurst Press.

Gilchrist, G., Mallory, M. & Merkel, F. (2005) Can local ecological knowledge contribute to wildlife management? Case studies of migratory birds. *Ecology and Society.* 10 (1).

Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Peterson, A. T., Loiselle, B. A. & Nceas Predect Species Working Gro. (2008) The influence of spatial errors in species occurrence data used in distribution models RID A-9560-2011. *Journal of Applied Ecology.* 45 (1), 239-247.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993–1009.

Harington, C.R. (1998) Govt. of Yukon. URL: http://www.beringia.com/research/saiga.html Accessed 2 Apr, 2011.

Hernandez, P. A., Graham, C. H., Master, L. L. & Albert, D. L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods RID A-9560-2011. *Ecography.* 29 (5), 773-785.

Hirzel, A. & Guisan, A. (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling.* 157 (2-3), 331-341.

Hölzel, N., Haub, C., Ingelfinger, M. P., Otte, A. & Pilipenko, V. N. (2002) The return of the steppe large-scale restoration of degraded land in southern Russia during the post-Soviet era. *Journal for Nature Conservation.* 10 (2), 75-85.

Kalmykia Govt. (2002) Mineral resources. URL: http://www.kalm.ru/en/resources.html Accessed 2 Dec, 2011.

Koubbi, P., Loots, C., Cotonnec, G., Harlay, X., Grioche, A., Vaz, S., Martin, C., Walkey, M. & Carpentier, A. (2006) Spatial patterns and GIS habitat modelling of Solea solea, Pleuronectes flesus and Limanda limanda fish larvae in the eastern English Channel during the spring. *Scientia Marina.* 70, 147-157.

Khludnev, A. (2010) Conservation of saigas is our common cause. Saiga News, issue 11.

Kuhl, A., Mysterud, A., Erdnenov,G.I., Lushchekina, A.A., Grachev, I.A., Bekenov, A.B. & Milner-Gulland, E.J. (2007).The ‘‘big spenders’’ of the steppe: sex-specific maternal allocation and twinning in the saiga antelope.

Kühl, A. (2008) "The Conservation Ecology of the Saiga Antelope" PhD Thesis Imperial College London. [online] URL: http://www.iccs.org.uk

Kühl, A., Mysterud, A., Grachev, Iu.A., Bekenov, A.B., Ubushaev, B.S., Lushchekina, A.A., Milner-Gulland, E.J., (2009). Monitoring population productivity in the saiga antelope. *Animal Conservation* 12, 355–363.

Kühl, A. (2010) Saiga antelopes in peril again: Call for Action. Saiga News, issue 10.

Lepretre, M. (2001) The promotion of minority languages in Russian Federation and the prevention of interethnic conflicts: the case of Kalmykia. URL: http://www.eric.ed.gov Accessed on 24 Jun, 2011.

Leon, J. (2009) Evaluating the use of local knowledge in species distribution studies – A case study of Saiga antelope in Kalmykia, Russia. MSc thesis, Imperial College, London.

Lushchekina, A. & Struchkov, A. (2001) The saiga antelope in Europe: Once again on the brink? The Open Country No 3.

Mallon, D.P., (2008) *Saiga tatarica*. In: IUCN 2010. IUCN Red List of Threatened Species. Version 2010.4. URL: www.iucnredlist.org. Accessed on 2 Apr, 2011.

Manel, S., Dias, J., Buckton, S. & Ormerod, S. (1999) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology.* 36 (5), 734-747.

Manel, S., Williams, H. & Ormerod, S. (2001) Evaluating presence-absence models in ecology: the need to account for prevalence RID A-4326-2010. *Journal of Applied Ecology.* 38 (5), 921-931.

McConville, A. J.*,* Grachev, I. A.*,* Keane, A.*,* Coulson, T.*,* Bekenov, A. B. *&* Milner-Gulland, E. J. (*2009*) Reconstructing the observation process to correct for changing detection probability of a critically endangered species*.* Endangered Species Research*,* 6*,* 231–237*.*

Milner-Gulland, E. J. 1994 A Population-Model for the Management of the Saiga Antelope. Journal of Applied Ecology 31, 25-39.

Milner-Gulland, E.J., Kholodova, M.V., Bekenov, A.B., Bukreeva, O.M., Grachev, Iu.A., Amgalan, L. & Lushchekina, A.A. (2001) Dramatic declines in saiga antelope populations. Oryx, 35, 340-345.

Milner-Gulland, E.J., Bukreeva, O.M., Coulson, T.N., Lushchekina, A.A., Kholodova, M.V., Bekenov, A.B. & Grachev, Iu.A. (2003) Reproductive collapse in saiga antelope harems. Nature, 422, 135.

Milner-Gulland, E.J & Rowcliffe, J.M (2007). Conservation and sustainable use. A handbook of techniques. Oxford University Press

Milner-Gulland, E.J. (2009) Report on Progress Towards the CMS Medium-Term International Work Programme for the Saiga Antelope MOU in the period September 2006 - June 2009. UNEP/CMS Secretariat.

Minoransky, V.A. (2009) Saiga News, Saiga Conservation Alliance, Issue 10.

Namrueva, L. (2004) Nature-preserving aspects of ethnic culture. In *Traditional knowledge and modern technology for the sustainable management of dryland ecosystems*. Proceedings of the International Workshop, Elista, Republic of Kalmykia, Russian Federation, 23–27 June. UNESCO–MAB Drylands Series, 4, 108-110.

Oba, G., Byakagaba, P. & Angassa, A. (2008) Participatory Monitoring of Biodiversity in East African Grazing Lands. *Land Degradation & Development.* 19 (6), 636-648.

Ochirova, N. G. 2004 Traditional knowledge as a factor of ecological conciousness: formation and conservation of the environment. In Combating Desertification. Traditional knowledge and modern technology for the sustainable management of dryland ecosystems: UNESCO-MAB Drylands Series No.4.

Phillips, S. (no date) A Brief Tutorial on Maxent. URL: www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc. Accessed: 10 Oct, 2011.

O'Neill, H. (2008) Designing Robust Ranger Based Monitoring Strategies for the Saiga Antelope *Saiga tatarica tatarica*. MSc Thesis, Imperial College, London.

Rao, M. & Ginsberg, J. (2010) From conservation theory to practice: crossing the divide, *in* Sodhi, N. S. & Ehrlich, P. R. (Eds.) Conservation Biology for All. Oxford University Press (2010).

Rodda, G. H., Jarnevich, C.S. & Reed, R.N. (2011) Challenges in Identifying Sites Cilmatically Matched to the Native Ranges of Animal Invades. PLOS ONE, 6, 2.

Saiga News (2010) Saiga Conservation Alliance, Issue 11.

Saiga News (2011) Saiga Conservation Alliance, Issue 12.

Shipley, B. (2010) Inferential permutation tests for maximum entropy models in ecology. *Ecology.* 91 (9), 2794-2805.

Singh, N.J. , Grachev, I.A., Bekenov, A.B. & Milner-Gulland, E.J. (2010a) Tracking greenery across a latitudinal gradient in central Asia – the migration of the saiga antelope. Diversity and Distributions, (Diversity Distrib.) 16,663–675

Singh, N.J. et al (2010b) Saiga antelope calving site selection is increasingly driven by human disturbance. Biological Conservation, 143, 1770–1779

Singh, N. J. & Milner-Gulland, E. J. (2011) Monitoring ungulates in Central Asia: current constraints and future potential. *Oryx.* 45 (1), 38-49.

Solano, R. Didan, K., Jacobson, A. & Huete, A. (2010) MODIS Vegetation Indices (MOD13) C5 User's Guide. University of Arizona.

Strand, H., Höf, R., Strittholt, J., Miles, L., Horning, N., Fosnight, E., Turner, W., eds. (2007). Sourcebook on Remote Sensing and Biodiversity Indicators. Secretariat of the Convention on Biological Diversity, Montreal.

UNDP (2010) National Human Development Report in the Russian Federation 2010. Millennium Development Goals in Russia: Looking into the Future. UNDP, Moscow.

Warren, D. L. & Seifert, S. N. (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications.* 21 (2), 335-342.

Whitebread, E. (2008) Evaluating the potential for participatory monitoring of saiga antelope by local villagers in Kalmykia, Russia. MSc Thesis, Imperial College, London.

Young, J. K., Murray, K. M., Strindberg, S., Buuveibaatar, B. & Berger, J. (2010) Population estimates of Endangered Mongolian saiga Saiga tatarica mongolica: implications for effective monitoring and population recovery. *Oryx.* 44 (2), 285-292.

ZAAB Invest (no date) Republic of Kalmykia Southern-Ermolinskiy license block. URL: http://www.zaabinvest.com/kalmykiaeng Accessed 1 Dec, 2011.

Appendix

Table A.1 Summary of models for November-February with description and AUC scores. Note: models 1-9 not listed as they were based on incomplete data.

Model	Model description	AUC
10	Snow, EVI & distance to roads excluded, Beta lqp: 2.3, all other parameters used default settings.	0.887
11	Snow, EVI & distance to roads excluded, all parameters used default settings.	0.899
12	Snow, EVI, distance to water & distance to roads excluded, Beta lqp: 2.3, all other parameters used default settings.	0.908
13	Snow, EVI & distance to roads excluded, Beta lqp: 2.3, all other parameters used default settings. Test data set to 33% (7 monitors), rather than default 25% (6 monitors).	0.86
14	Beta lqp: 2.3, all other parameters used default settings. Only distance to water & distance to PA included.	0.821
15	Beta lqp: 2.3, all other parameters used default settings. All variables included.	0.875
16	All variables used. Default parameters.	0.894
17	All variables used. Hinge set to 0.9, all other parameters used default settings.	0.906
18	All variables used. Hinge set to 0.1, all other parameters used default settings.	0.828
19	All variables used, with 2-way interactions between variables, all other parameters used default settings.	0.889
20	Only distance to water & distance to PA used, with 2-way interactions between variables, all other parameters used default settings.	0.837
21	All variables used. Only Hinge & Linear features were used.	0.875
22	Only distance to water & distance to PA used. Hinge set to 0.9, all other parameters used default settings.	0.859
23	Only distance to PA used. Hinge set to 0.9, all other parameters used default settings.	0.913
24	Only distance to water & distance to PA used, 2-way interactions between variables, hinge set to 0.9, all other parameters used default settings.	0.862
25	Only distance to water & distance to PA used, default params.	0.837

Table A.2 Summary of models for March-June with description and AUC scores. Note: models 1-9 not listed as they were based on incomplete data.

Model	Model description	AUC
10	All variables used. Beta lqp: 2.3 all other parameters used default settings.	0.876
11	Excluded snow, EVI & roads. Beta lqp: 2.3, all other parameters used default settings.	0.876
12	All variables, default settings.	0.898
13	Excluded snow, EVI & roads. Default settings.	0.889
14	Excluded snow, EVI, roads & water. Default settings.	0.909
15	Only distance to water & distance to PA included. Beta lqp: 2.3, all other parameters used default settings.	0.822
16	All variables used. Hinge set to 0.9, all other parameters used default settings.	0.908
17	All variables used. Hinge set to 0.1, all other parameters used default settings.	0.846
18	All variables used, with 2-way interactions between variables, all other parameters used default settings.	0.877
19	Only distance to water & distance to PA used, with 2-way interactions between variables, all other parameters used default settings.	0.831
20	Only distance to water & distance to PA used. Hinge set to 0.9, all other parameters used default settings.	0.856
21	Only distance to water & distance to PA used, 2-way interactions between variables, hinge set to 0.9, all other parameters used default settings.	0.858
22	Used distance to water, distance to PA, EVI and distance to town , hinge set to 0.9, all other parameters used default settings.	0.917

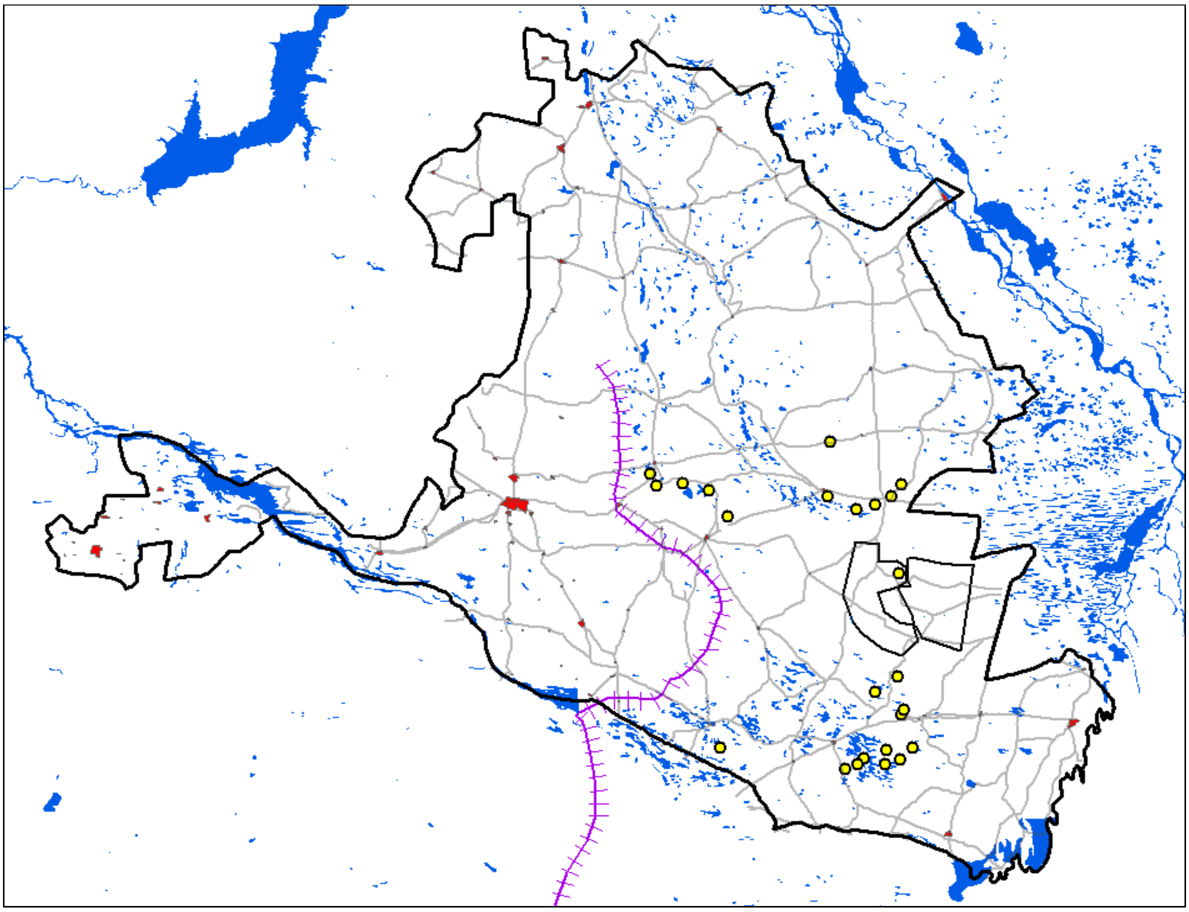

Figure A.1 Map of Kalmykia showing monitor locations (yellow dots), towns/settlements (red), water (blue), rail/canal (violet), PA (black polygon) and roads (light grey).