Piecing together the biogeographic history of *Chenopodium vulvaria* L. using botanical literature and collections

This study demonstrates the value of legacy literature and historic collections as a source of data on environmental history. *Chenopodium vulvaria* L. has declined in Northern Europe and is of conservation concern in several countries, whereas in other countries it has naturalised and is considered an alien weed. It is hypothesised that much of its former distribution was the result of repeated introductions from its native range in southern Europe and that its decline in northern Europe is the result of habitat change and a reduction in number of propagules imported to the north. An historical analysis of its ecology and distribution was conducted by mining legacy literature and historic botanical collections. Text analysis of habitat descriptions written on specimens and published in botanical literature covering a period of more than 200 years indicate that the habitat and introduction pathways of *C. vulvaria* have changed with time. Using the naturalised alien range in a climate niche model it is possible to project the range in Europe. By comparing this predicted model with a similar model created from all observations it is clear that there is a large discrepancy between the realised and predicted distributions. It is concluded that if *C. vulvaria* was native to northern Europe, then it was only ever a rare species, however it was more common in the 18th and 19th centuries due to a combination of repeated introductions and the creation of suitable habitats by people.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.536v1 | CC-BY 4.0 Open Access | rec: 13 Oct 2014, publ: 13 Oct 2014

Author: Quentin J. Groom
Affiliation: Botanic Garden Meise, Bouchout Domain, Nieuwelaan 38, 1860 Meise, Belgium
Email: quentin.groom@br.fgov.be

Introduction

Legacy biodiversity literature is a potential mine of useful information on the past distributions of organisms. While these texts have always been available in academic libraries, their accessibility and discoverability has been significantly enhanced by projects such as the Biodiversity Heritage Library (www.biodiversitylibrary.org) and other online digital libraries. The ability to search a whole corpus of historical literature for a Latin name of an organism dramatically increases the accessibility of this scientific information and makes literature searches possible that once would have been unfeasible. In parallel, the widespread digital imaging of herbarium specimens and transcription of their labels has also made these data considerably more accessible, which, combined with historic literature, has created a large pool of information from which the phytogeographic historian can draw evidence (Vellend et al., 2013).

Chenopodium vulvaria L., a small, inconspicuous species that grows largely in places disturbed by mankind. It is not remarkable morphologically, but it is nonetheless distinctive due to its foul smell, which is described as similar to that of rotten fish. Its distinctiveness makes it particularly suited to a study using historic literature, because there is less concern that published accounts refer to other species as a result of misidentification.

C. vulvaria is currently a red-listed species in several countries including Sweden (www.artfakta.se), the United Kingdom (Cheffings *et al*., 2005), Belgium (Kestemont, 2010), Luxembourg (Colling, 2005), Czech Republic (Grulich, 2012) and some regions of France (Ferrez, 2005). In contrast, it has naturalised in California (Calflora, 2014), Argentina (Planchuelo, 1975; Giusti, 1997), Chile (Boelcke et al., 1985) and Australia (Atlas of Living Australia, 2014). *C. vulvaria* is widespread in countries bordering the Mediterranean and eastward to Afghanistan and Mongolia (Jalas & Suominen 1980; Meusel, Jäger & Weinert, 1992). Yet it is clear from historical literature that it was common in parts of northern Europe during the 18th and 19th centuries. Turner (1548) wrote "*It groweth muche aboute the walles in Bon in Germany*"; Bucher (1806) wrote in the Flora of Dresden "*An den strassen der vorstadt und sonst gemein*" translated as "By the streets of suburbs and usually common"; Curtis (1777) stated "*This species is very common in the neighbourhood of London*…" and Hooker (1821), in his flora of Scotland, describes it as "*frequent*".

The native distribution of *C. vulvaria* is unknown and its long association with man-made disturbance makes this impossible to determine. Floras in Northern and Central Europe variously describe it either as a native or an archaeophyte, though the evidence for categorizing it in either category is slim and is probably based on the anthropogenic habitats that *C. vulvaria* often inhabits.

Many other members of the Amaranthaceae live in disturbed, nutrient rich habitats and may be halophytic. *C. vulvaria* itself is often found in disturbed, eutrophic and coastal habitats. In general, species of such habitats are increasing and spreading in northern Europe (Wróbel, Tomaszewicz & Chudecka, 2006; Van Landuyt et al., 2008; Smart et al., 2003; Šerá, 2011; Groom, 2013). So at face value, *C. vulvaria* appears well adapted to modern habitats in Europe and yet it has declined.

One possible explanation for its apparent decline in northern and central Europe may be a misunderstanding of its former occurrence, its presence in the north being the result of propagule pressure from its heartland in southern Europe, constantly reinforcing the introduced populations in the north. One or many introduction pathways may have existed that delivered *C. vulvaria* seed outside of its normal range and these pathways have since reduced in importance, causing a collapse in the population. Another possible explanation is change to its former habitat, though the details of its ecology are too poorly known to understand what these changes may have been.

For non-woody plants there are few sources of data to examine recent biogeographic change. Palynology and the study of archaeological remains can be useful, but many species do not have a sufficiently distinctive anatomy to identify them from their remains. In these cases, historical literature and collections may be the only source of data on their former habitats and locations. Given the shortage of data an alternative approach, widely used to model the potential distribution of organisms, is bioclimatic modelling. Many studies have used observations from the known native range of a species to extrapolate its potential invasive range (e.g. Macfadyen & Kriticos, 2012). In ecological theory the potential bioclimatic range is generally considered to be larger than the realised distribution as a consequence of additional non-climatic limitations to distribution, such as edaphic factors (Araujo & Peterson, 2012). However, in the case of *C. vulvaria* the native range is not known and frequent non-persistent introductions mean that the realised distribution predicted from observations may be larger than its true bioclimatic range. For *C. vulvaria* the location of naturalisation in Australia, North America and South America might be a clearer indication of its bioclimatic range than within Europe, where it is hard to distinguish established from casual occurrences. Assuming that this species is well established and stable in its alien range, we can use the known naturalised range to model the climate envelope and extrapolate this to Europe to identify the areas where the climate is suitable for *C. vulvaria*. In this manner we can indicate those places where this species has been observed but is unlikely to be persistent.

My hypothesis is that *C. vulvaria* was formally more abundant in northern Europe and its current decline is the result of changes in the introduction pathways and loss of habitat. In this study I draw on botanical literature and specimens to identify habitat change and historic introduction pathways. I use text analysis of habitat descriptions to demonstrate how its habitat has changed over the past 200 years and I use bioclimatic niche modelling to contrast the realised range within Europe with the projected range based upon naturalised occurrences outside Europe.

Methods

Observation and specimen details were collected in a Common Data Model (CDM) database which is the central component of the EDIT Platform for Cybertaxonomy (Ciardelli et al., 2009; Berendsohn et al., 2011). Two methods were used to extract observations from literature, either XML markup or direct data entry. Digitised treatments were marked up with XML using the GoldenGate editor (http://plazi.org/?q=GoldenGATE, Sautter, Böhm & Agosti, 2007); uploaded to the PLAZI taxonomic treatment repository (plazi.org) and imported to the CDM database. Alternatively the observation details were copied from the treatment and entered manually into the CDM database using the EDIT Taxonomic Editor (Ciardelli et al., 2009). Observations where gathered from the biodiversity literature by reading the BHL corpus systematically after searching for *Chenopodium vulvaria* L. and its synonym *Chenopodium olidum* Curt. Other published observations were gathered from publications in the Library of the Botanic Garden, Meise. A list of the sources of observations of *C. vulvaria* is available in supplementary file S3. A complete survey of non-digitised literature is unfeasible, but there was an effort to check multiple floras of every European country and any other country with a temperate climate suitable for *C. vulvaria*.

Digitised observation data was also gathered from databases, primarily from GBIF (data.gbif.org, accessed 08 Nov 2013; see appendix), but also from the Atlas of Living Australia (2014); the Botanical Society of Britain and Ireland (2013) and Herbaria@home (2013). Scientific articles and websites containing observations were also discovered using search engines (scholar.google.be; google.be). Data from databases were standardised and imported directly into the CDM database.

Specimen data were gathered from herbaria by transcription of label information. Specimens from the following herbaria are included in the study, their names and abbreviations follow the Index Herbariorum (http://sciweb.nybg.org/science2/IndexHerbariorum.asp). Botanical Garden Meise (BR); Botanical Museum Berlin-Dahlem (B); Botanische Staatssammlung München (M); Bulgarian Academy of Sciences (SOM); Bulgarian Academy of Sciences (SOMF); Charles University in Prague (PRC); Herbier J.H. Fabre (FABR); Institut Botànic de Barcelona (BC); Institute of Biodiversity and Ecosystem Research, Nationaal Herbarium Nederland (L); Moscow State University (MW); Museu Nacional de História Natural e da Ciência (LISU); Museum National d'Histoire Naturelle (P); National Academy of Science, Kyrgyzstan (FRU); Natural History Museum, London (BM); Natural History Museum of Denmark (C); New York State Museum (NYS); Reading University (RNG); Royal Botanical Gardens, Kew (K); Sapienza University of Rome (HFLA), Sofia University (SO); South London Botanical Institute (SLBI); Universidad Nacional del Sur Herbario (BBB); Universität Wien (WU); Universidad de Concepción, Chile (CONC); University of Alaska Herbarium (ALA); University of Birmingham (BIRM); University of California (UC); University of British Columbia (UBC); University of Manchester (MANCH); University of Wales (ABS); Wageningen University (WAG) and others contributing data to Global Biodiversity Information Facility (GBIF) (supplementary files S2). Many other herbaria and herbarium catalogues were searched without finding specimens and several herbaria were contacted and either contained no specimens or did not respond. Undoubtedly there are more specimens and observations of *C. vulvaria* to be discovered, but I believe these to be a representative sample and a large proportion of those that exist. Undated specimens were not used in the study; however, it is usual for published observations to be undated. Therefore the publication date was used for undated observations in literature. Studying biographical information of collectors it is clear that most undated observations in old floras are within 35 years of the publication date and author tend to provide dates when they are not recent. In total 2456 observations were collected from specimens and literature. These data span 465 years from 1548 to 2013, though there are only two observations from the 16th century, two from the 17th century and nineteen from the 18th century.

Text analysis of habitats

The text describing the habitat of *C. vulvaria* was collected from 104 floras, 33 scientific articles, 119 specimens and 5 websites, covering the years 1787 to 2014. The texts were written in 12 languages, English (35%) German (20%), French (17%), Latin (12%), Dutch (4%), Italian (3%), Portuguese (3%), Spanish (3%), Hungarian (1%), Danish (1%), Catalan (1%) and Czech (<1%). Each description was broken down into tokens consisting of either single words or short phrases describing a single aspect of the habitat. Thus the description "*In Straßen, an Häusern, Stallungen, Düngerstätten*" was broken down into the tokens "*Straßen*" (roads), "*an Häusern*" (near houses), "*Stallungen*" (stables) and "*Düngerstätten*" (mature heaps). This process created 475 habitat tokens. These tokens were then translated to English using native speakers of English, German, French and Dutch and for other languages a combination of Google Translate (translate.google.be) and the multilingual collaborative dictionary Wiktionary (wiktionary.org). To conduct the analysis it was necessary to reduce the number of habitat terms, which was done in two stages. The anglicized tokens were first simplified to closely related terms. Thus the terms "*by foot of the city walls*", "*along walls*", "*under walls*", "*mud walls*", "*foot of walls*", "*under walls*", "*under a wall*" and "*foot of the church yard wall*" were all replaced by "*by walls*". This process reduced the number of habitat words to fifty. These fifty words were then arranged into logically related categories. Thus "*by walls*", "*by fences*" and "*by hedges*" were grouped together under the term "*boundaries*". This reduced the number of habitat categories to fifteen (animal waste, boundaries (including walls), coastal, disturbed and grazed land, dry & bare soil, habitation, hills, horticulture, industry, rail, roads, sand and rock, shipping, waste, wetland. A full list of the tokens contributing to each category is provided in the appendix (Table S1). Throughout the process the tokens were kept associated with the date; either the year the specimen was collected, observed or the year of publication. To analyse the use of habitat words in the collected corpus the simplified habitat terms were pooled into 20 year periods from 1780 onwards. The proportional use of each habitat term was then calculated for each period. Statistical analysis was conducted in R (version 2.15.2) using generalized linear modelling with binomial errors, weighted with the number of tokens contributing to each pool. All models were checked for overdispersion using the ratio of the residual deviance and the degrees of freedom, but none were found to be overdispersed.

Analysis of distribution

Except for the rare occasions when coordinates where available with the specimen or observation, georeferencing was carried out manually according to best practise (Chapman & Wieczorek, 2006). Error radii for coordinates were not available for most records in databases, but they were estimated for the coordinates georeferenced in this study, however, they were not used to select data for the analysis. The average error radius was 11 km and the mode and median were both 10 km. *C. vulvaria* is a largely lowland species and errors in georeferencing of these magnitudes are insignificant for bioclimatic modelling at a global scale compared to the other inherent biases in these data.

Species distribution modelling was conducted using the BioVel Ecological niche modelling workflow and services (www.biovel.eu). The ecological niche modelling workflows were run on 6th Aug 2014. BioVeL is funded by the EU's Seventh Framework Program, grant no. 283359. The workflow uses the Maxent method based upon Phillips, Dudík & Schapire (2004) and using the openModeller web service (de Souza Muñoz et al., 2011). Models were created using the default parameter and all 19 layers of the WorldClim global climate layers 10 arc minutes, version 1.4, release 3 (Hijmans, 2005).

Non-European observations used for modelling were only those locations where it was clear, either from the notes on the specimens or from floras, that the species forms persistent population at these sites. If there was any doubt to the status modern floras were consulted to ascertain the persistence of the species in the area. The locations with non-native populations outside Europe were Southern Argentina; California; Chile; South Australia; Tasmania; Tierra del Fuego (Argentina) and Victoria (Australia). *C. vulvaria* is also recorded from South Africa and New Zealand, but its status there is not clear. It is also believed to occur natively in Mongolia but only one observation was found. A total of 42 observations from the naturalised range were used to model the range. However, weeding of duplicates during the workflow reduced the number to 32. The dates of these records were from 1863 to 2012, though 86% dated from 1950 onward. For modelling the realised range, all global observations where used which resulted in 1894 observations after weeding of duplicates.

Results

Text analysis

Four habitat categories were notably more frequent than the others (Fig. 1). These categories are firstly waste, including rubbish piles, rubble, ruins and waste places of all kinds; secondly, boundaries, mainly at the base of walls; thirdly, roads and roadsides, including streets and farm tracks; and fourthly horticulture, such as gardens and other cultivated places. The habitat categories in Fig. 1 are not mutually exclusive, but often describe different aspects of the same habitat such as the proximity to landscape features, soil type, nutrient status and moisture.

In summary, the habitat analysis underscores several aspects. *C. vulvaria* is strongly associated with mankind, natural habitats such as coastal and wetlands are mentioned infrequently. It is intolerant of competition; none of the habitats are defined by other vegetation, such as meadows, woodland or heaths. It is frequently associated with transport routes and it is usually associated with some form of soil disturbance.

When the use of these terms was compared over time, no significant change was found for the use of terms relating to animal waste, coastal, dry & bare soil, habitation, hills, horticulture, industry, rail, roads, shipping and waste. Figure 2 shows the changes of eight of these categories, including the only four where there were significant changes. The significant changes were increases in the proportion of the terms related to wetland ($p < 0.01$, DF=11), sand and rock ($p < 0.05$, DF=11) and disturbed and grazed land ($p < 0.05$, DF=11), whereas there has been a significant decrease in the proportion of terms related to boundaries ($p < 0.001$, DF=11). Of these significant changes only terms relating to boundaries were also highly frequent in the corpus (Fig. 1).

Introduction vectors, pathways and origins

Clear expressions of the introduction vector were rare on specimens and in publications. Where introduction vectors are evident they are summarised in Table 1. Ballast soil at ports was the earliest vector mentioned in the corpus and it was also most frequently mentioned. However, it stops being mentioned in the early 20th century. Several specimens and observations implicate the transport of ore. *C. vulvaria* was reported on Chromite in Baltimore, USA between 1953–1958 (Reed, 1964); in Norway in 1954 (Uotila, 2001); on manganese ore in Norway between 1931–1935 and near an ore crushing plant in Kyrgyzstan in 1961 (Lazkov, Sennikov & Naumenko, 2014). Various agricultural products are mentioned as vectors such as grain, but no mention of its introduction as an herbal medicine or other produces commonly imported from the Mediterranean such as tobacco, even though *C. vulvaria* is frequently associated with waste.

Evidence for the pathways of introductions is scant, but shipping and railways are mentioned. Although roads are the most frequently mentioned transport system (Fig. 1), it is unclear if the presence of this species on roads relates to the introduction pathway or whether roads just provide suitable habitat.

Evidence for the origin of introductions is also slim, though where the origin is mentioned it is always from a country in the Mediterranean region (Uotila, 2001). There is no evidence of return introductions from naturalised populations outside Europe.

Comparing actual climatic niche and realised distribution

The observations of *C. vulvaria* within Europe are from an inseparable mixture of stable populations and casual occurrences. It is therefore impossible to validate a model for the true climatic niche of *C. vulvaria*. For this reason I have not attempted to refine the output of the models by adjusting their default parameters or eliminating climate layers. It is nevertheless informative to contrast models created from the known naturalised range outside Europe with the realised range within Europe (Fig. 3, Fig. 4). The actual climatic niche, predicted from observations from the naturalised range outside Europe predicts the presence of *C. vulvaria* in southern and western Europe, North Africa and the Middle-east, notably, Spain, western France and Turkey (Fig. 3). The actual observations and the climate niche model created from them show a much wider distribution, which extends much further north and eastward than the niche model created from the naturalised range (Fig. 3, Fig. 4). Figure 5 shows the locations of actual observations and the dates they were made. It demonstrates that there has been a general decline in the number of observations from northern Europe, but it also suggests unevenness in surveying effort between different countries and different time periods.

Discussion

This study tracks the distribution and habitat changes of *C. vulvaria* over more than 200 years. Over this period botanical literature becomes more common and sufficiently abundant for analysis. Simultaneously botanical specimens became more frequently collected and better documented, further adding to the analysable corpus of historical documents.

Over the past two centuries many social, economic and technological changes have occurred that may have influenced the abundance and distribution of *C. vulvaria*. Some key events in this period are the expansion of the railway network in the 19th century, the adoption of motorised road transport in the early 20th century, the decline in the uses of horses for transport and agriculture in the 20th century; the transition from sail to steam powered ships at the turn of the 20th century; the discovery of herbicides in the mid-20th century and the Green Revolution in the latter half of the 20th century. *C. vulvaria* is an anthropophilic species and to some extent benefits from this association, however, for the same reason it will be more acutely affected by changes in human culture than many other species.

Text analysis was able to identify key habitat features of *C. vulvaria*. This species has been, and still is, strongly associated with mankind, both as a weed of cultivation and as a ruderal plant. The analysis identifies habitat traits such as its avoidance of competition and the association with waste. The genus *Chenopodium* is considered to be nitrophilous, indeed *C. vulvaria*, is sometimes associated with habitats linked to animal dung, however, it is much more commonly associated with other types of waste or cultivated place (Fig. 1).

The temporal analysis of habitat change indicates that *C. vulvaria* is still associated with many of the same habitats it was in the past, such as agriculture, transport and waste (Fig. 2). However, in the 20th century habitat descriptions have included proportionally more words related to natural or semi-natural habitats, such as grazing, sand and wetland.

The reference to wetland amongst the habitats needs further explanation, because *C. vulvaria* is not a typical wetland plant. It does not grow in water, but colonises bare soil exposed in the summer at the margins of rivers, ditches and lakes. Thus its association with wetland is of an opportunistic colonizer of habitats free from competition, rather than a true wetland plant.

The habitat where *C. vulvaria* has declined is along boundaries, particularly along walls, which contributed 80% of the boundary terms. *C. vulvaria* does not grow on walls, but beside them, which appears at first sight to be a rather non-specific habitat description. However, the margins of walls have changed considerable in the past 200 years. Walls were once built using lime mortar, rather than cement, and were frequently painted with whitewash, a mixture of calcium hydroxide and chalk. Whitewash gave the traditional white or pink colour to houses throughout Europe. Consequently, the soil in the immediate vicinity of walls would have been alkaline. *C. vulvaria* is not known as an alkaliphile, however it is clearly tolerant of high pH as it has been found on the ultrabasic rock chromite (Reed, 1964). Furthermore, because horses were used for transport and farm animals were driven along roads, the base of walls would have been strewn with animal waste. Such fertile alkaline habits do not occur by walls in modern towns and we can speculate that the technical changes in building practises and changes to transportation have contributed to the decline of *C. vulvaria*.

Text analysis is clearly a useful tool for environmental historians, nevertheless, it is susceptible to the fallibility and biases of authors, who may uncritically follow their forbears or write from hearsay rather than experience. Also, botanical activity is spatially and temporarily biased. For example, British and German botanical literature has, and continues to be, more abundant than for other countries in Europe.

Compared to the analysis of habitat, evidence for introduction vectors, pathways and origins was limited. The results show that there were multiple vectors introducing *C. vulvaria* to northern Europe, but particularly as a grain contaminant and in ship's ballast. The frequent occurrence of *C. vulvaria* in waste perhaps indicates that its seeds were contaminants of many crops. Indeed, different specimens mentioned *C. vulvaria* in crops of lentils and potatoes. Unfortunately, the source of a casual introduction is rarely obvious by the time the plant is mature. Weed species that are dispersed as seed contaminants have declined throughout Europe in the 20th century; this is, in part, a consequence of improved seed cleaning methods (Hilbig, 1987; Sutcliffe & Kay, 2000; Lososová, 2003). Most of these species are considered archaeophytes to northern Europe.

Soil was used as ballast on sailing ships during the 18th and 19th centuries to provide stability to cargo ships when not carrying heavy loads. In ports, where heavy materials were loaded, ballast was removed and replaced by cargo. Large hills of ballast soil where a common feature of busy ports, particularly in areas of mining and heavy industry, such as in northern Europe. These ballast hills were a large reservoir of propagules for many species (Carlton, 2011). The large number of specimens and observations reflects the importance of this invasion pathway, but might be somewhat over-represented because botanists were attracted to ballast heaps as a source of novel species and because the vector of the propagules is clear in this case.

Ore is also mentioned as an introduction vector to the USA and Norway. Chromium processing began in Baltimore, USA in 1822, at which time only local chromium ore deposits were processed (Newcomb, 1994). However, by the end of the 19th century local chromium deposits were exhausted and processing continued with imported ore until the end of the 20th century. Similarly, Norway is also a large processor of imported chromium ore, for example in 1992 the country imported 187,965 tonnes of chromite ore from Turkey (Plachy, 1992). Indeed, it is likely that some of the chromite imported into Baltimore was also from Turkey where chromite was first mined in 19th century (Zengin, 1957). Therefore, it seems that exports of chromite from Turkey could have been a pathway for dispersal of *C. vulvaria* during the 20th century.

Animal dung is often mentioned as a growing medium for *C. vulvaria*, which is indirect evidence for endozoochory. Certainly, other *Chenopodium* species are dispersed in this manner and *C. vulvaria* is eaten by ruminants despite its smell (Withering 1776; Haarmeyer et al., 2010). In the 21st century yet another vector of *C. vulvaria* introduction has been created, that of imported Olive trees (Hoste et al., 2009). These mature trees are extracted from olive groves with a large amount of soil and are sold in northern Europe as horticultural novelties.

Though dispersal vectors are rarely mentioned in the corpus, it is clear that *C. vulvaria* has been dispersed by a wide variety of vectors and through a number of pathways (Table 1). There are historic periods associated with each vector and if this analysis was extended to more species, one would be able to further refine the time frames during which these pathways were operating. From the diversity of distribution vectors it is clear that *C. vulvaria* has been widely introduced outside its natural climatic range and it grows often temporarily. However, with the exception of horticultural imports, introduction pathways of *C. vulvaria* ended midway through the 20th century.

The sporadic occurrence of *C. vulvaria* presents a problem for the selection of occurrences for distribution modelling. Unless all casual occurrences are eliminated from the data before fitting, the model would indicate a much broader climatic range. Separating permanent populations from casual occurrences is impossible for Europe where anthropogenic disturbance and trade have confused the quasi-natural distribution. However, in the naturalised range the situation is much clearer. Most, if not all, modern observations of *C. vulvaria* in California, Australia and South America appear to be from naturalised populations, that is to say, the associated meta-data indicates the presence of a population and there no indication of a recent introduction. Therefore, the naturalised distribution outside Europe should reflect the true climatic niche of the species, as long as the distribution is at equilibrium. This assumption seems reasonable since old casual records of *C. vulvaria* occur throughout the world, but naturalised populations persist in only a few of those places. Clearly, introduction events were occurring all over the world for several hundred years of international trade, but *C. vulvaria* only naturalised in a few of those places where the habitat and climate suited it.

Projecting the bioclimatic range in Europe from naturalised alien populations elsewhere predict a much more southern and western distribution of *C. vulvaria*, than the modelling using all occurrences (Fig. 3, Fig. 4). Yet, these rather crude models indicate that the naturalised distribution of *C. vulvaria* has a climate much closer to that of southern Europe and North Africa than to northern and central Europe. The distribution models are consistent with my hypothesis that historically *C. vulvaria* was only present in parts of Northern Europe because of repeated introductions, and that, in these places, the climate is unsuitable for lasting populations to exist. Discrepancies between the projected model and the realised distribution could be the result of several factors, either an incorrect model; lack of suitable habitat; spatial variations in surveying effort, or plants growing outside their actual climatic niche due to local factors. The model projecting distribution from the naturalised range is based on relatively few observations and could be improved by more data. Nevertheless, any distribution model of this species has to address the problem of casual occurrences. The shortage of observations from countries such as Turkey and Morocco, in apparent contrast to the models, are at least in part due to lack of collecting in these regions, but also due to the inaccessibility of the data from these countries. These results are a good reminder to those who would extrapolate native ranges onto potentially invasive ranges. Clearly, it is not always possible to predict the naturalised distribution from the native range due to the lack of data and indistinct range boundaries.

Conclusions

Text analysis is a useful technique to study recent ecological and distributional change. Despite its limitations it provided information, which would be difficult, if not impossible to obtain from other sources. As a larger volume of semantically enhanced biodiversity literature becomes available it will allow much more sophisticated habitat analysis covering many more species. The ability to contrast data from different species will strengthen results and allow correction for some of the biases. Furthermore, the development of environmental ontologies and thesauri will simplify the method and improve repeatability (Buttigieg et al., 2013). This will allow over-representation analysis of ontological terms from one species in comparison to these terms in the whole corpus.

Analysis of these descriptions indicates that the habitat of *C. vulvaria* has changed over the past two centuries, particularly next to walls. Multiple vectors and pathways have been involved in the human mediated dispersal of *C. vulvaria*, but different vectors and pathways were active in different periods. In the past *C. vulvaria* would have been dispersed to many places outside of its climatic niche. It is reasonable to believe that many of the observations of *C. vulvaria* in northern Europe were the result of introductions and that a reduction in the propagule pressure in recent years has consequently lead to a decline in observations of this species. It is concluded that mankind spread *C. vulvaria* to northern Europe and created habitat for it to grow and then inadvertently removed the habitat and the introduction pathways causing a decline.

References

Araujo MB, Peterson AT. 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology* 93:1527–1539.

Atlas of Living Australia. 2014. *Available at* http://bie.ala.org.au/species/Chenopodium+vulvaria (accessed 25 February 2013).

Berendsohn WG, Güntsch A, Hoffmann N, Kohlbecker A, Luther K, Müller A. 2011. Biodiversity information platforms: From standards to interoperability. *ZooKeys* 150:71–87.

Boelcke O, Correa NM, Moore DM, Roig FA. 1985. Catálogo de las Plantas Vasculares. In: Boelcke O, Moore DM, Roig FA, eds., *Transecta Botánica de la Patagonia Austral. Proyecto Internacional*. Buenos Aires: Consejo Nacional de Investigaciones Científicas y Técnicas, 129–255.

Botanical Society of Britain and Ireland. 2013. Distributions Database. *Available at* http://bsbidb.org.uk. (accessed 21 February 2013).

Bucher, C.T. (1806) *Florae Dresdensis nomenclátor*. Dresden: Walther.

Burk I. 1877. List of plants recently collected on ships' ballast in the neighborhood of Philadelphia. *Proceedings of the Academy of Natural Sciences of Philadelphia* 29:105–109.

Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, and the ENVO Consortium. 2013. The environment ontology: contextualising biological and biomedical entities. *Journal of Biomedical Semantics* 4:43.

Calflora. 2014. Information on California plants for education, research and conservation, based on data contributed by dozens of public and private institutions and individuals, including the Consortium of Calif. Herbaria. Berkeley (USA). *Available at* www.calflora.org (accessed 3 March 2014).

Ciardelli P, Kelbert P, Kohlbecker A, Hoffmann N, Güntsch A, Berendsohn WG. 2009. The EDIT Platform for Cybertaxonomy and the Taxonomic Workflow: Selected Components. In: Fischer S, Maehle E, Reischuk R, eds. *INFORMATIK 2009, Im Focus das Leben, Beiträge der 39. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 28.9. - 2.10. in Lübeck*. Lecture Notes in Informatics (LNI) 154 - S. 28: 625–638.

Carlton JT. 2011. Ballast. In: Simberloff D, Rejmanek M, eds. *Encyclopedia of Biological Invasions*. Berkeley and Los Angeles: University of California Press, 43–49.

Chapman AD, Wieczorek J. 2006. Guide to best practices for georeferencing. Copenhagen: Global Biodiversity Information Facility, 1–77.

Cheffings CM, Farrell L, Dines TD, Jones RA, Leach SJ, McKean DR, Pearman DA, Preston CD, Rumsey FJ, Taylor I. 2005. The vascular plant red data list for Great Britain. *Species Status* 7:1–116.

Colling G. 2005. Red list of the vascular plants of Luxembourg. *Ferrantia* 42:80.

Curtis W. 1777. *Flora Londinensis vol. 5*. London: William Curtis.

de Souza Muñoz M, De Giovanni R, de Siqueira M, Sutton T, Brewer P, Pereira R, Canhos DAL, Canhos VP. 2011. openModeller: A generic approach to species' potential distribution modelling. *GeoInformatica* 15:111-135.

Ferrez Y. 2005. Liste rouge de la flore vasculaire menacée ou rare de Franche-Comté Proposition. *Les nouvelles archives de la flore jurassienne* 3:217–229.

Groom QJ. 2013. Some poleward movement of British native vascular plants is occurring, but the fingerprint of climate change is not evident. PeerJ 1:e77.

Grulich V. 2012. Red list of vascular plants of the Czech Republic: 3rd edition. *Preslia* 84:631–645.

Giusti L. 1997. Chenopodiaceae. In: Hunziker AT, ed. *Flora fanerogámica argentina.* Córdoba: Proflora-Conicet: 40: 1–53.

Haarmeyer DH, Bösing BM, Schmiedel U, Dengler J. 2010. The role of domestic herbivores in endozoochorous plant dispersal in the arid Knersvlakte, South Africa. *South African Journal of Botany* 76:359–364.

Herbaria@home. 2013. *Available at http://herbariaunited.org/atHome/* (accessed 21 February 2013).

Hilbig W. 1987. Changes in segetal vegetation under conditions of industrialized agriculture. *Archives of Nature Conservation & Landscape Research* 27:229–249.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25:1965–1978.

Hjelt H. 1906. Caryophyllaceae-Resedaceae. In: *Conspectus Florae Fennicae in Acta Societatis pro Fauna et Flora Fennica 30*. Finland: Helsingforsiae.

Hooker WJ. 1821. *Flora Scotica*. London: Archibald Constable and Co. & Hurst, Robinson, and Co.

Hoste I, Verloove F. Nagels C, Andriessen L, Lambinon J. 2009. De adventievenflora van in België ingevoerde mediterrane containerplanten. *Dumortiera* 97:1–16.

Jalas J, Suominen J. 1980. *Atlas Florae Europaeae, Vol. 5*. Helsinki: The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo.

Kestemont B. 2010. A red list of Belgian threatened species. Brussels: Statistics Belgium.

Lazkov G, Sennikov A, Naumenko A. 2014. New records in vascular plants alien to Kyrgyzstan. *Biodiversity Data Journal* 2:e1018.

Lososová Z. 2003. Estimating past distribution of vanishing weed vegetation in South Moravia. *Preslia* 75:71–79.

Macfadyen S, Kriticos DJ. 2012. Modelling the Geographical Range of a Species with Variable Life-History. *PLoS ONE* 7: e40313.

Meusel H, Jäger E, Weinert E. 1992. *Vergleichende Chorologie der Zentraleuropäischen Flora, Vol. 1*. Jena: Gustav Fischer.

Mohr C. 1901. Plant Life of Alabama. In: *Contributions from the United States National Herbarium*. Washington: US Department of Agriculture.

Newcomb S. 1994. A History of Chromite and Copper in Maryland: The Tyson Years. *Matrix* 3:84–92.

Phillips SJ, Dudík M, Schapire RE. 2004. A maximum entropy approach to species distribution modeling. In: *Proceedings of the Twenty-First International Conference on Machine Learning. ACM (New York)*. 655–662.

Plachy J. 1992. *The Mineral Industry of Norway in Minerals yearbook mineral industries of Europe and central Eurasia*. U.S. Bureau of Mines.

Planchuelo AM. 1975. Study of the fruits and seeds of the genus Chenopodium in -Argentina.

Darwiniana 19:528–565.

Reed CF. 1964. A flora of the chrome and manganese ore piles at Canton, in the Port of Baltimore, Maryland and at Newport News, Virginia, with descriptions of genera and species new to the flora of eastern United States. *Phytologia* 10:324–406.

Sautter G, Böhm K, Agosti D. 2007. Semi-Automated XML Markup of Biosystematics Legacy Literature with the GoldenGATE Editor. In: *Proceedings of Pacific Symposium on Biocomputing 2007, Wailea (USA), January 2007*. International Society for Computational Biology: 391–402.

Šerá B. 2011. Stress tolerant plant species spread in the road-net. *Ecological Questions* 14:45–46.

Smart SM, Robertson JC, Shield EJ, van de Poll HAM. 2003. Locating eutrophication effects across British vegetation between 1990 and 1998. *Global Change Biology* 9:1763–1774.

Sutcliffe OL, Kay QON. 2000. Changes in the arable flora of central southern England since the 1960s. *Biological Conservation* 93:1–8.

Turner W. 1548. The names of herbes. (with an introduction, an index of English names, and an identification of the plants enumerated by Turner) by James Britten. Published for the English Dialect Society by N. Trübner 1881, London.

Uotila P. 2001. Chenopodium L. In: Jonsell B, ed. *Flora Nordica*. Stockholm: Bergius Foundation, 4–31.

Van Landuyt W, Vanhecke L, Hoste I, Hendrickx F, Bauwens D. 2008. Changes in the distribution area of vascular plants in Flanders (northern Belgium): eutrophication as a major driving force. *Biodiversity and Conservation* 17:3045–3060.

Vellend M, Brown CD, Kharouba HM, McCune JL, Myers-Smith IH. 2013. Historical ecology: using unconventional data sources to test for effects of global environmental change. *American Journal of Botany* 100:1294–1305.

Withering W. 1776. *A botanical arrangement of all the vegetables naturally growing in Great-Britain Vol. 1*. London: Cadel and Elmsley.

Wróbel M, Tomaszewicz T. Chudecka J. 2006. Floristic Diversity and Spatial Distribution of Roadside Halophytes along Forest and Field Roads in Szczecin Lowland (West Poland). *Polish Journal of Ecology* 54:303–309.

Zengin Y. 1957. The Mode of Distribution of Chrome-Ores in Peridotites in Turkey. *Bulletin of* *the Mineral Research and Exploration Institute of Turkey* 49:84–92.

Acknowledgements

Thanks to Ana Isabel D. Correia of the Faculdade de Ciências da Universidade de Lisboa; Prof. Dr. Roberto Rodríguez Ríos from the Universidad de Concepción, Chile; Noemí Montes and Neus Ibáñez Cortina of the Institut Botànic de Barcelona; Sabrina Eckert, Anton Güntsch, Patricia Kelbert, Andreas Müller and Robert Vogt of the Botanischer Garten und Botanisches Museum Berlin-Dahlem; Vladimir Vladimirov from the Bulgarian Academy of Sciences; Michal Štefánek of Charles University in Prague; Alan Paton from the Royal Botanic Gardens, Kew; Leni Duistermaat & Dr. Hubert Turner of Naturalis Biodiversity Center and the specimen digitization team at the Botanic Garden Meise and Sabine Metzger.

Table 1(on next page)

Introduction vectors gleaned from historical sources

The vectors stated or implied from specimens and publications, including the range of dates that vectors were mentioned either on specimens or in publications.

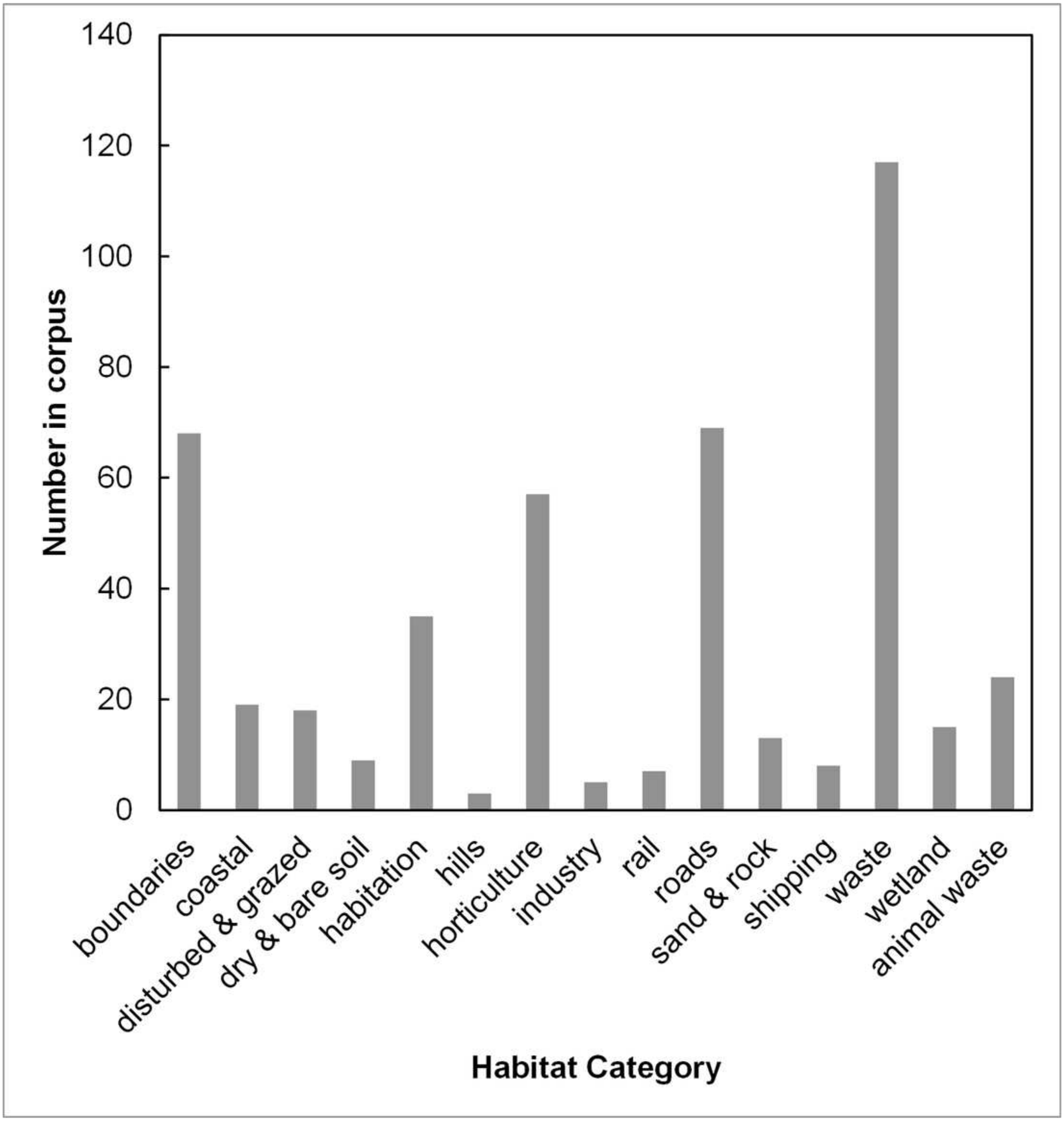
Table 1. The introduction vectors stated or implied from specimens and publications, including the range of dates that vectors were mentioned on specimens and in publications.

Vector	**Dates**	**number**	**Example references and specimens**
Ballast	1870 –1912	13	Publications: Burk, 1877; Mohr, 1901; Hjelt, 1906 Specimens: BIRM 032912; MANCH.94943.Kk803; S-H-2810
Grain	1936 –1964	3	Uotila, 2001; Unaccessioned specimens from Nationaal Herbarium Nederland (L);
Wool	1909	1	Observation by I.M. Hayward, Selkirkshire in database of the Botanical Society of Britain and Ireland (2013)
Ore	1931 –1961	4	Uotila, 2001; Reed, 1964; Lazkov, Sennikov & Naumenko, 2014; Specimen S-H-2141
Cork	1956 –1966	1	Uotila, 2001
Horticultural imports	2008	1	Hoste *et al.*, 2009

Figure 1

The use frequency of words in the collected corpus of *Chenopodium vulvaria* habitat descriptions

The frequency of each habitat category in the corpus of habitat descriptions from literature and specimens. The word and phrase tokens contributing to each category are presented in Table S1.



140
120
100
80
60
40
20
0
Number in corpus
boundaries
coastal
disturbed & grazed
dry & bare soil
habitation
hills
horticulture
industry
rail
roads
sand & rock
shipping
waste
wetland
animal waste
Habitat Category

Figure 2

The change with time of habitat categories from the collected corpus of *Chenopodium vulvaria* habitat descriptions

A temporal analysis of the corpus of habitat descriptions from publications and specimens of *Chenopodium vulvaria*. The graphs show the proportion of token usage related to each habitat category for periods of 20 years. The words contributing to each habitat category are listed in Table S1. The best fit lines are from generalised linear models of the data weighted with the number of tokens contributing to each proportion. The categories wetland ($P < 0.01$, DF=11), sand and rock ($P < 0.05$, DF=11), and disturbed and grazed ($P < 0.05$, DF=11) all significantly increased with time. Only the term boundaries decreased with time ($P < 0.001$, DF=11). All other categories shown in Fig. 1 did not show significant variations with time.

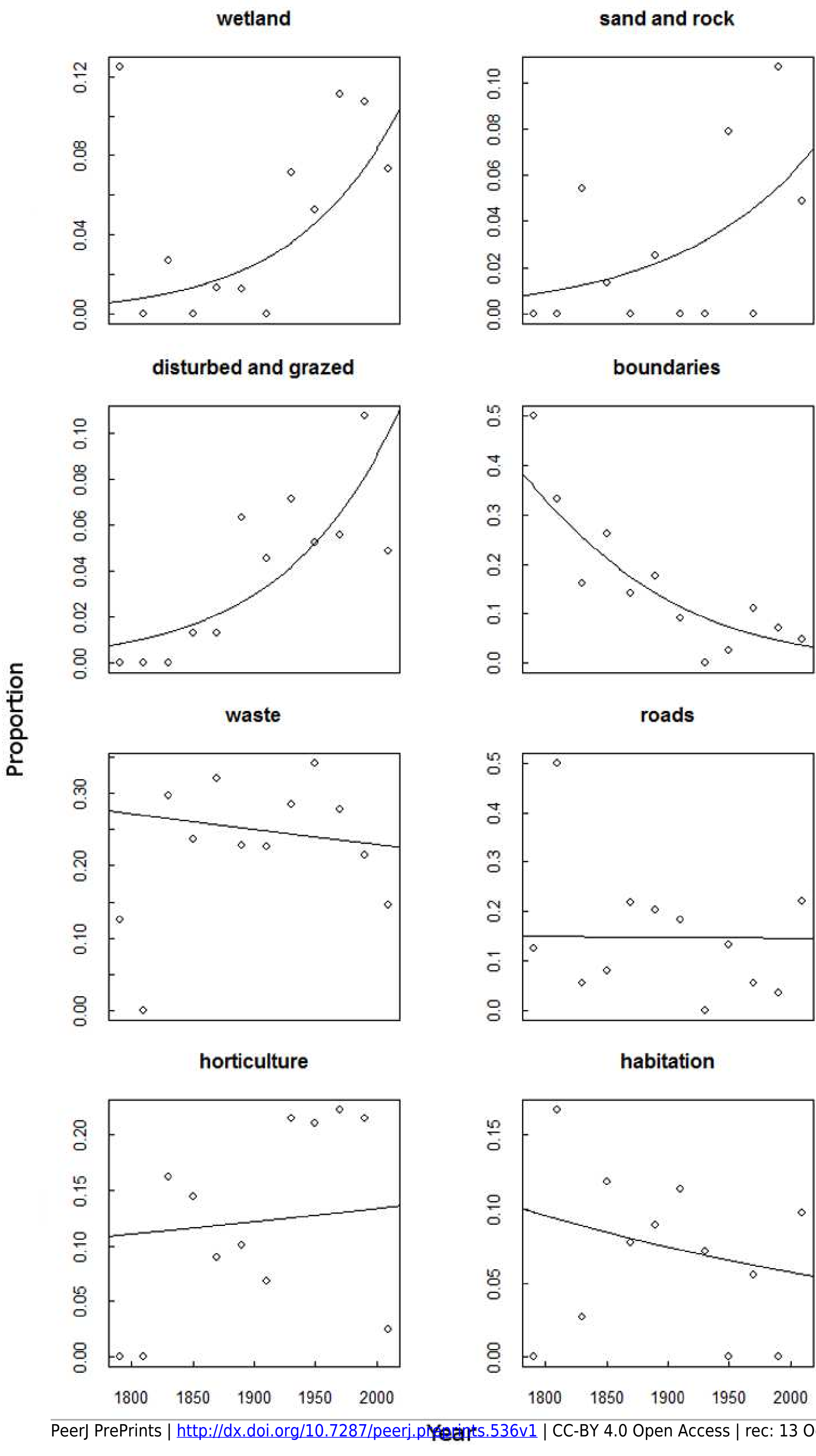

wetland
sand and rock
disturbed and grazed
boundaries
waste
roads
horticulture
habitation
Proportion
Year
1800
1850
1900
1950
2000

Figure 3

A distribution model created from the naturalised range of *Chenopodium vulvaria* outside Europe and extrapolated back to Europe.

A distribution model of Chenopodium *vulvaria* in Europe, North Africa and the Middle-east projected from its naturalised range in California, South America and Australia. This model aims to predict where, according to the naturalised range, the climate is suitable for persistent populations in Europe as opposed to casual occurrences. The map uses a Mollweide equal area projection.

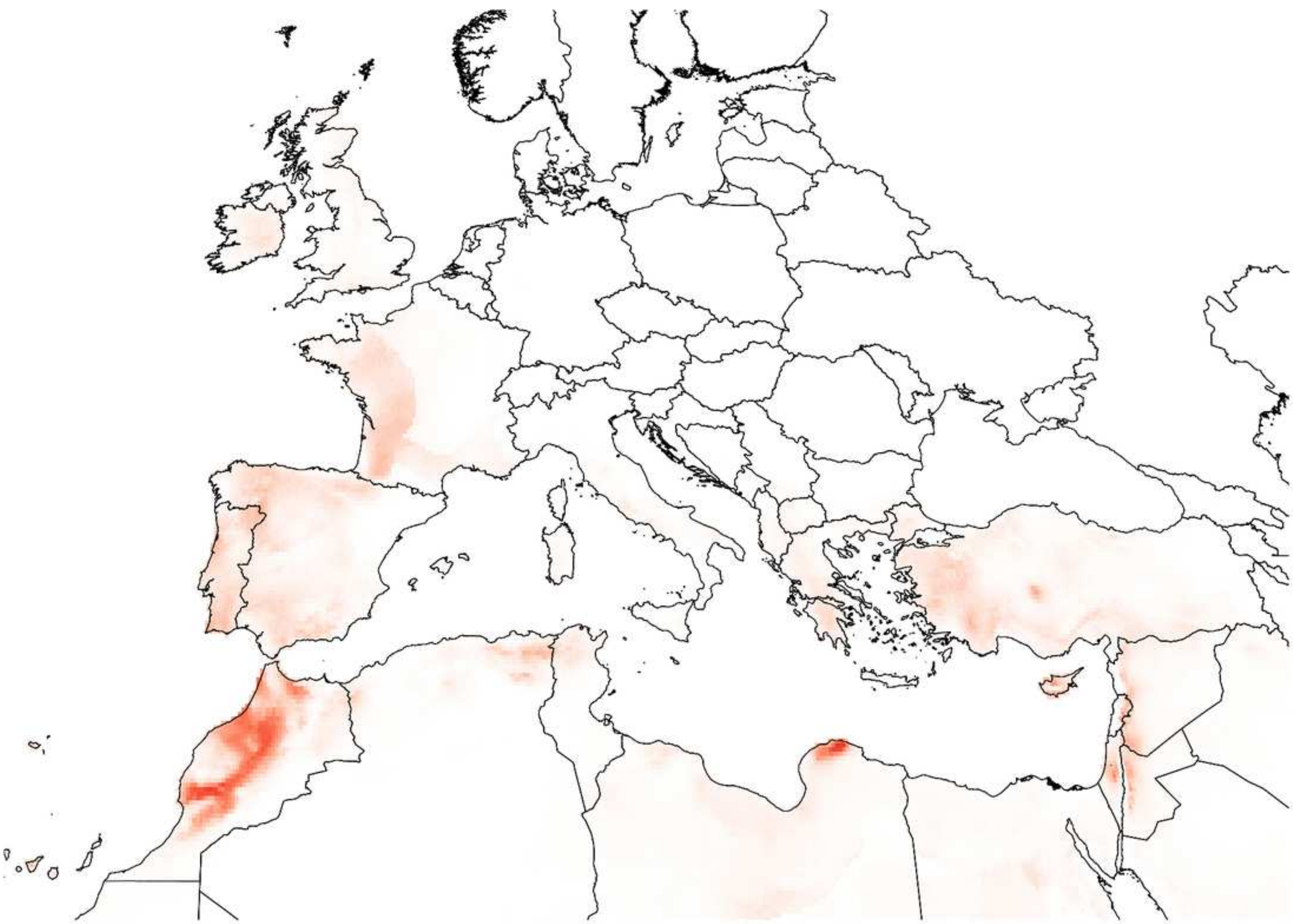


Figure 4

A distribution model of *Chenopodium vulvaria* created from all known locations

A distribution model of *Chenopodium vulvaria* in Europe, North Africa and the Middle-east created from all observations globally. This model aims to delimit the area where the climate is suitable for both stable populations and casual occurrences to occur. The map uses a Mollweide equal area projection.

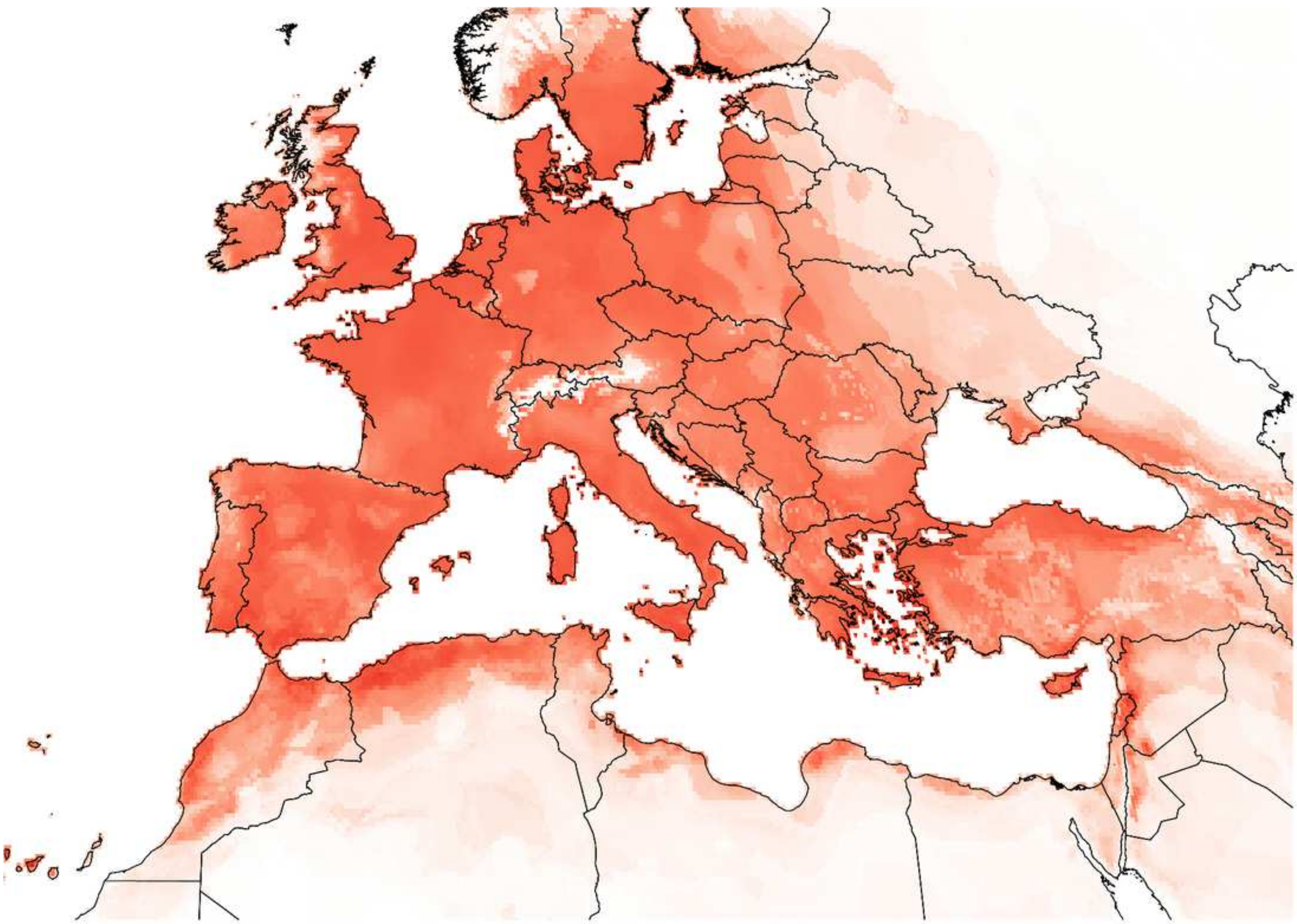


Figure 5

A dated distribution map of *Chenopodium vulvaria* observations from Europe, North Africa and the Middle-east

A distribution map of *Chenopodium vulvaria* in Europe, North Africa and the Middle-east. Circles represent georeferenced observations either from specimens or from the literature. The colour of the points denotes the date of observation, yet to emphasise the scarcer old records the date ranges are not equal, but the data is divided into equal-sized subsets. The map uses a Mollweide equal area projection.