British & Irish Botany 5(2): 180-195, 2023

Predicting the current and future distribution of pine woodland specialist plants in the Cairngorms National Park

Joshua Evans[1*], Andrew Carr[2]
[1]Global Wildlife Health and Conservation MSc candidate, University of Bristol
[2]Bristol Veterinary School, University of Bristol

*__Corresponding author__: Joshua Evans: joshe321@gmail.com

This pdf constitutes the Version of Record published on 12th June 2023

Abstract
The Caledonian pine forests of Scotland have declined significantly over the past century largely due to clearance for timber, fuel and grazing land. Approximately 1% of these native pinewoods remain from their historic extent, with a stronghold in the Cairngorms National Park. *Goodyera repens* (L.) R. Br. (Creeping Lady's-Tresses), *Moneses uniflora* (L.) A. Gray (One-Flowered Wintergreen) and *Linnaea borealis* L. (Twinflower) favour this habitat, meaning that the clearance of pine woodland has resulted in widespread declines. There are multiple plantation schemes underway in the Cairngorms to reforest the Highlands, such as Cairngorms Connect, partly with the intent of assisting the recovery of pine woodland specialists. This study uses Maximum Entropy modelling (MaxEnt) to create species distribution models (SDMs), firstly to predict the current distribution for these species; thus, enabling land managers in the Cairngorms to target surveys with the hope of finding new populations. Secondly, future distribution models are created using different emissions scenarios to predict the regions of maximum habitat suitability for the study species, with the aim of visualising where pine woodland should be planted to ensure long-term habitat viability. The results showed that there are regions in the north and east of the Cairngorms which have the highest habitat suitability, currently and in the future. The creation of habitat corridors between these two regions should be a priority to prevent isolation between populations of pine woodland specialists. The high emissions pathway could increase the likelihood of isolation by significantly decreasing the habitat suitability between the most suitable areas.

Key words: species distribution modelling; pine woodland; climate change; habitat restoration

Introduction
The Cairngorms National Park in the Scottish Highlands contains the largest extent of Caledonian forest in Britain. It consists largely of native pine woodland, mainly *Pinus sylvestris* (Scots Pine), with a scattering of *Alnus glutinosa* (Alder), *Betula* spp. (birches), *Sorbus aucuparia* (Rowan) and *Salix* spp. (willows) (The National Archives, 2011). This habitat has existed in Scotland since the first *P. sylvestris* trees arrived after the Late Glacial era *c.*7000 BC (Smout, 2014). The ancient Caledonian forests

once covered approximately 1.5 million hectares of Scotland, but the Forestry Commission estimated that this was reduced to 17,900 hectares in 1998. This is partly down to climate, with the wetter and windier conditions reducing the extent from *c.*5000 BC onwards (Steven & Carlisle, 1959). However, the forests have since come under increasing human-induced pressures. Beginning in the early Bronze Age, around 2000-1800 BC, land was extensively cleared for farmland. In present times, forestry has become one of the most prevailing threats (Tipping, 1995).

Caledonian pine forests host a unique ecosystem of relatively low diversity but a high number of rare species (The National Archives, 2011), due to multiple causal factors such as their isolation, fertile soils and wet microclimate (Featherstone, 1996). The forests are home to plant species that have become specialised to grow in pine woodland, including *Goodyera repens* (L.) R. Br. (Creeping Lady's-Tresses), *Moneses uniflora* (L.) A. Gray (One-Flowered Wintergreen) and *Linnaea borealis* L. (Twinflower), which are the focus of this study. These species have declined significantly (Online Atlas of British Flora and Fauna, 2023; Long and Scott, 2003; Wilcock, 2002), with *M. uniflora* and *L. borealis* being listed on the Highland Biodiversity Action Plan. Their reliance on native pine woodlands means that the clearance of this habitat will likely reduce their distribution and potentially cause localised extinctions.

Currently, 35 remnants of native pine woodland exist in Scotland (The National Archives, 2011). These are isolated and often of such a small size that they are at risk of being lost through grazing and forestry (Hobbs, 2009). Therefore, it is in the interest of conservation to increase the extent of pine woodland through large-scale plantation schemes to link up the remaining patches of forest. Cairngorms Connect is a collaboration of land managers, with the aim of enhancing and restoring habitats within the Cairngorms National Park over an area of 600 km^2 (Cairngorms Connect, 2023). It is a long-term vision, over a 200-year period, and the expansion of native pine woodland is one of the purposes of this project. However, this pine woodland must be within the climate envelope for the study species and our current knowledge of the best areas to target is lacking.

Maximum Entropy modelling (MaxEnt) is a machine-learning method which predicts species occurrence based on the environmental conditions in regions where there are presence records and a random sample of background points in the study area. It creates two probability densities, the first describes the contribution of each environmental variable in predicting presence. The second is a probability density for all the random background points, and the ratio between these two probability densities provides a relative habitat suitability for every point in the study area (Elith *et al.*, 2011). Therefore, MaxEnt creates a prediction for the current distribution under the environmental conditions in which a species exists, making it a proxy for the realised niche (Randin *et al.*, 2006). It can also make predictions on future distribution when future climate projections are included in the model input. MaxEnt is used in this study owing to its high predictive accuracy (Elith *et al.*, 2006) for small sample sizes, which is important when modelling rare and declining species (Pearson *et al.*, 2007).

In this study, the models predicting the current distribution for *G. repens*, *M. uniflora* and *L. borealis* will allow land managers to target future surveys in regions of high likelihood, potentially finding new populations of the study species. Using future climate projections, this study also endeavours to assess how climate change will impact pine woodland specialists and predict where regions of long-term habitat viability are likely to be found in the Cairngorms. These areas should be the focus for pine plantation schemes such as Cairngorms Connect.

Materials and methods

Modelling method

MaxEnt has previously been found to maintain a high predictive performance with species for which there are few known occurrences (Hernandez *et al.*, 2006). *G. repens*, *M. uniflora* and *L. borealis* have all been documented as having restricted distributions (Wilcock and Jennings, 1999**;** Online Atlas of the British and Irish Flora, 2023; Preston *et al.*, 2002), making MaxEnt a good choice of model. At the time of data collection, the study species were not known to have been the focus of planned surveys and the reliability of absence records on open-source databases can be low if survey efforts are insufficient (Lobo *et al.*, 2010). This was another justification for using MaxEnt which is a presence-only model.

Individual models for each study species were created for the current distribution, as well as combined models involving all three species. The *G. repens* model was ground-truthed during a visit to the Cairngorms in mid-August. This was the only model that was tested since it was the only plant which was flowering at the time. 1 km grid squares were randomly selected, with surveys being conducted based on the following categories: high likelihood, found; high likelihood, not found; low likelihood, found; low likelihood, not found. High likelihood was defined as a probability of presence >0.7, whereas low likelihood was <0.3. In total, 12 squares were surveyed, three for each category and 100 m squares were surveyed using the scan sampling technique. To ensure the independence of occurrences, which MaxEnt assumes (Phillips *et al.*, 2006), survey locations were at least 1 km apart.

Future prediction models are created for *G. repens*, *M. uniflora* and *L. borealis* for different emissions scenarios (SSP126, SSP245 and SSP585) (Carbon Brief, 2019) and over a range of time periods (2021-2040, 2041-2060 and 2061-2080). Only combined models were produced for the future predictions, this allows the planting of pine woodland to be directed to areas where all three species have a high likelihood of being found; thus, maximising the value of a restoration exercise. These future models had a spatial resolution of ~5 km^2 because this was the highest resolution available from WorldClim. The data were cleaned and processed in QGIS v3.16.8 (https://qgis.org/en/site/forusers/download.html) and models were created using MaxEnt v3.4.4 (https://biodiversityinformatics.amnh.org/open_source/maxent).

Jackknife analysis was used to determine which variables are the most important in terms of predicting species presence. First, each variable is used in isolation to show which ones provide the most useful information by themselves. Then, each variable is omitted in turn and the remaining variables are used to build models, showing which have the most information that is not present in other variables. A high Jackknife score suggests that a variable is highly correlated with at least one other variable. MaxEnt also provides a percent contribution for each variable, allowing the modeller to exclude variables which contain little information. Response curves are also created in MaxEnt which show how the likelihood of species presence varies against an increasing value for each variable. Two different sets of response curves are produced, the first of which involves changing each environmental variable whilst keeping all others at their average value. A level response curve at 0.5 suitability indicates that the variable is highly correlated with at least one other variable. For the second set, a separate model is produced in turn using each environmental variable which allows the degree of overfitting to be judged. In the case of a jagged curve, the regularisation parameter (a feature of MaxEnt which attempts to avoid overfitting) may need to be adjusted to create smooth response curves.

Model building
Species occurrence records were obtained from the Global Biodiversity Information Facility (GBIF) website (Global Biodiversity Information Facility, 2021). This covered the period from 1970-present, selected on the basis that it is concurrent with the historic climate data that was measured from 1970-2000. It was judged that the climate will not have changed significantly enough from the period 2000-2020 to influence the distribution of the study species. Climate data were downloaded from WorldClim (WorldClim, 2021) which included 19 different environmental variables. The 30-second spatial resolution (~1 km^2) dataset was clipped to the Cairngorms boundary. A correlation test involving the 'vifstep' function was performed using the R statistical computing package. A threshold VIF value of 10 was used to exclude highly correlated variables, as a value greater than 10 has been found to be problematic in multiple papers (Naimi *et al.*, 2014; Mela & Kopalle, 2002). Pine woodland distribution (Scottish Forestry, 2021), soil pH/type (James Hutton Institute, 2021) and elevation (WorldClim, 2021) were included in the correlation test, and the output found 11 out of the 23 variables to not have collinearity issues. A description of each variable used in the models is shown in Table 1.

Table 1. Description of each predictor variable used in the MaxEnt species distribution models for *G. repens, L. borealis and M. uniflora* in the Cairngorms

Variable	Type	Description	Source
Pine woodland	Categorical	Binary dataset of native pine woodland in the Cairngorms National Park.	https://opendatascottishforestry.hub.arcgis.com/
Soil pH	Continuous	Soil pH coverage for the Cairngorms, taken from a map of scale 1:250,000 for the entirety of Scotland.	https://www.hutton.ac.uk/learning/naturalresourcedatasets/soilshutton/soilsmaps-scotland/download
Soil type	Categorical	10 different soil types for the Cairngorms from the National Soils Inventory for Scotland)	https://www.hutton.ac.uk/learning/naturalresourcedatasets/soilshutton/soilsmaps-scotland/download

Isothermality	Continuous	How large the day-to-night temperatures oscillate relative to the summer-to-winter (annual) oscillations. Unit: %	https://pubs.usgs.gov/ds/691/ds691.pdf
Maximum temperature of the warmest month	Continuous	The maximum monthly temperature occurrence over a given year (time-series) or averaged span of years (normal). Unit: °C	https://pubs.usgs.gov/ds/691/ds691.pdf
Minimum temperature of the coldest month	Continuous	The minimum monthly temperature occurrence over a given year (time-series) or averaged span of years (normal). Unit: °C	https://pubs.usgs.gov/ds/691/ds691.pdf
Mean temperature of the wettest quarter	Continuous	Approximates mean temperatures that prevail during the wettest season. Unit: °C	https://pubs.usgs.gov/ds/691/ds691.pdf
Mean temperature of the driest quarter	Continuous	Approximates mean temperatures that prevail during the driest quarter. Unit: °C	https://pubs.usgs.gov/ds/691/ds691.pdf
Precipitation driest month	Continuous	The total precipitation during the driest month. Unit: mm	https://pubs.usgs.gov/ds/691/ds691.pdf
Precipitation seasonality	Continuous	A measure of the variation in monthly precipitation totals over the course of the year. Unit: %	https://pubs.usgs.gov/ds/691/ds691.pdf

Precipitation of the warmest quarter	Continuous	Approximates total precipitation that prevails during the warmest quarter. Unit: mm	https://pubs.usgs.gov/ds/691/ds691.pdf

Model evaluation

The variables with the most useful information were selected using the percent contribution for each variable. Variables which contributed less than 0.5% to the model were excluded. Jackknife analysis was used to remove variables which contributed little useful information to the models. Furthermore, Jackknife analysis was used alongside an inspection of the response curves to remove highly correlated variables. The variable with the most biological significance to the study species was then selected. With the remaining variables, the models were assessed on their ability to accurately predict presence against the random background points using the Area Under the Curve (AUC) statistic from the Receiver Operating Characteristic (ROC) curve. This is a plot of the sensitivity (the proportion of presence predicted correctly) against 1-specificity (the inverse of proportion of absence predicted correctly). Therefore, the AUC statistic is a measure of how accurately a model discriminates observed presence from absence (Jiménez-Valverde, 2012). An AUC score of 0.5 means that a model performs no better than random, whereas a score of 1 signifies that a model perfectly discriminates presence from absence. A model with a value of 0.7 is judged as having a 'good' level of discrimination and 0.9 suggests a model discriminates excellently. Therefore, the aim is for models to have an AUC statistic of at least 0.7, ideally closer to or greater than 0.9.

Results

Models of current distribution

The individual models of current distribution for *G. repens*, *M. uniflora* and *L. borealis* (Fig. 1) have high AUC scores (0.894, 0.975 and 0.877, respectively) which mean that the models can reliably distinguish presence from absence. For all individual models, the Jackknife analysis shows that the variable which has the highest gain when used in isolation is the precipitation of the driest month. This variable also has the most information that is not present in other variables, i.e. the most uncorrelated. Precipitation of the driest month is also the variable with the highest percent contribution to all three models. Therefore, it is an important factor in determining the presence for pine woodland specialists. For all individual models, probability of presence is close to 1 when the precipitation of the driest month is approximately 48 mm, before declining rapidly towards 0 when in excess 100 mm.

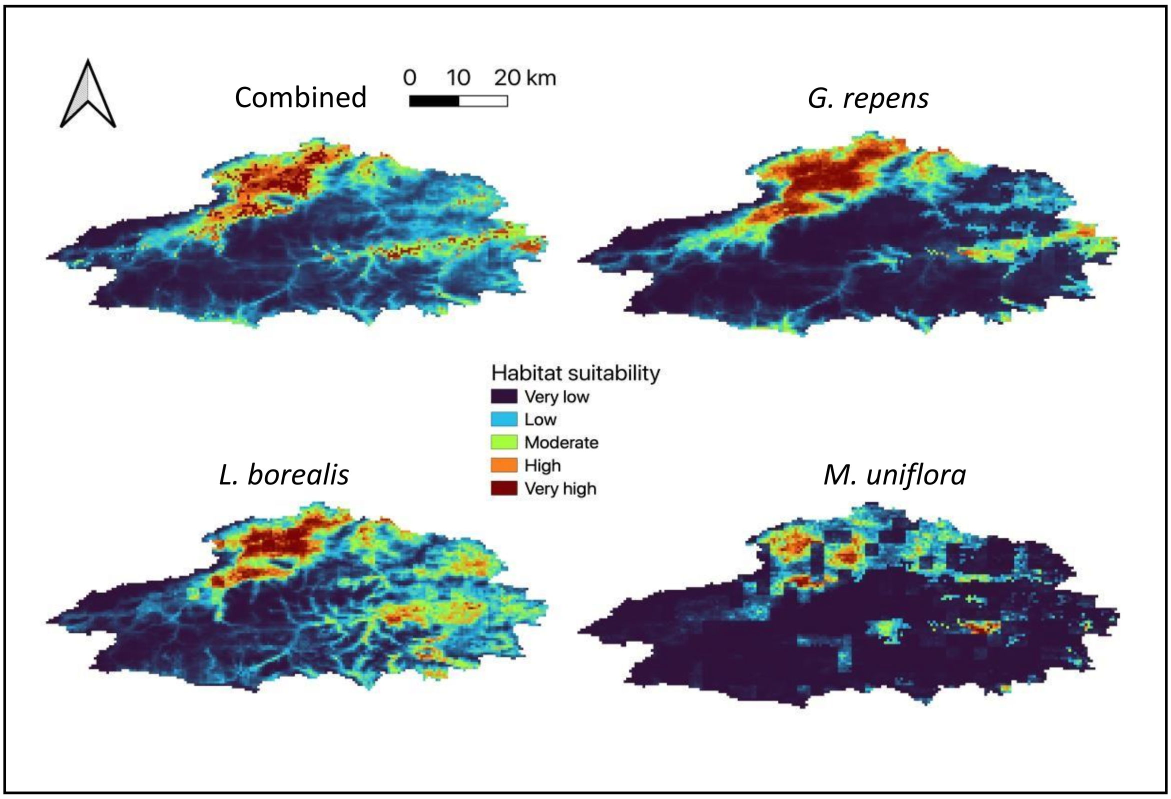


Figure 1. Current predicted distribution for all three species, i.e. combined (AUC=0.855, standard deviation=0.029); *Goodyera repens* (AUC=0.894, standard deviation=0.035); *Linnaea borealis* (AUC=0.877, standard deviation=0.011) and *Moneses uniflora* (AUC=0.975, standard deviation=0.007).

The results of the ground-truthing for the *G. repens* model are shown in Table 2 (see Appendix 1.0 for the precise locations), with sensitivity (the proportion of presences predicted correctly) and specificity (the proportion of absences predicted correctly) having values of 0.625 and 0.75, respectively. For a random model which predicts presence and absence correctly 50% of the time, these values would be 0.5. Therefore, the *G. repens* model performed better than random, and this increases the reliability of using the model to direct additional surveys.

Table 2. Confusion matrix showing the number of squares for which *G. repens* was recorded present or absent and the model prediction. Sensitivity is calculated by the following equation: true positive/(true positive + false negative). Specificity is the proportion of observed absences predicted correctly and is calculated by the following equation: true negative/(true negative + false positive)

	Recorded present	**Recorded absent**
Predicted present	5 (true positive)	1 (false positive)
Predicted absent	3 (false negative)	3 (true negative)
	Sensitivity = 0.625	Specificity = 0.75

Pine woodland, which was included because of its importance for species presence according to the literature, was omitted from the individual models due to low test gain and low percent contribution. However, it was included in the combined model and had a percent contribution of 40.1%. Additionally, it was the variable that decreases gain most when omitted and therefore appears to have the most information that is not present in other variables (Appendix 2.0). The combined model again shows precipitation of the driest month to have the most useful information by itself, according to the Jackknife analysis, as well as having the highest percent contribution (44.9%).

The combined and individual models for the current distribution show the north of the Cairngorms to have the highest habitat suitability, whereas the southwest has a very low habitat suitability. Interestingly, the *L. borealis* model and the combined model appear to have linear regions of higher suitability (signified by the light blue branched regions) which closely follow rivers (Fig. 1).

Models of future distribution

Models to predict future distributions were produced using eight climate variables from the current models, i.e. they excluded the pine woodland, soil pH and soil type variables. Mean temperature of the driest quarter and isothermality were removed from the models due to their low percent contribution. Jackknife analysis also shows that training gain increases when these variables are not included which suggests they are highly correlated with at least one other variable. The regularisation parameter was set to 3 from the default of 1. This is because the response curves are jagged, which is symptomatic of an overfit model.

The combined models for a low and high emissions scenario (SSP126 and SSP585, respectively) have a good AUC score of 0.736 for the short, medium and long term. The predicted habitat suitability for a low emissions scenario changes very little across all time periods, with two regions of high likelihood occurring to the north and east of the Cairngorms (Fig. 2). However, over time, the high emission models show that the habitat suitability will increase in the north of the Cairngorms and across a small region to the east (Fig. 3). In concurrence with the individual models, the variable with the highest percent contribution is precipitation for the driest month as well as having the highest training gain when used in isolation. Therefore, it contains the most useful information by itself. The response curves also show the highest probability of presence for the study species to occur when the precipitation for the driest month is approximately 42-48 mm.

Discussion

The current distribution models show that surveys should be targeted to the north for *G. repens*, *M. uniflora* and *L. borealis*, since this is where the highest level of predicted habitat suitability exists. There is also a smaller region of high habitat suitability in the east of the Cairngorms, and these two areas closely match up with the current extent of native pine woodland in the national park. The variable of pine woodland distribution is not included in the individual models, unlike for the combined model, which suggests that most of the information for pine woodland is captured in the remaining variables. These findings confirm the previous research and surveys of the study species, which are known to be largely confined to native pine woodland (Downie, 1943; Long & Scott, 2003). The ground-truthing surveys for

G. repens also suggests the importance of this habitat, with almost every plant found growing at the base of *P. sylvestris*. Another observation made during these surveys is that *G. repens* appears to be highly dependent on older pine stands. In the single case when the plant was not found where it was predicted to be present (false positive) it was in a young pine plantation. This corroborates previous findings, with *G. repens* generally appearing 60-80 years after a pine wood is planted (NBN Atlas, 2021). Similarly, *L. borealis* and *M. uniflora* are largely confined to old growth forest (Worrell & Dunlop, 2003; Wright & Lusby, 1999), and it has been found that *L. borealis* is three times more frequent in ancient semi-natural pine woods compared to new pine plantations. This underlines the importance of maintaining the remnant ancient Caledonian pine woodland but it means that when planting new pine woods, long-term habitat viability is key for the future of pine woodland specialist plants.

Pine woodland is omitted from the individual current models due to the low percent contribution and low test gain when it is included. However, in the combined model, the percent contribution increases to 40.1%, and it also contains the most useful information by itself. This does not necessarily mean that pine woodland is an unimportant variable, in terms of probability of presence for the study species individually, but it is relatively less important than other variables. When all three species are assessed together, pine woodland becomes very important which reinforces the argument for planting native pine woodland to increase the distribution for *G. repens*, *M. uniflora* and *L. borealis*.

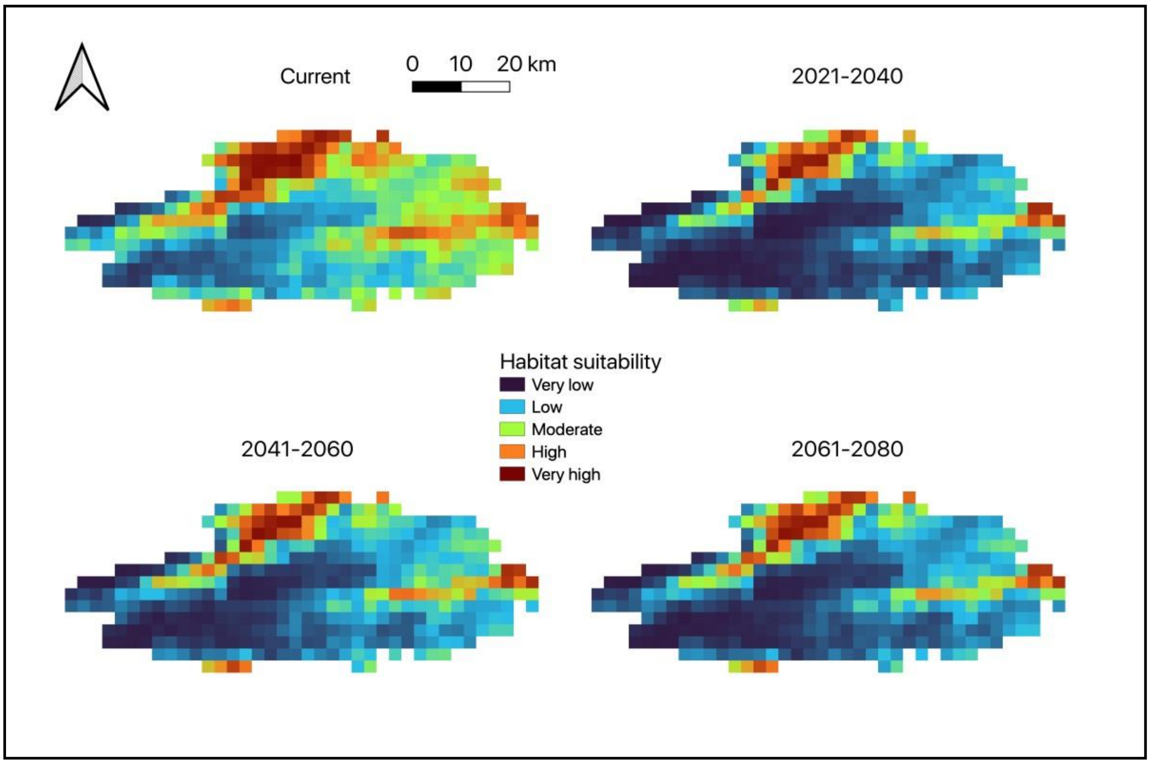


Figure 2. Current and future predictions for the combined distribution of *Goodyera repens, Moneses uniflora and Linnaea borealis* under a low emissions scenario (SSP126). AUC = 0.736, Standard deviation = 0.054.

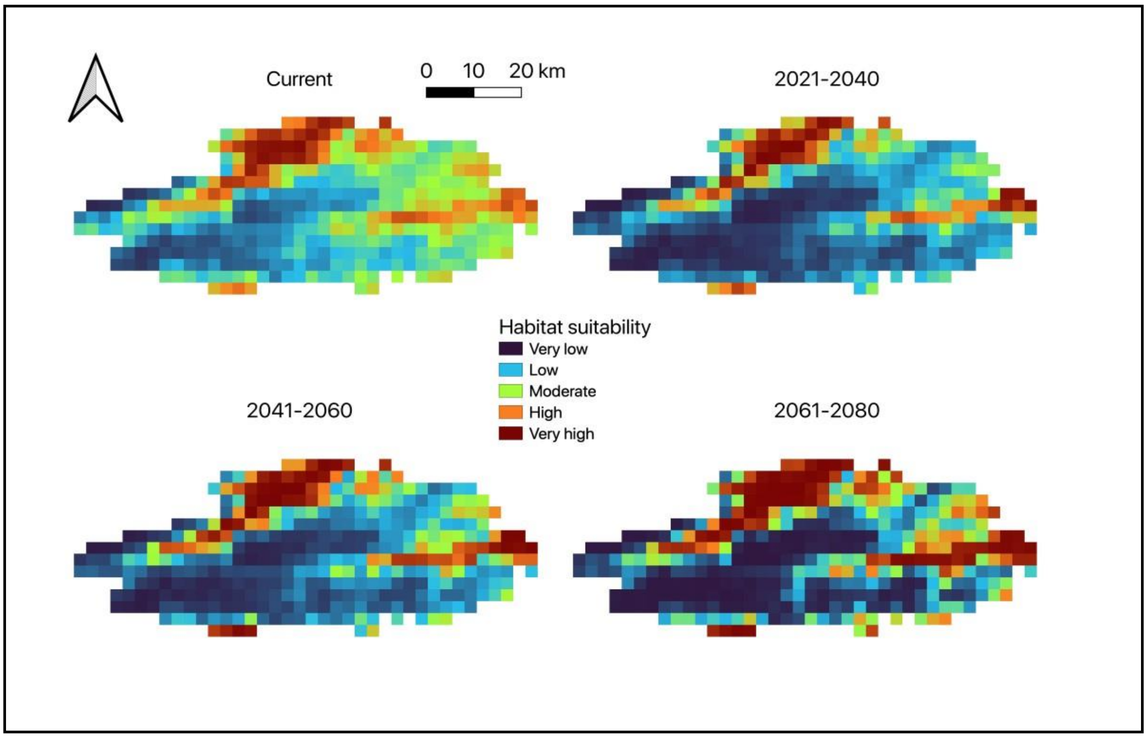


Figure 3. Current and future predictions for the combined distribution of *Goodyera repens, Moneses uniflora and Linnaea borealis* under a high emissions scenario (SSP585). AUC = 0.736, Standard deviation = 0.054.

Outside of the northern and eastern regions of high suitability for *G. repens* and *M. uniflora*, the habitat suitability is very low. However, there are linear regions of higher suitability for *L. borealis* and the combined model which follow a branching pattern, and they closely follow the rivers in the Cairngorms. The combined model could show a similar predicted distribution to the *L. borealis* model because they use six of the same variables (compared with five variables in common with the *G. repens* and *M. uniflora* models). Therefore, the predicted habitat suitability could be slightly more influenced by the *L. borealis* distribution than for the other study species, although it is not expected that this would be significant enough to negate the reliability of the models. The presence of *L. borealis* growing alongside rivers is not well documented in the literature, and thus neither are the mechanisms for this. However, it is suspected that soil moisture could be part of the reason, with *L. borealis* preferring wet soils (Kohn & Lusby, 2004). These fluvial habitats could form useful corridors with the aim of conjoining the existing populations.

The optimal precipitation level for the driest month was found to be relatively low, between 42 mm and 48 mm, for all current and future models. This does explain the finding that the habitat suitability increases over time in regions of high likelihood for the high emissions scenario, when dry conditions will likely become normal in spring. This variable also had the most useful information by itself which suggests it is of high importance for the growth of pine woodland specialists. This seems at odds to previous findings which suggest that moist soils are necessary for all three species (Kohn & Lusby, 2004; Tsiftsis *et al.*, 2012; Dorji & Wangyal, 2021). The driest month for the Cairngorms is April according to multiple sources (Climate Data, 2021; Statista, 2021), and the average rainfall in April is 65.4 mm (World Weather Online), which is significantly more than the optimal precipitation according to the models. April is at least a few months before the late summer flowering times for these species, when it is anticipated that dry conditions could be most harmful to the plants. Dry conditions tend to occur when temperatures are higher, which may be expected to negatively impact plants which grow at high elevations. However, high temperatures after snowmelt, when these species are germinating, may be beneficial (Billings & Mooney, 1968). This could be an adaptation to prevent germination in frosty conditions, when the success for seedlings will be much lower (Cavieres & Arroyo, 2000). Therefore, a warming climate could increase seedling survival initially. Although, if these high temperatures extend into late summer during flowering time, then seed production may be negatively impacted (Briceño *et al.*, 2015).

Goodyera repens and *L. borealis* are both self-incompatible (Liu *et al.*, 2020; Neiland & Wilcock, 1995) and need to be cross-pollinated by insect pollinators (Scobie & Wilcock, 2009), whereas *M. uniflora* can self-pollinate (Kevan *et al.*, 1993). However, cross-pollination is needed to maintain the genetic diversity of a population. Their decline due to the loss of pine woodland has caused populations to become highly fragmented, and in some cases, they may be too distant to allow cross-pollination. Over time, this could cause genetic barriers to dispersal because of inbreeding depression and limit their ability to adapt to a changing climate. *L. borealis* must differ by at least one allele at the S-locus (De Nettancourt, 1977) for successful fertilization to occur. The reduction in S-alleles leads to less mate compatibility amongst the isolated populations, such that reproductive failure is inevitable. This further reinforces the importance of linking up the remnant pine woodlands in the Cairngorms to facilitate the recovery of *G. repens*, *M. uniflora* and *L. borealis*.

The predicted habitat suitability does not differ significantly from the current distribution of pine woodland (Figure 1). However, these woodlands are sporadic at present which suggests that the optimal strategy is to target the linking up of the pine stands in a joined-up approach, rather than focusing on extensive plantation schemes away from the current pine woods. For long-term population viability, joining up the high likelihood regions in the north and east of the Cairngorms National Park may be necessary to maintain enough genetic diversity to allow for adaptation to a changing climate. There is a region of moderate habitat suitability stretching along the northeast edge of the park boundary which could form a corridor between the two areas of highest suitability.

For these habitat corridors to become a long-term option, there needs to be a band of moderate to high suitability between the north and east of the Cairngorms. Although the predicted habitat suitability increased further in the high likelihood regions to the north and east for the high emissions scenario (SSP585), it decreased to very low likelihood in other areas of the Cairngorms. Therefore, the study species may proliferate in two principal regions, but most of the national park will become unsuitable and the populations could become isolated. This will limit the genetic flow between the populations and reduce their ability to adapt in the face of climate change. In comparison, the habitat suitability changes relatively less from 2020-2080 in the low emissions scenario (SSP126), maintaining regions of higher suitability between the north and east. This is particularly evident in the southwest, which will become almost entirely unsuitable for pine woodland specialists by 2080 in the high emissions scenario. Similarly, for the business-as-usual emissions pathway (Appendix 3.0), the habitat suitability declines significantly in the southwest up to 2080.

Therefore, this study suggests that for maximum long-term habitat viability, the lowest emissions pathway is optimal.

Abies spp. (Fir) and *Picea* spp. (Spruce) trees are commonly included in new plantations; however, in Scotland, *G. repens* is often associated with *Pinus sylvestris* (Tsiftsis *et al.*, 2012). Therefore, the propensity for *G. repens* to grow alongside conifers, other than Scots Pine, needs further investigation. During the ground-truthing surveys for *G. repens*, it was observed that this species tended to be found at the base of older Scots Pine trees. In cases where the species was found when predicted absent (a frequency of 50%), this happened in new pine plantations. This suggests that woodland management is important for *G. repens* and probably for the other study species. Therefore, species composition and age of the woodland should be included in the models.

Acknowledgements
The data for this study were provided by Global Biodiversity Information Facility (species records), WorldClim (climate data), Scottish Environment Protection Agency, SEPA (pine woodland cover) and the James Hutton Institute (soil data).

References

Billings, W.D. & Mooney, H.A. 1968. The ecology of arctic and alpine plants. *Biological reviews* 43: 481-529.

Briceño, V.F., Hoyle, G.L. & Nicotra, A.B. 2015. Seeds at risk: how will a changing alpine climate affect regeneration from seeds in alpine areas?. *Alpine Botany* 125: 59-68.

Cairngorms Connect, 2023. *Cairngorms Connect Restoration Projects*. Available at: https://cairngormsconnect.org.uk/projects/restoration-projects

Carbon Brief, 2019. *CMIP6: the next generation of climate models explained*. Available at: https://www.carbonbrief.org/cmip6-the-next-generation-of-climate-models-explained/

Cavieres, L.A. & Arroyo, M.T. 2000. Seed germination response to cold stratification period and thermal regime in *Phacelia secunda* (Hydrophyllaceae)–altitudinal variation in the Mediterranean Andes of central Chile. *Plant Ecology* 149: 1-8.

Climate Data, 2021. *Climate Aviemore*. Available at: https://en.climatedata.org/europe/united-kingdom/scotland/aviemore-7406

De Nettancourt, D. 1977. *Incompatibility in angiosperms* (pp. 28-57). Berlin: Springer.

Dorji, G. & Wangyal, J.T. 2021. *Moneses uniflo*ra (Ericaceae): A new record to Bhutan. *Bhutan Journal of Natural Resources and Development* 8: 5.

Downie, D.G. 1943. Source of the symbiont of *Goodyera repens*. *Transactions of the Botanical Society of Edinburgh* 33: 383-390.

Elith, J., Phillips S.J., Hastie T., Dudík M., Chee Y.E. & Yates C.J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43-57.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J. *et.al.* 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.

Featherstone, A.W. 1996. Regenerating the Caledonian Forest. *Journal of Wildnerness* 2(3),p.36.

Global Biodiversity Information Facility (2021). *Occurrence Search*. Available at: https://www.gbif.org/occurrence/search
Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29(5): 773-785.
Highland Council (2021). *Highland Nature Biodiversity Action Plan*. Available at: https://www.highland.gov.uk/download/meetings/id/78100/item_11_highland_biodiversity_a ction_plan_-_council_commitments
Hobbs, R. 2009. Woodland restoration in Scotland: ecology, history, culture, economics, politics and change. *Journal of Environmental Management* 90(9): 2857-2865.
James Hutton Institute 2021. *Soil Maps of Scotland*. Available at: https://www.hutton.ac.uk/learning/natural-resource-datasets/soilshutton/soils-mapsscotland/download
Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography* 21(4): 98-507.
Kevan, P.G., Tikhmenev, E.A. & Usui, M. 1993. Insects and plants in the pollination ecology of the boreal zone. *Ecological Research* 8(3): 247-267.
Kohn, D. & Lusby, P. 2004. Translocation of twinflower (Linnaea borealis L.) in the Scottish Borders. *Botanical Journal of Scotland* 56(1): 25-37.
Liu, H., Xia, M., Xiao, Q., Fang, J., Wang, A., Chen, S. & Zhang, D. 2020. Characterization of the complete chloroplast genome of *Linnaea borealis,* a rare, clonal selfincompatible plant. *Mitochondrial DNA Part B* 5(1): 200-201.
Lobo, J.M., Jiménez-Valverde, A. & Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33(1): 103-114.
Long, D. & Scott, M. 2003. Action for biodiversity priority species in Scotland. *Botanical Journal of Scotland* 55(1): 65-76.
Mela, C.F. & Kopalle, P.K. 2002. The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlations. *Applied Economics* 34(6): 667.
Naimi, B., Hamm, N.A., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. 2014. Where is positional uncertainty a problem for species distribution modelling?. *Ecography* 37(2): 191-203.
NatureScot, 2020. *Woodland Plants*. Available at: http://nature.scot/plants-animals-andfungi/flowering-plants/woodland-plants
NBN Atlas, 2021. *Creeping Lady's Tresses*. Available at: https://species.nbnatlas.org/species/NHMSYS0000459112
Neiland, M.R.M. & Wilcock, C.C. 1995. Maximisation of reproductive success by European Orchidaceae under conditions of infrequent pollination. *Protoplasma* 187(1-4): 39-48.
Online Atlas of the British and Irish Flora. 2023. Available at: https://plantatlas2020.org/atlas
Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Townsend Peterson, A. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of biogeography* 34(1): 102-117.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological modelling 190*(3-4): 231-259.
Preston, C. D., Pearman, D. A. & Dines, T. D. (eds.) (2002) *New Atlas of the British and Irish Flora*. Oxford: Oxford University Press.
Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. 2006. Are niche-based species distribution models transferable in space?. *Journal of biogeography* 33(10): 1689-1703.
Scobie, A.R. & Wilcock, C.C., 2009. Limited mate availability decreases reproductive success of fragmented populations of *Linnaea borealis*, a rare, clonal self-incompatible plant. *Annals of Botany* 103(6): 835-846.
Scottish Forestry, 2021. Scottish Forestry Open Data. Available at: https://open-datascottishforestry.hub.arcgis.com/
Smout, C. 2014. The history and the myth of Scots pine. *Scottish Forestry* 68(1): 9-12.
Statista, 2021. *Monthly amount of rainfall in Scotland from 2015 to 2021*. Available at: https://www.statista.com/statistics/610092/monthly-rainfall-in-scotland
Steven, H.M. & Carlisle, A., 1959. The native pinewoods of Scotland. *The native pinewoods of Scotland.*
The National Archives, 2011, *Habitat Action Plan*. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20110303150022/http://www.ukbap.org.u k/UKPlans.aspx?ID=6
Tipping, R., 1995, November. The form and the fate of Scotland's woodlands. In *Proceedings of the Society of Antiquaries of Scotland* 124: 1-54.
Tsiftsis, S., Tsiripidis, I. & Papaioannou, A. 2012. Ecology of the orchid *Goodyera repens* in its southern distribution limits. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology* 146(4): 857-866.
Wilcock, C.C. & Jennings, S.B. 1999. Partner limitation and restoration of sexual reproduction in the clonal dwarf shrub *Linnaea borealis* L.(Caprifoliaceae). *Protoplasma* 208(1): 76-86.
Wilcock, C.C., 2002. Maintenance and recovery of rare clonal plants: the case of the twinflower *(Linnaea borealis* L.). *Botanical Journal of Scotland* 54(1): 121-131.
WorldClim, 2021. *Global Climate and Weather Data*. Available at: https://worldclim.org/data
World Weather Online, 2021. *Cairngorm Weather Averages*. Available at: https://www.worldweatheronline.com/ski/cairngorm-weather-averages/gb.aspx
Worrell, R. & Dunlop, B. 2003. The influence of past management of pinewoods on the occurrence of twinflower. *Unpublished PLANTLIFE Report*.
Wright, J.A. & Lusby, P.S. 1999. The past and present status of *Moneses uniflora* (L.) Gray (Pyrolaceae) in Scotland. *Watsonia* 22(4): 343-352.

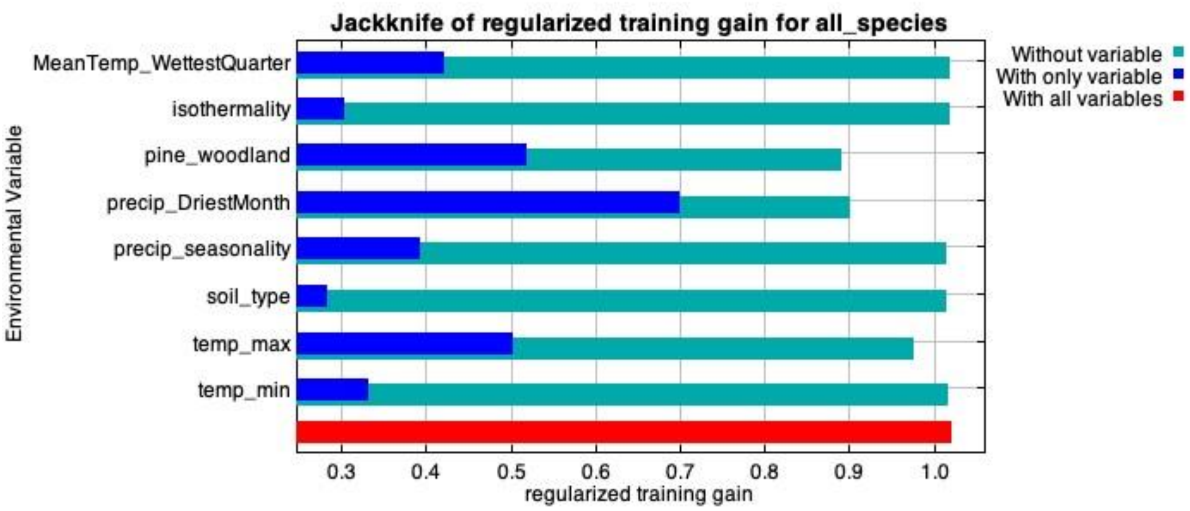
ISSN: 2632-4970

https://doi.org/10.33928/bib.2023.05.180

Appendix 1. Grid references and geographic coordinates for ground-truthed locations in the Cairngorms National Park

	High likelihood, found	High likelihood, not found	Low likelihood, found	Low likelihood, not found
Grid reference	NH 88942 06955; NH 97324 09251; NJ 01783 14897	NH 90470 07678; NH 96196 10222; NJ 03173 16536	NH 97043 07614; NH 95118 11499; NJ 02796 12165	NH 91555 06843; NH 99061 07521; NJ 00784 14019
Geographic coordinates	N57.14002, W-3.83678; N57.16257, W-3.70078; N57.21426, W3.62783	N57.14688, W-3.81187; N57.17105, W-3.71834; N57.22927, W3.60547	N57.14783, W-3.70327; N57.18227, W-3.73670; N57.18995, W3.61000	N57.13964, W-3.79359; N57.14744, W-3.66989; N57.20616, W3.64402

Appendix 2. Jackknife analysis of regularised training gain for the combined, current model. Pine woodland has the shortest light blue bar, which suggests it has the most useful information not present in other variables. Precipitation of the driest month has the longest dark blue bar, meaning it has the most useful information by itself.



Appendix 3. Current and future predictions for the distribution of *Goodyera repens, Moneses uniflora and Linnaea borealis* under a business-as-usual emissions scenario (SSP245). AUC = 0.736, Standard deviation = 0.054.