## Abstract:

## Introduction:

The functional ecology of arable weeds offers a means of comparing present-day and past farming regimes, of disentangling the ecological variables that characterise them and of exploring ancient regimes that have no modern analogue (REFS). Functional ecology identifies a series of traits (or ‘attributes’) that reflect species’ ecological potential in relation to environmental variables such as soil productivity and disturbance (REFS). These traits are morphological or behavioural characteristics that bear a functional relationship to species’ responses to these ecological parameters (REFS).

In this paper we present a functional trait dataset for nearly 1000 arable weed species [OR is this a separate data-in-brief paper? Check with JH] occurring in Europe and North Africa, together with tools to compare present-day and past weed flora on the basis of these traits. The traits presented are those identified in previous studies as useful for discriminating present-day farming regimes that differ primarily in terms of soil productivity and/or disturbance (REFS). The weed survey studies undertaken to characterise the present-day regimes in Europe and North Africa have been published over the past two decades (REFS). The underlying weed survey data will be released in a future publication. Further survey studies are desirable to broaden the range of scenarios and comparisons, including other world regions.

Below we introduce previously published models for comparing present-day and past weed flora on the basis of relevant functional traits. We also present the R package WeedEco, which allows users to run these models. The R package WeedEco has been designed for users wishing to compare unknown data against the previously published models, providing functions for data organisation, classification and visualisation. The functional trait data published here [OR in accompanying paper] can also be used in other, exploratory ways: for example, to assess the relevance of the traits to compositional trends among archaeobotanical samples (REFS), or to characterise the ecology of newly surveyed crop fields.

The models

The R package presented below provides the tools needed to take a raw archaeobotanical weed dataset and conduct linear discriminant analysis to understand where the archaeobotanical samples fall in comparison to the three presented modern models.

Three models are presented here, and while these models have been previously published, this paper for the first time provides the data required to run them (see supplementary data, and data section). For ease of use the different models are referred to as model 1, 2, or 3 based on order of publication and are briefly summarised below.

Model 1: High- versus low-input farming (Asturias and Provence)

In 2016 Bogaard et al. published the results of modern field surveys in Haute Provence, the low input farming regime side of model 1. These data were combined with data from Asturias, Spain which provided the high input farming regime side of a newly constructed model (Jones et al 1995, Charles et al 2002). Bogaard et al. (2016) conducted discriminant analysis of these two datasets, identifying which functional traits best separated the two groups. The resulting discriminant model separated the two regimes based on five functional traits (or attributes): canopy height, canopy diameter, leaf area per node/leaf thickness, mean specific leaf area, and length of flowering period. The results indicated that differences in fertility and disturbance were the driving ecological processes separating the two regimes, with the highly fertile, highly disturbed Asturias plots contrasting with the low fertility, low disturbance Haute Provence fields. This model has been used in other research to explore the intensity of cultivation in multiple time periods in Germany (Bogaard et al. 2016, Styring et al. 2017, Hamerow et al., 2021), as well as medieval England (Hamerow et al., 2020) and Iron age France (Alagich et al. 2018).

Model 2: High- versus low-input farming (Asturias, Evvia, Provence and Morocco)

Arable field surveys in Morocco provided data from oases and rain-fed fields in a semi-arid region, allowing for the construction of model 2 (Bogaard et al. 2018). This model incorporated data from model 1 (Haute Provence and Asturias) and from Evvia in Greece (Jones et al. 1999, 2000). Model 2 uses data from three different locations to represent high-intensity cultivation (Moroccan oases, Evvian gardens and Asturian plots) and low-intensity regimes (Moroccan rain-fed terraces, Evvian fields and Haute Provence fields). The model places the emphasis on fertility rather than soil disturbance, since fertility-related traits successfully separated high- and low-intensity regimes and the inclusion of disturbance attributes did not improve this separation further. This model has been applied to semi-arid locations, in particular in western Asia (see Bogaard et al. 2018, Green et al. 2018, Stroud et al. 2020, Diffey et al. 2020, Maltas et al. 2022).

Model 3: High versus low disturbance (Highgrove and Laxton)

Model 3 uses data from two locations within the UK to distinguish levels of soil disturbance irrespective of fertility. The model uses modern botanical survey data from Laxton in Nottinghamshire and Highgrove’s Duchy Home Farm in Gloucestershire. The Laxton data include surveys of unploughed but periodically grazed and cut hay meadow areas on the edges of the open fields (called ‘sykes’), as well as unsprayed edges of strip cereal fields and fallow fields, managed within an open-field system, while the Highgrove data are from cereal fields cultivated in different a rotation system (see Bogaard et al. 2022 for full details). The combination of these data provides a comparison of highly disturbed arable fields with unploughed meadow areas, creating a model with which to explore disturbance. This model has been used to investigate levels of disturbance expected under mouldboard plough tillage, through comparison with experimental ridge-and-furrow fields at Lorsch, Germany, and to assess how arable disturbance levels developed through the English medieval period, when the mouldboard plough is hypothesised to become widely used (Bogaard et al. 2022).

Data collection and functional traits

The data for the models comes from botanical surveys of modern fields; in each field the species present within one metre square quadrats were recorded. These surveys were conducted in five to ten 1m2 quadrats along a transect from one end of the field to the other. For use in the models and comparison with archaeobotanical data, the botanical survey data were converted to species’ presence/absence per field. The average score of the functional traits per field was then calculated as the sum of the attribute value for the species divided by the number of species in the field (see Bogaard et al 2016, Bogaard et al 2018, Bogaard et al 2022 for full details).

The averaged functional trait data per survey field is published with this paper and is required to run the models: in the case of the R package these data are written into the functions (Supplementary file 1). Furthermore, the functional trait data of 929 archaeobotanically relevant species are also published with this paper (Supplementary file 2). Only species that have values for all required functional traits are provided.

The values of five functional traits are provided for the 929 species: SLA, ARNODE, LOGCANH, LOGCAND and VEGPROP (see Table 1 for full details). An additional functional trait, FLOWPER, is not provided and requires users to obtain these data from relevant Floras that cover their study region. The trait values provided represent the average of multiple specimens of that species from different geographic locations. Only plants growing in optimal conditions were measured to capture the species’ potential, rather than an individual plant’s performance. The six functional traits are used in varying combinations within the three presented models (Table 1). Further details on the measurements of the attributes can be found in Jones et al. (2000) and Bogaard et al. (2001).

Table 1. The six functional traits used within the different models, and whether the data is provided.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Trait code | Meaning | Calculation | Provided? | Model 1 | Model 2 | Model 3 |
| SLA | Mean specific leaf area | Leaf area per unit of dry leaf weight | Yes | Yes | Yes |  |
| ARNODE | Leaf area per node/ leaf thickness |  | Yes | Yes | Yes |  |
| LOGCANH | Log of canopy height | Max | Yes | Yes | Yes |  |
| LOGCAND | Log of canopy diameter | Max | Yes | Yes | Yes |  |
| FLOWPER | Flowering period | Length in months | No | Yes |  | Yes |
| VEGPROP | Vegetative propagation | Yes or no | Yes |  |  | Yes |

Functionality of the R package

The R package WeedEco allows users to organise raw archaeobotanical data and then conduct linear discriminant analysis to classify that data against the supplied modern models. WeedEco also has functions to produce plots of the output of the linear discriminant analysis. WeedEco can be broken up into three different groups of functions: data organisation, classification, and visualisation. The package can be downloaded into R from github using the devtools package. The package WeedEco can be manually downloaded from the archaeobotanyatoxford github account or the code below can be used to download it within R using the devtools package’s function install\_github: install\_github("archaeobotanyatoxford/WeedEco\_Package")

### Data organisation

The functional trait database is organised by species, with each species designated a four-three species code (henceforth called species code). Where possible a species code is made up of the first four letters of the genus name and the first three letters of the species name. There are exceptions to this rule such as when a code is not unique, or the species does not have a genus name four letters long. For taxa that have an identical four-three code, commonly the next letter is used. It should be noted that species with genus names less than four letters long are separated in the code with a “\_”; e.g., poa\_ann for *Poa annua*. These four-three species codes are used to extract functional data from the database by the R package WeedEco. It is therefore essential that the species codes used are the correct codes for the species required. The four-three codes of the species currently available in the database can be found here (Supplementary file 3).

The function **wdata\_org** organises a raw archaeobotanical spreadsheet into the format required for linear discriminant analysis. It changes the archaeobotanical data into presence/absence data and then extracts the functional trait values, based on the species codes, of the species with each sample. Finally, the averaged functional attribute values for each trait per sample is calculated, returning a data frame[[1]](#footnote-1) which is suitable for use within the classification functions of the package.

WeedEco’ two other data organisational functions are **weed\_data** and **ave\_wdata.** **weed\_data** allows users to extract functional trait data based on the entered species codes and is useful if the data are required for alternative uses or validity checking. **ave\_wdata** was created to deal with occurrences of specimens which cannot be identified to species, allowing users to average the functional trait data of a number of species to form a composite value. If the genus has limited species, or the specimen is one of only a few species, then it is possible to average the multiple species’ trait data to produce trait values which can be used in **wdata\_org**. Users are recommended to be cautious in averaging more than three or four species, or in averaging values that are very different.

### Classification

WeedEco provides a function to conduct linear discriminant analysis called **wmodel.LDA**. It uses in part the MASS package’s **lda** function, but wraps it, allowing for the comparison of the inputted archaeobotanical data against one of the three included modern crop regime models. This makes use simple with the only inputs needed being the dataset produced from **wdata\_org**, as well as instructions on which model is required. If the output of **wdata\_org** is not used then the function requires specific column names and order for the averaged functional trait data for each archaeobotanical sample (SLA, ARNODE, LOGCANH, LOGCAND, VEGPROP, FLOWPER).

**wmodel.LDA** conducts discriminant analysis on the selected comparative modern model: such data are stored within the R package but are also included in supplementary file 1. **wmodel.LDA** creates a discriminant model that is used to classify the entered archaeobotanical data as either regime 1 or regime 2, with interpretation dependent on the selected model (column called Class). **wmodel.LDA** also provides the posterior probability of each archaeobotanical sample falling within group 1 or group 2 (Prob.1 and Prob.2). The samples’ linear discriminant scores (LD1) are also calculated, and they are later used within the plotting functions. The function provides both standardised and un-standardised data, as well as the unstandardised centroid values for group 1 and group 2 (for further detail on the linear discriminant analysis procedure see MASS package help file, or Venables and Ripley (2002)). Further information on the returned data is provided in the WeedEco help file.

### Visualisation

The final group of functions provide options for plotting the output of **wmodel.LDA**. The four options vary in the amount of detail shown of the classifying model used. For example, **wplot\_arch** plots the archaeobotanical samples’ discriminant scores, while comparing them with the two centroids of the modern model selected (Figure 1a). **wplot\_basic** plots the archaeobotanical samples’ discriminant score against the model’s samples and centroids (Figure 1b). **wplot\_geog** separates out the model’s samples into their different geographical locations – this is particularly relevant for model 2 and model 3 which have modern data from multiple geographical locations (Figure 1c). The final plotting option, **wplot\_phase,** produces a stacked graph of up to five subplots allowing the display of multiple sites or different phases (Figure 1d). **wplot\_phase** is more complex than the other plotting function which only require the model number and the column with the LDA data from the **wdata.LDA** output; **wplot\_phase** requires an additional grouping variable.

The plotting functions use the Beeswarm package’s function **swarmy**, and some of that function’s arguments can be used to modify the arrangement and order of the points (compact and priority) (see Beeswarm help documents). Further modification is possible of the axes, limits, symbols, colour, and legend location (see below and the WeedEco help documentation for more details).

## The case study site of medieval Stafford

#### Model 1

Archaeobotanical data from the early medieval site of Stafford was analysed by Hamerow et al. (2020) and Bogaard et al. (2022) to understand the evolving nature of the crop production system and to assess a hypothesised trend towards extensification (expansion through low-input farming) during this period of agronomic reorganisation. The Stafford data are re-analysed below using WeedEco to demonstrate the utility of the R package. The archaeobotanical samples come from three excavations in the centre of Stafford: St Mary’s Grove, Bath Street and Tipping Street South (Carver 2010, Dodd et al. 2014, Hamerow et al. 2020). The data spans four broad phases of occupation from the late 9th to 13th centuries.

A flow diagram of the steps required to take a raw archaeobotanical dataset, conduct discriminant analysis, and then produce plots is shown in supplementary file 4. Each step required for the Stafford data is described in detail below. However, it should be noted that this does not include the use of model 2, due to its geographical/environmental unsuitability for Stafford. Furthermore, an R script containing the step-by-step processes involved within R to analyse and plot the Stafford data using WeedEco is included in supplementary data.

The raw archaeobotanical data from Stafford, a total of 46 samples, was ‘cleaned’, removing any crop species or non-arable taxa. Any tentative identification – for example c.f. identifications – were either removed or added to positively identified tallies of that species, as appropriate. Taxa that could not be identified to species were either excluded or, if thought to be one of a small number of species, included as a composite taxon, with occurrences summed into one row (Table 2). A column was added to the datasheet which included the species code of each taxon obtained from supplementary file 3 (Table 2, column 1; “species.code”). For composite taxa, a species code was constructed using something unique and memorable. If a species is not in the supplied spreadsheet, this indicates that full functional trait data is not available for it: the species either needs to be removed or functional trait data obtained from elsewhere. The only species which did not have trait data, *Carex flava,* was retained, as trait data was obtained from an alternative source (see below). Only archaeobotanical samples which contain at least ten seeds of weed species included in the final cleaned dataset were kept, resulting in a total of 45 samples. The final cleaned dataset is shown in supplementary data 5.

Table 2 The format of the raw dataset from Stafford showing the first 8 samples and first 12 taxa. Full dataset in supplementary data. The rows in bold are composite species and the first column shows the species codes for each taxon obtained from X.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| species.code | Species | 2228 | 2323 | 2223 | (1071A) | 2136 | 1682 | 2222 | 068/076 |
| agrogit | Agrostemma githago | 115.4 | 144 | 50 | 450 | 23.1 | 12 | 12 | 100 |
| anetgra | Anethum graveolens |  |  |  |  |  | 1 |  |  |
| anthcot | Anthemis cotula | 9092.3 | 1456 | 5430 | 2000 | 3630.8 | 2065 | 2268 | 100 |
| arrhels | Arrhenatherum elatius |  | 8 |  |  |  |  |  |  |
| avenstfa | **Avena sterilis / Avena fatua** |  |  |  |  |  |  |  |  |
| brasnig | Brassica nigra |  | 4 |  |  |  | 1 |  |  |
| bromhose | **Bromus secalinus / Bromus hordeaceus** | **15.4** | **64** | **30** | **340** | **23.1** | **18** | **12** | **100** |
| buglarv | Lithospermum arvense |  |  |  |  |  |  |  |  |
| buplrot | Bupleurum rotundifolium |  |  |  |  |  |  |  |  |
| carefla | Carex flava |  |  |  |  |  |  |  | 200 |
| carenig | Carex nigra |  |  |  |  |  |  |  |  |
| carepna | Carex panicea |  |  |  | 2 |  |  |  |  |

In addition to the cleaned raw dataset, a spreadsheet detailing the flowering periods of each of the 80 species in the Stafford dataset is required. Flowering period (FLOWPER) data are not provided within the trait dataset due to geographical differences and such data needs to be collated from relevant Floras and imported into R. A spreadsheet was constructed using flowering data from a UK Flora (Clapham et al. 1987) supplemented by data from a German Flora (Rothmaler 1995), containing a column of species codes and a column with the flowering duration in months (Table 3, supplementary data 6). Note that FLOWPER is only needed for model 1 and model 3. For composite taxa an averaged flowering period of the multiple species was entered.

Table 3. The format of the second spreadsheet required: flowering periods. The first 12 species from Stafford are shown with composite species in bold, species missing trait data underlined.

|  |  |
| --- | --- |
| species.codes | FLOWPER |
| agrogit | 2 |
| anetgra | 3 |
| anthcot | 5 |
| arrhels | 2 |
| avenstfa | **3** |
| brasnig | 4 |
| bromhose | **2** |
| buglarv | 4 |
| buplrot | 2 |
| careflv | 5 |
| carenig | 2 |
| carepna | 2 |

For the Stafford dataset composite taxa were created when an item could have derived from one of two or three species. Before constructing the averaged functional traits of the composite species, **weed\_data** was used to examine the species, confirming that their functional traits were not widely divergent. For species which have very different growing conditions or functional traits it is suggested that the models be run with and without the composite taxa to see the impact of their inclusion.

To obtain the composite taxa’s functional trait data the function **ave\_wdata** was used. The Stafford data contained 12 composite taxa (Table 4) and **ave\_wdata** was used to average the different species’ functional trait values, creating an averaged SLA, ARNODE, LOGCANH, LOGCAND and VEGPROP for each composite taxon. For example, using ave\_wdata the functional trait values of the species Ranunculus acris (ranuacr), Ranunculus bulbosus (ranubul) and Ranunculus repens (ranurep) were averaged together to create the composite taxon ranuabr. This was done for all composite species to create a data frame containing those taxa and their averaged functional trait data (supplementary code line 8; Figure 2a).

Table 4 The new species codes for the composite taxa from Stafford and their included species

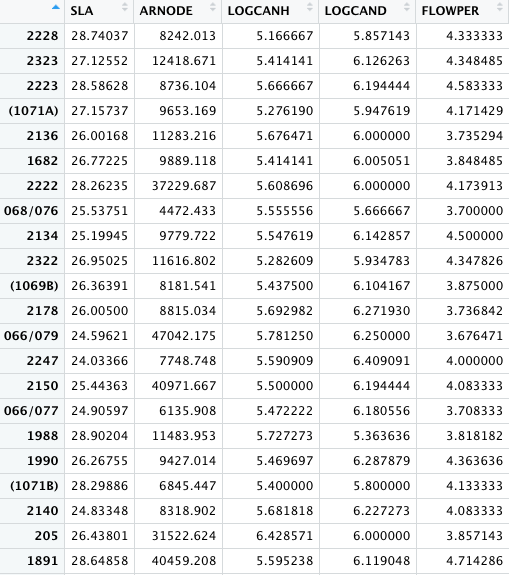
|  |  |  |  |
| --- | --- | --- | --- |
| Species code | Species 1 | Species 2 | Species 3 |
| avenstfa | avenste | avenfat |  |
| bromhose | aromsec | bromhor |  |
| galetesp | galetet | galespe |  |
| eleounpa | eleouni | eleopal |  |
| galiapsp | galiapa | galispu |  |
| juncefco | Junceff | junccon |  |
| polypela | polyper | polylap |  |
| ranuabr | ranuacr | ranubul | ranurep |
| ranurefl | ranurpt | ranufla |  |
| stelpagr | stelplu | stelgra |  |
| veropoag | veropol | veroagr |  |
| vicihite | vicitet | vicihir |  |

*Carex flava* was identified within the Stafford assemblages, but trait data are not available for this species within the released dataset. There are two options when this occurs: exclude this species during data cleaning or manually add data from alternative sources. As we have data for this species from alternative sources it was added to the composite taxa data frame as a new row (supplementary code line 26; Figure 2b). The data from *Carex flava* came from partial trait data for that species already collected as well as data from alternative databases (REF?).

At this point three datasets have been created: the cleaned raw Stafford data (DF1), the composite taxa trait values (including the additional *Carex flava data*) (DF2) and the flowering period data for all 80 of Stafford’s weed taxa (FLOWPER). As shown in supplementary file 4’s flow diagram, DF1 and FLOWPER can be created outside R, while DF2 requires R and the functions within the WeedEco package.

The next steps within R combine the three created datasets in the correct format for linear discriminant analysis. The **wdata\_org** function combines the three datasets, FLOWPER, composite taxa (DF2) and the cleaned dataset (DF1), to produce a new data frame that contains the averaged functional trait values of the taxa within each archaeobotanical samples from Stafford (Table 5, supplementary file 7). Each species code in the Stafford data (DF1) is used by **wdata\_org** to extract the functional trait data for the species from the trait database. It is then averaged with the other species trait values within that sample. The resultant data frame produced was saved as an R object for use when conducting discriminant analysis (supplementary code 30).

Table 5. the averaged functional trait data of the species within the archaeobotanical samples from Stafford. Note this is output of wmodel\_org when the model assigned is model 1.



As explained above, WeedEco’ linear discriminant analysis function **wdata.LDA** takes the output of **wdata\_org** and conducts discriminant analysis – in particular the classification stage on the averaged trait values of the samples. The results of running the Stafford data though **wdata.LDA** is shown in Table 6, which gives the classification (group 1 vs group 2)(Class\_std\*), the posterior probability of the sample being in group 1 (Prob.1\_std\*) or 2 (Prob.2\_std\*), and the linear discriminant scores (LD1\*) of each archaeobotanical sample. The full results of the analysis can be view when the wdata.LDA is save as a data frame and will shows both standardised and unstandardised LDA scores, classifications and probability (see the WeedEco help document or Venables and Ripley (2002) for full details). The unstandardised linear discriminant scores (LD1\*) are used in the plotting functions and is what is graphically represented in Hamerow et al. 2020. Interpretation of the results is done visually by plotting the linear discriminant scores against the discriminant scores of the model’s samples or against the model’s centroids. The results of the Stafford discriminant analysis were saved as a data frame to be used in the plotting functions (supplementary code line 33).

The different plotting functions in WeedEco provide a range of options as explained above and shown in Figure 1. **wplot\_basic** was used for the Stafford data and shows that the majority of the Stafford samples plot around the group 2 centroid, indicating that the samples more closely resemble the Haute Provence data, indicating low-input cultivation with relatively low fertility and disturbance, and consistent with the hypothesis of extensification. To present the data in the same way as Hamerow et al. (2020) **wplot\_phase** was used, with the data separated by four broad phases (Figure 3). A column delineating the phasing of all the samples was added to the output of **wdata.LDA** (supplementary code line 38). Each phase was then graphed on separate subgraphs, each with different colours or symbols (gcol, gbg and gpch, supplementary code line 45). By separating the results into broad phases, change over time can be examined, with the Stafford data showing that the earliest phase has variable scores while from the mid tenth century onwards the samples are more regularly around the centroid of group 2. This indicates a trend towards extensification over time at Stafford, with little to no manuring or hand weeding from the mid tenth century onwards (see Hamerow et al. 2020 for full details).

#### Models 2 and 3

The WeedEco functions can also be used to compare archaeobotanical data against model 2 and model 3, with only a few differences in usage. For model 2 **wdata\_org** only requires the raw archaeobotanical datasheet and the composite taxa data frame. All other functions are used in the same way as model 1, except for entering model 2 whenever the model number is required.

Model 3 differs from the two other models as it only uses two functional attributes: FLOWPER and VEGPROP. As this model is applicable to temperate datasets, the Stafford data will be compared to model 3 to show the usage. Bogaard et al. (2022) conducted the same analysis using SPSS. The data cleaning steps are the same as with model 1, and if already completed can follow on from model 1. Differences between model 1 and model 3 are due to the functional trait VEGPROP.

When averaging composite taxa **ave\_wdata** is again used to produce the dataset, but for the Stafford taxa there were uncertainties as to the VEGPROP value of composite taxa ranuabr and ranurefl. It was not certain that the items within those two composite taxa necessarily came from perennial species, nor that they displayed vegetative propagation, so the VEGPROP value of these two composite taxa were modified to show no data, excluding them from any VEGPROP calculations (supplementary code line 51).

It is recommended that when using model 3 users confirm their understanding of the life history of each species and whether the typical life history of that species within their geographical region is reflected (see flow diagram section X supplementary data). Using the **weed\_data** function, species found within the Stafford data were entered and the trait “VEGPROP” selected. This produced a table of the vegetative propagation ability of the entered species, as well their life history (Table 7a). In the Stafford data there was uncertainty as to whether *Silene nutans* (silenut) should be counted within the VEGPROP calculations – in the dataset it shows as a perennial and therefore a 0 within the VEGPROP column (Table 7b). To remove it from any averaged calculation of VEGPROP the silenut’s VEGPROP value was changed from a 0 to an NA (supplementary code, line 60). Note that it is the VEGPROP value, not the LIFEHIST value, that is changed as it is the VEGPROP trait which is directly used in model 3. The altered data frame was then entered into the **wdata\_org** function using the argument vg\_pr, along with the other required datasets, raw data (df1), composite species (DF2) and flowering period, as well as the other associated instructions (model 3 etc).

The output of **wdata\_org** was then analysed with **wmodel.LDA** using the model argument of model 3. This produced the linear discriminant result, which, as with model 1, can then be plotted using one of the functions. For example, **wplot\_geog** produces Figure 4a which can be further modified including changing the symbols, colour, and axis limit. One helpful aspect of the plotting functions is that the colours and symbols of different subsets of the data can be changed – allowing the visualisation of multiple phases on one graph (Figure 4b). However, to more easily see change through time **wplot\_phase** allows comparison of separate phases within sub graphs (see Figure 4c). By calling on the newly created phase column each phase is plotted as a subplot. To make the symbols the same as Bogaard et al. (2022)(Figure 4d), the gpch, gcol and gbg arguments need to be modified from the defaults (supplementary code, line 99) (Figure 4c).

Comparing model 3 to the Stafford data shows a tendency through time for the samples to occur more towards centroid 1, the high disturbance end of the graph, indicating that the samples increasingly resemble the conditions of the modern arable fields. Figure 5c shows that from the 12th century onwards no samples resemble the undisturbed model samples. Before the 12th century there are samples which are similar to the undisturbed model samples and may represent the arable-grassland interface or the cultivation of land normally used as pasture (Bogaard et al. 2022). Consistent and comprehensive disturbance seen in the 12th century samples has been interpreted as more the systematic use of the mouldboard plough (Bogaard et al. 2022).

Discussion

All previously published functional weed ecology analysis has been conducted in SPSS but to enable accessibility, an equivalent process in R was perceived as important when these data were released. In the transition from SPSS to R small differences were noted in how each statistical program ran linear discriminant analysis. The primary difference is that, due to being purely arbitrary, the positive and negative signs for the group 1 and group 2 were reversed between SPSS and R. To make the graphs similar to the published output from SPSS, the R package WeedEco formally made all group 1 linear discriminant values positive and all group 2 linear discriminant values negative. This should be noted should the raw model data be run in alternative statistical programs.

It is strongly recommended that the version of the R package, R and trait dataset are explicitly detailed within the method section of any articles. Furthermore, users should indicate if R is not used for the linear discriminant analysis as some statistical programs have slight variations in how they are conducted. This does not prevent alternative statistical programs being used but for ease of comparison between different archaeobotanists’ results, explicitly stating what has been used in the methods is best practice. This leaves a path for further updates to the R package, as well as the functional trait dataset, facilitating traceability. Furthermore, the sources of the functional trait data used within the models needs to be referenced….

MORE?

### Conclusion

The released functional trait data, model data and R package provide a resource allowing archaeobotanists to investigate the crop husbandry regimes and practices represented within the weed flora of crops. Model 1 and model 2 provide an understanding of the degree of high or low input cultivation conducted – fertility and disturbance for model 1 and fertility for model 2. Model three allows users to disentangle fertility and disturbance, examining only the nature of disturbance within crop fields.

References:

Alagich, R., Gardeisen, A., Alonso, N., Rovira, N. and Bogaard, A., 2018. Using stable isotopes and functional weed ecology to explore social differences in early urban contexts: The case of Lattara in mediterranean France. *Journal of Archaeological Science*, *93*, pp.135-149.

Bogaard, A., Hodgson, J., Nitsch, E., Jones, G., Styring, A., Diffey, C., Pouncett, J., Herbig, C., Charles, M., Ertuğ, F. and Tugay, O., 2016. Combining functional weed ecology and crop stable isotope ratios to identify cultivation intensity: a comparison of cereal production regimes in Haute Provence, France and Asturias, Spain. *Vegetation History and Archaeobotany*, *25*(1), pp.57-73.

Bogaard, A., Styring, A., Ater, M., Hmimsa, Y., Green, L., Stroud, E., Whitlam, J., Diffey, C., Nitsch, E., Charles, M. and Jones, G., 2018. From traditional farming in Morocco to early urban agroecology in northern Mesopotamia: combining present-day arable weed surveys and crop isotope analysis to reconstruct past agrosystems in (semi-) arid regions. *Environmental Archaeology*, *23*(4), pp.303-322.

Bogaard , A., Hodgson, J., Kropp, C., McKerracher, M. and Stroud, E., 2022. Lessons from Laxton, Highgrove and Lorsch: Building arable weed-based models for the investigation of Early Medieval agriculture in England. In McKerracher, M. and Hamerow, H., (eds). *New Perspectives on the Medieval ‘Agricultural Revolution’: Crop, Stock and Furrow*. Liverpool University Press p 25

Clapham, A. R., D. M. Moore, and T. G. Tutin. 1987 Flora of the British Isles. 3rd Ed. Cambridge: Cambridge UP

Charles, M., Bogaard, A., Jones, G., Hodgson, J. and Halstead, P., 2002. Towards the archaeobotanical identification of intensive cereal cultivation: present-day ecological investigation in the mountains of Asturias, northwest Spain. *Vegetation History and Archaeobotany*, *11*(1), pp.133-142.

Carver, M.O.H., 2010. *The Birth of a Borough: an archaeological study of Anglo-Saxon Stafford*. Woodbridge: Boydell Press.

Diffey, C., Neef, R., Seeher, J. and Bogaard, A., 2020. The agroecology of an early state: New results from Hattusha. *Antiquity*, *94*(377), pp.1204-1223.

Dodd, A., Goodwin, J., Griffiths, S., Norton, A., Poole, C. & Teague, S. 2014. Excavations at Tipping Street, Stafford, 2009–10: Possible Iron Age Roundhouses, Three Stafford-type Ware Kilns, and Medieval and Post-Medieval Urban Remains. Staffordshire Archaeological and Historical Society Transactions, 47: 1–114

Green, L., Charles, M. and Bogaard, A., 2018. Exploring the agroecology of Neolithic Çatalhöyük, Central Anatolia: an archaeobotanical approach to agricultural intensity based on functional ecological analysis of arable weed flora. *Paléorient*, *44*(2), pp.29-44.

Hamerow, H., Zerl, T., Stroud, E. and Bogaard, A., 2021. The cerealisation of the Rhineland: Extensification, crop rotation and the medieval ‘agricultural revolution’in the longue durée. *Germania: Anzeiger der Römisch-Germanischen Kommission des Deutschen Archäologischen Instituts*, pp.157-184.

Hamerow, H., Bogaard, A., Charles, M., Forster, E., Holmes, M., McKerracher, M., Neil, S., Ramsey, C. B., Stroud, E. and Thomas, R. (2020) “An Integrated Bioarchaeological Approach to the Medieval ‘Agricultural Revolution’: A Case Study from Stafford, England, c.ad 800–1200,” *European Journal of Archaeology*. Cambridge University Press, 23(4), pp. 585–609. doi: 10.1017/eaa.2020.6.

Jones, G., Bogaard, A., Halstead, P., Charles, M. and Smith, H., 1999. Identifying the intensity of crop husbandry practices on the basis of weed floras1. *Annual of the British School at Athens*, *94*, pp.167-189.

Jones, G., Bogaard, A., Charles, M. and Hodgson, J.G., 2000. Distinguishing the effects of agricultural practices relating to fertility and disturbance: a functional ecological approach in archaeobotany. *Journal of Archaeological Science*, *27*(11), pp.1073-1084.

Maltas, T., Şahoğlu, V., Erkanal, H. and Tuncel, R., 2022. From horticulture to agriculture: New data on farming practices in Late Chalcolithic western Anatolia. *Journal of Archaeological Science: Reports*, *43*, p.103482.

Rothmaler, W. 1995. *Exkursionsflora von Deutschland 3*. Stuttgart, Gustav Fischer Verlag.

Stroud, E., Bogaard, A. and Charles, M., 2021. A stable isotope and functional weed ecology investigation into Chalcolithic cultivation practices in Central Anatolia: Çatalhöyük, Çamlıbel Tarlası and Kuruçay. *Journal of Archaeological Science: Reports*, *38*, p.103010.

Styring, A., Rösch, M., Stephan, E., Stika, H.P., Fischer, E., Sillmann, M. and Bogaard, A., 2017, December. Centralisation and long-term change in farming regimes: comparing agricultural practices in Neolithic and Iron Age south-west Germany. In *Proceedings of the Prehistoric Society* (Vol. 83, pp. 357-381). Cambridge University Press.

Venables, W.N. and Ripley, B.D., 2013. *Modern applied statistics with S-PLUS*. Springer Science & Business Media.

1. R uses a wide range of data formats. The package WeedEco was written assuming the data will be entered as a data frame – which is a table-like two-dimensional array in which there are columns with an equal number of rows [↑](#footnote-ref-1)