**Brief summary of species distribution modelling applied to Marsh Fritillary in Scotland**

**Emily Dennis – Feb 2020**

We use species distribution models (SDMs) to predict the potential distribution (in terms of suitable habitat) of Marsh Fritillary in Scotland, to help inform the production of a risk/opportunity layer for tree planting in Scotland.

**Marsh Fritillary data preparation**

Presence records of Marsh Fritillary for Scotland were obtained from the BC Scotland dossier for 2000-2017. The spatial extent for the modelling was defined by a 20km buffer applied to the min/max easting and northing of Marsh Fritillary records (i.e. to form a square).

Absence records were produced using records from the Butterflies for the New Millennium (BNM) database for 2000-2017 made within May - July when Marsh Fritillary adults are most likely to be recorded. The data were limited to records of adults with at least 1km resolution, falling within the spatial extent defined above, and with at least 3 records during 2000-2017.

**Environmental variable**

We used land cover data from the CEH land cover map 2015 at 1km resolution (Rowland et al. 2017) which consists of 21 land cover classes. We removed several classes which had low coverage in the region of interest, hence we initially considered the following classes: broadleaf woodland, coniferous woodland, arable, improved grassland, acid grassland, heather, heather grassland, bog, and urban (combined from the urban and suburban target classes). Elevation data was sourced from OS Terrain 50 in gridded format.

Average mean summer temperatures (for 2000-2017) were derived from HadUK-Grid data (Hollis et al 2019). Average temperatures for other seasons were also considered, but not used in the analyses due to the high correlation with summer temperature. Rainfall was also considered but highly correlated with elevation hence also ignored.

**Modelling**

We applied species distribution models (SDMs) to the data for Marsh Fritillary at 1 km square resolution using the Biomod2 package in R (Thuiller et al 2009, 2019). Seven different individual model types were applied: GLM, GBM, GAM, RF, MARS, CTA and ANN. The data were randomly split into training/calibration (70%) and testing/evaluation (30%) data for 10 runs of each model (70 models fitted in total). Average variable importance (across models and randomised data) was assessed and broadleaf woodland, arable, acid grassland, heather and urban were removed in subsequent model runs (thus retaining coniferous woodland, improved grassland, heather grassland, bog, summer temperature and elevation). Based on having the lowest average AUC values, RF, CTA and ANN were excluded from the model ensembles.

Ensemble modelling was applied using a weighted average of the individual model evaluation metrics (alternative approaches were also briefly checked and showed minimal effects on the evaluation metrics). Models were primarily evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC), as well as sensitivity (the proportion of correctly predicted presences) and specificity (the proportion of correctly predicted absences). AUC measures the performance of a model to discriminate between sites where a species is present versus those where it is absent. AUC values of 1 represent a perfect fit whereas a value of 0.5 indicates a model that is as good as random. The final maps presented are based on an ensemble of the 4 best individual models applied to the full data set. Maps could also be produced for the coefficient of variation of the predictions from the 4 individual models, to describe some of the underlying uncertainty in the predictions.

**References**

Hollis, D.; McCarthy, M.; Kendon, M.; Legg, T.; Simpson, I. (2019): HadUK-Grid Gridded Climate Observations on a 1km grid over the UK, v1.0.1.0 (1862-2018). Centre for Environmental Data Analysis, 14 November 2019. doi:10.5285/d134335808894b2bb249e9f222e2eca8

Rowland, C.S.; Morton, R.D.; Carrasco, L.; McShane, G.; O'Neil, A.W.; Wood, C.M. (2017). Land Cover Map 2015 (1km percentage target class, GB). NERC Environmental Information Data Centre. [**https://doi.org/10.5285/505d1e0c-ab60-4a60-b448-68c5bbae403e**](https://doi.org/10.5285/505d1e0c-ab60-4a60-b448-68c5bbae403e)

Wilfried Thuiller, Damien Georges, Robin Engler and Frank Breiner (2019). biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.3-7.1. https://CRAN.R-project.org/package=biomod2

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD–a platform for ensemble forecasting of species distributions. Ecography, 32, 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x