July 2018 v. 1.4

A short guide to using British and Irish plant occurrence data for research

O.L. Pescott[1], T.A. Humphrey[2], & K.J. Walker[3]

1. Biological Records Centre, CEH Wallingford, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK; olipes@ceh.ac.uk
2. Botanical Society of Britain and Ireland c/o CEH Wallingford, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK; tom.humphrey@bsbi.org
3. Botanical Society of Britain and Ireland, Room 14, Bridge House, 1-2 Station Bridge, Harrogate, UK; kevin.walker@bsbi.org

The following is a check list of things to ask when using British and Irish plant occurrence data for analyses. Note that many of the issues discussed here may also apply to datasets of animal occurrence data, and to data from other territories.

This document has been prompted by several recent incidences of poor practice in the analysis of BRC/BSBI plant data in published papers and reports, and is intended to make researchers stop and reflect on what they might be doing by plugging datasets into analyses without a full understanding of both data structures and the underlying biological (and recording scheme-related) information contained therein.

In general, those undertaking analyses without a knowledge of the databases they are analysing should consult those who do understand them before undertaking the analysis; they should also ask an expert to check their conclusions once an analysis is nearing completion, or at appropriate points along the way. All three authors are available to advise on issues relating to the vascular plant and bryophyte databases discussed in this document.

Note that we do not cover every possible error or assumption that an uninformed user might commit or make in analysing species occurrence data originating from biological records. For example, issues such as developing surrogates for abundance, or the expectation that biological records will show no spatial or temporal bias, are considered either too involved or too simplistic, respectively, to merit space here.

1. Which dataset should I choose for my analysis?

For vascular plant (and charophyte) data there are (at least) four options to consider: the entire BRC dataset (known as the VPDB); the *New Atlas* (Preston et al., 2002) dataset (which excludes charophytes); the BSBI 'Distribution Database' (DDb); and, the NBN (and related '*Atlas of Living* –' websites).

The VPDB is the historic home for BSBI data, and contains plant records collated up to the late-2000s. The *New Atlas* dataset is the subset of the BRC VPDB that was used in the production of the *New Atlas of the British and Irish Flora* (Preston et al., 2002); although this is outdated (both in terms of changes to records since that date and in terms of temporal coverage), these data were checked as carefully as possible, distinguish native/alien (10 × 10 km) occurrences, and are relatable to a published source. It may therefore be a better resource for studies where up-to-date information is not essential, but rigour and accuracy are paramount. The *New Atlas* tables are available separately from the general VPDB tables within the BRC Oracle database.

Since the late-2000s, most new plant data have gone directly to the BSBI DDb. The DDb should contain all of the data that the VPDB contains, but much new data as well. Newer data in the DDb

Cite as: Pescott, O.L., Humphrey, T.A. & Walker, K.J. 2018. *A short guide to using British and Irish plant occurrence data for research. v. 1.4.* Unpublished report. CEH and BSBI: Wallingford.

have not been subject to such rigorous quality checks as the VPDB (which has undergone extensive checks for various projects over the years), but BSBI vice-county recorders do oversee the verification and acceptance of data in the DDb. The DDb will be the main home for (BSBI) plant data in the coming years.

The NBN Living Atlas (NBNLA; https://nbnatlas.org/) hosts datasets from numerous organisations, including the BSBI; however, at the current time where the BSBI dataset is concerned, only data for Scotland are available. It is likely that increasing amounts of high resolution vascular plant data will be added to the NBNLA for other countries in the near future. The NBNLA includes plant datasets from a wide range of sources, with data of varying degrees of quality. Using data from the NBNLA may make it harder to deal with some of the issues described under point (**5**) below.

Note that considerable amounts of older BRC/BSBI plant data are available through all three databases – however, there may be differences in species names and definitions due to the use of different species dictionaries and concepts between the different databases (this is particularly the case for the NBNLA versus the VPDB and DDb). Again, see (**5**) below for more on this topic.

2. What time scale(s) should I use, and why might this affect data availability?

Data in all three databases have associated dates – however, an important point to grasp is that many older records will not be resolved to a single day. All three databases allow records to have start dates and end dates. For a record resolved to a single day, the start date will equal the end date. For a record known to have been made during a particular month, the start date and end dates will be the first and last days of the month. For a record known to have been made during a particular year, the start date and end dates will be the first and last days of the year etc.

Of particular significance for BSBI plant data are so-called ‘date classes’, which relate to periods of time (on the order of years) into which records are grouped to reflect periods of recording for national atlases. Many historic plant records are only resolved to these date classes. For the BSBI data these are: before 1930; 1930-1969; 1970-1986; 1987-1999; 2000-2009; and 2010-2019. For example, the date class that covered fieldwork for the BSBI *New Atlas* (Preston et al., 2002) was 1987-1999. For reasons of both contemporary resourcing and data-basing, many records for this period have a start date of 01/01/1987 and an end date of 31/12/1999. Depending on how one subsets (or extracts) plant data for analyses, you may therefore end up excluding large amounts of data from your analyses. For example, contrasting plant data for 1960-1980 with 1990-2010 would mean ignoring all of the data resolved to the 1987-1999 date class, because this straddles the second period. Likewise, analyses that attempt to derive annual or decadal indices will ignore large volumes of data within older date classes. This issue appears to affect Wales and Scotland far more than England, due to the fact that the historic lower recording effort in these areas has meant that atlas projects, with their concomitant bursts of data entry and broad temporal resolutions, have provided most of the older data.

Note that is easy to make mistakes with dates if one does not keep the data structure in mind – for example, for the 1960-1980 v. 1990-2010 comparison mentioned above, taking an end date as the actual date of a record would mean that the broad resolution 1987-1999 data would be included in the second period, an assumption which will not be justified for thousands of records.

Related to this issue is the fact that, for these broad date classes, it will usually be the rarer species that have the more highly resolved records (both in space and time). So, by choosing time periods that drop large amounts of broadly resolved data, you are likely to end up biasing your analyses in ways that you will have little hope of understanding. Recall that Frescalo (Hill, 2012) works by estimating species relative frequency curves for multiple neighbourhoods of 100 ten × ten km squares – is it likely that these will be well estimated for time periods that drop the most common species?

Some older records have been digitised with only the end-date of the range specified. This was often done where the precise dates of records were unknown. Date estimates have also sometimes been based on the date of publication, date of death of the recorder, or the date when the record was digitised. The assigned date may be therefore sometimes be highly misleading. For most date-sensitive analyses, records with an incomplete or excessively broad date range should probably be excluded, but see Hill & Preston (2015) for an example where seemingly odd-looking time periods have been chosen to get around such an issue. Analyses in Preston et al. (2002) are also based on BSBI date classes. Carefully chosen sensitivity analyses could also be performed to test whether conclusions are the result of data quirks relating to broad date classes.

Note also that analyses using more finely resolved spatial data (e.g. 2 × 2 km records) may have fewer records resolved to broad date classes at the national level – this is likely to be due to the fact that historic data resolved to this finer spatial scale will probably have been entered separately from records made during a period of intense atlas recording (see Groom, 2013 for one example of the use of tetrad data). However, at the vice-county level, tetrad data may also be resolved to broad data classes if these were used for historic county floras – such issues may then lead to uneven spatial coverage at fine spatial scales if the time periods used in analyses exclude tranches of data from particular vice-counties. Pescott et al. (2015) discuss the BSBI Monitoring Scheme (Rich and Woodruff, 1996)/Local Change (Braithwaite et al., 2006) projects which have contributed much of the older tetrad data in the vascular plant databases discussed here.

3. What spatial resolution should I use, and why might this affect data availability?

As stated above, records of rarer species, or records made outside of a period of intense recording for an atlas, are likely to be at a finer spatial scale (i.e. finer spatial grain size) than historic atlas records. This is changing, and the vast majority of new records are resolved to the 2 × 2 km (tetrad) or 1 × 1 km (monad) level. The DDb will therefore have many more records at a finer grain size than the VPDB.

Since the inauguration of grid-based (lattice) data collection for atlasing in the 1960s, the hectad (10 × 10 km) grid has been the unit most often in recorders' minds in terms of making lists of species; this has now been largely replaced by the tetrad. For example, recent BSBI guidance (Walker et al., 2010) on recording for an atlas for 2020 focuses on recording at the tetrad scale, although it is also emphasised that records are perhaps best (and most easily) made at the monad scale (given that this is the finest grid on OS 1:25,000 and 1:50,000 leisure mapping). Even where the monad is the targeted sampling unit, some recorders may collect data at the tetrad level to save time in the field, so that species already observed on a previous visit are not re-recorded (i.e. the final list is collated at the tetrad level, despite repeated visits to subunits of the area). This means that reducing datasets to only include monad records will result in the loss of considerable amounts of occurrence data; in addition, reducing datasets to the monad level may result in odd inferences where Bayesian occupancy modelling is utilised: If a recorder does most of their (tetrad level) recording in one of the four monads within a tetrad, but then makes shorter lists (of additional species only) in bordering monads, metrics of 'list lengths' at the monad level, sometimes used as indices of recording effort (Isaac et al., 2014; van Strien et al., 2013), will not actually contain much information on detectability. Note then, that the assumptions of modellers and the cultural practices of recording at any one time and place may be at odds in ways that are not clearly evident from the data themselves. This may be increasingly evident in the future as more metadata are collected from recorders.

In general terms, analyses at the tetrad and monad levels will also limit the temporal and spatial (and, by the same token, taxonomic/habitat) breadth of any analyses, or, to put it another way, using these data will introduce significant biases to an analysis that will need to be considered or accounted for.

This is because tetrad/monad surveys have generally taken place in lowland regions with high population densities; furthermore, very few counties have been resurveyed at these scales. Various modelling approaches are being developed to address these variations in recording effort/detectability but, as yet, it is far from clear how effective they are at detecting real change.

For many data sets it will be important to assess whether the records are likely to have been georeferenced accurately at appropriate precision. There are several widespread sources of geo-referencing error that may distort the analysis of gridded plant records. A frequent mistake by recorders or digitisers is to assign a large set of records to an arbitrary centroid point, related to a named place or to the centre of a survey. The can give the records a false precision and masks the extent of the recording area. A related problem is that records from historical sources (literature reports or herbarium specimens) often provide only vague locality information but are often later geo-referenced at spurious precision.

A second geo-referencing problem that may be peculiar to gridded occurrence data from the British Isles is the misrepresentation of tetrad or quadrant (5 × 5 km) data using 1 km precision grid-references (often specifying the south-west corner of the larger grid-unit). This type of error is readily apparent if the mapped monad data appear to show a checkerboard pattern. This error is related to the practice of representing grid-based data as points (e.g. centroids or corners) with attached uncertainty; this is standard practice on GBIF, and, unfortunately, now also on the NBNLA. Again, checkerboard patterns of highly precise species occurrence data on GBIF or the NBNLA will almost certainly relate to grid-based data that have been spuriously reduced to points. Note that in these cases the associated precision or uncertainty information (if available) should not be interpreted as a radius; if it is, then the resulting circle will exclude some areas that could have been covered by the original (grid-based) record, and include others that could not have been. Re-gridding such point data to a different projected coordinate system than the one originally used will lead to similar errors of omission and commission (see Fielder et al., 2015 for an example of such an error).

4. What about status? Natives and aliens

The decision whether to deal with native species only, native plus alien, or aliens only, may seem straightforward – it is not quite as simple as this. First, recognise that there are two categories of alien, archaeophyte and neophyte (see Hill et al., 2004 for definitions). So, the choice becomes native taxa, native and archaeophytes, archaeophytes and neophytes, or just neophytes. Note also that for some species there is doubt about native status (Hill et al., 2004), and that not all authors agree on the list of species about which there is doubt (Stace and Crawley, 2015, their Appendix 1).

If native taxa are chosen, recognise that even for these there is a division between occurrences (at the 10 × 10 km level) that are considered to be native and those that are thought to be recent introductions (Preston et al., 2002). This information is recorded in both the VPDB, the *New Atlas* database, and the BSBI DDb. The BSBI DDb contains additional information on hectad statuses for native taxa that were decided upon during the recent English Red Listing exercise (Stroh et al., 2014), and more recently as part of the verification process for Atlas 2020 which is currently in preparation. It may be relevant to your project to consider whether the datasets you are using take a consistent approach to status, both at the species level and at the record (i.e. occurrence) level. See Preston (2002) for an interesting review of the topic of alien species in the UK.

A related issue, which overlaps in part with question (**5**) below, is the actual choice of species. Particularly for studies using multi-species information in analyses (e.g. measures of richness, Frescalo, classification or ordination algorithms), the question of which species are used may be very important: e.g. only ‘well-defined’ species which have long been recognised (i.e. those without any subtle taxonomic issues relating to them – those known from Britain by 1850 has been used as one criterion (Hill and Preston, 2015), or all species except microspecies of apomictic genera (in *Hierarcium*, *Rubus*, *Taraxacum* and *Sorbus*, for example)). Whether to include or exclude hybrids is another question of importance (Preston and Pearman, 2015), as is the inclusion of subspecies or aggregates where the recording of a particular taxon has not been consistent in the past (see **5** below).

Finally, the culture of recording aliens has changed significantly through time: prior to the publication of the 1st edition of Stace's *New Flora* in 1991 the recording of alien species was much less even than today (e.g. see contemporary reviews of Stace, 1991). Apparent increases in the spread of aliens may then have as much to do with recording practice as reality in some cases. Whether or not such changes invalidate analyses of alien spread is no doubt something that must be evaluated on a case-by-case basis; we leave it to the reader to consider whether aggregate analyses across alien taxa are therefore likely to be reliable (e.g. Doxford and Freckleton, 2012).

5. What do I need to understand about the hierarchical structure of the species data? Or, what's the difference between a species and species aggregate?

a. General points (VPDB and DDb)

Nomenclature and taxonomy can be esoteric and bewildering at the best of times, and, unfortunately, the data-basing of biological records adds an additional layer of complexity to the topic. Readers who want a more complete overview of the topic and its implications for analysis are referred to Franz et al. (2008) and Jansen & Dengler (2010).

There are two key things to understand:

(i) Individual biological records can be assigned to different levels in a taxonomic hierarchy; and,

(ii) the taxonomic hierarchy in a database is not limited to those categories (i.e. species, subspecies, varieties etc.) normally recognised by taxonomists, but is also likely to include 'database concepts'[1]. These are a way of indicating that a particular (historic) observation was made according to a species concept that has been superseded, or was originally made with a degree of precision lower than that of the level of species (that is, the observation could represent one of several species as currently recognised).

Both of these points are important, and both can, and have, led to errors in data compilation and analysis. The implications of the two points above are expanded with examples below. Note that the two points overlap: the separation is slightly artificial, and was done to try to separate out 'standard' features of taxonomy from those that apply especially to databases of biological records (which, as noted above, attempt to deal with the changing interpretation of names over time, as well as field identifications which may contain ambiguities in relation to any particular taxonomic list).

(i) Individual biological records can be assigned to different levels in a taxonomic hierarchy

This means that records can be assigned to species, subspecies, varieties etc. However, some analyses will want to count (or model) species (or some close approximation), rather than the sum of all the specific and infraspecific names made by recorders during a time period.

---

[1] Note that the use of the word 'concept' in this context, and in what follows, is not synonymous with the use of the word in a taxonomic context. That is to say, a 'species concept' in taxonomic and evolutionary terms is a theory on what should be represented by designating a group of individuals as a species. The classic example being the biological species concept (BSC) of Ernst Mayr. This states that species in general are defined by their ability to interbreed, and therefore form an evolutionary unit. Species concepts in the BRC database may represent the embodiment of an application of some theoretical species concept such as the BSC, but there is no guarantee that this is the case. See Knapp (2008) for a discursive, if slightly inconclusive, overview of this area, and Avise (2000) for an entertaining and acerbic look at phylogenetic species concepts versus the BSC.

One recent analysis took data directly from the bryophyte tables in the BRC database, and produced SDMs for each name – this meant that for a species containing several varieties, species distribution models (SDMs) were produced which (a) missed relevant data relating to infraspecific taxa when modelling at the species level and (b) produced SDMs for varieties for which there was a distinct lack of information (because varieties are not, generally, as consistently recorded as species). This produced a contradiction in the results, because the SDMs were radically different between a species and its most common variety, even where the distribution of the common variety is known to match that of the species very closely, and when the SDMs would therefore be expected to be very similar (see Pearce-Higgins et al., 2017, 2015).

The solution to this problem depends on which database is being used. The VPDB and the bryophyte tables in the BRC database both have 'aggregate' tables, which specify the hierarchical relationship between taxa – this allows one to perform an SQL join on the aggregate table, aggregating records below the species level to that level for analysis. Analyses below the species level should only be performed if one is confident that the taxon is likely to be well-sampled (or if one has a plan for dealing with this in analyses). For the BSBI DDb, the web interface allows one to 'group by species', which is again performing the relevant SQL operation in the background.

Note that for some analyses, for example Bayesian occupancy modelling, where 'list length' is a desired parameter to derive from recording data, then aggregating records at different stages of the analysis may produce different list lengths. For example, aggregating before any analyses are performed will produce shorter list lengths because infraspecific taxa will have been aggregated to the species level.

(ii) The taxonomic hierarchy in a database is not limited to those categories (i.e. species, subspecies, varieties etc.) recognised by taxonomists, but is also likely to include 'database concepts'

This issue can be rather bewildering for the beginner (e.g. see Franz et al., 2008), but is of particular importance for studies of biodiversity that seek to compare time periods. The recording of particular species may have changed over time – that is, the exact meaning of names may have changed. Database managers, and national recording schemes and societies, try and manage this by keeping the organising taxonomy and database concepts up-to-date (e.g. maintaining the taxon register and the aggregates table in the BRC databases). This may also involve re-determining records in the database – for example, records in the BRC databases point to the relevant (species or aggregate) concept, so, a record of a species that becomes ambiguous due to a change in taxonomy (e.g. a species being split), will be updated so that the record now points to a new concept, which will be an aggregate concept (unless the record is supported by an actual specimen and can be assigned to one of the new segregated species).

In the BRC databases aggregate concepts are labelled as 'species aggregates' in the 'RANK' field of the taxon register. For the BSBI DDb, the relevant aggregate names need to be specified directly in any query that wants to use some species aggregates alongside actual species names when extracting data from the DDb (using the 'group by [supplied] names' option). The alternative is to aggregate data subsequent to extracting information from a database.

Errors have been made in analyses by dropping 'species aggregates' (including the microspecies mentioned above under point (**4**)) – this is very undesirable, because species within these aggregates may show patterns of change over time that reflect recording effort and focus, rather than true change. For any analysis you should consult with experts to determine which aggregates are likely to be appropriate for the time period(s) that you seek to analyse or compare.

b. The NBNLA

The same issues apply as for the two databases above, with the exception that we are currently not aware that it is easy to produce automated exports from the NBNLA that aggregate records to any particular level. Perhaps it is possible to aggregate to the species level using the new *Atlas of Living* - websites, but we have not investigated this. Even if this is possible, or becomes possible, the issue of aggregation above the species level is still relevant, and would presumably require manipulation of datasets after retrieval. The UK Species Inventory (i.e. the taxonomy and database concepts used by the NBN) may be used to assist with this processing step – note, however, that the UKSI is not currently identical to the taxonomy and database concepts used within the BSBI DDb or the BRC VPDB, although attempts to reconcile these species lists are currently nearing completion.

Final remarks

Much time can be wasted through a lack of consideration of these issues at an early stage of analysis; just as a statisticians frequently remark that they are most often called upon to diagnose the cause of death of an experiment, so might a taxonomist or field ecologist comment on the zombie status of numerous papers in the literature that fail to adequately state what choices were made in assembling biodiversity data to answer an ecological question of interest. Considering the choices, and clearly stating decisions, will remove the need to redo analyses, and improve the quality (including the reproducibility) of your science.

We hope that increasing understanding of the structure and quirks of the datasets being used by researchers will lead to more confidence in conclusions. It bears repeating that those undertaking analyses without a knowledge of the databases they are analysing should consult those who do understand them before starting, and again ask an expert to check their conclusions once the analyses are completed.

Acknowledgements

We thank Chris Preston for comments on, and contributions to, this guidance. Chris is also to be thanked for initially suggesting that we write such a guide. OP's contribution to this work was supported by the Natural Environment Research Council award number NE/R016429/1 as part of the UK-SCaPE programme delivering National Capability.

References

Avise, J.C., 2000. Cladists in wonderland. Evolution 54, 1828–1832.

Braithwaite, M.E., Ellis, R.W., Preston, C.D., 2006. Change in the British flora 1987-2004. Botanical Society of the British Isles, London.

Doxford, S.W., Freckleton, R.P., 2012. Changes in the large-scale distribution of plants: extinction, colonisation and the effects of climate. J. Ecol. 100, 519–529.

Fielder, H., Burrows, C., Woodman, J., Ford-Lloyd, B.V., Maxted, N., 2015. Enhancing the conservation of crop wild relatives in Wales. New J. Bot. 5, 178–191.

Franz, N., Peet, R.K., Weakley, A.S., 2008. On the Use of Taxonomic Concepts in Support of Biodiversity Research and Taxonomy, in: Wheeler, Q.D. (Ed.), The New Taxonomy, Systematics Association Special Volume. CRC Press, Boca Raton, FL.

Groom, Q.J., 2013. Some poleward movement of British native vascular plants is occurring, but the fingerprint of climate change is not evident. PeerJ 1, e77. https://doi.org/10.7717/peerj.77

Hill, M.O., 2012. Local frequency as a key to interpreting species occurrence data when recording effort is not known. Methods Ecol. Evol. 3, 195–205.

Hill, M.O., Preston, C.D., 2015. Disappearance of boreal plants in southern Britain: habitat loss or climate change? Biol. J. Linn. Soc. 115, 598–610. https://doi.org/10.1111/bij.12500

Hill, M.O., Preston, C.D., Roy, D.B., 2004. PLANTATT- attributes of British and Irish plants: status, size, life history, geography and habitats. Centre for Ecology & Hydrology.

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P., Roy, D.B., 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. Methods Ecol. Evol. 5, 1052–1060. https://doi.org/10.1111/2041-210X.12254

Jansen, F., Dengler, J., 2010. Plant names in vegetation databases–a neglected source of bias. J. Veg. Sci. 21, 1179–1186.

Knapp, S., 2008. Species concepts and floras: what are species for? Biol. J. Linn. Soc. 95, 17–25. https://doi.org/10.1111/j.1095-8312.2008.01090.x

Pearce-Higgins, J.W., Ausden, M.A., Beale, C.M., Oliver, T.H., Crick, H.Q.P., 2015. Research on the assessment of risks & opportunities for species in England as a result of climate change (Natural England Commissioned Reports No. Number 175).

Pearce-Higgins, J.W., Beale, C.M., Oliver, T.H., August, T.A., Carroll, M., Massimino, D., Ockendon, N., Savage, J., Wheatley, C.J., Ausden, M.A., Bradbury, R.B., Duffield, S.J., Macgregor, N.A., McClean, C.J., Morecroft, M.D., Thomas, C.D., Watts, O., Beckmann, B.C., Fox, R., Roy, H.E., Sutton, P.G., Walker, K.J., Crick, H.Q.P., 2017. A national-scale assessment of climate change impacts on species: Assessing the balance of risks and opportunities for multiple taxa. Biol. Conserv. 213, 124–134. https://doi.org/10.1016/j.biocon.2017.06.035

Pescott, O.L., Walker, K.J., Pocock, M.J.O., Jitlal, M., Outhwaite, C.L., Cheffings, C.M., Harris, F., Roy, D.B., 2015. Ecological monitoring with citizen science: the design and implementation of schemes for recording plants in Britain and Ireland. Biol. J. Linn. Soc. 115, 505–521. https://doi.org/10.1111/bij.12581

Preston, C.D., 2002. Approaches to native and alien species. Suffolk Nat. Hist. 38, 37–48.

Preston, C.D., Pearman, D.A., 2015. Plant hybrids in the wild: evidence from biological recording. Biol. J. Linn. Soc. 115, 555–572. https://doi.org/10.1111/bij.12577

Preston, C.D., Pearman, D.A., Dines, T.D. (Eds.), 2002. New Atlas of the British and Irish Flora. Oxford University Press, Oxford, England.

Rich, T.C.G., Woodruff, E.R., 1996. Changes in the vascular plant floras of England and Scotland between 1930–1960 and 1987–1988: The BSBI Monitoring Scheme. Biol. Conserv. 75, 217–229. https://doi.org/10.1016/0006-3207(95)00077-1

Stace, C., 1991. New Flora of the British Isles, 1st ed. Cambridge University Press, New York, USA.

Stace, C.A., Crawley, M.J., 2015. Alien Plants (Collins New Naturalist Library, Book 129). HarperCollins UK.

Stroh, P.A., Leach, S.J., August, T.A., Walker, K.J., Pearman, D.A., Rumsey, F.J., Harrower, C.A., Fay, M.F., Martin, J.P., Pankhurst, T., Preston, C.D., Taylor, I., 2014. A Vascular Plant Red List for England. Botanical Society of Britain and Ireland, Bristol.

van Strien, A.J., van Swaay, C.A.M., Termaat, T., 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. J. Appl. Ecol. 50, 1450–1458. https://doi.org/10.1111/1365-2664.12158

Walker, K.J., Pearman, D., Ellis, B., McIntosh, J., Lockton, A., 2010. Recording the British and Irish flora 2010-2020. BSBI, London.