Weedmodels:

The r package weedmodels was constructed to help archaeobotanists understand the intensity of crop husbandry etc.

## Introduction:

The package uses linear discriminant analysis which is a statistical method which can be used to understand the differences between groups, and then based on that those calculated differences can then classify datasets using the previously constructed model. Within weed functional ecological models, discriminant analysis is used to find the best separation between the two crop husbandry regimes of the model dataset based on the load functional ecological traits. Once the equations that best separates the two regimes is found, it can then be used to classify archaeobotanical data as either regime 1 or regime two.

## Background?

Data set – functional traits, how they are collected,

X number of species relevant to archaeobotanical assemblages are published with this paper, providing the functional attribute data (whole or in part) required to run the models. The functional data of the species presented here have been found within archaeobotanical assemblages (and identified in studies assessing functional weed ecology).

The models - Link to the three papers- brief description of each one and when they have been used?

Three models are presented here. While these models have previously been published; this paper for the first time provides the data required to run them. Within the R package, for ease of use, the different models have been names, model 1, 2, or 3 based on order publication.

Model 1 was published in Bogaard et al 2016, and detailed the results of modern field studies in Haute Provence, which makes up the low input farming regime, while data from Asturias Spain, provided the high input farming regime. Bogaard et al. 2016 presented the discriminant analysis of these two datasets, discriminant which functional attribute best separated the two groups. The resulting discriminant model (figure 1) separated the two regimes based on five functional attributes: canopy height, canopy diameter, leaf area per node/leaf thickness, Mean species leaf area, length of flowering period.

Model 2

Model 3

Functionality of the package

Weedmodels can be broken up in to three different groups of function: data organisation, classification, and visualisation. This paper will go through these different groups separately however the Weedmodels R help file also provides examples of the use of each function.

### Data organisation:

The functional ecology trait database is organised by species with each species having a three-four species code. Each species has a code maded up of the first four letters of its genus name and the first three letters of the species name. these codes are used to extract the functional data of that species. The four-three codes of the species currently availability in the database can be found here. It should be noted that species with genus names less than 4 letters long are separated in the code with a “\_”; e.g. poa\_ann for *Poa annua*.

To conduct functional weed ecology discriminant analysis on an archaeobotanical dataset the data needs to be in a specific format. The function data\_org organises a raw archaeobotany spreadsheet to the require format – it changes the archaeobotanical data into presence/absence data. Then it extracts the functional trait data based on the four-three species codes of the species identified in column x. Finally, the averaged functional attribute values for each trait per sample is calculated, returning a dataframe? Which is then suitable for the classification functions within the package.

Two other data organisation functions are weed\_data and ave\_sp. Weed\_data allows users to extract trait data based on the four-three species codes, allowing uses to assess specific data if they want to use alternative statistical programs. Ave\_sp is a function which allows users to ave\_sp the functional trait data of a set of species, with can then be added into the data\_org function. This function was created due to the fact that in some cases specimens cannot be identified to species. If the genus has limited species, or the specimen could be one of only a couple of species, then it is possible to average the multiple species, trait data to produce values which can be used in data\_org. CAVEATS????

### Classification:

Weedmodels provides a function for linear discriminant analysis. Wmodel.LDA uses the MASS packages LDA function but wraps in so that it compares the imputed archaeobotanical data against one of the three included modern models of WHAT. This makes use simple as the only inputs required are the dataset produce from data\_org, as well as instructions on which model the use wishes to compare the dataset against. If the output of data\_org is not used then the model requires specific column names for the averaged functional trait data for each archaeobotanical sample (See table 1).

This function conducts discriminant analysis on the selected model (data stored in r package) and then uses the discriminant model to classify the entered archaeobotanical data as either a 1 or 2. (EXPLAIN). It also provides the posterior probability of each archaeobotanical sample falling within group 1 or group 2. This function provides information about how similar the functional trait values of the species within a sample are to regime 1 or regime 2 of the selected model.

Table 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Trait code | meaning | Provided? | Model 1 | Model 2 | Model 3 |
| SLA | Specific leaf area | Yes | Yes | Yes |  |
| ARNODE | Leaf area: leaves per node | Yes | Yes | Yes |  |
| LOGCANH | Log of canopy Height | Yes | Yes | Yes |  |
| LOGCAND | Log of canopy diameter | Yes | Yes | Yes |  |
| VEGPROP | Vegetative propagation | Yes |  |  | Yes |
| FLOWPER | Flowering period | No | Yes |  | Yes |

### Visualisation

The final group of functions provide options for plotting the output of wmodel.LDA. there are three options provided for plotting the archaeobotanical data, each varying in how much detail is show of the classifying model used. For example, wplot\_arch plots the archaeobotanical samples’ discriminant scores, while comparing them with centroids of the two modern models (figure X). wplot\_basic plots the archaeobotanical samples’ discriminant score against the model’s dataset and centroids while wplot\_geog separates out the model’s samples into their different geographical locations – this is particularly relevant for model 2 which has modern data from multiple geographical locations.

The plotting function use the beeswarm package’s function swarmy, and a couple of that function’s arguments can be used to modify the arrangement and order of the points (compact and priority).

Example of use: