A century of social wasp occupancy trends from natural history collections: spatiotemporal resolutions have little effect on model performance

GALINA M. JÖNSSON, 1,2,3,4 GAVIN R. BROAD, 1 SEIRIAN SUMNER 4 and NICK J. B. ISAAC 3
1Department of Life Sciences, The Natural History Museum, London, UK, 2Grantham Institute, Imperial College London, London, UK, 3UK Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, UK and 4Centre for Biodiversity and Environment Research, University College London, London, UK

ABSTRACT

1. The current dearth of long-term insect population trends is a major obstacle to conservation. Occupancy models have been proposed as a solution, but it remains unclear whether they can yield long-term trends from natural history collections, since specimen records are normally very sparse. A common approach for sparse data is to coarsen its spatial and/or temporal resolution, although coarsening risks violating model assumptions.

2. We (i) test whether occupancy trends of three social wasp (Hymenoptera: Vespidae: Vespinae) species – the common wasp (Vespula vulgaris), the German wasp (Vespula germanica) and the European hornet (Vespa crabro) – have changed in England between 1900 and 2016, and (ii) test the effect of spatiotemporal resolution on the performance of occupancy models using very sparse data. All models are based on an integrated dataset of occurrence records and natural history collection specimen records.

3. We show that occupancy models can yield long-term species-specific trends from very sparse natural history collection specimens. We present the first quantitative trends for three Vespinae species in England over 116 years. Vespula vulgaris and V. germanica show stable trends over the time series, whilst V. crabro’s occupancy decreased from 1950 to 1970 and increased since 1970. Moreover, we show that spatio-temporal resolution has little effect on model performance, although coarsening the spatial grain is an appropriate method for achieving enough records to estimate long-term changes.

4. With the increasing availability of biological records, the model formulation used here has the potential to provide novel insights by making use of natural history collections’ unique specimen assemblages.

Key words. Bayesian occupancy models, biodiversity, distributions, imperfect detection, long-term trends, natural history collections, presence-only, species occurrence, Vespinae.

Introduction

Globally declining insect populations have been flagged as a cause for concern by scientists for over a decade (e.g., Conrad et al., 2006; Franzén & Johannesson, 2007; Harmon et al., 2007; Goulson et al., 2008; Shortall et al., 2009), but have only recently caught the public’s attention due to worldwide media coverage in response to alarming new studies. Such studies include Hallmann et al. (2017), suggesting that German insect biomass declines exceed 75% over 27 years, and Sánchez-Bayo & Wyckhuys (2019), who warn that 40% of global insect species may become extinct within a few decades. However, although global insect declines are widely acknowledged, such magnitudes are unlikely to represent all insect
species (Cardoso et al., 2019; Goulson, 2019; Hambler & Henderson, 2019; Thomas et al., 2019; Saunders et al., 2020; van Klink et al., 2020; Wagner, 2020). Responding to the continued degradation of insect populations and ecosystem services necessitates robust population trends over long temporal scales as well as large spatial scales, since drivers of change are more likely to be confounded for short-term local changes (Parmesan & Yohe, 2003; Root et al., 2003).

Methods for estimating insect trends traditionally relied on systematically collected data from extensive monitoring schemes, whilst opportunistic presence-only records are the most abundant type of existing biodiversity data (Soberón & Peterson, 2005; Elith et al., 2006; Kissling et al., 2018). Presence-only records detail where and when species were present but lack information about species’ absences. As the recording method is rarely known, presence-only records are subject to inherent biases including: irregular recording intensity over time, uneven spatial coverage, varied sampling effort per sampling occasion, and uneven detectability among species and habitats (Isaac & Pocock, 2015). Nonetheless, since presence-only records cover longer time periods than systematically collected data (Troia & McManamay, 2016), they have great potential for estimating long-term changes.

Two types of presence-only records are particularly abundant: firstly, those collected in citizen science projects and species recording schemes, and secondly, specimen records held by natural history collections (Boakes et al., 2010; Kissling et al., 2018). The number of such records is rapidly growing due to natural history collections’ increasing commitment to specimen digitisation projects (Graham et al., 2004; Page et al., 2015; Paterson et al., 2016) and a surge in citizen science projects over recent decades (Pocock et al., 2017). The majority of UK species recording schemes were set up after 1970 (Pocock et al., 2015), which postdates by several decades many large-scale transformations of the British landscape (Spencer & Kirby, 1992; Dallimer et al., 2009; Atkinson & Townsend, 2011; Loft et al., 2019). On the other hand, the majority of specimens in museum collections date from the late 1800s to the 1980s (Winker, 2004; Boakes et al., 2010; Brooks et al., 2014; Speed et al., 2018). As different sources of presence-only data vary in temporal coverage, multiple sources can potentially be integrated to produce comprehensive, long-term datasets (Tingley & Beissinger, 2009; Jetz et al., 2012).

Bayesian occupancy models have been shown to produce robust and reliable species trends from presence-only records, even when the data are biased (van Strien et al., 2013; Isaac et al., 2014). Bayesian occupancy models consist of two sub-models: a state sub-model estimating occupancy and a detection sub-model estimating detection probability to account for imperfect species detection and inconsistent recording (MacKenzie et al., 2002; Royle & Kéry, 2007). Occupancy models are now routinely used to describe national-scale trends in biodiversity (Cruickshank et al., 2016; Powney et al., 2019; Termaat et al., 2019; van Strien et al., 2019), as well as multispecies biodiversity indicators (Outhwaite et al., 2015; van Strien et al., 2016). To date, occupancy models have rarely been applied to periods before 1970, reflecting that whilst records from recording schemes are often sparse (Outhwaite et al., 2018), natural history collections’ specimen records are even sparser (Boakes et al., 2010; van Strien et al., 2019). Indeed, van Strien et al. (2019) argued that most decades of the 20th century had insufficient records of Dutch butterflies to estimate the parameters of an occupancy model. However, Outhwaite et al. (2018) showed that certain model specifications make it possible to use sparse data in occupancy models; although the method has not yet been applied to natural history collection specimens.

A common approach to sparse data is to coarsen the spatial and/or temporal resolution of species trends and thereby gain sufficient data for estimating model parameters (e.g., Heessen, 1996; Klapwijk et al., 2013; van Strien et al., 2019). For occupancy models, coarsening either resolution increases the number of records per grid cell and closure period combination (the area and time period over which an occupancy state is estimated; Fig. 1), which theoretically results in more precise estimates of detection probability. However, spatiotemporal coarsening increases the risk of violating the key assumption of occupancy models: no change in occupancy (e.g., through immigration and emigration) within grid cells and closure period combinations (the closure assumption). Hence, there is a trade-off: short closure periods and small grid cells are theoretically best for meeting the closure assumption as they account for fine-scale (meta)population dynamics, whilst long closure periods and large grid cells ensure enough records for estimating detection probability. A remaining question is whether, and to what extent, the closure assumption can be violated by coarsening spatiotemporal resolutions to allow for sparse datasets to be used in occupancy models.

UK social wasps (Hymenoptera: Vespidae: Vespinae) are an ideal model system for exploring questions of long-term change and integrating sparse datasets. There are two main sources of Vespinae presence-only records in the United Kingdom. First, the Bees Wasps and Ants Recording Society (BWARS) collect records based on volunteer and taxonomic expert observations. The majority of records date from 1970 onwards (Fig. 2). Secondly, the Natural History Museum in London (NHM) holds the largest collection of UK Vespinae specimens (see Appendix S1 in the Supporting Information: Table S1.1 for specimen numbers of six British collections) with the majority dating pre-1980 (Fig. 2). BWARS records are considered sparse (Outhwaite et al., 2018) and the NHM records are certainly sparser (Fig. 2), making them suitable for testing whether Bayesian occupancy models can estimate long-term species changes from sparse data.

The common wasp (Vespula vulgaris, Linnaeus 1758), the German wasp (Vespula germanica, Fabricius, 1793) and the European hornet (Vespa crabro, Linnaeus, 1758) all perform important ecosystem services (Sumner et al., 2018), but long-term UK occupancy trends are not well known, and studies of their long-term population dynamics are limited to a few locations in southern England (Archer, 1985, 2001, 2015, 2018; Lester et al., 2017). Through their foraging activity, social wasps contribute to pest control (Richter, 2000; Southon et al., 2019), pollination (Carreck & Williams, 2002) and seed dispersal (Jules, 1996). However, they face numerous threats, including loss of habitat suitable for nesting and agricultural intensification, particularly increased pesticide use (Archer, 2001; Ollerton et al., 2014). Vespa crabro has undergone a northwards range expansion in Britain since the 1970s (Bees,
Wasps & Ants Recording Society, 2019) that has been attributed to climatic warming (Burton, 2003), and anecdotal evidence suggests populations in south-east England declined around the 1950s (Bees, Wasps & Ants Recording Society, 2019). Whilst population sizes of *V. vulgaris* fluctuate markedly from year to year, including periods of sustained population increase (Lester et al., 2017), the distribution of this species is believed to have changed little in England (Edwards & Telfer, 2002). *Vespula germanica* has similar population dynamics in England, although this species is thought to have declined in the 1970s and 1980s (Archer 2001, 2018). Long-term understanding of these species’ occupancy is important for maintaining ecosystem services and for predicting how anthropogenic change will affect populations of these apex predators. This is especially important given the success of social wasps as invasive species in parts of the world: *V. germanica* and *V. vulgaris* have been highly successful colonisers of South America, New Zealand and Australia where they are causing ecological and economic damage (Lester & Beggs, 2019).

Here, we estimate occupancy trends for three Vespinae species in England from 1900 to 2016 and compare multiple Bayesian occupancy models based on an integrated BWARS-NHM dataset with several combinations of spatial resolution (1 × 1 km, 2 × 2 km, 5 × 5 km or 10 × 10 km grid cells) and temporal resolution (1 year, 5 years or 10 years). Note that we are interested in trends rather than absolute occupancy estimates. To compare models, we quantify how the trends, precision and measures of model fit are affected by varying the closure period-length and spatial grain. We aim to (i) test whether occupancy trends of *Vespula vulgaris*, *Vespula germanica* and *Vespa crabro* have changed in England between 1900 and 2016 and (ii) test the effect of varying spatial and temporal resolutions on model performance when the data are very sparse, such as specimen records. Our general hypothesis is that coarser

---

**Figure 1.** Schematic illustrating how (a) sparse presence-only records from nine grid cells over two closure periods can be coarsened to increase the number of records per combination of grid cell and closure period. Data can be (b) spatially coarsened resulting in fewer and larger grid cells, (c) temporally coarsened resulting in fewer and shorter closure periods, or (d) both spatially and temporally coarsened resulting in fewer grid cells and closure periods. Squares indicate grid cells. Wasp heads indicate single records. Source: Reproduced with permission from Big Wasp Survey. 2017. *Vespula germanica* head icon. [Image]. Retrieved from https://www.bigwaspsurvey.org/the-wasps/.

---

**Figure 2.** The number of *Vespula vulgaris* (*n* = 2053), *Vespula germanica* (*n* = 1089) and *Vespa crabro* (*n* = 935) records used in analyses as well as the total number of records used in analyses (*n* = 4077). All records are from England, date between 1900 and 2016, and are held by the Natural History Museum, London (NHM; grey; *n* = 566) or the Bees, Wasps and Ants Recording Society (BWARS; red; *n* = 3511). Note that the y-axis is plotted on a square root scale.

---

© 2021 The Authors. *Insect Conservation and Diversity* published by John Wiley & Sons Ltd on behalf of Royal Entomological Society. *Insect Conservation and Diversity*, doi: 10.1111/icad.12494
resolutions will contain less information but allow model parameters to be estimated more precisely. Our hypotheses for the Vespinae trends are based on previous work demonstrating that ranges (occupancy) of V. vulgaris and V. germanica have both changed relatively little in England (Archer, 2001, 2018; Edwards & Telfer, 2002), and that V. crabo occupancy decreased in the 1950s and increased since 1970 (Bees, Wasps & Ants Recording Society, 2019).

Materials and methods

Data collection and standardisation

We digitised all UK V. vulgaris, V. germanica and V. crabo specimens held by the NHM, totalling 1830 specimens collected between 1884 and 2016. Digitisation involved recording all information on specimen labels including collection date and location to the most detailed level available, as well as georeferencing by assigning point coordinates to all collection locations and estimating associated error radii (for protocol, see Blagoderev et al., 2017). We assigned all point coordinates to 1 x 1 km grid cells and removed records with location error radii greater than 10 km. The identifications of NHM Vespinae specimens were recently verified by M. Archer and GRB.

BWARS granted the use of their V. vulgaris, V. germanica and V. crabo records for these analyses, totalling 1744 records from between 1840 and 2016. These records are UK species observations that specify the observation date and location as grid cells. All records are subject to validation and verification by experts. We removed any records whose locations were less precise than 1 x 1 km grid cells.

For both NHM and BWARS datasets, we only included records from England that dated from 1900 onwards in analyses because the number of records from other countries and earlier years was felt to be insufficient (Supporting Information Fig. S1.1). We included only those record with precise dates.

To test the effect of varying spatiotemporal resolution on model performance, our starting point was a combined NHM + BWARS dataset with record locations and dates defined to 1 x 1 km and 1-day precision, respectively. We coarsened all locations to three additional spatial resolutions: 2 x 2 km, 5 x 5 km and 10 x 10 km grid cells and assigned all record dates to 1-, 5- and 10-year closure periods. The 1 x 1 km grid cell precision provides the greatest number of spatial replicates, 10 x 10 km grid cell precision increases the number of visits (defined as a unique combination of record date and grid cell throughout) per grid cell:closure period combination, whilst 2 x 2 and 5 x 5 km grid cells represents intermediates between the two. Vespa workers normally forage within 400 m of their nest (Akre et al., 1975) and foraging distances of Vespa workers is assumed to be less than 1 km (Edwards, 1980; Matsuura & Yamane, 1990); hence, all grid cell resolutions capture a colony’s range and the spatial closure assumption is reasonable. A 1-year closure period was chosen as the finest temporal scale of estimates. Closure periods of 5 and 10 years were chosen to increase the number of visits within closure period:grid cell combinations. One year is an appropriate temporal unit for meeting the assumption of temporal closure, since UK Vespine populations have one generation per year (Spradbery, 1973). However, 5- and 10-year closure periods likely violate this assumption, which can lead to inflated occupancy estimates (Steenweg et al., 2018). As such, only models at temporal resolutions of 1 year meet the closure assumption.

Since the model formulation used in these analyses discards grid cells that were only visited in a single closure period throughout the time series, we removed all records from grid cells that had not been visited in at least two 10-year closure periods. This ensures that all models, including those at finer temporal resolutions, are based on the same visits and allows for the comparison of models at different spatiotemporal scales.

The standardisation process resulted in 566 and 3511 records from NHM and BWARS, respectively (Fig. 2; see Supporting Information Fig. S1.1 for a workflow of the standardisation process and the number of records removed at each stage). Both the percentage of grid cells that were visited two or more times within closure periods, and the number of repeat visits to such sites increased with coarser spatiotemporal resolutions, such that there is a 16-fold and a two-fold increase, respectively, when comparing the finest and the coarsest spatiotemporal resolutions (1 km + 1 year vs. 10 km + 10 years; Supporting Information Table S1.2). The geographical coverage of records is biased towards south-east England, indicating that the result of this study may not be representative for all of England (Fig. 3).

Model choice

Our occupancy model builds on the work of Outhwaite et al. (2018) that enabled the use of sparse data by including a random walk in the state sub-model, allowing information to be shared across closure periods. We use a detection sub-model based on that of Van Strien et al. (2013), which estimates separate detection probabilities for three categories of visits as independent realisations of the same underlying state. These authors account for varying sampling effort by categorising visits based on the number of species observed per visit. We modified this approach by dividing the NHM + BWARS dataset into three categories based on data source: (i) BWARS 1 sp. visits, that is, BWARS visits with one observed species; (ii) BWARS 2-3 spp. visits, that is, BWARS visits with two or three observed species; and (iii) NHM visits, that is, all NHM visits regardless of the number of observed species. This classification accounts for potential differences in the recording process between the NHM and BWARS data, and for inconsistent sampling effort among two categories of the BWARS data (see Supporting Information Appendix S2 for details on the model formulation).

Model fitting and evaluation

We fitted Bayesian occupancy models for every combination of grid cell size (1 x 1 km, 2 x 2 km, 5 x 5 km or 10 x 10 km grid cells) and closure period length (1 year, 5 years or 10 years), totalling 12 models per species (see Table 1 for sample
All models were fitted in a Bayesian mode of inference using JAGS (version 4.3.0; Plummer, 2009) through the R package ‘sparta’ (version 0.2.06; August et al., 2019) in R (version 3.5.2; R Core Team, 2018) and we selected the random walk model with half-Cauchy hyperpriors. All models were fitted with three Markov chain Monte Carlo chains for 50 000 iterations each with the first 25 000 iterations being discarded as “burn-in”. We used the Gelman–Rubin statistic (Rhat; Gelman & Rubin, 1992) to determine convergence of occupancy estimates, where Rhat-values below 1.1 were considered satisfactory (Kéry & Schaub, 2012). The four models estimating V. germanica occupancy at the 1-year temporal resolution did not converge and are therefore not presented here.

We used three metrics to evaluate model performance: occupancy trends, detection probability ($\alpha_t$ in Eqn. S2.5) trends and the deviance information criterion (DIC; Spiegelhalter et al., 2002). DIC was chosen over methods that penalise the deviance by the absolute number of parameters as opposed to the number of effective parameters, since these can suffer from boundary effects (Bolker et al., 2009). As the number of parameters increases with finer spatiotemporal resolutions (Supporting Information Fig. S3.2), we also compared deviance and the effective number of parameters, as calculated using the R package ‘AICcmodavg’ (version 2.2-2; Mazerolle, 2019).

Results

Varying spatiotemporal resolutions in Bayesian occupancy models showed little effect on trends of occupancy and detection probability for all three species, as evident from overlapping 95% credible intervals (CIs) throughout the time series (Figs. 4 and 5). Similarly, 95% CIs show that spatiotemporal resolution has little effect on the precision of occupancy and detection probabilities.

Figure 3. The distribution and source of English (a) Vespula vulgaris, (b) Vespula germanica and (c) Vespa crabro records used in analyses. Each square represents a 10 × 10 km grid cell (n = 282) coloured according to (d) the total number of species recorded per grid cell between 1900 and 2016, and (e) the total number of records per grid cell between 1900 and 2016 (n = 4077). All records are from the Natural History Museum, London (NHM; n = 566) or the Bees, Wasps and Ants Recording Society (BWARS; n = 3511).
probability estimates; however, precision of both estimates increases throughout the time series, particularly after 1970 (Figs. 4 and 5). Longer closure periods tend to smooth out trends and are unable to detect short-term changes in occupancy as well as detection probability (Figs. 4 and 5).

All models show that neither *V. vulgaris*, *V. germanica* nor *V. crabro* have experienced statistically significant changes in occupancy trends when comparing 1900 to 2016 (Fig. 4; Supporting Information Table S3.3). Nonetheless, all models apart from those at a 10 × 10 km spatial resolution show a statistically significant increase in *V. vulgaris* occupancy of 26% (95% CI: 3.2–64%) from 1940 to 1980 when averaged across models (Fig. 4a; Supporting Information Table S3.4). Seven of 12 *V. crabro* models indicate an absolute occupancy increase from 0.29 in 1920 to 0.61 (95% CI: 0.36–0.85) in 1950, when averaged across models (Fig. 4c; Supporting Information Table S3.5). Since 1950, *V. crabro* shows an occupancy decline followed by an increase, which are both statistically significant for all models: the occupancy decreased from 0.61 (95% CI: 0.36–0.85) in 1950 to 0.18 (95% CI: 0.08–0.37) in 1970, and increased to 0.66 (95% CI: 0.45–0.84) in 2016, when averaged across models (Fig. 4c; Supporting Information Table S3.5).

None of the three species showed statistically significant change in detection probability between 1900 and 1990, whilst models at coarse temporal resolutions generally showed decreasing detection probability among *Vespa* species since 1990 (Fig. 5; Supporting Information Table S3.6). Between 1990 and 2016, all models for *V. vulgaris* with closure periods greater than 1 year, and models for *V. germanica* with 10-year closure periods on average showed decreasing detection probabilities of 29.9% (95% CI: 54.6–15.1) and 52.8% (95% CI: 64.2–39.3), respectively (Fig. 5a,b; Supporting Information Table S3.6). No statistically significant change was observed for *V. crabro* between 1990 and 2016 (Fig. 5c; Supporting Information Table S3.6).

For *V. vulgaris* and *V. germanica*, models with coarse spatio-temporal resolutions result in lower DIC-values, whereas for *V. crabro*, finer spatio-temporal resolutions give lower DIC-values (Fig. 6). For all species, finer spatio-temporal resolutions increase the number of effective parameters (Supporting Information Fig. S3.2), whilst decreasing model deviance (Fig. S3.3).

### Discussion

Bayesian occupancy models can yield long-term occupancy trends by integrating sparse presence-only data from natural history collections and species recording schemes. Our findings add to a nascent literature on long-term (>50 year) change in insect biodiversity from museum data (Habel et al., 2019; Soroye et al., 2020; Zattara & Aizen, 2021), although to our knowledge this is the first to present century-long trends at the species level. As evident from the congruence across models, coarsening the spatio-temporal resolution of model estimates has little effect on overall occupancy trends throughout species’ time series although information regarding short-term changes are lost. According to DIC-values, the optimal spatio-temporal resolution is species-specific.

We present the first quantitative *V. vulgaris*, *V. germanica* and *V. crabro* trends in England before the 1970s (Fig. 4). Our models suggest that *V. vulgaris* and *V. germanica* trends have remained stable in England from 1900 to 2016, apart from a previously undocumented 26% (95% CI: 3.2–64%) increase in *V. vulgaris* occupancy from 1940 to 1980. This increase was detected in all models at a finer spatial resolution than 10 × 10 km grid cells, further suggesting that coarse spatial resolutions hold less information and tend to smooth out trends. These results provide no support for the suggestion of differential changes in abundance of *V. vulgaris* (stable) and *V. germanica* (declining) in the United Kingdom during the 1970s and 1980s (Archer, 2001). One explanation is that Archer may have observed a decline in local abundance without a loss of range size. On the other hand, *V. crabro* occupancy trends show considerable fluctuations throughout the time series. Two such
fluctuations are especially prominent: a decrease from 1950 to 1970, and an increase from 1970 to 2016 (Fig. 4c), which are both in line with previous anecdotal evidence describing the same trend (Bees, Wasps & Ants Recording Society, 2019). At coarse temporal resolutions, *Vespula* species show decreasing detection probabilities since 1990 that is not observed for *V. crabro* (Fig. 5; Supporting Information Table S3.6). This trend is consistent with the hypothesis that naturalists are less likely to record relatively common species, as suggested elsewhere from results of citizen science projects (Isaac & Pocock, 2015), and is worthy of further investigation.

The contrasting patterns between the *Vespa* and *Vespula* occupancy trends may be explained by differences in diet and habitat. A possible explanation for why *V. crabro* declined between 1950 and 1970 whilst *V. vulgaris* and *V. germanica* largely remained stable is that both *V. vulgaris* and *V. germanica* are habitat and dietary generalists (Edwards & Telfer, 2002; Harris, 1991), whereas *V. crabro* is associated with ancient woodlands and farmland, where it nests in mature tree hollows (Edwards, 1997). These land-cover types were disproportionately affected by anthropogenic pressures in the mid-20th century (Peterken & Harding, 1975; Spencer & Kirby, 1992; Robinson & Sutherland, 2002; McKerman & Goldberg, 2011), conceivably causing the observed *V. crabro* decline. An alternative, but not mutually exclusive, explanation is that *V. crabro* was more severely affected by a series of cold and wet springs between 1950 and 1970 (Mayes & Wheeler, 1997). Since *V. crabro* generally has a more southerly distribution than the two *Vespula* species (Edwards, 1980; Pekkarinen, 1989), it is potentially more negatively affected by cold temperatures; however, these same variables have been identified as having significant effects on *Vespula* populations, with a particular strong negative effect of cold springs on three populations of *V. vulgaris* studied over 39 years (Lester et al., 2017). Extending our approach to more species would permit tests of hypotheses about habitat and diet as factors mediating responses to environmental change over centennial time scales, and allow us to address questions about collection behaviours, e.g., whether the recording of common species is decreasing.

Our models suggest particularly interesting pattern of change for the hornet, *V. crabro*. *Vespa crabro* was largely restricted to southern England when its occupancy plummeted in the mid-20th century (Phillips & Roberts, 2010) and the subsequent increase in occupancy from 1970 returned occupancy to approximately that of the 1950s (Fig. 4c). This period of expansion is believed to be associated with climate warming (Burton, 2003). Supporting this hypothesis, we found a northward shift in the range margin of *V. crabro* during recent decades that is not

matched by Vespula species (Supporting Information Fig. S3.4), suggesting that the majority of grid cells colonised during the expansion period were not the same grid cells that were lost during the earlier decades. This result is consistent with the notion that, since 1970, V. crabro’s range has been expanding, whilst simultaneously thinning, as reported in some butterfly and moth species (Wilson et al., 2004; Mair et al., 2012; Fox et al., 2014).

It is therefore surprising that we observe almost no difference in the occupancy trends of V. crabro when measured at different spatial grains (Wilson et al., 2004). These results raise questions that could be addressed using a spatially explicit model of changes in the hornet’s distribution during the last century.

Northward shifts in range are also expected for Vespula populations (Sorvari, 2018). As our analysis was limited to records from England, we are unable to suggest whether there have been occupancy changes at the northern edges of the ranges of V. germanica or V. vulgaris.

For the three species tested here, we demonstrate that occupancy models can yield fine-grained long-term species-level trends by integrating records from natural history collections and monitoring schemes. Moreover, spatiotemporal resolution has little effect on occupancy trends as well as associated precisions regardless of whether the closure assumption is met.
However, violating the closure assumption by coarsening the temporal resolution, smooths short-term occupancy variation. Our models make contrasting assumptions about the sampling process and employ different strategies for the trade-off between meeting the closure assumption and having sufficient records for estimating detection probability. Hence, the observed congruence across models’ occupancy trends serves as an internal validation indicating sufficient power to estimate the detection probability, despite models having up to a 16-fold difference in the percentage of grid cells visited twice or more within closure periods (Supporting Information Table S1.2). Uncertainties with regard to the sampling protocols underlying natural history collections have long been considered a major obstacle to understating long-term species change (Elith et al., 2006; Tingley & Beissinger, 2009; Kissling et al., 2018), but our results demonstrate the potential of detection sub-models to satisfactorily account for such uncertainties. Moreover, as a result of data deficiency, studies estimating long-term species trends from natural history collections often make estimates over aggregated multi-decade periods (e.g., Davidson et al., 2001; Myers et al., 2009; van Strien et al., 2019). We produced long-term trends at the finest spatiotemporal scale tested (1 x 1 km grid cells and 1-year closure period) for two of the three species (V. germanica trends showed poor convergence at 1-year closure periods), demonstrating that Bayesian occupancy models can yield fine-grained annual occupancy trends from natural history collections. However, estimates of occupancy and detection probability for all species are more precise after the 1970s, as evident from narrower CIs (Figs. 4 and 5). Since only 15% of the total number of records pre-date 1970, but comprise 60% of the total time series, this result is hardly surprising and further suggests that reasonably precise trends can be obtained from sparse specimen data.

The congruence among models highlights that occupancy trends alone cannot single out an objectively superior spatiotemporal resolution. According to DIC-values, coarse spatiotemporal resolutions are favoured for estimating V. vulgaris and V. germanica occupancies, whilst the opposite is true for V. crabro (Fig. 6). DIC components show the same trend for all three species such that spatiotemporal coarsening decreases the models’ number of effective parameters and increases deviance (Supporting Information Figs. S3.2–3). Hence, for species with relatively stable occupancies throughout the time series (i.e. V. vulgaris and V. germanica), increasing the number of parameters by coarsening the spatiotemporal resolution does not provide additional information to counter the decreased deviance. Conversely, coarsening the spatiotemporal resolution for V. crabro occupancy results in a measurable loss of information as evident from these models being least preferred by DIC-values. In contrast with V. vulgaris and V. germanica, the occupancy of V. crabro shows considerable year-to-year variation throughout the time series so additional effective parameters allow interannual variation to be described by models. As a result of additional parameters adding information, the deviance of V. crabro models at fine temporal resolutions is penalised to a lesser extent by an increase in the number of effective parameters than for species with stable occupancies.

In order to identify a single superior spatiotemporal resolution for estimating long-term occupancy trends, it is important to differentiate between coarsening spatially versus temporally. According to DIC-values, the optimal spatiotemporal resolution depends on the degree to which occupancy varies throughout species’ time series and can therefore only be identified by running multiple models, which can be time-consuming and computationally expensive. We advise against using DIC-values as an objective measure of model fit and instead propose an approach to identify the ideal spatiotemporal resolution that is based on the research question. For answering questions concerning trends over time with sparse data that requires coarsening to estimate model parameters, we suggest that models should maximise the number of closure period parameters by using short closure periods to avoid concealing potential short-term variation. Reducing the number of grid cell parameters by coarsening the spatial resolution does not incur the same cost of concealing short-term variation, although it might overestimate occupancy (Steenweg et al., 2018).

Our results indicate that when very sparse data are used to estimate long-term trends, coarsening the spatial grain is a promising method for achieving sufficient records to estimate parameters in Bayesian occupancy models. However, the model formulation used here assumes that relative detection probabilities, and in effect the record collection processes, are constant among the three data type categories throughout the time series. As previous work indicates that record collection processes may differ between natural history collections and species monitoring schemes (Boakes et al., 2010; Speed et al., 2018), testing these assumptions is a priority for future research.

Our results show that Bayesian occupancy models can yield long-term species trends by spatiotemporally coarsening very sparse data, regardless of whether the closure assumption is met. However, a trade-off exists where coarsening the temporal resolution risks losing information about short-term interannual variation. We argue that one should not blindly follow DIC-values and propose that the most appropriate spatiotemporal resolution will depend on the question asked and data available. Very sparse data should preferentially be spatially coarsened when estimating long-term species trends. The current dearth of such long-term trends is considered a major cause of the observed gradual shift in the accepted norm of the state of biodiversity. This phenomenon, termed shifting baseline syndrome (Pauly, 1995), is considered a major obstacle to conserving biodiversity (Papworth et al., 2009; Soga & Gaston, 2018). For example, our results imply that climate change and large-scale land-use change do not need to be invoked as drivers of change in V. vulgaris and V. germanica population trends, which contrasts with previous studies (Archer, 2001, 2018; Lester et al., 2017; Bees Wasps & Ants Recording Society, 2019), whilst both drivers seemingly affect V. crabro occupancy trends. Given that these drivers are expected to intensify in the future, understanding their separate and interacting effects on social wasps is crucial for maintaining the ecosystem services they provide. Moreover, expanding populations may come with economic and social costs if species become invasive, as evident from expansive, high-density Vespa populations in New Zealand (Lester & Beggs, 2019). Hence, the effective control of invasive species also requires understanding of drivers of change, ideally on a national or global scale.
which in turn calls for appropriate statistical integration of multiple datasets such as specimen data and citizen science data presented here. With the increasing availability of presence-only records (including digitised natural history collection specimen records; Graham et al., 2004; Page et al., 2015) from international databases such as the Global Biodiversity Information Facility (GBIF; Troudet et al., 2017), the Bayesian occupancy model formulation used here has the potential to provide novel insights into the numerous species and time periods that were previously considered data deficient.

ACKNOWLEDGEMENTS

The authors would like to acknowledge all volunteers and collectors who collected the data used within this research. The authors thank the committee and members of BWARS (the UK Bees, Wasps and Ants Recording Society) for access to their data holdings, and the trustees of the Natural History Museum, London, for their support in maintaining the invaluable collections the museum holds. G.M.J. is supported by a Natural Environment Research Council studentship, award number: NE/L002515/1. N.J.B.I. is funded by Natural Environment Research Council award NE/R016429/1 as part of the UK-SCAPE programme delivering National Capability.

Data availability statement

All scripts used to generate the analyses presented in this paper are available via https://github.com/galinajonsson/HistoricalWaspOccupancies. The georeferenced Natural History Museum specimen data used in analyses is available via https://doi.org/10.5519/0082541. The model outputs are available via https://doi.org/10.5519/0055095. The Bees, Wasps and Ants Recording Society (BWARS) data is available on request from www.bwars.com/wasp/vespidae/vespinae/vespa-crabro> 1st November 2018.


Accepted 16 March 2021

Editor: Simon Leather, Associate Editor: Sandra Rehan