School of GeoSciences

DISSERTATION

For the degree of

MSc in Geographical Information Science

Student Name: Christine Ellen Ratcliffe

August 2015
Statement of Copyright

Copyright of this dissertation is retained by the author and The University of Edinburgh. Ideas contained in this dissertation remain the intellectual property of the author and their supervisors, except where explicitly otherwise referenced.

All rights reserved. The use of any part of this dissertation reproduced, transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise or stored in a retrieval system without the prior written consent of the author and The University of Edinburgh (Institute of Geography) is not permitted.

Statement of Originality

I declare that this dissertation represents my own work, and that where the work of others has been used it has been duly accredited. I further declare that the length of the components of this dissertation is 4,983 words for the Research Paper and 8,058 words for the Technical Report.

Signature __________________

Date ________________
Acknowledgments

This study was made possible through the kind contributions and support of a number of people and organisations. I would first like to thank Alice Broome, Chloe Bellamy, Shelley Barbour and Vanessa Burton at Forest Research for their help in refining the topic and methods of this project, especially Alice and Chloe for their on-going support and advice. The aims of this project were further refined from the contributions of Kenny Kortland. I would also like to thank Kenny for his efforts in securing permission for the use of Capercaillie location data and putting me in touch with Robert Moss, a renowned academic and Capercaillie expert who also kindly offered his thoughts and suggestions for this study.

John Tullis from the Forestry Commission has provided advice throughout this project and a number of others including Phil Baarda, Susan Haysom and Duncan Blake (Scottish National Heritage) have contributed along the way. I would like to thank Louise Sing and Willie Towers at the James Hutton Institute for both their advice and provision of key data. I would also like to thank Gareth Marshall from the RSPB for providing the Capercaillie presence data essential to this study and information on the methods of collection.

The guidance and support provided by my supervisor Iain Woodhouse has been invaluable. In addition, David Cooper and Owen Macdonald as well as other members of the department have been more than happy to help with any questions I had. Finally I would like to thank all the staff on the M.Sc. on GIS course, my friends and family for their on-going support and all my class mates for being such a positive influence, even the ones who are sadly no longer with us.
Dissertation Contents

**Part I** – Research Paper

**Part II A** – Supplementary Information

**Part II B** – Technical Report

**Appendix**
Modelling habitat suitability for capercaillie (*Tetrao urogallus*) in Scotland under future climate and forest scenarios: using Maxent to model the combined influence of biotic and abiotic factors affecting capercaillie distribution.

"Tetrao urogallus 01" by H. Zell - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons
Part I: Research Paper
Modelling habitat suitability for capercaillie (*Tetrao urogallus*) in Scotland under future climate and forest scenarios: using Maxent to model the combined influence of biotic and abiotic factors affecting capercaillie distribution

Abstract

Capercaillie (*Tetrao urogallus*), a large charismatic woodland grouse native to Scotland, are at risk of extinction. Their population has rapidly declined since the 1970s and the reasons for this are not clearly understood. Loss of habitat and climate change are two of the main theories explored in this paper.

This study uses a software called Maxent, a probabilistic modelling approach that employs machine learning techniques to create a Habitat Suitability Map (HSM). The results identify which climatic and physical environmental characteristics most influence the current distribution of capercaillie. This is used to evaluate how much of the current landscape meets the needs of capercaillie based on their habitat preferences. The HSM for capercaillie is then ‘projected’ onto future climate in order to assess how robust capercaillie are to climate change. Two future climate scenarios are considered for climate in 2070 to investigate the effect of high and low emissions of greenhouse gases, as defined by the Intergovernmental Panel on Climate Change (IPCC) Representative Concentration Pathways.

The influence of future forest scenarios are considered jointly with climate change. Often bioclimatic variables are solely used to assess habitat suitability as there is good evidence that vegetation cover is strongly determined by climate. The results here show that explicitly including vegetation cover, especially for a species such as capercaillie that are heavily depended on a single species of tree (Scots Pine, *Pinus sylvestris*), significantly improves present day HSMs and drastically alters the predicted effects of climate change.

Introduction

Climate change and loss of biodiversity are critical issues globally. The effects of climate change are already showing on natural systems worldwide (Parmesan and Yohe, 2003). Predicting the pattern of change in habitat can help us produce effective conservation schemes (Li et al., 2015). Spatial datasets available at an increasing spatial and temporal resolution enable us to model and track changes in ecological systems and predict the effect
of climate change (Phillips et al., 2006, Seoane et al., 2004). Conservation efforts have to be targeted in order to balance the needs of human populations with the ecosystems we live in. Conservation helps protect the vital ecosystem services humanity relies on both directly and indirectly (Chan et al., 2006). Moreover, predicting and managing for changes in habitat, for an umbrella species such as capercaillie benefits countless other species (Braunisch and Suchant, 2007). The fascination with nature is demonstrated by the increasing international popularity of movements such as re-wilding (Navarro and Pereira, 2015). The movements demonstrate the intrinsic worth of unique species that provide valuable cultural ecosystem services.

Capercaillie, an iconic and rare forest grouse native to Scotland are the focus of many studies in Scotland and internationally (e.g., Braunisch and Suchant, 2007, Quevedo et al., 2006, Piccozzi et al., 1992). Their charismatic nature and role as an umbrella species has made them the focus of conservation initiatives in Scotland (Marshall and Edwards-Jones, 1998). Originally native to Scotland, they became extinct in 1785 (Moss and Piccozzi, 1994). After their reintroduction in 1837 they established a successful and stable population in Scotland to the extent to which they were once considered pests to commercial forest plantation saplings (Moss and Piccozzi, 1994). However, since the 1970s, when their population was around 20,000 (Marshall and Edwards-Jones, 1998) their population has severely contracted and they are now a protected species (Moss and Piccozzi, 1994). Their population is now thought to be around 1,285 (Ewing et al., 2012). Although there has been a 20% decline in woodland specialist birds since 1970s (Forestry Commission, 2013), capercaillie are the only conifer specialist that is declining (Fuller et al., 2005).

While there are a number of possible causes, the mechanisms behind Capercaillie decline are not well understood. There are five main theories: (i) reduction in suitable habitat, including increases competition with deer (Kortland, 2003, Wilkinson et al., 2002), (ii) increased predation, particularly from crows and foxes (Kortland, 2003), (iii) higher mortality as the result of increased deer fencing (Moss, 2011, Kortland, 2003), (iv) disturbance from commercial and recreational use of forest (Ruddock and Whitfield, 2007) and (v) increase in adverse weather conditions that effect breeding success (Moss et al., 2001, Kortland, 2003).

With evidence that climate has had a significant role in capercaillie decline, concern about their prospects under further climate change is an important conservation consideration.
exploration of the effects of climate change is needed in order to establish what impacts it will have, and whether any negative impact could be mitigated by an increase in Scots Pine forest. Therefore the main research questions are: what are the most important factors driving current capercaillie distribution? What effect will climate change have on the amount of suitable habitat? What is the effect of including Scots Pine forests in future climate scenarios and what effect would an increase Scots Pine forest area have on the prediction? This is a national scale analysis of habitat suitability for capercaillie in Scotland at a resolution of 1km.

**Current Understanding of Capercaillie Habitat**

Adult capercaillie primarily feed on Scots Pine, preferentially the needles of mature trees, though they also eat Sitka Spruce needles (Summers et al., 2004). In Scotland Sitka Spruce is non-native therefore the protection of capercaillie is interconnected with the preservation and restoration of native Scots Pine forests. In addition to being a food source, Scots Pine forests are essential for breeding because display grounds, called leks, are established in the forest with each different male’s territory extending around the display area (Picozzi et al., 1992). Native pine forests in Scotland are highly fragmented and the present day extent of the forest is far smaller than any historical coverage (Hobbs, 2009) therefore limiting the available habitat for capercaillie.

There are a number of studies on capercaillie habitat in Scotland, most of which have primarily focused on the role of forest type and structure on their abundance and distribution (Saniga, 2003, Wilkinson et al., 2002 and Picozzi et al., 1992). Several important habitat characteristics have been identified by previous studies. Aside from Scots Pine, one of the most prevalent of these is the importance bilberry cover as a food source for adults and chicks alike (Summers et al., 2004), especially when found in old forest (Gregersen, and Gregersen, 2008). However, the majority of habitat suitability studies carried out in Scotland are focussed on the scale of the forest. Some studies of capercaillie in mainland Europe have conducted more of a landscape scale approach where forest fragmentation and configuration was shown to be important a factor (Saniga, 2003, Quevedo et al., 2006).
Modelling Climate Change and Habitats

Habitat suitability modelling is a well-established practice and there are numerous studies that investigate the effects of climate change on a range of species (e.g., Li et al., 2015, Meyer et al., 2014, Silva et al., 2013). However, despite the known importance of biotic (non-climatic) variables they are rarely included in assessments of change in future habitat (Triviño et al., 2011). There are two main reasons for this. Firstly, climate has been shown to strongly influence vegetation, reducing the need to explicitly model vegetation (Araújo and Peterson, 2012). Secondly, data of predicted changes in land cover and vegetation are less freely available (Triviño et al., 2011).

Concern over increased chick mortality since the 1970s prompted questions of the influence of climate, particularly changing weather patterns in April, May and June on breeding success (Moss et al., 2001). While future changes in weather patterns in these months have not yet been explored, two multi-species studies carried out in the UK reported significant loss in capercaillie habitat under climate change (Walmsley et al., 2007, Fuller et al., 2005). However, these studies were climate focused and the influence of forest was not taken into account. A similar study carried out in the Swiss Alps reported an increase in capercaillie habitat under climate change but was dependent on increase in forest extent (Li et al., 2015).

While some studies have shown that vegetation makes no difference to HSMs the results appear to be highly dependent on the species in question and vegetation appears to be a key consideration for birds (Heikkinen et al., 2006). Where vegetation is determinant of a species presence, as is the case with the giant panda, the effect of climate change on bamboo rather than the panda was modelled to determine future suitable habitat (Li et al., 2015).

Habitat Suitability Models and key concepts

A key concept for species distribution modelling, especially in the context of climate change, is that of ‘extinction vortex’ (Fagan and Holmes, 2006). An extinction vortex occurs if a population falls below a certain threshold and cannot survive without human intervention. In order to assess a population’s resilience to climate change niche theory must be considered.

Niche theory is a central theme when assessing habitat in a landscape, though in spite of this Hirzel and Le Lay (2008) suggest there is a lack of understanding and application of niches in
habitat suitability modelling. A fundamental niche has all the conditions to support a population however due to human influence and competition with other species the true area, or realised niche a species can occupy is much smaller (Phillips et al., 2006).

The proportional availability of realized niches in the landscape is influenced by how connected suitable habitat is. Habitat can be connected structurally meaning it is physically connected or functionally meaning the species can freely move between habitats (Ziolkowska et al., 2014). Connectivity is an important consideration especially in a model such as the one presented here where a maximum dispersal is assumed (Li et al., 2015). Assessing connectivity and dispersal is beyond the scope of this research but should be kept in mind when interpreting the results.

Approach, Data and Methods

*Current understanding of Capercaillie Habitats*

Capercaillie are only found on mainland Scotland and their range does not extend further south than the Trossachs National Park (figure 1). However the potential for reintroduction into southern Scotland has been explored (Marshall and Edwards-Jones, 1998) despite being outside the range identified as native to Scots Pine (O’Sullivan, 1977).
Current Distribution of Capercaillie Leks in Scotland

Figure 1 Current distribution of capercaillie leks (white squares) in Scotland. The largest population is found in the Cairngorm National Park. The majority of the leks are within the native pinewood zone: the extent where Scots Pine are deemed to be native.
Maxent takes a maximum (information) entropy approach that uses probability theory and machine learning to establish the relationship between presence points and environmental variables (Elith et al., 2011). Maxent requires presence only data, an alternative to presence-absence models that requires information about where a species is not found in order to determine unsuitable habitat (Yackulic et al., 2013). Maxent makes a comparison of the values at the presence sites to that of the wider study area to fit a probability distribution, and the model ends its iterative approach when probability of occurrence at presence sites is maximised (Phillips and Dudík, 2008). The main benefit of this approach over more traditional methods is that it is less susceptible to human error and any assumptions about the importance of certain variables over others and the relationships between variables. Maxent performed well with default settings. An in-depth review of the methods and settings can be found in the technical report.

Scale

The scale of analysis can dictate the importance of variables and the varying importance of drivers at different scales is well-established (Schneider, 2009). Ideally a multi-scale analysis would be conducted to fully ascertain the most suitable areas however that is beyond the scope of this study. Ecologically 1km grid cell size is appropriate as males stay within 1km of the lek for the majority of the year (Picozzi et al., 1992).

WorldClim climate data was used at the highest available resolution of 30 arc seconds which approximates to 1km² [http://www.worldclim.org/]. Higher resolution does not necessarily equate with higher accuracy and the methods of interpolation can affect the derived values (Daly, 2006). However, lower resolution climate datasets can mask substantial variation within the grid cells (Hijmans et al., 2005). All other data used in this study was available at a 1km² or finer resolution, however due to the requirements of Maxent all data was resampled to a 1km².
**General Circulation Model (GCMs) selection for future climate**

Various General Circulation Models (GCMs) are available via WorldClim. The predicted effects of climate vary depending on the methods used (Daly, 2006). The HadGEM2-ES model was chosen because of its performance at latitudes appropriate to Scotland (Bellouin et al., 2011). Multiple emission scenarios as defined by the IPCC are available and this model focuses on the lowest and highest in 2070. 2070 was chosen over 2050 projections as any young current native pine is likely to be mature by 2070.

**Presence data**

The presence data is of leks in Scotland and was made available to this study at an accuracy of 1km. Sampling bias is a critical consideration that can significantly influence models (Kramer-Schadt et al., 2013, Phillips and Dudík, 2008). It can result in overfitting of the model and erroneous estimation of the relationship between environmental variables and species presence (Phillips and Dudík, 2008). A number of solutions have been suggested to moderate the effects of sampling bias (Boria et al., 2014, Kramer-Schadt et al., 2013). Here the presence data was spatially filtered to reduce the effects of sampling bias and spatial autocorrelation (Boria et al., 2014, Radosavljevic and Anderson, 2014). A 10km filter was selected after testing environmental heterogeneity in the study region (Ratcliffe, 2015).

**Selecting variables**

The climatic data was obtained from WorldClim [http://www.worldclim.org/]. An exploration of other climate surfaces, base line climates and future projections were explored, however WorldClim was most appropriate for this study (Ratcliffe, 2015).

A literature review was used to establish which variables should be considered for inclusion in the model. Table 1 shows the Bioclimatic variables that are ubiquitous in habitat suitability models (Hijmans and Graham, 2006), which were included to assess what influence if any they had on capercaillie habitat. Tables 2 and 3 show additional climatic and non-climatic variables that were selected for testing and the reasons why. The initial variable selection
requires careful consideration because if a factor critical to capercaillie habitat selection was omitted, the model results would be of limited use.

Highly correlated variables (> 0.75) were removed but with the variable of most ecological relevance to capercaillie retained. Intermediate models were run and the variables reduced until the final selection was reached where removal of any variable cause a drop in performance.

*Table 1* 'Bioclim' variables from WorldClim [http://www.worldclim.org/] at a 30 second (~1km) resolution. They are derived from precipitation and rainfall climate data averaged from 2050-2000 (Hijmans et al., 2005.)

<table>
<thead>
<tr>
<th>Name</th>
<th>Bioclimatic description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio 1</td>
<td>Annual Mean Temperature</td>
</tr>
<tr>
<td>Bio 2</td>
<td>Mean Diurnal Range (Mean of monthly (max temp - min temp))</td>
</tr>
<tr>
<td>Bio 3</td>
<td>Isothermality (BIO2/BIO7) (* 100)</td>
</tr>
<tr>
<td>Bio 4</td>
<td>Temperature Seasonality (standard deviation *100)</td>
</tr>
<tr>
<td>Bio 5</td>
<td>Max Temperature of Warmest Month</td>
</tr>
<tr>
<td>Bio 6</td>
<td>Min Temperature of Coldest Month</td>
</tr>
<tr>
<td>Bio 7</td>
<td>Temperature Annual Range (BIO5-BIO6)</td>
</tr>
<tr>
<td>Bio 8</td>
<td>Mean Temperature of Wettest Quarter</td>
</tr>
<tr>
<td>Bio 9</td>
<td>Mean Temperature of Driest Quarter</td>
</tr>
<tr>
<td>Bio 10</td>
<td>Mean Temperature of Warmest Quarter</td>
</tr>
<tr>
<td>Bio 11</td>
<td>Mean Temperature of Coldest Quarter</td>
</tr>
<tr>
<td>Bio 12</td>
<td>Annual Precipitation</td>
</tr>
<tr>
<td>Bio 13</td>
<td>Precipitation of Wettest Month</td>
</tr>
<tr>
<td>Bio 14</td>
<td>Precipitation of Driest Month</td>
</tr>
<tr>
<td>Bio 15</td>
<td>Precipitation Seasonality (Coefficient of Variation)</td>
</tr>
<tr>
<td>Bio 16</td>
<td>Precipitation of Wettest Quarter</td>
</tr>
<tr>
<td>Bio 17</td>
<td>Precipitation of Driest Quarter</td>
</tr>
<tr>
<td>Bio 18</td>
<td>Precipitation of Warmest Quarter</td>
</tr>
<tr>
<td>Bio 19</td>
<td>Precipitation of Coldest Quarter</td>
</tr>
</tbody>
</table>

*Table 2* Other climate variables found to be important in previous studies. Data layers from WorldClim [http://www.worldclim.org/] at a 30 second (~1km) resolution. (Hijmans et al., 2005)

<table>
<thead>
<tr>
<th>Month</th>
<th>Climate Variable</th>
<th>Influence on capercaillie</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>Minimum and maximum temperature</td>
<td>Influence on egg viability (Wegge et al., 2010, Moss et al., 2001, Slagsvold and Grasaas, 1979)</td>
</tr>
<tr>
<td>May</td>
<td>Minimum and maximum temperature</td>
<td>Influence on egg viability and chick mortality (Wegge et al., 2010, Moss et al., 2001)</td>
</tr>
<tr>
<td>June</td>
<td>Minimum and maximum temperature and precipitation</td>
<td>Influence on chick mortality (Wegge and Kastdalen, 2007, Moss et al., 2001)</td>
</tr>
<tr>
<td>All</td>
<td>Snow cover</td>
<td>Number of days with &gt;10cm snow found to be important (Braunischand and Suchant, 2007)</td>
</tr>
</tbody>
</table>
Table 3 various non-climatic variables used in building the model. They were selected based on findings from previous studies that suggest they influence capercaillie habitat preference.

<table>
<thead>
<tr>
<th>Variable (distance to)</th>
<th>Description</th>
<th>Influence on capercaillie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scots Pine forest</td>
<td>Forest areas where the Scots Pine is dominant. Further classed according to stand age.</td>
<td>Scots pine are a food source for adults and the forest areas are used for mating displays (Picozzi et al., 1992)</td>
</tr>
<tr>
<td>Bog in forest</td>
<td>Bog areas that are within forests</td>
<td>A source of high protein food (Summers et al., 2004)</td>
</tr>
<tr>
<td>Roads</td>
<td>Distance to all roads in Scotland</td>
<td>Disturbance can displace capercaillie (Ruddock and Whitfield, 2007)</td>
</tr>
<tr>
<td>Foot paths</td>
<td>Distance to core footpaths</td>
<td>Disturbance can displace capercaillie (Ruddock and Whitfield, 2007)</td>
</tr>
<tr>
<td>Broadleaf</td>
<td>Broadleaf forest, Birch and Oak are of particular importance</td>
<td>Food source from common ground cover (bilberry) (Summers et al., 2004)</td>
</tr>
<tr>
<td>Non-native coniferous</td>
<td>Non-native coniferous forest, Sitka Spruce is of particular importance</td>
<td>Alternative food source for adults and breeding ground (Picozzi et al., 1996)</td>
</tr>
<tr>
<td>Bog</td>
<td>Land classified as bog (not in forest)</td>
<td>A potential source of high protein food (Summers et al., 2004)</td>
</tr>
<tr>
<td>Agricultural land</td>
<td>Agricultural and pastoral farm land combined</td>
<td>A proxy for influence of predators (Huhta et al., 1996)</td>
</tr>
</tbody>
</table>

**Future forest**

Input features for future forest were obtained from a previous study and clipped to the native pine zone. Figure 2 shows the present day Scots Pine forest and the areas identified as suitable for native pine regeneration in the native pine zone (Sing et al., 2013). Other native trees could potentially grow in these areas so this is a best case scenario for Scots Pine, although there is some evidence to suggest that capercaillie benefit from mixed forests (Summers et al, 2004).
Figure 2 the map shows present day native pine extent in dark green and areas suitable for regeneration in light green. The potential future pine areas have been clipped to the native pine zone extent. The map also shows that the Cairngorm National Park has a large proportion of the present day native pine forest and visually inspecting the map it also appears to be the least fragmented.
Validation and performance

The model was validated using a separate set of presence data. This method is preferred over sub-sampling and cross validation where possible (Radosavljevic and Anderson, 2014). The model was assessed using the area under the curve (AUC) of the receiver operating characteristic (ROC). The main advantage of this approach is it is threshold independent (Jiménez-Valverde, 2012) although it can be susceptible to artificial inflation (Yackulic et al, 2013).
Results

The results below detail the graphical and numerical outputs of Maxent. For further information and a review of the results please see the technical report. A review of capercaillie response curves to the climate variables and their predicted surfaces under climate change is presented in the supplementary information.

The present day model is presented and compared with future predictions to examine changes in habitat suitability. The model findings reveal the relative importance of different variables to the distribution of suitable capercaillie habitat. The models performance is measured using a threshold independent area under receiver operator curve (AUC) (Jiménez-Valverde, 2012). A jacknife test is used to measure the predictive ability of individual variables and the loss of the models AUC when the variable is excluded (figure 3 and table 4) (Phillips, 2005). Permutation importance measures the decrease in model performance when the values of a selected variable are randomised (Phillips, 2005). The permutation importance of each variable varies between the models and is presented for each model (table 4).

A variables performance on an individual basis does not change between the models as it is a measure of how well the variable can explain capercaillie distribution when used in isolation. Each individual variable is measured using a jacknife test. The jacknife AUC values for the final selected variables are shown in the figure 3. Distance to young native pine and distance to mature and veteran native pine have an AUC value of over 0.9, therefore they can efficiently discriminate suitable habitat even when used in isolation.
Performance of variables

Permutation values vary between the different models as their effect is dependent upon both the number of other variables and the nature of the other variables (Phillips, 2005). For example, June precipitation has the highest permutation importance in the climate only model but has the lowest importance in the final present day model suggesting that other variables in the final model mitigate the effect of June precipitation on predictive performance. However, June precipitation still has the fourth highest AUC loss of predictive performance therefore the variable contains information that is not present in other variables. The climate variables with the greatest permutation importance were annual mean temperature and mean diurnal range. Annual mean temperature has a negative AUC loss in the final model which means that the models ability to predict presence in the validation data improves when that variable is removed. The variables with negative AUC loss were retained because of their contribution in permutation importance.

* April minimum temperature was removed from the final model as it had no permutation importance in the context of the final model.
Table 4 contribution to the model as measured by permutation importance and loss in AUC value when tested against validation data. The table is ordered by permutation importance for each model. This measured the drop in AUC when the values for that environmental layer are randomised while all other layer values retain their values. The AUC loss is a measurement of how much the model lost in its predictive performance when the variable was removed from the model. The cells highlighted orange show the variable with the highest drop in AUC suggesting this variable has the highest information that is not present in the other variables. The negative values in the final model show an improvement in the models ability to predict presence in the validation data when that variable is removed. These variables were retained because of their contribution in permutation importance.

<table>
<thead>
<tr>
<th>Permutation Importance</th>
<th>Test AUC Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Climate only</strong></td>
<td></td>
</tr>
<tr>
<td>June Precipitation</td>
<td>38.81</td>
</tr>
<tr>
<td>April minimum temperature</td>
<td>18.55</td>
</tr>
<tr>
<td>June maximum temperature</td>
<td>18.13</td>
</tr>
<tr>
<td>Mean annual temperature (Bioclim 1)</td>
<td>14.34</td>
</tr>
<tr>
<td>Mean diurnal range (Bioclim 2)</td>
<td>10.17</td>
</tr>
<tr>
<td><strong>Non-climatic</strong></td>
<td></td>
</tr>
<tr>
<td>Mature and veteran native pine distance</td>
<td>76.00</td>
</tr>
<tr>
<td>Younger native pine distance</td>
<td>15.61</td>
</tr>
<tr>
<td>All roads distance</td>
<td>5.22</td>
</tr>
<tr>
<td>Non-native coniferous</td>
<td>2.15</td>
</tr>
<tr>
<td>Core paths distance</td>
<td>0.57</td>
</tr>
<tr>
<td>Bog in forest distance</td>
<td>0.35</td>
</tr>
<tr>
<td>Broadleaf distance</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>Final model – combined climatic and non-climatic</strong></td>
<td></td>
</tr>
<tr>
<td>Mature and veteran native pine distance</td>
<td>49.08</td>
</tr>
<tr>
<td>Younger native pine distance</td>
<td>21.22</td>
</tr>
<tr>
<td>Mean Diurnal range (Bioclim 2)</td>
<td>13.36</td>
</tr>
<tr>
<td>Mean annual temperature (Bioclim 1)</td>
<td>5.36</td>
</tr>
<tr>
<td>Non-native coniferous</td>
<td>3.08</td>
</tr>
<tr>
<td>Bog in forest distance</td>
<td>3.01</td>
</tr>
<tr>
<td>June maximum temperature</td>
<td>1.75</td>
</tr>
<tr>
<td>Broadleaf distance</td>
<td>1.07</td>
</tr>
<tr>
<td>All roads distance</td>
<td>0.82</td>
</tr>
<tr>
<td>Core paths distance</td>
<td>0.74</td>
</tr>
<tr>
<td>June precipitation</td>
<td>0.49</td>
</tr>
</tbody>
</table>
Logistic output

The resulting graphical output surface from Maxent is a probability surface. Maxent assigns a probability to each cell based on the environmental characteristics of that cell (Elith et al., 2011). The sum of the cells equals 1, though for study areas with varying sizes the results are not easily comparable using the original raw output (Phillips and Dudík, 2008). The logistic output is proportional to the raw output and is monotonically related so areas have the same relative importance (Phillips and Dudík, 2008). However some have cautioned the use of logistic output because of the assumptions used to scale the values to presence probability (Yackulic et al, 2013). For this reason, and other ecological reasons, the logistic output is more appropriately interpreted as suitability of habitat with red areas of the map being highly suitable habitat, rather than presence probability. Therefore in the results the logistic output is referred to as ‘approximate presence probability’.

Performance

The model’s performance is assessed using the AUC value. An AUC value over 0.75 is deemed to be of significant predictive performance (Elith, 2000). Both the climatic (training AUC = 0.915, test AUC = 0.891) and non-climatic (training AUC = 0.950, test AUC = 0.958) models had significant predictive performance, with the non-climatic model performing slightly better. The non-climatic model performs better with the validation data showing a higher AUC with the test data compared with the training data, indicating that it is a better model as it transfers well.

The final model, with combined climatic and non-climatic variables, had slightly higher AUC values (training = 0.962, test = 0.965) than the non-climatic model. Despite all models having a high predictive performance the distribution of suitable habitat varies somewhat between them. The climate only model suggests very low suitability across the west of Scotland, whereas the non-climatic model extends further west. The final present day model has the greatest AUC but the omission-predicted area graph, shown on the bottom right hand side of each figure, shows a slight decrease in fit for the omission in the final present day model (figure 6). However the fractional predicted area of background cells, an alternative method of assessing commission errors (Reddy et al., 2015), shows a close to ideal shape.
Figure 4 the model for present day climate only shows the highest suitability in and surrounding the Cairngorm National Park. This is where the majority of the Capercaillie population are found. The East of Scotland generally has a higher suitability that the west. The model shows good fit in both the omission and predicted area graph (right) and similarly for the ROC AUC graph on the right. However the shape of the test curve in the left hand graph is further than the ideal fit and the AUC value for the test data also reports a drop in predictive performance.
Figure 5 the model for present day non-climatic only shows the high suitability in the Cairngorm National Park but also extends around the North East more than the climate only mode and extends to the West considerably more than the climate only model. The model shows good fit in both the omission and predicted area graph (right) and the ROC AUC graph on the left. The shape of the test curve in the left hand graph is a closer to the ideal fit than the climate only model. In addition this model performs better with the validation data showing a marginally higher AUC with the test data compared with the training data indicating that it is a better model as it transfers well.
Figure 6: The present day final model again shows high suitability in the Cairngorm National Park. Combining the climatic and non-climatic data results in a very similar distribution to the non-climatic model but has reduced the amount of area with high suitability. The model shows a reasonable fit in the omission and a very good fit in the fractional predicted area in the right hand. The AUC graph on the left also shows a very good fit. The shape of the test curve in the left hand graph is a similarly good fit to the non-climatic model but has a slightly higher AUC value. This model also performs well with the validation data showing a marginally higher AUC with the test data compared with the training data indicating that it is a better model as it transfers well.
2070 Scenarios

The different models result in varied predictions of suitable habitat for capercaillie in Scotland by 2070. The magnitude of difference between climate only predictions and the final models that include forest is striking (figure 4). When modelled with climate variables only, the area and suitability decrease substantially in both the low and high emission scenarios (figure 7 1a and 1b). The climate only model predicts a significant increase in area with little or no suitable habitat in both emission scenarios. Furthermore there is a large decrease in the value of approximated presence probability value from 0.945 in the present day climate only model to 0.495 in the high emission scenario. However, climate change under both low and high emission scenarios appears to have very little effect on suitable habitat when forest at its present day extent was included (figure 7 2a and 2b). The predicted suitable habitat figure in this model is very similar to the present day final model and shows a small increase in the highest value in the low emission scenario (0.984) compared with present day full model (0.958). When Scots Pine was increased to simulate the effect of afforestation in all suitable areas (figure 7 3a and 3b) the suitable habitat increased compared to present day for both emission scenarios, although the low emission scenario prediction has higher maximum suitability.
Figure 7 shows the various future habitat scenarios. 1a and 1b show the future predicted suitable area when forecasting using only climate. Both show a reduction in suitable area but the high emission scenario shows a dramatic reduction in area and suitability with the maximum value being only 0.495. 2a and 2b include non-climatic variables with no change in forest extent but the predicted changes in climate. Comparing the high and low emission scenarios the extent is very similar though low emission has higher suitability as can be seen from the more intense red areas. The scenarios in 3a and 3b show the model with non-climatic variables with increase in forest extent. Both the low and high scenarios show increase in habitat suitability from present day however the low emission scenario has more areas with high suitability.
Discussion

The output shows the distribution of suitable habitat rather than explicitly the distribution of capercaillie. Capercaillie may only able to access a limited area of suitable habitat due to barriers to dispersal and issues of connectedness (Pascual-Hortal and Saura, 2008). This is assuming capercaillie that presence means that the area can support a self-sustaining a population which is a critical conjecture (Poiani et al., 2000). In addition, whether a species is present only provides some information about suitability and abundance and mortality are not explicitly accounted for. Moreover, there may be suitable habitat that the model has been unable to account for because of the limitations of male dispersal and the fragmented nature of the forest (Hjeljord et al., 2000). Although these issues are not directly considered here they are a vital consideration for interpreting the effects of climate change.

Present day model – key variables

The focus of the discussion here will be the influence of forest and climate change. The strongest influence on capercaillie habitat suitability in this model is distance to Scots Pine with mature pine having the strongest influence. This finding agrees with many previous studies which have found that although capercaillie use Sitka spruce forests, younger scots pine forests and have also been found in broadleaf forests, they mostly favour mature Scots Pine (e.g., Summers et al, 2004, Picozzi et al., 1992, Picozzi et al., 1996). The weaker relationship with non-native coniferous and broadleaf forest was also reflected in the model results (figure 3 and table 4). Broadleaf forests are most likely to be appealing to capercaillie because of the ground vegetation cover; bilberry Vaccinium myrtillus is often a significant cover in broadleaf forests, especially Oak and Birch and are therefore attractive to hens with broods (Summers et al, 2004).

The biological reasons for favouring mature pine forests are the nutritional value of the needles of older trees and the structure of the forest. Mature forests are more open, allowing capercaillie, a large bird, a suitable flight path (Summers et al, 2004).). Additionally, mature trees have larger branches that can support the large adult males when feeding (Summers et al, 2004). Mature pine forests can also support a higher density of capercaillie (Rolstad et al., 2009). This is because the male’s territory is smaller compared to younger Scots Pine forests or plantations where habitat is less suitable resulting in a larger territory being needed to meet
their needs. Moreover, mature Scots Pine forests structure increases irradiance to the forest floor as the tree density is lower and crown height is higher (Parlane, 2006). Intermediate levels of irradiance favour the growth of bilberry over other common forest shrubs (Parlane, 2006) increasing suitability for both adult capercaillie and chicks (Baines et al., 2004).

The many ecological reasons for capercaillie favouring Scots pine supports the finding that Scots Pine forest are the strongest influence on capercaillie distribution. The importance of forest is reinforced by the nature of the area with the largest remaining capercaillie population, the Cairngorms National Park (CNP). Despite forest cover in Scotland increasing in area during the 20th Century (Forestry Commission, 2013) it is thought that ancient native forests have declined by up to 14% (Patterson et al., 2014). The relative success of capercaillie in the CNP is thought to be the result of the relatively large tracts of ancient native pine forest (Cairngorms National Park Authority, 2015).

Availability of suitable forest is one of the main theories behind capercaillie decline though changes in climate are also thought to be important (Baines et al., 2004). Previous studies have identified April, May and June weather as influencing chick survival (Wegge et al., 2010, Wegge and Kastdalen, 2007, Moss et al., 2001, Slagsvold and Grasaas, 1979). These studies found a correlation between cold wet weather in these months and chick mortality. However it was also reported that predation and food availability were the dominant cases and that climate caused additional attrition (Wegge and Kastdalen, 2007, Baines et al., 2004, Moss et al., 2001). The model presented here did not find any importance in April or May climate and the influence of June precipitation and June maximum temperature was low. The approach of Maxent is perhaps not the most appropriate technique to investigate mortality issues.

The projected climate in 2070 in both high and low emission scenarios potentially benefits capercaillie chicks with generally warmer, dryer summers (Ratcliffe, 2015). However changes in climate could affect the phenology of plants and although there is evidence that capercaillie alter mating in response to the climate of that year there is no synchronicity with the hatching of arthropods, the food source of chicks, which determines the survival rate (Wegge et al., 2010). The interaction of multiple species should also be considered in the context of climate change. For example the grazing of Bilberry by deer, is thought to reduce food availability for capercaillie (Wilkinson et al., 2002) and could therefore amplify any negative effects of changes in plant phenology.
The results of influential climatic variables may be the result of the environment where Scots Pine in Scotland grow. For example a study in Germany found that the number of days of in which snow cover was greater than 10cm was positively correlated with capercaillie presence (Braunischand and Suchant, 2007) suggesting that they prefer colder climates, which fits with their evolution (Gjerde and Wegge, 1987). It is possible that capercaillie favour different climates but are limited by forest location, which in turn have a number of climatic and none climatic restrictions to range and extent (Chytrý et al., 2008).

**Predicted effects of climate change on suitable habitat**

The results show a perhaps counterintuitive effect of climate change when forest is included. A study by Harrison et al. (2003) looking at the effect of climate change in 2050 on British birds showed that capercaillie habitat reduced considerably. This finding was further supported by another study that reported extreme loss, by 2080 under high emission scenarios (Walmsley et al., 2007). While this supports the findings for the climate only model the results for the model including are strikingly different.

The model revealed that forest was the most important factor and furthermore that its influence was much greater than climate. Suitable habitat in this model is not identified by an intersection of suitable vegetation cover and climate but the relative importance of elements of the habitat. A study that simulated the effects of climate change and increase in forest in the Swiss Alps also predicted an increase in suitable capercaillie habitat (Grêt-Regamey et al., 2008). Additionally a study by Virkkala et al. (2005) of boreal birds found that habitat was significantly more important than climate. However the results of this study were not used for predicting the effects of climate change so it unclear as to whether their findings also apply to future climates.

There is an abundance of studies that investigate vegetation change in a changing climate and similarly so for investigating the effects of climate change on species distribution but they are rarely integrated (Triviño et al, 2011). The majority of climate change studies focus solely on climate, omitting vegetation cover and other biotic factors such as interactions with competitor or predator species and barriers to movement (Heikkinen et al., 2006). Biotic interactions are difficult to model as they are both extremely complex and restricted by data availability (Triviño et al, 2011). Factors such as interspecies competition are not taken into
account in this study, however the results suggest that given Scots Pine forest has such a strong influence on the current distribution of capercaillie. Moreover the inclusion of Scots Pine considerably modifies the predicted effects of climate change, suggesting that vegetation is key to modelling capercaillie habitat. Heikkinen et al. (2006) suggested that vegetation be included in habitat suitability models for birds as climate models can result in incomplete detection of a species needs. A study in 2011 found that the inclusion of vegetation when modelling the impact of climate change on Iberian birds improved the fit of half of the models (Triviño et al, 2011). While climate only modelling is sufficient for many species, a consideration of the species ecology is necessary (Triviño et al, 2011).

Pearson et al. (2004) have also found that some species respond more strongly than others with the inclusion of land cover data. However they also found greater performance when using continental rather than regional climate data supporting the argument for a hierarchical multi-scale approach to species distribution modelling (Schneider, 2009).

Dominant processes depend on scale (Hijmans and Graham, 2006). The dominance of climate at a large scale is widely acknowledged (Araújo and Peterson, 2012). For this reason bioclimatic modelling is a popular approach in the field of habitat suitability modelling and climate change predictions. This is because climate drives the main processes relevant to ecological modelling as it delineates biomes (Araújo and Peterson, 2012). The effect of climate on current Scots Pine populations and future regeneration is a key consideration given the critical importance to capercaillie’s resilience to climate change shown in the results presented here.

Climate may be of lesser importance to capercaillie but if Scots Pine forests are highly sensitive to the predicted changes the effects would cascade and the results suggest negative changes in Scots Pine would have a severe impact on capercaillie. The response of Scots Pine forests to climate change in Scotland are mixed (Ray, 2008). An increase in the number of days with very high temperatures (>20 degree) have been attributed to the recent decline of Scots Pine forest in Sweden due to the increased suitability for trees parasites (Rebetez and Dobbertin, 2004). The correlation of trees with temperature is only a weak one (Grace et al., 2002) and is dependent on latitude. However, another study in Sweden attributed recent increase in altitude of tree line to warmer temperatures (Kullman, 2007), although this is only a possibility if the geology is suitable (Grace et al., 2002).
There is evidence to show that the vigour of Scots Pine has increased in the past 20 years and that disturbance and land management are the causes of low success of Scots Pine in Scotland over the past 1000 years (Grace et al., 2002). Further evidence that Scots Pine shrubs are extending into higher latitudes in Scotland, thought to be the result reduced browsing pressures, also bodes well for the future of Scots Pine forests and therefore capercaillie (Miller and Cummins, 1982). A HSM specific to Scots Pine would be extremely useful and perhaps reveal more about the impact of climate change on capercaillie than a model focused solely on capercaillie.

These results are reliant on the GCM selected, other GCMs could predict changes which influence capercaillie differently. In addition, other biotic interactions such as competition with other species and predation should also be considered as climate change may favour these species causing an increased pressure on capercaillie.
Conclusions

Climate change is thought to be one of the main contributory factors in the rapid fall in capercaillie numbers in Scotland, through the impact on chick survival (Moss et al., 2001). However, the literature also suggests that the effect of predators is dominant and weather simply causes further attrition (Wegge and Kastdalen, 2007, Moss et al., 2001). The results here suggest climate is of secondary importance to the presence of Scots Pine, which is the most important factor driving current and future capercaillie distribution.

Bioclimatic modelling dominates the field in habitat suitability modelling and climate change predictions (Araújo and Peterson, 2012). When non-climatic variables are included the model of present day capercaillie improves in predicted performance under a scenario of increased Scots Pine, suitable habitat for capercaillie in Scotland shows a slight increase in comparison to present day despite changes in climate. The results here suggest vegetation should be considered when assessing the vulnerability of a species to climate change. This supports other studies that have found vegetation is crucial for reliably modelling certain species (Grêt-Regamey et al., 2008, Heikkinen et al., 2006).

However, from a hierarchical perspective if climate changes to such an extent as to change the entire ecology of a biome then this change has the dominant control on habitat distribution (Araújo and Peterson, 2012). In this model the change in climate did not reach a level that influenced the distribution of suitable capercaillie habitat. A sensitivity analysis is required to establish what levels of climate change would significantly alter the predicted capercaillie habitat and is a suggested extension of the results presented here. The results are heavily dependent on the assumption that climate change will not negatively affect Scots Pine in Scotland. Though the evidence is mixed, Scots Pine forests could potentially respond positively to climate change (Ray, 2008).

Encouraging establishment of more Scots Pine forests will make the capercaillie population in Scotland more robust to the effects of climate change. The Scots Pine forests in Scotland are highly fragmented and increasing area and connectivity could benefit not only capercaillie but a wide range of species and ecosystem services (Wilkinson et al., 2002). The findings have been encouraging for capercaillie in Scotland under climate change because of the overriding importance of Scots Pine, though vegetation changes could compound as well as mitigate the effects of climate change for many other species worldwide.
References


Cairngorms National Park Authority (2015) *Cairngorms Capercaillie Framework*


Patterson, G., Nelson, D., Robertson, P. & Tullis, J. (2014). *Scotland's native woodlands. Results from the native woodland survey of Scotland*. Forestry Commission Scotland


Part II A: Supplementary Information
Introduction

Climate change has already resulted in the range shifts of species around the world (Parmesan and Yohe, 2003). The focus here is on an exploration of how the variables in this model change under high and low emission scenarios. Additionally, a consideration of how these changes might influence the availability of suitable capercaillie habitat is made. Climate variables in this study were of secondary importance in the full final model however when climate was considered in isolation the results show dramatic reduction in areas suitable for capercaillie (Ratcliffe, 2015).

The possible ecological reasons for capercaillie’s response to bog in forest is also considered here. Response to bog in forest in addressed here because it is thought to be of ecological importance rather than one of an incidental correlation. Distance to roads and paths is discussed in the technical report as the relationship is thought to be the result of factors affecting both the distribution of roads, paths and capercaillie rather than a direct relationship between the two.

Response to bog

There is some evidence to suggest that bogs surrounded by forest might be important to capercaillie, in particular to hens with broods as the vegetation and arthropods associated with bogs are high sources of protein (Summers et al., 2004, Stuen and Spidso, 1988). Two different bog classifications were tested; distance to bog in forest and land classified as bog (not in forest). The model results show two very different responses of capercaillie to these two classifications of bog. Bog that is not in forest had a very low predictive performance AUC = 0.55 however bog in forest has a high predictive performance AUC = 0.87.

The response curves show an increase in presence probability in areas close to bog within forest (figure 1). The response curve is almost identical to that of native pine although forest bog is only somewhat correlated with mature scots pine (0.72) and less so with younger pine (0.62) (Ratcliffe, 2015). In the context of the full model there is a very slight drop in AUC when forest bog is omitted, which when considered with the degree to which it is correlated with forest suggests there is possibly some information in this variable not accounted for by
the forest alone. However the average value for distance from bog in forest to presence points (leks) is 3764m therefore considering this without the response curve it would appear capercaillie do not favour forest with bog. However this distance is most likely the result of the dataset showing only very small areas of forest bog and its distribution is highly sporadic in the landscape. Other studies have suggested bog type areas benefit capercaillie due to the high number of invertebrates (Stuen and Spidso, 1988) though it is difficult to draw conclusions in this study about the extent to which forest bog is important to capercaillie due to the nature of the data.

Figure 1 shows capercaillie response to different classifications of bog. The responses shown here are when the variable is modelled in isolation from other variables.
Response to climate

Capercaillie response curves for different climate variables are shown in figure 2. The optimum values for climatic variables could be a function of the location of pine forests. Forests in Scotland are not found at higher altitudes therefore capercaillie can only select habitat that includes Scots Pine or Sitka Spruce forests, which potentially creates bias in their response to climate. If the forest cover in Scotland was continuous the climatic response curves for capercaillie might be more informative as to their preferences rather than the location of suitable vegetation. Therefore it is perhaps possible that capercaillie’s climate range is greater than the model suggests. However responses to changes in climate, for example greater temperatures than capercaillie have ever experienced in Scotland, have an unknown effect, making predictions and extrapolations problematic.
Figure 2 shows the response curves for the environmental variables included in the final model. They show the predicted probability of presence of capercaillie (y axis) against the environmental variable (x axis). The response curves are effected by covariance with other variables (1a, 2a, 3a and 4a).
Spring weather (April, May and June) was explored due to the reported influence on chick mortality (Wegge et al., 2010, Wegge and Kastdalen, 2007, Moss et al., 2001, Slagsvold and Grasaas, 1979). The full final model includes precipitation and maximum temperature for June only. June precipitation has an AUC of 0.78 when tested in isolation suggesting reasonable capacity to predict capercaillie distribution. June maximum temperature has an AUC of only 0.58 however in the context of the final model it has a higher permutation importance; the models predictive capacity lowers when June maximum temperature values are randomised. However when each of the variables is excluded from the model in turn, the exclusion of June precipitation results in the greater loss in predictive performance.

The response curves of this model show that capercaillie generally prefer warmer, drier areas. Tables 1 and 2 show the predicted changes in precipitation and temperature under high and low emission scenarios. Under the HadGEM2-ES projected climate model June weather is predicted to be generally warmer and drier in the high emission scenario though precipitation is very similar to present day in the low emission scenario. A study in the Swiss Alps reported an increase in capercaillie habitat under a simulated increase in average temperature of 2.4 °C though this was dependent on an increase in forest area (Grêt-Regamey, 2008).

Table 1 June Maximum Temperature for present day and future scenarios in Scotland (Mean at presence sites = 16.2 °C). Values at present day leks sites used for testing are shown, which is subset of the know lek sites.

<table>
<thead>
<tr>
<th>Mean temperature (°C)</th>
<th>Value at present day lek sites</th>
<th>Present day climate</th>
<th>2070 climate (low emission)</th>
<th>2070 climate (high emission)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>15 - 17.80</td>
<td>10.80 - 18.40</td>
<td>11.80 - 21.10</td>
<td>12.70 - 23.10</td>
</tr>
<tr>
<td>Mean</td>
<td>16.23</td>
<td>16.10</td>
<td>18.05</td>
<td>19.50</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.69</td>
<td>1.21</td>
<td>1.54</td>
<td>1.78</td>
</tr>
</tbody>
</table>

Table 2 June precipitation for present day and future scenarios in Scotland (Mean at presence sites 6.4mm) Values at present day leks sites used for testing are shown, which is subset of the know lek sites.

<table>
<thead>
<tr>
<th>Precipitation (mm)</th>
<th>Value at present day lek sites</th>
<th>Present day</th>
<th>2070 low emission</th>
<th>2070 high emission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>56.0 - 85.0</td>
<td>48.0 - 127.0</td>
<td>45.0 - 135.0</td>
<td>47.0 - 120.0</td>
</tr>
<tr>
<td>Mean</td>
<td>63.90</td>
<td>74.3</td>
<td>74.3</td>
<td>71.4</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>7.20</td>
<td>17</td>
<td>18.9</td>
<td>15.6</td>
</tr>
</tbody>
</table>
A study by Moss et al. (2001) found a climate change trend from 1975 that indicated the cooling of mid-April temperatures relative to the rest of the month. However, while they saw no change in trends in May or June, they did find higher chick survival in forest with lower June rainfall. The model here found no explanatory value of April temperature however the relative coolness of mid-April was not assessed.

The importance of availability of large arthropods and can offset the influence of suboptimal climate conditions (Wegge et al., 2010, Moss et al., 2001). Indeed, it is suggested that the size of arthropods rather than the abundance has more of an influence on chick survival (Picozzi et al, 1999). If this is the case and changes in climate effect the timing arthropod hatching this could impact the availability of food sources for capercaillie chicks. As they require a high protein diet only in the first few weeks of their life, changes in the synchronicity of capercaillie hatching and arthropod hatching in the future could have substantial effects on chick survival (Wegge et al., 2010).

**Bioclim 1: Mean Annual Temperature**

In the context of the full present day model capercaillie’s response to mean annual temperature (Bio 1) is uniform up to ~ 6.5 degrees centigrade. The response curve falls quite sharply for temperatures above this. When considering Bio 1 in isolation a temperature range of ~6 to 6.5°C has the highest predictive performance. The temperature associated with the highest presence probability is surprisingly high given that capercaillie are well adapted to low temperatures and preferentially root in snow burrows in northerly populations (Thiel et al., 2007 and Gjerde and Wegge, 1987). However burrowing is a response of very cold overnight temperatures and is an adaptation to conserve energy and escape extreme overnight lows (Marjakangas et al., 1984).

Bio 1 has an AUC of 0.76 when modelled alone so has reasonable capacity to predict capercaillie distribution. The average temperature for capercaillie locations used for training the model was 6.64°C. This is within one standard deviation of the mean temperature for Scotland in present day (7.13 °C) showing that capercaillie do not occupy significantly colder areas despite evolving in cold habitats. Table 3 shows the values for B1. For future scenarios the current climatic range of capercaillie is still available however the average temperature in the high emission scenario has increased by over 3 degrees. The maximum average
temperature for present day locations used in training the model is 8.4°C however considering the maximum temperature at present day is 9.2°C (standard deviation = 1.01°C) it does not appear that capercaillie avoid warmer areas of Scotland. Capercaillie can perhaps survive in climates that they have not yet been exposed to since the current optimum temperature is a function of availability as it so closely matches the mean of Bio 1.

Table 3 Mean annual temperature for present day and future scenarios in Scotland (Mean at presence sites 6.64). Values at present day leks sites used for testing are shown, which is subset of the know lek sites.

<table>
<thead>
<tr>
<th>Mean temperature (°C)</th>
<th>Value at present day lek sites</th>
<th>Present day climate</th>
<th>2070 climate (low emission)</th>
<th>2070 climate (high emission)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>5.6 - 8.6</td>
<td>1.30 - 9.20</td>
<td>3.2 - 10.80</td>
<td>5.2 - 12.9</td>
</tr>
<tr>
<td>Mean</td>
<td>6.62</td>
<td>7.13</td>
<td>8.85</td>
<td>10.60</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.68</td>
<td>1.08</td>
<td>1.10</td>
<td>1.17</td>
</tr>
</tbody>
</table>

**Bioclim 2: Mean Diurnal Range**

Mean diurnal range (Bio 2) is a measure of the changeability in temperature. Areas with a high diurnal range tend to be more continental in character. This is because the high specific heat capacity of water modulates temperatures in coastal areas and extremes in temperature are less severe. The response curves show a greater probability of presence in areas with a higher mean diurnal range (figure 2 (2a)). This can be attributed to the preferences of Scots Pine which have a high affinity with continental areas (Chytrý et al., 2008).

Bio 2 has a high AUC (0.80), the third highest permutation importance of all the variables in the full final model, and shows the highest loss in AUC when it is omitted from the model (table 4). Bio 2 changes the least under climate change compared with the other climate variables. The small changes predicted in this value and the high importance is perhaps the reason climate change has little effect in the models future scenarios that include forest (Ratcliffe, 2015). The future model projections based only on climate show a dramatic change in suitable capercaillie because removing the non-climatic variables forces the other climate variables to gain a higher importance in the model.
Table 4 Mean diurnal range for present day and future scenarios in Scotland (Mean at presence sites 7.03). Values at present day leks sites used for testing are shown, which is subset of the know lek sites.

<table>
<thead>
<tr>
<th>Mean diurnal range ( ºC )</th>
<th>Value at present day lek sites</th>
<th>Present day climate</th>
<th>2070 climate ( low emission)</th>
<th>2070 climate (high emission)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>6.4 - 7.3</td>
<td>4.0-7.5</td>
<td>3.2-8.00</td>
<td>3.8-8.2</td>
</tr>
<tr>
<td>Mean</td>
<td>7.04</td>
<td>6.63</td>
<td>7.00</td>
<td>7.01</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.19</td>
<td>0.61</td>
<td>0.80</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Conclusions

The explorations of climate variables here support the future climate model predictions. Bio 2 is the most important climate variable in the full final model and changes very little under climate change. Although these predictions bode well for capercaillie the effects of changes in climate on Scots Pine and arthropods should be considered. Any negative change in arthropod hatching times and population could potentially be mitigated for with an increase in bog areas within forest.
References


Part II B: Technical Report
## Contents

List of figures .................................................................................................................. iii
List of tables.................................................................................................................... iii
Introduction....................................................................................................................... 1
Habitat suitability modelling............................................................................................. 3
Maxent modelling ............................................................................................................. 3
  * Maximum entropy ........................................................................................................ 3
  * Maxent software .......................................................................................................... 4
  * Maxent features ........................................................................................................... 5
Resolution and scale ........................................................................................................ 5
  * Extent of study area ..................................................................................................... 6
Temporal resolution .......................................................................................................... 6
  * Time period ................................................................................................................ 7
Data sources and methods ............................................................................................... 7
  * Climate data sources ................................................................................................. 7
  * Presence data .............................................................................................................. 10
Initial selection of variables to test ................................................................................ 10
Alternative climate variables ......................................................................................... 12
Forests .............................................................................................................................. 12
  * Native Pine .................................................................................................................. 12
  * Non-native coniferous and broadleaf ....................................................................... 12
  * ‘Future forest’ ............................................................................................................ 13
Roads and paths .............................................................................................................. 13
  * Response to roads and paths ..................................................................................... 14
Consideration of omitted variables ............................................................................... 17
Modelling process ........................................................................................................... 17
  * Selecting variables for the final model ..................................................................... 17
  * Accounting for sampling bias .................................................................................. 19
  * Spatial filtering .......................................................................................................... 20
  * Correlated variables ................................................................................................. 22
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maxent settings and requirements</td>
<td>23</td>
</tr>
<tr>
<td>Relative importance of variables</td>
<td>23</td>
</tr>
<tr>
<td>Model performance and validation</td>
<td>24</td>
</tr>
<tr>
<td>Validation points</td>
<td>24</td>
</tr>
<tr>
<td>Output surfaces</td>
<td>24</td>
</tr>
<tr>
<td>Logistic output</td>
<td>24</td>
</tr>
<tr>
<td>Caveats to model interpretation</td>
<td>25</td>
</tr>
<tr>
<td>Thresholds</td>
<td>25</td>
</tr>
<tr>
<td>Logistic scores at presence points</td>
<td>26</td>
</tr>
<tr>
<td>Projecting and clamping</td>
<td>28</td>
</tr>
<tr>
<td>Assessment of AUC as measure of performance</td>
<td>30</td>
</tr>
<tr>
<td>Appendix</td>
<td>37</td>
</tr>
</tbody>
</table>
List of figures

Figure 1 An overview of the modelling approach ................................................................. 2
Figure 2 Performance of selected and alternative climate baselines ................................ 9
Figure 3 Capercaillie response curves for distance to roads and footpaths ...................... 14
Figure 4 Distances to roads Scotland and locations of native pine forests .................... 19
Figure 5 Individual performance of each of the variables selected for modelling ............ 19
Figure 6 Climate heterogeneity in Scotland ........................................................................ 21
Figure 7 Logistic values of presence probability for training and test data .................... 27
Figure 8 Variables subject to clamping in future model projections ............................... 29
Figure 9 Assessing the models performance based on omission and pseudo commission errors ................................................................................................................................................................................................. 31

List of tables

Table 1 Variables considered for the model based on previous studies ............................. 11
Table 2 Bioclimatic variables considered ............................................................................. 11
Table 3 Contribution of the variables in each of the models ............................................. 10
Table 4 Correlations between variables in the final model ................................................ 22
Introduction

The aim of this research was to produce a Habitat Suitability Model (HSM) for capercaillie in order to determine the most important elements of the habitat and therefore the drivers behind their distribution. An overview of this process can be found in figure 1. This information was used to address the further research questions of what effect climate change would have on capercaillie and how do different forest scenarios affect the predicted changes? To address these questions and produce a reliable model, a number of methodological and data considerations had to be made. A critical review of the model and methods is made here and possible sources of error are considered. Incompleteness, spatial, temporal and classification errors are the main considerations. The probability modelling was carried out using a software called Maxent 3.3.3k (Philips et al., 2004), while the pre-processing was carried out in ArcGIS 10.3.1 (and 10.3.3) and made use of the tools available from the free package SDMtoolbox compatible with ArcGIS 10.3.1 (Brown, 2014).
Capercaillie presence location data

Climate data

Non-climatic data

Apply 10km filter to mitigate for sampling bias

Test for correlation

Create continuous data layers (Euclidean distance)

Standardise according to Maxent’s specifications

Run Maxent with climate and non-climatic date separately

Establish most important climate and non-climatic

Combine climatic and non-climatic variables

Model assessment and interpretation

Project the model onto future scenarios

Create projection files based on future climate and forest data

Figure 1 an overview of the modelling approach.
Habitat suitability modelling

The terminology in the field varies somewhat when referring to the model technique itself. Terms such as Habitat Suitability Modelling (HSM), Species Distribution Modelling and Ecological Niche Modelling are interchangeable in the literature, though some have recently attempted to define and differentiate the terms based on methods used (Sillero, 2011).

A key consideration in HSM is niche theory (Hirzel and Le Lay, 2008). Species often only inhabit part of the suitable habitat in a region (realised niche) because of restrictions to occupying the full fundamental niche (Phillips et al., 2006). Metapopulations are made up of interdependent sets of populations, and how connected these areas are partly determines how much of the fundamental habitat can be occupied (Nicholson et al., 2006). For example, when considering source-sink populations, an area may not be able to support a population (sink) without migration from other connected areas with high productivity (source) (Poiani et al., 2000). Suitable areas of habitat can be either structurally connected: i.e. physically connected by wildlife corridors or riparian habitat, or functionally connected, the areas are not physically connected but given suitable land cover a species can freely move between the areas (Tischendorf and Fahrid 2000).

Although HSMs have been used to make inferences about connectivity this is based on the assumption that animals will travel through the landscape selecting pathways in a similar way to which they select habitat (Gonzales and Gergel, 2007). This study assumed maximum dispersal when interpreting available suitable habitat (Li et al., 2015). Barriers to dispersal could mean that only a subset of the areas identified as suitable are accessible to capercaillie therefore the suitable habitat could be much smaller than the results suggest (Phillips et al., 2006).

Maxent modelling

*Maximum entropy*

Maximum information entropy is a measure of how much information an event contains. A highly probable event has a low information content (Usher, 1984). For instance, in the case
of the relative entropies of a fair die and a weighted die, the fair die has the highest entropy as all number have an equal probability of outcome (Usher, 1984). Moreover, an eight sided fair die has a higher entropy than a six sided fair die (Usher, 1984). Maximum entropy is used as a measure of biodiversity and is based on the relative abundance and rarity of each species (Harte, 2015). In the context of Maxent software the algorithm is iterated until it maximises the probability of occurrence at presence sites (Phillips and Dudík, 2008). This is achieved by producing a probability distribution (one that represents the true distribution) based on a set of constraints. These constraints are based on the relationship between occurrence points with environmental variables (Phillips and Dudík, 2008). Many functions can be fit to the relationship between occurrence points and the variables. The chosen function is the one closest to the variance of the variable across the study region, constrained by the mean of the variable across the presence sites (Elith et al., 2011). The model is parameterised to prevent over-fitting by using the error bounds of the variable responses.

**Maxent software**

Maxent software (Phillips et al., 2004) rapidly became one of the most popular methods for modelling habitat suitability and species distribution, and is now one of the most commonly used methods along with Generalized Linear Models (GLMs) and Generalized Additive Models (GAMs) (Anderson et al., 2006). Maxent uses a maximum entropy approach in contrast to the statistical methods of GLMs and GAMs.

Besides usability, one of the main attractions of Maxent software is its performance using presence only data (Halvorsen, 2013). Traditional HSMs require both presences points, where a species is known to be found, and absence points, where a species is known not to occur (Phillips and Dudík, 2008). As many studies are only able to collect presence data this is a major advantage of using Maxent (Yackulic et al, 2013). In addition presence only models remove the risk of classifying an area as unsuitable because of absence when the absence could be the result of dispersal limitations (Phillips and Dudík, 2008). However, presence only models have other limitations which are addressed in the following sections.
Maxent features

There are six types of features that fit Maxent uses to fit probability distributions (Phillips and Dudík, 2008). Many functions can be fit to the relationship curves between occurrence points and the variables. The chosen feature is the one closest to the variance of the variable across the study region, constrained by the mean of the variable across the presence sites (Elith et al 2011). By default Maxent selects the most appropriate combination of features for the model. The more features used the greater the complexity of the model. A model that is too complex will overfit affecting transferability which is important in climate change modelling. Where bias is accounted for simpler feature types can be applied (Syfert et al., 2013).

The features Maxent employs for continuous data are as follows:

**Linear** – pattern of values at presence locations is taken to be the same as the variation in the variable (Elith et al., 2011).

**Quadratic** – similar to the linear feature but takes into account the variance of the variables (Elith et al., 2011).

**Product** – uses linear and quadratic functions to fit function based on the covariance of two variables (Merow et al., 2013).

**Threshold** – a threshold is applied to make a continuous predictor binary (Merow et al., 2013). Multiple thresholds can be used to create a stepwise relationship (Elith et al., 2011).

**Hinge** – below a certain value the effect of the variable is held constant, above this value the relationship takes a linear form (Elith et al., 2011).

Resolution and scale

In habitat suitability modelling species respond differently to an environmental variable depending on scale (Hijmans and Graham, 2006). For example at the landscape scale areas of forest fragments have been shown to be one of the most important factors for capercaillie (Braunischand and Suchant, 2007). At the local scale studies have found that there is an optimum percentage of bilberry cover (Baines et al., 2004). A forest with ideal bilberry cover could not support a Capercaillie population if it was too small therefore at the landscape scale area of forest is dominant and a study at this scale could overlook the importance of percentage bilberry cover (Schneider, 2009). There is therefore a hierarchy of factors that
determine suitable habitat. Because of this multi-scale analyses that take into account the most important factors from each scale are extremely informative.

A multi-scale analysis was beyond the scope of this study. A study scale of 1km was chosen for both ecological and data availability reasons. Male capercaillie spend most of year within 1km of the lek despite having a range of up to ~ 5km (Picozzi et al., 1992). Females have a range of up to 30km (Cairngorms National Park Authority, 2015) however as the presence data was of male display grounds (leks) the males range was deemed more appropriate.

The climate data was available at a 30 second resolution (approximately 1km) from WorldClim [http://www.worldclim.org/]. Although resolution does not determine accuracy (Daly, 2006) low resolution climate datasets can mask substantial variation within the grid cells (Hijmans et al., 2005). WorldClim data at a 1km resolution has been used to create habitat models for other rare Scottish species (Silva et al., 2013).

**Extent of study area**

The extent of the study area is important because of the way Maxent utilises background points. In Scotland the extent is mainly restricted by natural features, the coastline for example, and the border with England is the only human defined boundary in this context. The extent chosen is an important consideration as inappropriate area coverage can skew Maxent’s reported performance (Guisan and Thuiller, 2005). For example, because the performance is measured based on the ratio of presence points to background cells (Yackulic et al., 2013), which are pseudo absences rather than true absence data, the performance can be inflated by increasing the number of background points or choosing an extent that includes large areas of inappropriate habitat such as large areas of desert when modelling a riparian species (Yackulic et al., 2013). Therefore the background sample should be chosen from the sampling distribution (Phillips and Dudík, 2008).

**Temporal resolution**
One of the main limitations of HSMs is that they provide only a snapshot of a dynamic system (Guisan and Thuiller, 2005) as both species populations and habitat fluctuate. The suitability of the temporal resolution of datasets was considered here and the presence data used was leks that are currently active. All of the non-climatic data is from the past 10 years. The WorldClim climate data is averaged from 1950-2000 however it had a better performance compared to more recent MetOffice climate data (1981-2010) when tested at the same resolution (see section ‘Climate data sources’).

**Time period**

Capercaillie favour mature Scots Pine forests therefore conservation schemes planting trees at present would take some time to reach their full ecological potential for capercaillie. Pine trees maturity does not linearly increase with age, however age is a good proxy for maturity (Kärenlampi and Riekkinen, 2004). The model uses data available through WorldClim which is available for two future time periods: 2050 (average for 2041-2060) and 2070 (average for 2061-2080). 2070 was chosen as any young current native pine will potentially be mature by this time.

**Data sources and methods**

The section below provides information on the nature of the data and any alterations made before entering it into the model. Once pre-processing was complete all data was converted to a continuous surface by applying a Euclidean distance. The model performed better with continuous data, which could be the result of moderating the effects of data inaccuracies which resulted in capercaillie leks falsely being located outside forest.

**Climate data sources**

Various present day climate surfaces and climate change projections are available (e.g., [http://www.ualberta.ca/~ahamann/data/climateeu.html], [http://www.cru.uea.ac.uk/data]). They vary in time period used to calculate current climate, how far they project into the future and the resolution at which they are available.
Habitat Suitability Models used to predict climate change are sensitive to baseline climate used (Bedía Jiménez et al., 2012). The presence data used to create the HSM were leks that are currently active or have become ‘extinct’ this year. Therefore it would be expected that the leks presence is more appropriately modelled with more recent climate data. However, as capercaillie numbers have declined so rapidly in the past 40 years a more historic climate average might be expected to show more suitable habitat. MetOffice base line climates from 1961-1991 and 1981-2010 (available at a 5km resolution [http://www.metoffice.gov.uk/climatechange/science/monitoring/ukcp09/]) were compared with WorldClim baseline (1950-2000) at a 5km resolution. Additionally a comparison between the WorldClim 1km and 5km climate data was made to investigate the effects of scale. Figure 2 show the predictive performance of each of the baselines.

The WorldClim data has a better predictive performance than the MetOffice data when assessed using the area under the curve (AUC) of the receiver operating characteristic (ROC). Despite the slightly better performance of WorldClim data at a 5km resolution a 1km resolution was chosen because of the suitability of the resolution to this study. Multiple General Circulation Model (GCM) are available from WorldClim however the HadGEM2-ES was chosen for this study because of its performance at latitudes appropriate to Scotland (Bellouin et al., 2011). Different GCMs generate different predictions according to methods used (Daly, 2006). GCMs are sometimes combined to average out the differences for large continental scale analysis however it has been shown that some individual models perform best at regional scale (Fordham et al., 2011).
Performance of different climate baselines in predicting capercaillie presence illustrated by the relative predictive performance as measured by Area under the curve (AUC) of the receiver operating characteristic (ROC)

<table>
<thead>
<tr>
<th>Climate Baseline</th>
<th>Training AUC</th>
<th>Test AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MetOffice climate 1961–1991 (5km resolution)</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>MetOffice climate 1981–2010 (5km resolution)</td>
<td>0.89</td>
<td>0.86</td>
</tr>
<tr>
<td>WorldClim climate 1950–2000 (5km resolution)</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>WorldClim climate 1950–2000 (1km resolution)</td>
<td>0.92</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Red Line = Training data  
Blue line = Test data  
Black line = Random prediction (AUC = 0.5)

Figure 2: The various predictive performance of MetOffice and WorldClim climate baselines using the climate variables selected for the final climate model (see table 3). The WorldClim data has the highest AUC values and most ideal shape with the peak of the curve being close to 0.1.
Presence data

The presence data shows locations of known leks (male capercaillie display grounds) across Scotland. The data is distributed by the Royal Society for the Protection of Birds (RSPB) and publicly available at a 10km resolution but was made available to this study at a 1km accuracy. The following information regarding the presence data was provided by via the RSPB (Gareth Marshall, personal communication, 26th June 2015). The presence location data comes from annual lek surveys carried out in the spring (March, April and May) of each year across Scotland. The presence only data in this study shows sites that are currently occupied and includes sites that are have been occupied since 2005 but have recently become extinct. These sites are checked for re-colonisation for around 5 years after extinction. Every known lek is visited each spring. A lek is deemed occupied if there is clear evidence of presence, for instance fresh dropping or feathers. There is no standardised method for monitoring leks and the surveys rely on full time staff, game keepers and volunteers. ‘Cold searches’ are carried out to search for new leks in suitable areas.

This model used 34 presence sites that were validated against 25 test sites. Other methods are affected by insufficient sample size and there is no method to determine the minimum sample size (Guisan and Thuiller, 2005). However Maxent is not as sensitive and has shown good performance regardless of sample size (Anderson and Gonzalez 2011).

Initial selection of variables to test

A literature review was used to establish which variables should be considered for inclusion in the model. Additionally, a model was run with categorical land cover to establish any other land cover variables that might be important. This resulted in some erroneous results, for example water was of the third highest importance as presence points had been located in lochs. This is possibly the result of inaccuracies in the data or the result of the 1km resolution presence points or a combination. Table 1 shows the variables included for testing and the reasons for inclusion (excluding bioclimatic variable, see table 2).
Table 1 shows variables potentially influential in Capercaillie distribution as suggested by findings in the literature:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Influence on Capercaillie</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>Minimum and maximum temperature</td>
<td>Influence on egg viability (Wegge et al., 2010, Moss et al., 2001, Slagsvold and Grasaas, 1979)</td>
</tr>
<tr>
<td>May</td>
<td>Minimum and maximum temperature</td>
<td>Influence on egg viability and chick mortality (Wegge et al., 2010, Moss et al., 2001)</td>
</tr>
<tr>
<td>June</td>
<td>Minimum and maximum temperature and precipitation</td>
<td>Influence on chick mortality (Wegge and Kastdalen, 2007, Moss et al., 2001)</td>
</tr>
<tr>
<td>Snow cover</td>
<td>Number of days with snow cover</td>
<td>Number of days with &gt;10cm snow found to be important (Braunischand and Suchant, 2007)</td>
</tr>
<tr>
<td>Scots Pine forest</td>
<td>Forest areas where the Scots Pine is dominant. Further classed according to stand age.</td>
<td>Scots pine are a food source for adults and the forest areas are used for mating displays (Picozzi et al., 1992)</td>
</tr>
<tr>
<td>Bog in forest</td>
<td>Bog areas that are within forests</td>
<td>A source of high protein food (Summers et al, 2004)</td>
</tr>
<tr>
<td>Roads</td>
<td>Distance to all roads in Scotland</td>
<td>Disturbance can displace capercaillie (Ruddock and Whitfield, 2007)</td>
</tr>
<tr>
<td>Foot paths</td>
<td>Distance to core footpaths</td>
<td>Disturbance can displace capercaillie (Ruddock and Whitfield, 2007)</td>
</tr>
<tr>
<td>Broadleaf</td>
<td>Broadleaf forest, Birch and Oak are of particular importance</td>
<td>Food source from common ground cover (bilberry) (Summers et al, 2004)</td>
</tr>
<tr>
<td>Non-native coniferous</td>
<td>Non-native coniferous forest, Sitka Spruce is of particular importance</td>
<td>Alternative food source for adults and breeding ground (Picozzi et al., 1996)</td>
</tr>
<tr>
<td>Bog</td>
<td>Land classified as bog (not in forest)</td>
<td>A potential source of high protein food (Summers et al, 2004)</td>
</tr>
<tr>
<td>Agricultural land</td>
<td>Agricultural and pastoral farmland combined</td>
<td>A proxy for influence of predators (Huhta et al., 1996)</td>
</tr>
</tbody>
</table>

Table 2 shows bioclimatic variables ubiquitous in habitat suitability modelling:

<table>
<thead>
<tr>
<th>Name</th>
<th>Bioclimatic description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio 1</td>
<td>Annual Mean Temperature</td>
</tr>
<tr>
<td>Bio 2</td>
<td>Mean Diurnal Range (Mean of monthly (max temp - min temp))</td>
</tr>
<tr>
<td>Bio 3</td>
<td>Isothermality (BIO2/BIO7) (* 100)</td>
</tr>
<tr>
<td>Bio 4</td>
<td>Temperature Seasonality (standard deviation *100)</td>
</tr>
<tr>
<td>Bio 5</td>
<td>Max Temperature of Warmest Month</td>
</tr>
<tr>
<td>Bio 6</td>
<td>Min Temperature of Coldest Month</td>
</tr>
<tr>
<td>Bio 7</td>
<td>Temperature Annual Range (BIO5-BIO6)</td>
</tr>
<tr>
<td>Bio 8</td>
<td>Mean Temperature of Wettest Quarter</td>
</tr>
<tr>
<td>Bio 9</td>
<td>Mean Temperature of Driest Quarter</td>
</tr>
<tr>
<td>Bio 10</td>
<td>Mean Temperature of Warmest Quarter</td>
</tr>
<tr>
<td>Bio 11</td>
<td>Mean Temperature of Coldest Quarter</td>
</tr>
<tr>
<td>Bio 12</td>
<td>Annual Precipitation</td>
</tr>
<tr>
<td>Bio 13</td>
<td>Precipitation of Wettest Month</td>
</tr>
<tr>
<td>Bio 14</td>
<td>Precipitation of Driest Month</td>
</tr>
<tr>
<td>Bio 15</td>
<td>Precipitation Seasonality (Coefficient of Variation)</td>
</tr>
<tr>
<td>Bio 16</td>
<td>Precipitation of Wettest Quarter</td>
</tr>
<tr>
<td>Bio 17</td>
<td>Precipitation of Driest Quarter</td>
</tr>
<tr>
<td>Bio 18</td>
<td>Precipitation of Warmest Quarter</td>
</tr>
<tr>
<td>Bio 19</td>
<td>Precipitation of Coldest Quarter</td>
</tr>
</tbody>
</table>
Alternative climate variables

In this study the weather conditions in April, May and June were considered with the final full model including only June precipitation and maximum temperature. Previous studies modelling bird habitats have suggested the use of summer averages (April, May, June and July) (Harrison et al. 2003). This data had a very similar performance and very highly correlated with April and June climate. April and June were selected over summer averages because of their ecological relevance as shown by previous studies (Wegge et al., 2010, Wegge and Kastdalen, 2007, Moss et al., 2001, Slagsvold and Grasaas, 1979). However some studies suggest July temperatures are important for Scots Pine growth, which could be relevant in this model although a HSM explicitly for Scots Pine would be more appropriate to test this (Grace et al., 2002).

Forests

Native Pine

It is only very recently that a full and comprehensive survey of native woodland in Scotland was completed. The Native Woodland Survey of Scotland (NWSS) records areas of native forest >0.5ha at a 1:10000 resolution. Previous to this data was collated from various surveys and was temporally disjointed (Patterson et al., 2014). This data was available in vector format and was highly spatially and temporally accurate. Forest is deemed native where >50% of the trees were native. The fuzziness of the real world can make classification problematic and result in over simplification resulting in an incorrect approximation of suitable habitat. The areas of Scots Pine were extracted from NWSS where native pine was the dominant habitat. The dominant stand age was also extracted where mature and veteran pine were combined and all other age classes made up the ‘Young Native Pine’ dataset.

Non-native coniferous and broadleaf

Non-native coniferous and broadleaf forests were extracted from the 25m raster land cover survey 2007 (Morton, 2011). Where there was a disagreement between classification for this data and the native pine the native pine took precedence and the area was subtracted from the non-native pine dataset. Non-native coniferous was included as capercaillie occasionally
occupy Sitka Spruce plantations. No distinctions were made between types of broadleaf forests although female capercaillie have been found in Oak and Birch forests where the understory is often dominated by bilberry. Additionally there is some evidence to suggest that capercaillie benefit from mixed forests, though this was not explicitly investigated here (Summers et al, 2004).

‘Future forest’
Areas suitable for future forest regeneration were required to model an increased forest area scenario in 2070. Data available from the Forestry Commission that identifies core and target regeneration areas was investigated [http://www.forestry.gov.uk/forestry/infdg5bzf]. However, the final dataset used to model the effects of future native pine forests was from a study carried out by Sing et al. (2013). This study was carried out specifically to identify suitable area for regeneration. Areas that are currently forested, protected, occupied by other important ecological land cover types, or biophysically unsuitable were removed. The remaining areas constitute areas for potential woodland expansion. This data was further modified for this study by clipping the area to the native pine extent. The native pine extent is a shapefile publically available from the Forestry Commission and is based on previous work (O’Sullivan, 1977).

When projecting onto future scenarios, what is currently classed as young pine was merged with mature pine, assuming present day young pine will be mature in 2070. The suitable areas identified by Sing et al. (2013) were entered into the future model as young pine.

Roads and paths
Roads and paths were included to investigate the effects of disturbance. Capercaillie are known to be sensitive to disturbance and forests are increasingly used for recreational activities. The paths dataset is from 2014 and is available from Scottish Natural Heritage [http://gateway.snh.gov.uk/natural-spaces/index.jsp]. It is a collection of core paths reported by different district councils. As this is a compilation of data sources and only including core paths means this dataset is somewhat incomplete so has limited function in the model.

The road dataset is a topological network of all the roads in Scotland at a 1:25000 scale and is updated twice a year (OS Open Roads, 2015). The resulting surface of all the merged roads
created a dense network of roads. This is perhaps one of the reasons that the response curve of capercaillie for roads was not as expected.

Response to roads and paths
Forest has increasingly used for recreational (Cairngorms National Park Authority, 2015). More footpaths and mountain bike tracks and increased use of cars in the same period could have negatively affected capercaillie (Ruddock and Whitfield, 2007). The distance to core footpaths and distance to all roads was included in the final model. The methods of this model cannot test or quantify the effects of disturbance however there does not appear to be significant preference to forests far from roads and footpaths. The response curve shows a positive relationship between Capercaillie and proximity to roads and paths (figure 3). The mean distance to roads was 330m and 2641m for foot paths.

There are a number of possible reasons for the apparent relationship. For instance, road networks could be close to forest for access for commercial and recreational uses. Forests and roads could also be strongly related because of their position in the landscape. Forests don’t tend to grow successfully over around 500m in Scotland (Miller and Cummins, 1982) and it is likely that roads are also located at lower altitudes due to the high topographic variation of

Figure 3 shows a positive relationship between distance to roads and paths and probability of capercaillie presence. The response curves shown here are for the variables used in isolation. Covariance results in slightly different response curves in the context of other variables (see appendix)
Scotland. The results of a test of correlation using Pearson’s Coefficient suggest that distance to roads and distance to forest are not highly correlated (table 4). However, visually inspecting figure 4 forests are not present in the whiter areas of the map where the distance to roads is greater. The area of Scotland far from roads is very small so there is limited probability that these areas fulfil all other habitat requirements for capercaillie and are accessible for colonisation.
Figure 4 shows Euclidean Distance to roads and areas of native pine. The pine forests tend to be located in areas with a higher road density. There are few forests in the white areas of the map, where the distance to roads is greatest.
**Consideration of omitted variables**

Among the many other possible elements there are a handful of variables that have not been included which are often included in HSMs. These include elevation, slope and aspect (Daly, 2006). Elevation is used to create the interpolated climate surfaces so does not need to be explicitly included (Hijmans et al., 2005). There is a limited number of variables that can be included (Kramer-Schadt et al., 2013). Although slope and aspect could have some importance they would perhaps be more appropriately included in a HSM to determine suitable areas for Scots Pine which could then combined with the results shown here.

**Modelling process**

**Selecting variables for the final model**

Once highly correlated variables had been removed the jacknife results, showing permutation importance and loss in AUC when a variable was removed were used in combination to select the final set of variables for the models (Table 3). Some of the relationships were most likely driven by factors not accounted for, for example the correlation with roads and capercaillie (see section Response to disturbance’). Figure 5 shows the AUC value for each of the variables and the response curves for all of the variables can be found in the appendix.
Table 3 contribution to the model as measured by permutation importance and loss in AUC value when tested against validation data. The table is ordered by permutation importance for each model. This measured the drop in AUC when the values for that environmental layer are randomised while all other layer values retain their values. The AUC loss is a measurement of how much the model lost in its predictive performance when the variable was removed from the model. The cells highlighted orange show the variable with the highest drop in AUC suggesting this variable has the highest information that is not present in the other variables. The negative values in the final model show an improvement in the models ability to predict presence in the validation data when that variable is removed. These variables were retained because of their contribution in permutation importance.

<table>
<thead>
<tr>
<th>Climate only</th>
<th>Permutation Importance</th>
<th>Test AUC Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>June Precipitation</td>
<td>38.81</td>
<td>0.0133</td>
</tr>
<tr>
<td>April minimum temperature</td>
<td>18.55</td>
<td>0.0001</td>
</tr>
<tr>
<td>June maximum temperature</td>
<td>18.13</td>
<td>0.0282</td>
</tr>
<tr>
<td>Mean annual temperature (Bioclim 1)</td>
<td>14.34</td>
<td>0.0068</td>
</tr>
<tr>
<td>Mean diurnal range (Bioclim 2)</td>
<td>10.17</td>
<td>0.0037</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-climatic</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mature and veteran native pine distance</td>
<td>76.00</td>
<td>0.0175</td>
</tr>
<tr>
<td>Younger native pine distance</td>
<td>15.61</td>
<td>0.0020</td>
</tr>
<tr>
<td>All roads distance</td>
<td>5.22</td>
<td>0.0002</td>
</tr>
<tr>
<td>Non-native coniferous</td>
<td>2.15</td>
<td>0.0023</td>
</tr>
<tr>
<td>Core paths distance</td>
<td>0.57</td>
<td>0.0054</td>
</tr>
<tr>
<td>Bog in forest distance</td>
<td>0.35</td>
<td>0.0008</td>
</tr>
<tr>
<td>Broadleaf distance</td>
<td>0.11</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Final model – combined climatic and non-climatic</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mature and veteran native pine distance</td>
<td>49.08</td>
<td>0.0067</td>
</tr>
<tr>
<td>Younger native pine distance</td>
<td>21.22</td>
<td>0.0042</td>
</tr>
<tr>
<td>Mean Diurnal range (Bioclim 2)</td>
<td>13.36</td>
<td>0.0087</td>
</tr>
<tr>
<td>Mean annual temperature (Bioclim 1)</td>
<td>5.36</td>
<td>-0.0043</td>
</tr>
<tr>
<td>Non-native coniferous</td>
<td>3.08</td>
<td>-0.0009</td>
</tr>
<tr>
<td>Bog in forest distance</td>
<td>3.01</td>
<td>0.0014</td>
</tr>
<tr>
<td>June maximum temperature</td>
<td>1.75</td>
<td>0.0031</td>
</tr>
<tr>
<td>Broadleaf distance</td>
<td>1.07</td>
<td>0.0004</td>
</tr>
<tr>
<td>All roads distance</td>
<td>0.82</td>
<td>-0.0016</td>
</tr>
<tr>
<td>Core paths distance</td>
<td>0.74</td>
<td>0.0019</td>
</tr>
<tr>
<td>June precipitation</td>
<td>0.49</td>
<td>0.0043</td>
</tr>
</tbody>
</table>
Accounting for sampling bias can substantially improve a model’s predictive capabilities (Phillips et al., 2006). The number of papers that focus on the effects of sampling bias and possible solutions illustrates its importance (Boria et al., 2014, Kramer-Schadt et al., 2013, Syfert et al., 2013 and Phillips et al., 2006). Correlations between the elements of the landscape and the environment can be imposed through sampling bias (Elith et al., 2011). A common source of bias and misleading correlation comes from sampling near access points such as roads (Yackulic et al., 2013). As well as the apparent correlation with the roads, in a landscape where roads follow geomorphological features such as river valley bottoms an environmental bias can also occur due to the lack of sampling on high ground, resulting in a bias in altitude and therefore temperature (Yackulic et al., 2013).

Sampling bias effort across a region can be difficult to account for as many studies use herbarium records and therefore metadata is rarely available (Yackulic et al., 2013). The presence data in this study is comprehensive and complete in comparison to many herbarium
records. Cold searches are carried out to look for new leks and there is a network of enthusiasts that could report sightings, therefore it is believed that the vast majority of leks are known. If all sites are known then sampling bias is not an issue. However because there is no standardised method for monitoring sites and the experience of the surveyors is varied, sampling errors could potentially arise.

*Spatial filtering*

The relative reliability of the presence data in this study means that the main source of sampling bias comes from over-representing certain climatic conditions. The majority of the capercaillie population in Scotland is found in the Cairngorm National Park (CNP) (Cairngorms National Park Authority, 2015). Over representing leks in this area could lead to overfitting the model to climate conditions of the CNP. Geographical clustering of presence points can falsely inflate model accuracy causing commission errors as a result of autocorrelation of models residuals (Kramer-Schadt et al., 2013). Although the climate conditions may be the reason they favour the CNP the other presence locations suggest that they are able to survive in areas with different climatic conditions figure 6. Moreover some evidence suggests that capercaillie populations in the CNP are relatively successful because of the relatively large tracts of ancient pine forest so over representing climate in this area could be misleading (Cairngorms National Park Authority, 2015).

Spatial filtering is a popular method to manage sampling bias caused by spatial autocorrelation (Boria et al., 2014, Radosavljevic and Anderson, 2014) and was used to prevent overfitting and environmental bias in this study. A spatial filter of 10km was chosen. This filtering resolution is recommended in heterogeneous mountainous areas (Boria et al., 2014) and a tool from SDMtool box (Brown, 2014) was used to calculate climate heterogeneity in Scotland. This tool uses the Eigenvalues of a Principle Component Analysis (PCA) of the climate variables to assess the differences in climate across Scotland, figure 6 shows the result. The results show mixed climate variability across Scotland with clear gradient of increasing homogeneity from west to east. Further filtering was considered but the number of presence points dropped to the point where the model over generalised.
Figure 6 shows the variation in climate as a measure of heterogeneity. The west of Scotland has a more variable climate than the East. The majority of capercaillie leks are located in the Cairngorm National Park where climate is relatively homogenous however leks are also located in areas with high climate heterogeneity.
Correlated variables

Maxent and other HSM software have an enhanced predicted performance when highly correlated variables are removed. Pearson’s Correlations Coefficient as calculated using the ArcMap tool Band Statistics was used to identify highly correlated variables where Pearson’s |r| > 0.75 (Kramer-Schadt et al., 2013). This process was aided by the use of an SDMtoolbox tool (Brown, 2014) for the large Bioclim dataset. The tool ‘Remove highly correlated variables’ was used though the results were interrogated and adjusted to retain the variables most ecological importance.

Table 4 shows the correlations between the final variables. Some of the variables were highly correlated but were retained because of their contribution to the model.

<table>
<thead>
<tr>
<th>Layer</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>-0.40</td>
<td>-0.21</td>
<td>0.02</td>
<td>0.10</td>
<td>0.29</td>
<td>-0.44</td>
<td>-0.07</td>
<td>0.36</td>
<td>-0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>2</td>
<td>-0.40</td>
<td>1.00</td>
<td>-0.04</td>
<td>-0.07</td>
<td>0.13</td>
<td>-0.31</td>
<td>0.63</td>
<td>0.19</td>
<td>-0.13</td>
<td>0.21</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>-0.21</td>
<td>-0.04</td>
<td>1.00</td>
<td>-0.45</td>
<td>-0.17</td>
<td>-0.28</td>
<td>0.53</td>
<td>-0.40</td>
<td>-0.30</td>
<td>-0.26</td>
<td>-0.49</td>
</tr>
<tr>
<td>4</td>
<td>0.02</td>
<td>-0.07</td>
<td>-0.45</td>
<td>1.00</td>
<td>-0.04</td>
<td>-0.12</td>
<td>-0.43</td>
<td>0.77</td>
<td>0.07</td>
<td>0.65</td>
<td>0.81</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
<td>0.13</td>
<td>-0.17</td>
<td>-0.04</td>
<td>1.00</td>
<td>0.13</td>
<td>-0.03</td>
<td>-0.01</td>
<td>0.08</td>
<td>-0.04</td>
<td>-0.03</td>
</tr>
<tr>
<td>6</td>
<td>0.29</td>
<td>-0.31</td>
<td>-0.28</td>
<td>-0.12</td>
<td>0.13</td>
<td>1.00</td>
<td>-0.20</td>
<td>-0.21</td>
<td>0.06</td>
<td>-0.18</td>
<td>-0.13</td>
</tr>
<tr>
<td>7</td>
<td>-0.44</td>
<td>0.63</td>
<td>0.53</td>
<td>-0.43</td>
<td>-0.03</td>
<td>-0.20</td>
<td>1.00</td>
<td>-0.23</td>
<td>-0.35</td>
<td>-0.11</td>
<td>-0.42</td>
</tr>
<tr>
<td>8</td>
<td>-0.07</td>
<td>0.19</td>
<td>-0.40</td>
<td>0.77</td>
<td>-0.01</td>
<td>-0.21</td>
<td>-0.23</td>
<td>1.00</td>
<td>0.09</td>
<td>0.94</td>
<td>0.72</td>
</tr>
<tr>
<td>9</td>
<td>0.36</td>
<td>-0.13</td>
<td>-0.30</td>
<td>0.07</td>
<td>0.08</td>
<td>0.06</td>
<td>-0.35</td>
<td>0.09</td>
<td>1.00</td>
<td>0.03</td>
<td>0.15</td>
</tr>
<tr>
<td>10</td>
<td>-0.12</td>
<td>0.21</td>
<td>-0.26</td>
<td>0.65</td>
<td>-0.04</td>
<td>-0.18</td>
<td>-0.11</td>
<td>0.94</td>
<td>0.03</td>
<td>1.00</td>
<td>0.62</td>
</tr>
<tr>
<td>11</td>
<td>0.03</td>
<td>0.00</td>
<td>-0.49</td>
<td>0.81</td>
<td>-0.03</td>
<td>-0.13</td>
<td>-0.42</td>
<td>0.72</td>
<td>0.15</td>
<td>0.62</td>
<td>1.00</td>
</tr>
</tbody>
</table>

1 = Roads, 2 = Mean Annual Temperature, 3 = Mean Diurnal Range, 4 = Broadleaf Forest, 5 = Core footpaths, 6 = June Precipitation, 7 = June Maximum Temperature, 8 = Mature Native Pine, 9 = Non-native coniferous, 10 = Young Native Pine, 11 = Bog in forest

Highly correlated variables were also used as substitute for variables with no prediction available. For example the number of snow days is positively correlated with capercaillie though no prediction was available in 2070. However the number of snow days was highly correlated with Bio 6 (Minimum temperature of coldest month), although in the final model bio 6 was not included as it had no contribution.

Table 3 shows the set of variables for the final model. Although mature pine, young pine and bog in forest are highly correlated they still contain relevant and important information needed to determine the most suitable habitat for capercaillie.
Maxent settings and requirements

Maxent requires all the data to be in the same format with the same projection and cell size. A WGS84 project was used. Although this is not geographically appropriate in the context of Maxent the importance of projection is more acute for large study areas where there are significant changes in cell size depending on the latitude. The environmental layers are required to be in ASCII format with the presence points in CSV format. Results summarized in HTML file and are available in graphical and CSV format for further analysis.

Maxent performed well with default settings. A systematic testing of regularisation parameters, as recommended by Anderson and Gonzalez (2011), did not result in any improvements in model performance. Regularisation parameters are applied to prevent overfitting by smoothing the model (Phillips and Dudík, 2008). They are a method to reduce overfitting thereby enabling the model to accurately predict independent validation data (Radosavljevic and Anderson, 2014). This is done by penalising models that are too complex as they include too many features (Anderson and Gonzalez 2011).

Changes to the default prevalence were considered but given that Capercaillie are not rare in favourable landscapes, the default prevalence of 0.5 was deemed suitable under the assumptions outlined by Elith et al. (2011) (see section ‘Output surfaces’).

Relative importance of variables

The jacknife test is a non-parametric test of bias and variance (Efron and Stein, 1981). Figure 5 shows the jacknife test results which identifies the importance of each variable by assessing its predictive capacity when used in isolation (Elith et al., 2011). This test is also used to measure any drop in model performance when a variable is excluded.

The jacknife test gave marginally different values for each variable in different model runs. This highlights one of the limitations of the jacknife test in that if the estimator is not sufficiently smooth this leads to inconsistent estimation of variance (Shao and Wu, 1989). An estimator is used to estimate unknown parameters for variables given a certain probability distribution (Hodge and Seed, 1977). This again illustrates the importance of adjusting for sampling bias in preventing overly complex relationships. Smoothing can be achieved by using regularisation (Anderson and Gonzalez, 2011).
Model performance and validation

Validation points

A separate set of 25 capercaillie lek locations was set aside for model validation. Using independent data is favoured over methods such as cross validation because spatially autocorrelated data can inflate the AUC. Ideally validation data would be from a different region or time period to ensure independence and reduce the effects of spatial autocorrelation (Radosavljevic and Anderson, 2014).

Output surfaces

The Maxent software offers three different graphical outputs of the habitat suitability map. The logistic output is the default graphical result for the model’s surface. These surfaces can be further modified to presence absence maps through the selection of a threshold to assess changes in area and distribution though this study focused on visual interpretation of habitat distribution.

Logistic output

The graphical results of Maxent display the distribution of suitable habitat. Maxent assigns a probability to each cell based on the environmental values of that cell (Elith et al., 2011). The sum of the cells equals 1, though for study areas with varying sizes the results are not easily comparable for the raw output (Phillips and Dudík, 2008). The logistic output is proportional to the raw output and is monotonically related so areas have the same relative importance (Phillips and Dudík, 2008). The logistic output is an approximation of probability of presence but because of the nature of Maxent, having no information about absence localities, the results should be interpreted with caution.

Some have recommended that the logistic output should not be used because it assumes a probability of presence of 0.5 at locations with suitable environment which for many rare species is unsuitable (Yackulic et al, 2013). For example, a Jaguar will not have a presence likelihood this high even in areas with suitable habitat (Elith et al., 2011). However given that Capercaillie are not inherently rare as they were seen in great numbers historically, this presence value used to scale the raw output is reasonably appropriate. For these reasons and
ecological reasons (see section ‘Caveats to model interpretation’) the logistic output is more appropriately interpreted as suitability of habitat with red areas of the map being highly suitable habitat, rather than presence probability. Therefore, for the results in this study the logistic output is referred to as ‘approximate presence probability’.

_Caveats to model interpretation_
Maxent identifies suitable habitat based on the known locations the species is currently found in. The output shows the distribution of suitable habitat rather than explicitly the distribution of capercaillie. The reasons for this distinction are many. Firstly a probability of presence interpretation of Maxent’s logistic output has been criticised because of the assumptions needed to scale the raw values (Yackulic et al, 2013). Secondly, there are many ecological reasons why probability of presence interpretation may be inappropriate. For instance, capercaillie may only able to access a limited area of suitable habitat due to barriers to dispersal and issues of connectedness (Phillips et al., 2006). This has an impact on both their distribution and abundance due to the role of source-sink populations which could have profound implications for the success of the species, especially in future climates (Pascual-Hortal and Saura, 2008). Assuming capercaillie presence means that the area can support and self-sustaining a population is a critical conjecture. In addition, whether a species is present at a location only provides some information about suitability; abundance, mortality and chick survival which not explicitly accounted for.

The results show suitability of habitat for adult capercaillie based on the presence of leks. Female capercaillie have a much greater range than male capercaillie and may be able to access and survive in different forest to the males (Cairngorms National Park Authority, 2015). Evidence suggests that males establish leks close to other existing leks rather than extending to their full potential natal range (Hjeljord et al., 2000). Therefore their distribution, and suitable habitat needed for a viable population, is somewhat restricted by male mobility. There may be suitable habitat that the model has been unable to account for because of the limitations of male dispersal and the fragmented nature of the forest.

_Thresholds_
A visual inspection and interpretation of the differences between the distributions of suitable habitat between models was made in this study. This makes it difficult to comprehensively
assess changes so conversion to a binary presence-absence map are often made, from which changes in area and distribution can be quantified. However this requires the selection of a threshold on the logistic values above which the cells are determined as present and below absent, where the habitat suitability is so low a capercaillie is extremely unlikely to be found. However the selection of a threshold can substantially change the apparent area of suitable habitat. It is crucial to select the optimum threshold as it contribute further bias to the model (Norris, 2014), although unfortunately there is no general purpose rule for the method of selection (Hu and Jiang, 2010). Two popular thresholds, Maximum Sensitivity and Specificity (MSS) and Equal Sensitivity and Specificity (ESS), have been employed by many studies following the results of Liu et al (2005), however the debate is ongoing (Norris, 2014). Nevertheless visual outputs can be misleading too. Map interpretation can be influence by map design, colour and thresholds in the legend (Chainey and Ratcliffe, 2005).

**Logistic scores at presence points**

Maxent runs the probabilistic method until presence probability at occupied cells is maximised. The logistic values at the training and test presence points can aid in assessing a models performance. The histograms for the logistic values of the presence points for each of the present day models are shown in figure 7. In addition to differences in the AUC, the distribution of logistic scores between the models varies. A high probability of presence is expected at presence sites and the default presence probability at suitable sites is 0.5 (Elith et al, 2005). Both the climate only model and non-climate model assigns 66% of the training points have a presence probability of >0.5. Additionally both models assign 71% of the test values a presence probability of >0.5. Figure 7a shows the logistic values of presence points for the final model and although the training data received a generally lower logistic value, 83% of the test data has a logistic presence probability of >0.5. However this assessment is applying a threshold, considering the distribution of the values is more important. The histograms should show a positive skew towards higher probability values. The final model showing a skew to the right as expected from a well performing model. The number of presence points assigned a low logistic value illustrates the potential errors in estimated suitable habitat that can arise from creating a binary presence-absence map.
Graphs to show the probability of presence at different logistic output bands

Figure 7 Logistic values of presence probability for training and test data
**Projecting and clamping**

Maxent uses the relationships between the present day variables and presence points to project onto future or historical climates. This feature of Maxent makes it a popular tool for modelling suitable habitat under climate change (e.g., Li et al., 2015, Meyer et al., 2014, Silva et al., 2013). In order to project onto future scenarios Maxent requires the future surfaces to be grouped in a single file and for all the present day variables to have corresponding future layers. Only the climate variables and the mature and young native pine had future climate layers with the remaining variables being held constant.

Maxent interrogates the future surfaces for suitable habitat based on current relationships between presence points and variables and the relative importance of each variable (Phillips, 2005). Where Maxent encounters environmental values outside of the range seen in present day it ‘clamps’ the new value to the maximum seen in the present day. Predictions of suitable habitat identified in these areas must be interpreted with extreme caution where clamping occurs (Phillips and Dudík, 2008).

Figure 8 shows the Multivariate Environmental Similarity Surface (MESS) output Maxent generates that identifies areas and variables that have been subject to clamping in the climate only model under the high emission scenario. These areas could have values that are incompatible with capercaillie biology but have been clamped to maximum present day values (Elith et al 2010). Therefore it is important to understand the biological responses to various climate variables and how the climate change.
Figure 8 shows areas that have been subject to clamping in the 2070 projection under a high emission scenario. Where more than one variable is subject to clamping the one with the values most different to present day are shown.
Assessment of AUC as measure of performance

The models predictive performance is measured using the area under the curve (AUC) of the receiver operating characteristic (ROC), which is a threshold independent measure of performance traditionally based on omission and commission errors (Jiménez-Valverde, 2012). The AUC measure is based on the comparison of presence probability at presence locations with the null model of random predictive performance where AUC = 0.5 (Slater and Michael, 2012). Because Maxent is a presence only modelling technique the background cells are used as pseudo-absences. As a result of this the interpretation of the AUC is the probability of that a presence location will receive a higher probability of presence than a background cell (Phillips and Dudík, 2008).

The AUC is widely used in assessing the predictive power of Maxent models however due to the use of pseudo-absences rather than true absences some have criticised its suitability (Peterson et al., 2008). Maxent’s AUC can be artificially inflated through unsuitable methods such as inappropriate extent of the study area, spatially autocorrelated presence data or simply increasing the number of background cells (Yackulic et al, 2013). Due to these shortcomings a number of amendments have been suggested such as Area Under Kappa (AUK) curve (Kaymak et al., 2012) and partial AUC (Peterson et al., 2008). Jiménez-Valverde (2012) suggests the AUC measure has been used in uncritically in the field of habitat modelling. However they suggest that the AUC measure can be interpreted more reliably when considered alongside an assessment of the ROC curve. Ideally the blue and red lines in the sensitivity-specificity graph (right hand graph in figure 9) would reach 0,1.

Additionally Maxent provides an omission and predicted area graph (left hand graph in figure 9). The ‘omission and predicted area’ graph is an alternative method of assessing omission and commission errors and uses fractional predicted area of total study area predicted present as a substitute for traditional absence point commission errors (Reddy et al., 2015). The training and test (blue) lines should be close to predicted omission (black) line and the red line in an ideal model would reach 0.0,0.0.
Graphs to show measure of performance based on omission errors and pseudo commission errors for the full final present day model

Figure 9 shows the graphical result of the assessment of maxent's predictive performance which strengthens the interpretation of the models reliability compared to using the measure of AUC alone.
References


Norris, D. (2014). Model thresholds are more important than presence location type: Understanding the distribution of lowland tapir (Tapirus terrestris) in a continuous Atlantic forest of southeast Brazil. Tropical Conservation Science, 7(3), 529-547

Norris, D. (2014). Model thresholds are more important than presence location type: Understanding the distribution of lowland tapir (Tapirus terrestris) in a continuous Atlantic forest of southeast Brazil. Tropical Conservation Science, 7(3), 529-547


Appendix

Response curves for variables - all other variables held constant

Response curves for variables - modelled in isolation