Dissertation Workflow: First Four Weeks

This document outlines the detailed technical steps for the first four weeks of the dissertation project: 'How do land use and climate change interaction affect the future spread of the Asian hornet (*Vespa velutina*) in England?'. The plan includes data acquisition, preprocessing, model training, and projection.

Week 1: Data Acquisition and Preparation

- 1.1 Download GBIF occurrence records for *Vespa velutina* in France, Spain, and Portugal:
  - Use GBIF.org or the `rgbif` R package in R.
  - Filter for records between 2010 and 2024.
  - Clean and spatially thin the data to 10 km using the `spThin` package.
- 1.2 Download WorldClim v2.1 climate data:
- Load full 19-variable raster stack Start with broad information
- Sample across study area Build a representative data table
- Calculate Pearson's r Detect collinearity
- Use caret::findCorrelation() Automatically filter down
- Justify final variables Biological + statistical rationale
  - Variables: Bio2 (mean diurnal range), Bio6 (min temperature of coldest month), Bio12 (annual precipitation).
  - Baseline (1970–2000) and future (2041–2060) under RCP 4.5 and RCP 8.5.
  - Resolution: 30 arc-seconds (~1 km).
- 1.3 Download CORINE Land Cover data (2000 and 2018):
  - Source: Copernicus Land Monitoring Service.
  - Reproject to EPSG:4326 (WGS84).
  - Prepare for reclassification into urban, forest, agriculture, and natural categories.
- 1.4 Acquire and prepare England boundary shapefile:
  - Use GADM or UK Data Service.
  - Clip all raster layers to the extent of England using this shapefile.

Week 2: Raster Preprocessing and Harmonisation

- 2.1 Reproject and resample CORINE land cover rasters:
  - Match resolution and CRS of WorldClim layers.
  - Use nearest neighbour resampling for categorical land cover data.

- 2.2 Generate additional predictor layers (optional):
  - Create distance-to-urban and distance-to-forest rasters.
  - Use QGIS or R with the `terra` or `gdistance` packages.
- 2.3 Stack and verify environmental layers:
  - Use `raster::stack()` or `terra::rast()` in R.
  - Ensure consistent resolution, extent, alignment, and CRS.
- 2.4 Begin drafting the Methods section:
  - Write up data sources, processing steps, spatial resolution, CRS, and rationale for variable selection.

Week 3: Model Training and Tuning

- 3.1 Set up training data:
  - Use cleaned and thinned GBIF occurrence points from Western Europe.
  - Extract environmental values at occurrence locations using R (`raster::extract`).
- 3.2 Run baseline MaxEnt model:
  - Use `dismo::maxent` or GUI-based MaxEnt software.
  - Train using current climate and land cover predictors.
- 3.3 Tune model with ENMeval:
  - Select best combination of feature classes and regularisation multiplier.
  - Evaluate models using AICc, AUC, and Boyce Index.
  - Select the best-performing model for projection.

Week 4: Scenario Projections and Visualisation

- 4.1 Prepare future projection layers:
  - Download or derive climate layers for RCP 4.5 and RCP 8.5 (2041–2060).
  - Extrapolate LULC change scenarios from CORINE (e.g., linear urban growth).
- 4.2 Project model to England under future scenarios:
  - Use the best MaxEnt model from Week 3.
  - Apply it to raster stacks representing each scenario (2 RCPs × 2 LULC).
- 4.3 Visualise outputs in QGIS:
  - Import MaxEnt output rasters (.asc or .tif).
  - Generate continuous and binary suitability maps.
  - Highlight areas of suitability gain/loss between scenarios.
- 4.4 Document preliminary findings:
  - Take notes on variable contributions, hotspot emergence, and suitability shifts.
  - Begin outlining Results and Discussion sections based on initial outputs.