Modelling the distribution of Pearl-bordered Fritillary, *Boloria euphrosyne* and Chequered Skipper, *Carterocephalus palaemon*

Stuart Ball, JNCC (stuart.ball@jncc.gov.uk)

# Introduction

|  |
| --- |
| PBFrit_BC_presences.png |
| Fig. : Supplied distribution of Pearl-bordered Fritillary in Scotland (Symbols are 5-km in order to show up). |
| CS_BC_Presences.png |
| Fig. : Supplied distribution of Pearl-bordered Fritillary in Scotland (Symbols are 5-km in order to show up). |

A species distribution model for the Pearl-bordered Fritillary in Scotland was requested by Dr Tom Brereton of Butterfly Conservation. He supplied a list of occupied squares (recent records, excluding singletons) at 1-km square resolution (264 squares, ).

He describes the habitat of the Pearl-bordered Fritillary as follows:

“*In Scotland Pearl-bordered Fritillary is chiefly found in mid-Scotland (Perthshire, Strathspey) and in the west (Argyll and the Highlands) on south-facing slopes of valley systems, in open woodland/woodland edge (Birch, Oak, Hazel) habitats or wayleaves/pylon lines through woodland (including conifer). Associated plants include Bracken and others of base-rich/moderately acidic soils e.g. Pignut, Common Dog Violet, Primrose, Wood Sage, Bugle, Wood Anemone, Sweet Vernal Grass and* Luzula multiflora*. The field layer is usually herb-rich, but almost always contains Bracken.*

*The Small Pearl-bordered Fritillary prefers wetter ground and breeds on Marsh Violets (Bracken not needed), though the two co-occur on many sites (Pearl-bordered on knolls and other drier micro-habitats)*

*In summary spatial datasets which would be good to include in the modelling (if available)*

1. *Altitude – generally there is an upper limit to distribution*
2. *Presence of south-facing slopes*
3. *Woodland cover*
4. *Soils*
5. *Plants listed above*
6. *Larger wayleaves (of the sort illustrated on 1:50 000 OS Maps) – Scottish Power may have this?*
7. *Climate data*

*A smaller number of colonies occur in conifer clearfell e.g. in the SW, and along coastal undercliff e.g. Moray Firth.*”

And the habitat of Chequered Skipper as:

“*In Scotland, it is now associated with areas of lightly grazed or ungrazed grassland, scrub and marsh around open broad-leaved woodland, usually oak or birch, on steep or gentle slopes, on wet but well-aerated soils (Ravenscroft 1994b). The general landscape is species-poor and dominated by the foodplant, Purple Moor-grass* Molinia caerulea*, Heather* Calluna vulgaris *and other plants characteristic of waterlogged soils and peats such as Cotton Grasses* Eriophorum *spp.. The larvae of the Chequered Skipper occur on patches of well-aerated soils (i.e. that allows water movement through the soil), often close to the richer soils of woodland, where the foodplant grows in a luxuriant form, typically in association with tall Bog Myrtle Myrica gale and Birch scrub* Betula pubescens *(Ravenscroft 1994b). These patches are only sometimes close to nectar sources along the shelter of the woodland edge or on patches of more base-rich soils.*

*The butterfly also occurs along wayleaves and in grassy/boggy areas within and on the edges of conifer plantations.I”*

# Environmental layers

An initial, Maxent model was developed using the following environmental layers:

|  |  |
| --- | --- |
| **ITE Land Cover Map 2007** | |
| Number of hectares of coverage in each 1km square for 26 cover classes. | |
| 01\_bwood | Coverage of broad-leaved plantations and woodland |
| 02\_conif | Coverage of coniferous plantations and woodland |
| 03\_arable | Coverage of arable farming |
| 04\_impgrss | Coverage of improved grassland |
| 05\_rghgrss | Coverage of rough grassland |
| 06\_neutgrss | Coverage of neutral grassland |
| 07\_calcgrss | Coverage of calcareous grassland |
| 08\_acidgrss | Coverage of acid grassland |
| 09\_fenmsh | Coverage of fen and marsh |
| 10\_heather | Coverage of ericaceous heath |
| 11\_heathgrss | Coverage of grass heath |
| 12\_bog | Coverage of bog |
| 13\_montane | Coverage of montane habitats |
| 14\_inlndrock | Coverage of inland rock and bare ground |
| 15\_saltwater | Coverage of saltwater |
| 16\_freshwater | Coverage of moving and still freshwater |
| 17\_sublitrock | Coverage of sublitoral rock |
| 18\_supralitsed | Coverage of supra-litoral sediments (included dunes) |
| 19\_litrock | Coverage of litoral rock |
| 20\_litsed | Coverage of litoral sediments |
| 21\_saltmsh | Coverage of saltmarsh |
| 22\_urban | Coverage of continuous urban areas |
| 23\_suburban | Coverage of sub-urban areas |

|  |  |
| --- | --- |
| **OS Topography data** | |
| AltMean | Mean altitude derived from DEM |

|  |  |
| --- | --- |
| **European Soils Database** | |
| Derived from ESDB v2 - 1km Raster Library - "dominant value" and "dominant STU" Rasters (STU = Soil Typological Units). The original grids are in the ETRS89 Lambert Azimuthal Equal Area (ETRS\_LAEA) co-ordinate system at 1-km resolution. Re-projected to OSGB and nearest-neighbour method used to assign categories. Of the very large number of potential variables (103), the following were found to be most useful for a sample of 15 species each of Hoverflies, Moths and Vascular plants. | |
| cat\_fao85fu | Full soil code of the STU from the 1974 (modified CEC 1985) FAO-UNESCO Soil Legend |
| cat\_hg | Hydro-geological class. |
| cat\_parmado | Major group code for the dominant parent material of the STU |
| cat\_parmase | Major group code for the secondary parent material of the STU. |

|  |  |
| --- | --- |
| **NASA Digital Elevation Model processed by International Institute for Applied Systems Analysis** | |
| The global terrain slope and aspect database has been compiled using elevation data from the Shuttle Radar Topography Mission (SRTM) by IISA, Vienna. The SRTM data is publicly available as 3 arc second (WGS84), i.e. approximately 90 meters resolution at the equator, Digital Elevation Models (DEM) (CGIAR-CSI, 2006). Because of the convergence of lines of longitude towards the pole, the x-resolution of the DEM is approx. 50m in Britain.  The high resolution SRTM data have been used for calculating:   * Terrain slope gradients and classes for each 3 arc-sec grid cell; * Aspect of terrain slopes for each 3 arc-sec grid cell;   The number of grid-cells (pixels) falling in each category was then counted for 30 arc second grid cells (approximately 1 km resolution at the equator).  These were re-projected to OSGB and counts assigned to 1km x 1km squares using bilinear interpolation. | |
| gb\_GloAspectClE\_30as | Number of pixels with slope >2% and facing East (45°–135°) |
| gb\_GloAspectClN\_30as | Number of pixels with slope >2% and facing North (315°–45°) |
| gb\_GloAspectClS\_30as | Number of pixels with slope >2% and facing South (135°–225°) |
| gb\_GloAspectClU\_30as | Number of pixels where aspect is either unknown or slope ≤ 2% |
| gb\_GloAspectClW\_30as | Number of pixels with slope >2% and facing West (225°–315°) |
| gb\_GloSlopesCl1\_30as | Number of pixels with slope 0–0.5% |
| gb\_GloSlopesCl2\_30as | Number of pixels with slope 0.5–2% |
| gb\_GloSlopesCl3\_30as | Number of pixels with slope 2–5% |
| gb\_GloSlopesCl4\_30as | Number of pixels with slope 5–8% |
| gb\_GloSlopesCl5\_30as | Number of pixels with slope 8–16% |
| gb\_GloSlopesCl6\_30as | Number of pixels with slope 16–30% |
| gb\_GloSlopesCl7\_30as | Number of pixels with slope 30–45% |
| gb\_GloSlopesCl8\_30as | Number of pixels with slope > 45% |

|  |  |
| --- | --- |
| **ITE Land Class 2000** | |
| lc\_diversity\_5 | Shanon-Weiner Diversity Index calculated using the ITE Land Class 2000 classes of 1km squares within 5km of the target square. |

|  |  |
| --- | --- |
| **UK Climate Impacts Programme** | |
| UK Climate Impacts Programme (UKCIP) “UKCP09: Access to gridded data”. The main input was a set of observations made at meteorological stations, irregularly distributed in time and space. This is gridded at 5km x 5k resolution. Data for 1997 – 2006 was averaged (2006 is the last year for which most data is currently available). The first 3 layers are annual summaries calculated by the Met Office. The function “biovars” from R package dismo was used to calculate the other raster layers from the 12 monthly layers for MaxTemp, MinTemp and Rainfall. 1km x 1km rasters were then created from the resulting 5km x 5km grid using the disaggregate function of R package “raster” using bilinear interpolation. | |
| ConsecutiveDryDays | Number of Consecutive Dry Days |
| GrowingDegreeDays | Growing Degree Days |
| GrowingSeasonLength | Growing Season Length |
| RainCoolestQ | Precipitation of coolest quarter |
| RainDriestMonth | Precipitation of driest month |
| RainDriestQ | Precipitation of driest quarter |
| RainSeasonality | Precipitation seasonality (coefficient of variation) |
| RainTotal | Total (annual) precipitation |
| RainWarmestQ | Precipitation of warmest quarter |
| RainWetestMonth | Precipitation of wettest month |
| RainWetestQ | Precipitation of wettest quarter |
| TAnnRange | Temperature annual range (bio5 - bio6) |
| TIsothermality | Isothermality (bio2/bio7) \* 100 |
| TmaxWarmestMonth | Max temperature of warmest month |
| TmeanAnn | Mean annual temperature |
| TmeanCoolestQ | Mean temperature of coldest quarter |
| TmeanDiurnalRange | Mean diurnal range (mean of max temp - min temp) |
| TmeanDriestQ | Mean temperature of driest quarter |
| TmeanWarmestQ | Mean temperature of warmest quarter |
| TmeanWetestQ | Mean temperature of the wettest quarter |
| TminCoolestMonth | Min temperature of coldest month |
| TSeasonality | Temperature seasonality (standard deviation \*100) |

# Results

An initial Maxent (version 3.3.3k) model was developed using a mask so that only 1km squares from Scotland (defined as 1kms falling within VCs 72 – 112) were used to select background data to train the model. Half of the available 1km occurrences were selected randomly for model training and the other half used for testing. The resulting model was projected for the whole of GB.

|  |
| --- |
| Pearl-bordered_fritillary_roc.png |
| Chequered_Skipper_roc.png |
| Fig. : ROC plot. |

shows Receiver Operating Characteristic (ROC) plots comparing the fit to the training data and test data. The Area Under the Curve (AUC) for the test data of 0.939 for Pearl-bordered Fritillary and 0.980 for Chequered Skipper. shows the predicted logistic probability distribution (note that this is an index of probability, not an estimate of likelihood of occurrence! Thus if the value is 0.5 then it is around twice as likely to occur as if the value was 0.25, but neither value indicate how likely the species is to occur).

shows the estimated contribution of environmental variables. Only those with a contribution of ≥1% are shown. These account for 85% in aggregate for the Pearl-bordered Fritillary and 86.5% for the Chequered Skipper.

For Pearl-bordered Fritillary, as expected, the presence of south facing slopes and woodland cover were the most important factors followed by factors relating to water and temperature.

For Chequered Skipper, climatic factors, especially rain, and altitude seem to be most important, with woodland coverage, rough grass and southern aspect coming in somewhat further down the list.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Pearl-bordered_fritillary_Layers.png | **Pearl-bordered Fritillary**   |  |  | | --- | --- | | **Variable** | **Contribution (percent)** | | gb\_GloSlopesCl7\_30as | 16.27 | | 01\_bwood | 16.26 | | 02\_conif | 10.19 | | gb\_GloAspectClS\_30as | 5.91 | | cat\_hg | 4.79 | | TmeanDiurnalRange | 4.68 | | 16\_freshwater | 4.60 | | 03\_arable | 3.43 | | TSeasonality | 2.76 | | TminCoolestMonth | 2.23 | | TmeanDriestQ | 1.91 | | gb\_GloSlopesCl4\_30as | 1.72 | | gb\_GloSlopesCl8\_30as | 1.67 | | ConsecutiveDryDays | 1.43 | | 11\_heathgrss | 1.32 | | 04\_impgrss | 1.31 | | 08\_acidgrss | 1.11 | | AltMean | 1.10 | | TIsothermality | 1.05 | |
| Chequered_Skipper_Layers.png | **Chequered Skipper**   |  |  | | --- | --- | | **Variable** | **Contribution (percent)** | | RainWetestMonth | 17.53 | | RainWetestQ | 10.74 | | AltMean | 10.61 | | RainDriestMonth | 9.52 | | TSeasonality | 9.48 | | 01\_bwood | 9.00 | | 05\_rghgrss | 6.05 | | RainCoolestQ | 3.01 | | gb\_GloAspectClS\_30as | 2.36 | | cat\_parmase | 2.25 | | TmeanAnn | 2.19 | | RainSeasonality | 1.42 | | cat\_parmado | 1.20 | | lc\_diversity\_5 | 1.15 | |
| Fig. : Model prediction predicted over GB. Red is high, light blue low. | Table : Variables ( for variables contributing ≥1%). |

## Predicting the distribution

To produce a prediction of occupied grid squares requires that a threshold is applied to the output probability surface. Determining the appropriate threshold is difficult! It cannot be done using the presence only data used to fit the model since that cannot contain any indication of where the species is absent (which is why the logistic probabilities have no absolute meaning !).

## GB distribution at 10km square resolution

The process I went through for Pearl-bordered Fritillary to come up with some sort of “guestimate” for a threshold was:

1. Download the known presences from the NBN Gateway. Tom Brereton does not say what date he considers “recent”, so I guessed 1990 onwards by comparing the presences supplied in Scotland with those available from the NBN. The NBN data included 76 10kms from Scotland, 61 from England and 59 from Wales from 1990 onwards.
2. The supplied data includes 111 Scottish 10km squares, i.e. 111/76 = 1.46x as many as were available from the NBN.
3. I therefore estimated the “real” number of occupied 10km squares in England and Wales using the same multiplier. This gives: 111 + (61 \* 1.46) + (59 \* 1.46) = 268 total occupied 10km squares. This is 9.8% of the total number of 10km squares containing land in GB.
4. I therefore chose a threshold which would predict that 9.8% of 10km squares are occupied (threshold = 0.74). (I term this an “Occupancy threshold”).

The situation is a bit easier for Chequered Skipper because it is extinct in England, therefore the Scottish presences constitute the whole of the GB distribution. The supplied occurrences are from 35 10km squares which is 1.22% of those containing land. This leads to an occupancy threshold of 0.7656.

Applying this threshold gives the predicted 10km square distribution shown in and (based on the 10km square being occupied if any 1km square within it is occupied). This is compared to the observed distribution obtained by combining the 1990 onwards 10km squares from the NBN with the supplied Scottish data. Interestingly, the model for Chequered Skipper predicts occurrence in North Wales (where, as far as I can see, it has never been found), but does not predict it in the East Midlands where it was previously resident.

|  |  |
| --- | --- |
| 10km_predict.EMF | observed.EMF |
| Fig. : Predicted, 10km distribution for Pearl-bordered Fritillary from the model fitted above for a threshold of 0.74 (see text) | Fig. : Observed, 10km distribution: data from 1990 onwards downloaded from NBN Gateway + supplied occurrences. |

|  |  |
| --- | --- |
| Cheq_Skip_predict10.EMF | skipper 10km.EMF |
| Fig. : Predicted, 10km distribution for Chequered Skipper from the model fitted above for a threshold of 0.766 (see text) | Fig. : Observed distribution from supplied occurrences. |

## Scottish distribution at 1km resolution

The models were further tested using both the random sample of testing data that was not used to train the models combined with absence data supplied by Tom Brereton. shows the number of 1km squares available for training and testing. Note that the total number of squares (Training + Testing) is less that the number of supplied occurrences. This is because some of the supplied squares lack environmental data in one or more of the layers used. This arises because of the remapping of datasets in different projections systems (i.e. the slope and altitude data arrived from DEM and the European Soils Database) to the OSGB grid. Some OSGB squares, especially on the coast, will not overlap with the centres any squares of the original data in the other projection.

Table : Numbers of 1km squares available for training and testing models.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Species** | **Supplied by Tom Brereton** | | **Random selection by Maxent** | | **Test set constructed by SGB** | |
| **Presence** | **Absence** | **Train** | **Test** | **Presence** | **Absence** |
| Chequered Skipper | 249 | 11,446 | 107 | 107 | 107 | 214 |
| Pearl-bordered Fritillary | 264 | 1,030 | 124 | 123 | 123 | 246 |

Although Tom Brereton does not say how his absences were derived, one thing that is immediately obvious is that there are orders of magnitude more of them than presences – especially the 50% sample of presences selected for testing. If all were used to test models, the results would be heavily dominated by cases where the model predicts absence and absence was “observed”. This is likely to suggest far greater model accuracy than is really warranted. Therefore I constructed a test dataset consisting of the test set selected by Maxent (so that the model was not tested against occurrence to which it was fitted) and a random sample of the absences consisting of twice as many squares as there were test presences.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | | Measure | Chequered Skipper | Pearl-bordered Fritillary | | AUC | 0.987617 | 0.875137 | | Max accuracy | 0.968641 | 0.808642 | | Max Kappa | 0.933573 | 0.600048 | | Max AC1 | 0.940804 | 0.640322 | | Max g-mean2 | 0.971157 | 0.804595 | | Max F1 | 0.958904 | 0.757813 | |
| Fig. : ROC plot for model tested against 1km dataset constructed by SGB consisting of the test occurrences selected by Maxent and a random sample of twice as many absences. | Table : Statistics from ROC analysis. |

shows ROC plots for the two species and the AUC is shown in . The model for Chequered Skipper fits the observed data extremely well (AUC = 0.987 – incredibly high in my experience. A result this good would normally suggest that you cheated!) and that for Pearl-bordered Fritillary very well (AUC = 0.875). The maximum accuracy of prediction was 97% for Chequered Skipper and 81% for Pearl-bordered Fritillary.

Various measurements have been proposed to quantify the accuracy of predictions from this type of model. A threshold is applied to turn the probability surface output by the model into presence/absence predictions and these are compared to the observations of presence/absence for particular grid squares (i.e. the test dataset). The simplest measure, “accuracy” is the proportion of cells that the model predicts correctly (i.e. presence predicted and observed or absence predicted and observed). This has often been criticised because it is dominated by predictions of absence for scarcer species and also because it does not take into account the number of correctly predicted cells you would expect to get by chance alone (i.e. randomly scatter the observed number of presences and see how many are predicted by the model). Kappa attempts to correct for the chance element, but has been criticised because the way it calculates how many correct predictions you would expect by chance alone is not correct. AC1 is a more modern measure that does calculate this number correctly. Finally, one can ignore the cases where absence is correctly predicted and only take into account correct predictions of presence. g-mean2 and F1 are measures which attempt to do this. In general accuracy is the most easily understood measure (it’s a direct measure of the proportion of grid cells in the test dataset that the model predicted correctly for a given threshold), AC1 provides the best correction for chance prediction rate and g-mean2 is the most frequently used “presence only” measure.

shows the maximum values of these measures for our two species and & show how some of these measures vary with threshold.

|  |  |
| --- | --- |
|  |  |
| Fig. : Variation of measures with threshold for Chequered Skipper | Fig. : Variation of measures with threshold for Pearl-bordered Fritillary |

& potentially provide another way of determining a suitable threshold. You can read off the threshold at which the model fits the test data “best” (i.e. at the maximum value of one of the measures of accuracy). I term this an “accuracy threshold”.