
Sánchez-Fernández, David; Fox, Richard; Dennis, Roger L.H.; Lobo, Jorge M. 2021. **How complete are insect inventories? An assessment of the British butterfly database highlighting the influence of dynamic distribution shifts on sampling completeness.** *Biodiversity and Conservation*, 30 (3). 889-902. https://doi.org/10.1007/s10531-021-02122-w

How complete are insect inventories? An assessment of the British butterfly database highlighting the influence of dynamic distribution shifts on sampling completeness.

David Sánchez-Fernández[1]. Richard Fox[2], Roger L.H. Dennis[3,4,5] & Jorge M. Lobo[6*]

[1]Departamento de Ecología e Hidrología, Universidad de Murcia, Murcia, 30100, Spain

[2]Butterfly Conservation, Manor Yard, East Lulworth, Dorset BH20 5QP, UK

[3]Centre for Ecology, Environment and Conservation, Department of Biological and Medical Sciences, Oxford Brookes University, Oxford OX3 0BP, UK

[4] The School of Life Sciences and Education at Staffordshire University, Staffordshire University, Stoke-on-Trent, ST4 2DF, U.K

[5]UK Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Wallingford, Oxon, OX10 8BB, UK

[6]Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain

*corresponding author: jorge.lobo@mncn.csic.es

Abstract

Much recent scientific, media and public attention has focussed on the evidence for and consequences of declines in insect biodiversity. Reliable, complete inventories can be used to estimate insect trends accurately, but incomplete data may distort assessments of biodiversity change. Thus, it is essential to understand the completeness of insect inventories. Assessing the database of Great Britain butterfly occurrences, likely the most complete database for any group of insects in the world (with 10,046,366 records for 58 butterfly species), we found that only 62% of the cells have complete inventories at the finest scale evaluated. The dynamic nature of butterfly distributions in response to climate change could explain this result, as the distribution of completeness values is related to the increasing occurrence of some species at higher latitudes as a consequence of recent range expansions. The exceptional quantity of information collected in Great Britain about this appealing group of insects is insufficient to provide a complete picture. Consequently, we cannot expect to build complete inventories for less popular taxa, especially in less comprehensively sampled countries, and will require other techniques to understand the full extent of global biodiversity loss.

Keywords Completeness · Great Britain · KnowBR · Climate change · Butterflies · Biodiversity inventory.

Introduction

The study of the distribution of biodiversity at different spatial and temporal scales has long been one of the main foci of disciplines such as ecology and biogeography (Cox and Moore 2004). Reliable descriptions of species distributions are critical for obtaining a better understanding of the causes underlying biodiversity patterns (Gaston 2000), for assessing the impact of global change (Powney and Isaac 2015), for producing efficient conservation plans (Meyer et al. 2015), and for directing action towards multiple Aichi targets (Pereira et al. 2013). Unfortunately, it is well known that our knowledge of the geographical distribution of biodiversity remains, in general, taxonomically and geographically biased (the so-called Linnaean and Wallacean Shortfalls; Brown and Lomolino 1998; Lomolino and Heaney 2004; Whittaker et al. 2005). The number of studies on large-scale diversity patterns has rapidly increased in the last two decades in order to overcome the Wallacean Shortfall (Beck et al. 2013), facilitated by the development of biodiversity information networks, such as the Global Biodiversity Information Facility (Edwards et al. 2000). However, despite most of these biological databases being taxonomically exhaustive, they lack comprehensive information on survey effort, making it difficult to determine which territories have reliable species inventories (Ball-Damerow et al. 2019). A direct consequence of the data limitations for biogeographical and conservation analyses is that it becomes extremely difficult to determine whether the apparent absence of a species in a spatial unit reflects its actual absence or is the result of insufficient survey effort. Data limitations also generate maps of observed species richness that often suspiciously resemble maps of the number of records per spatial unit (Hortal et al. 2007). Therefore, databases of point occurrences provide very limited and spatially-biased species inventories for most taxa, except in a few well-sampled regions and data limitations are rather the rule than the exception (Meyer et al. 2015). Only a few countries, typically with a longstanding tradition of natural history study by amateur enthusiasts and sufficient resources, such as Great Britain, are able to produce good distribution maps based on adequate sampling for a range of taxonomic groups, including invertebrates (Griffiths et al. 1999; Meyer et al. 2015; Pocock et al. 2015). However, even comprehensive databases of species occurrences in these intensively-surveyed regions are prone to geographic (Yang et al. 2013) and taxonomic biases (Soberón et al. 2007; Pyke and Ehrlich 2010).

Although insect populations are intrinsically dynamic (Hengeveld 1992), much scientific, media and public attention has focussed on recent declines in insect biodiversity (Goulson 2019; Habel et al. 2019; Cardoso et al. 2020), the consequences of with calls for immediate policy responses (Forister et al. 2019; Harvey et al. 2020). Others, however, have cautioned against overextrapolation from the limited current evidence (Saunders et al. 2020), noting the heterogeneity of insect responses (Wagner et al. 2021) and highlighting the need for more data, especially from the tropics (Montgomery et al. 2020). In order to estimate insect trends accurately and thus better to understand the full extent of global biodiversity loss, it is thus essential to assess the completeness of insect inventories. To date, most studies of faunistic databases have reported a dearth of complete and extensive inventories for insect taxa (e.g., Romo et al. 2006; Sánchez-Fernández et al. 2008; Santos et al. 2010; Bruno et al. 2012; Ballesteros-Mejia et al. 2013; Fattorini 2013; Lobo et al. 2018). Within insects, diurnal Lepidoptera are expected to be affected by under-sampling to a lesser degree than other taxa (Troudet et al. 2017), likely due to their relatively large size and aesthetic appeal. Butterflies have traditionally been popular subjects of study for the general public as well as for scientists, and are acknowledged as important flagships for insect conservation (Barua et al. 2012). Yet, a recent study using a dataset of over 19 million species occurrence records, identified knowledge gaps in butterfly biodiversity inventories at a global level (Girardello et al. 2019); while inventory completeness is relatively good for the Global North (especially some European countries), major gaps exist in the Southern Hemisphere (particularly in tropical areas). Furthermore, most of these completeness assessments have been conducted at a relatively coarse spatial resolution (at least 50x50 km). While of value in describing broad scale biodiversity patterns, the information obtained from these studies is too coarse for practical conservation planning applications (Hopkinson et al. 2000). High quality data from a broad spectrum of taxa and regions based on fine-grained and intensive collection and assessment of insect inventories are essential for reliable macroecological and conservation analyses (Beck et al. 2012). The challenge now is to be able to identify areas with reliable inventories for insects at fine spatial scales. Currently, only the combination of data on charismatic insects such as butterflies compiled in countries with a long tradition in natural history such as Great Britain would meet this objective. Despite the fact that the quality of data on Great Britain butterflies has been extensively studied during recent decades (e.g. Dennis et al. 1999; Dennis and Thomas, 2000), and measures adopted in analyses of the data to account for spatiotemporal variation in sampling effort (e.g. Carvalheiro et al. 2013; Macgregor et al. 2019), no fundamental assessment of sampling has been made to determine the completeness of this inventory.

In this study, we examine the database of Great Britain butterfly occurrences, likely the most complete dataset for any group of insects in the world and which has been widely used to examine the role of climate change on butterfly distributions (e.g. Warren et al. 2001; Hill et al. 2002; Menéndez et al. 2006). We explore the spatial distribution of completeness values at different spatial resolutions and identify those spatial units that can be considered as sufficiently surveyed. We additionally examine if spatial variation in the distribution of completeness values may be partially explained by the temporally dynamic nature of species distributions driven, in part, by anthropogenic climate change (Parmesan and Yohe, 2003; Chen et al. 2011; Ripple et al. 2020.

Material and methods

Species selection and data sources

We analysed a database containing Great Britain records of species from the families Hesperiidae, Lycaenidae, Nymphalidae, Papilionidae, Pieridae and Riodinidae for the period 1800-2014. Only resident species were considered, but *Phengaris arion* was excluded (due to a long-standing programme of reintroductions), while *Vanessa atalanta* was included as it now appears to be resident year-round in Great Britain (Fox and Dennis 2010). This database has been generated by the Butterflies for the New Millennium project operated by Butterfly Conservation. The project commenced in 1995, but has accumulated a substantial volume of historical records (Asher et al. 2001; Fox et al. 2015), notably from a previous recording scheme that led to the first butterfly atlas of Britain and Ireland (Heath et al. 1984). The Butterflies for the New Millennium database principally comprises butterfly occurrence records (unique combinations of species x recorder x location x date) from opportunistic, non-standardised sampling by citizen scientists, either submitted directly to the Butterflies for the New Millennium project or compiled from other sources including iRecord and iNaturalist. The Butterflies for the New Millennium also includes records from the structured sampling of the UK Butterfly Monitoring Scheme. Prior to being collated into the Butterflies for the New Millennium database, all records are verified by a network of expert volunteers. Most Butterflies for the New Millennium records, especially those since 1995, contain fine-scale spatial and temporal information, enabling their use in biodiversity conservation. However, for our analysis records were binned to specific combinations of species/10km x 10km grid cell/year. The centroid of each 10km x 10km cell based on the Great Britain Ordnance Survey National Grid was translated to geographic coordinates (latitude/longitude) following the World Geodetic System (WGS84).

Completeness of the database

KnowBR (Lobo et al. 2018; Guisande and Lobo 2019) was used to examine the accumulation in the number of species with the addition of database records. These accumulation curves were performed simultaneously for all the Great Britain cells for five different spatial resolutions (grid cells at 5’, 10’, 20’, 30’ and 60’) applying the exact estimator proposed by Ugland et al. (2003). KnowBR automatically estimates the final slope of the accumulation curve for each cell, the completeness of each cell inventory, and the ratio between the number of database records and the number of species. Completeness values (i.e. the percentage of species that have been inventoried) were calculated after adjusting the accumulation curves to a rational function (see Lobo et al. 2018 for details), extrapolating the asymptotic values of all these curves to estimate the probable number of species in each cell when the number of records tends towards infinity. In our case, well surveyed cells are considered as those that have a final slope in the accumulation curve ≤0.01 (one new species added to the inventory for each 100 new database observations), a completeness value ≥ 90%, and at least 20 times more database records than species. These values seem reasonable thresholds to consider a grid cell as complete (well surveyed cells). The results provided by KnowBR at the different resolutions considered are available as Supplementary Information.

Locally uncommon species and recent species dispersal

As the results of species accumulation curves are greatly influenced by the proportion of locally uncommon and/or transient species (Thompson et al. 2003) we estimate the number and proportion of species in each cell with five or fewer database records (hereafter “locally uncommon species”). The number of locally uncommon species was correlated against completeness values using the Spearman rank correlation coefficient (*rs*). The difference in the number and proportion of locally uncommon species among well surveyed cells and insufficiently surveyed cells was also assessed by using Mann-Whitney U tests. Additionally, as some Great Britain butterflies are expanding northwards in response to climate change (Hill et al. 2002; Hickling et al. 2006), these recent and quite limited occurrences may influence completeness estimates. To examine this question, the first year in which each species is observed in the northernmost cell was determined, and the average date of locally uncommon species computed for each cell (mean year at which these species attain their maximum latitude).

Results

The database contains a total of 10,046,366 records (species/locality-cell/year) for 58 species belonging to six families (Table 1). None of the species has been recorded less than 1000 times, with *Maniola jurtina* (Nymphalidae) and *Pieris rapae* (Pieridae) being the most recorded (with more than 700,000 records each; see Fig. 1). Fifty-four percent of total database records include species from the family Nymphalidae. The mean number (± SD) of species and database records per cell (5' resolution) is 26.0 ±11.0 and 3620 ± 6736, respectively. The geographical distribution of observed species richness is positively correlated with the number of database records ($rs = 0.869$, $p<0.001$) having a clear latitudinal pattern (Fig. 2).

At the highest resolution assessed (5' cells), 62% of total cells may be categorized as well surveyed cells (Table 1; Fig. 3). Completeness and location of well surveyed cells suggest that at the 5' resolution many cells in Scotland, as well as some in Wales and north-west England would be regarded as insufficiently surveyed (Fig. 3). An increasing percentage of cells can be considered well surveyed cells as the size of spatial units is increased (Table 2; Fig. 3). For example, almost 90% of Great Britain cells at 30' (≈60 x 60 km) can be considered as well surveyed cells. In the same way, a fit of the % of well surveyed cells against cell resolution using the Morgan-Mercer-Flodin growth function predicts that only 30% of total cells can be considered as well surveyed cells when the cell resolution is 1 minute (4 km^2, approximately; see Fig 4).

Both the number and the percentage of locally uncommon species in the cells are negatively correlated with completeness values (*rs* = -0.492 and -0.870, respectively; $p<0.001$) (Fig. 5). The number of locally uncommon species differs between well surveyed cells and insufficiently surveyed cells ($U = 15.41$, $n_1 = 1054$, $n_2 = 1721$, $p <0.0001$) as well as the proportion of locally uncommon species ($U = 39.17$, $n_1 = 1054$, $n_2 = 1721$, $p <0.0001$), so that well surveyed cells harbour a significantly lower proportion of locally uncommon species (mean ± sd; 19.5 ± 7.8 %) than the remaining cells (55.5 ± 24.8 %). On the other hand, completeness percentages are also negatively correlated with the values of the mean year at which locally uncommon species attain their maximum latitude (*rs* = -0.458, $p<0.001$). In the same way, the mean year at which locally uncommon species attain their maximum latitude significantly differ between well surveyed cells and insufficiently surveyed cells ($U = 26.33$, $n_1 = 1049$, $n_2 = 1716$, $p <0.0001$); insufficiently surveyed cells harbour locally uncommon species reaching their maximum latitude more recently (year 1978.4 ± 16.1) than well surveyed cells (1951.5 ± 26.2).

Discussion

The results clearly indicate that the Wallacean shortfall emerges even in one of the most complete distributional databases for insects in the world. This result outlines the difficulty of obtaining reliable inventories of insects at fine scales relevant to conservation, as less than 2/3 of the total territory can be considered complete even after more than 200 years of field sampling resulting in more than 10 million records for just 58 species. Of course, these results are dependent on the thresholds selected for establishing those cells recognized as well-surveyed. In our opinion, the criteria applied (one new species recorded when 100 database observations are added, completeness values ≥ 90%, and at least 20 times more database records than species) are reasonable and not excessively exigent. Similarly, the number and proportion of cells identified as being well-surveyed depends on the level of resolution selected; the proportion diminishes exponentially at cell resolutions finer than ≈40 x 40 km. Thus, the number of localities that can be recognized as harbouring reliable inventories diminishes drastically at resolutions closest to home ranges of butterflies, those at which management decisions are being made (Brakefield 1982; Mallet 1986). At the lowest considered resolution (≈ 100 km^2) more than one third of Great Britain territory does not meet our criteria for being considered as well-surveyed.

For other taxonomic groups in other regions worldwide, results obtained have been even worse, where only small fractions of the spatial units examined provide relatively complete inventories. For example, another relatively well surveyed area is the Iberian Peninsula. In this area, some studies have evaluated the completeness of the inventories of water beetles (Sánchez-Fernández et al. 2008), dung beetles (Lobo and Martín Piera 2002), and butterfly species (Romo et al. 2006), revealing that less than a third of the Iberian and Balearic 50 × 50 km grid cells can be considered as relatively well surveyed areas (90% completeness).

If, in an exceptional case such as Great Britain, the quantity of information collected about the species of an appealing group of insects is not able to provide a complete picture, there can be little expectation that biodiversity distribution patterns for less popular taxa can be resolved accurately, especially in countries with lower levels of sampling. Two complementary options exist to tackle this knowledge gap. First, sampling effort could be directed to increase the number of well surveyed cells, particularly in order to represent the full geographical extent and environmental heterogeneity of the chosen territory (Hortal & Lobo, 2005; Sánchez-Fernández et al.

2008; Varela et al. 2014). Techniques used in the current study are valuable for identifying geographical areas, at different scales, where attention is required. Second, modelling techniques can be used to determine the probability of presence/absence of species in absence of exhaustive information (Peterson et al. 2011). Species distribution modelling has been extensively used in ecology, biogeography and biodiversity conservation to predict occurrences using existing (albeit incomplete) records of taxa correlated with environmental variables such as land cover and climate data (Guisan and Thuiller 2005; MacKenzie et al. 2006; Norberg et al. 2019). Additionally, modelling approaches have been developed to predict distributions of communities of species, including Great Britain butterflies (Dapporto et al. 2015; Ovaskainen et al. 2016), and for assessing temporal trends in species distributions using incomplete (and spatiotemporally biased) occurrence data (e.g. Isaac et al. 2014; Dennis et al. 2017; Outhwaite et al. 2018).

The completeness pattern found in this study is simple and homogeneous, independent of spatial scale: northern cells appear recurrently as insufficiently surveyed at all resolutions. While likely causes include low human population density and mountainous terrain, this latitudinal pattern could be an artefact caused by the increasing occurrence of new species in cells due to climate-driven northward range expansions. Completeness values derived from accumulation curves are highly dependent on the number of locally uncommon species (Chao et al. 2009), and our results suggest that a significantly lower number of locally uncommon species appear in well-surveyed cells. Furthermore, the cell average year at which locally uncommon species attain their maximum latitude differs between well-surveyed and insufficiently surveyed cells. These two results, together with the statistically significant relationship between the mean year at which locally uncommon species attain their maximum latitude and completeness values, suggest that the dynamic nature of butterfly distributions in response to climate change (Hill et al. 2002; Hickling et al. 2006; Menéndez et al. 2006) could frustrate attempts to delimit complete inventories when the data are collected over a long period of time. Hence, the dynamic nature of species assemblages (Hengeveld 1992), particularly when habitats and climate are changing rapidly as a consequence of human actions (Ripple et al. 2020), and our incapacity to provide reliable snapshot estimations of the distribution of a large portion of biodiversity, could prevent the use of the available occurrence data for accurate research for basic and applied purposes.

Although the primary data of biodiversity databases may be used to improve the available distributional knowledge of a species (e.g., Dennis et al. 2002; Dennis and Shreeve 2003), the compilation of this information often encompasses such long time periods that the presence or absence of a species in a locality may not reflect current reality, thereby undermining the usefulness of the hard-earned biodiversity databases. Thus, the dynamic character of biological inventories makes it even more difficult to use them to derive useful patterns for applied uses such as in biodiversity conservation. Despite the undoubted interest and long history of trying to establish the distributions of species (Pocock et al. 2015), we suggest that conservation decisions should be complemented with modelling approaches (Guisan et al. 2013) and time-series data coming from systematic standardized surveys. In addition, simple steps can be taken to improve the usefulness of opportunistic distribution records (Altwegg and Nichols 2019; Callaghan et al. 2019a, b) while retaining the societal benefits of involving citizen scientists (Lewandowski and Oberhauser 2017; Turrini et al. 2018).

Acknowledgments

We are very grateful to all of the volunteers who have contributed to the Butterflies for the New Millennium project, which is run by Butterfly Conservation with support from Natural England. D. S-F was supported by a postdoctoral grant from the University of Murcia (Spain). We also thank Leonardo Dapporto and an anonymous reviewer for their helpful comments on an earlier draft of this manuscript

Data availability All the data used to obtain the results shown are available as supplementary material.

Conflict of interest The authors declared that, there is no conflict of interest with regard to this article

References

Altwegg R, Nichols JD (2019) Occupancy models for citizen-science data. Methods Ecol Evol 10:8– 21.

Asher J, Warren M, Fox R, Harding P, Jeffcoate G, Jeffcoate S (2001) The Millennium Atlas of Butterflies in Britain and Ireland. Oxford University Press, Oxford, UK.

Ball-Damerow JE, Brenskelle L, Barve N, et al (2019) Research applications of primary biodiversity databases in the digital age. PLoS ONE 14(9): e0215794.

Ballesteros-Mejia L, Kitching IJ, Jetz W, Nagel P, Beck J (2013) Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. Global Ecol Biogeogr 22:586–595.

Barua M, Gurdak DJ, Ahmed RA, Tamuly J (2012) Selecting flagships for invertebrate conservation. Biodivers Conserv 21:1457-1476.

Beck J, Ballesteros-Mejia L, Buchmann CM et al (2012) What's on the horizon for macroecology? Ecography 35: 673–683.

Beck J, Ballesteros-Mejia L, Nagel P, Kitching IJ (2013) Online solutions and the Wallacean shortfall what does GBIF contribute to our knowledge of species ranges? Divers Distrib 19: 1043–1050.

Brakefield PM (1982) Ecological studies on the butterfly *Maniola jurtina* in Britain: I. Adult behaviour, microdistribution and dispersal. J Anim Ecol 51:713-726.

Brown JH, Lomolino MV (1998) Biogeography. 2nd ed. Sinauer Associates, Inc., Sunderland, Massachusetts, 691 pp.

Bruno D, Sánchez-Fernández D, Millán A, Ros RS, Sánchez-Gomez P, Velasco J (2012). Assessing the quality and usefulness of different taxonomic groups inventories in a semiarid Mediterranean region. Biodiv Conserv 21: 1561–1575.

Callaghan CT, Rowley JJL, Cornwell WK, Poore AGB, Major RE (2019) Improving big citizen science data: Moving beyond haphazard sampling. PLoS Biology 17: e3000357.

Callaghan CT, Poore AGB, Major RE, Rowley JJL, Cornwell WK (2019) Optimizing future biodiversity sampling by citizen scientists. Proc. R. Soc. B 286: 20191487.

Cardoso P, Barton PS, Birkhofer K et al (2020) Scientists' warning to humanity on insect extinctions. Biol Conserv 9:108426.

Carvalheiro LG, Kunin WE, Keil P et al (2013) Species richness declines and biotic homogenization have slowed down for NW-European pollinators and plants. Ecol Letters 16: 870-878.

Chao A, Colwell RK, Chih-Wei L, Gotelli NJ (2009) Sufficient sampling for asymptotic minimum species richness estimators. Ecology 90:1125–1133.

Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD (2011) Rapid range shifts of species associated with high levels of climate warming. Science 333:1024–1026.

Cox CB, Moore PD (2004) Biogeography: An Ecological and Evolutionary Approach. Oxford: Blackwell.

Dapporto D, Ciolli G, Dennis RLH, Fox R, Shreeve TG (2015) A new procedure for extrapolating turnover regionalization at mid-small spatial scales applied to British butterflies. Methods Ecol Evol 6:1287–1297.

Dennis RLH, Thomas CD (2000) Bias in butterfly distribution maps: The influence of hot spots and recorder's home range. J Insect Conserv 4:73–77.

Dennis RLH, Sparks TH, Hardy PB (1999) Bias in butterfly distribution maps: The effects of sampling effort. J Insect Conserv 3:33–42.

Dennis RLH, Shreeve TG, Sparks TH L, Honore JE (2002) A comparison of geographical and neighbourhood models for improving atlas databases. The case of the French butterfly atlas. Biol Conserv 108:143-159.

Dennis RLH, Shreeve TG (2003) Gains and losses of French butterflies: tests of predictions, under-recording and regional extinction from data in a new atlas. Biol Conserv 110:131 - 139.

Dennis EB, Morgan BJT, Freeman SN, Ridout MS, Brereton TM, Fox R, Powney GD, Roy DB (2017) Efficient occupancy model-fitting for extensive citizen-science data. PLoS One 12:e0174433.

Disney RH (1986) Assessments using invertebrates: posing the problem, pp. 271–293. In Usher M.B.Wildlife conservation evaluation London, UK, Chapman & Hall.

Eldredge N (1992) Systematics, ecology, and the biodiversity crisis New York,Columbia University Press.

Edwards J L, Lane MA, Nielsen ES (2000) Interoperability of biodiversity databases: biodiversity information on every desktop. Science 289: 2312–2314.

Fattorini S (2013) Regional Insect Inventories Require Long Time, Extensive Spatial Sampling and Good Will. PLoS ONE 8:e62118.

Forister ML, Pelton EM, Black SH (2019) Declines in insect abundance and diversity: We know enough to act now. Conserv Sci Prac 1:e80.

Fox R, Dennis RLH (2010) Winter survival of *Vanessa atalanta* (Linnaeus, 1758) (Lepidoptera: Nymphalidae): a new resident butterfly for Britain and Ireland? Entomologist's Gazette 61:94-103.

Fox R, Bereton TM, Asher J et al (2015) The State of the UK's Butterflies 2015. Butterfly Conservation and the Centre for Ecology & Hydrology, Wareham, UK.

Gaston KJ (2000) Global patterns in biodiversity. Nature 405: 220-227.

Girardello M, Chapman A, Dennis R, Kaila L, Borges PAV, Santangeli A (2019) Gaps in butterfly inventory data: A global analysis of the completeness of butterfly inventory data. Biol Conserv 236:289-295.

Goulson D (2019) The insect apocalypse, and why it matters. Curr Biol 29: R967–971.

Griffiths GH, Eversham BC, Roy DB (1999) Integrating species and habitat data for nature conservation in Great Britain: data sources and methods. Glob Ecol Biogeogr 8:329–345.

Guisan A, Tingley R, Baumgartner JB et al (2013) Predicting species distributions for conservation decisions. Ecol Lett 16:1424-1435.

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8:993–1009.

Guisande C, Lobo JM (2019) Discriminating well surveyed spatial units from exhaustive biodiversity databases. R package version. 2.0. http://cran.r-project.org/web/packages/KnowBR.

Habel JC, Samways MJ, Schmitt T (2019) Mitigating the precipitous decline of terrestrial European insects: Requirements for a new strategy. Biodiv Conserv 28:1343–1360.

Harvey JA, Heinen R, de Kroon H et al (2020) International scientists formulate a roadmap for insect conservation and recovery. Nature Ecol Evol 4:174–176.

Heath J, Pollard E, Thomas JA (1984) Atlas of Butterflies in Britain and Ireland. Viking, Harmondsworth, UK.

Hengeveld R (1992) Dynamic biogeography. Cambridge University Press

Hickling R, Roy DB, Hill JK, Fox R, Thomas CD (2006) The distributions of a wide range of taxonomic groups are expanding polewards. Glob Change Biol 12:450-455.

Hill J, Thomas C, Fox R, Telfer M, Willis S, Asher J, Huntley B (2002) Responses of Butterflies to Twentieth Century Climate Warming: Implications for Future Ranges. Proc Biol Sci 269:2163-2171.

Hopkinson P, Evans J, Gregory RD (2000) National-scale conservation assessment at an appropriate resolution. Diver Distrib 6:195–204.

Hortal J, Lobo JM (2005) An ED-based protocol for optimal sampling of biodiversity. Biodiv Conserv 14:2913-2947.

Hortal J, Lobo JM, Jiménez-Valverde A (2007) Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife (Canary Islands). Conserv Biol 21: 853-863.

Isaac NJB, van Strien AJ, August TA, de Zeeuw MP, Roy DB (2014) Statistics for citizen science: extracting signals of change from noisy ecological data. Methods Ecol Evol 5:1052–1060.

Lewandowski EJ, Oberhauser KS (2017) Butterfly citizen scientists in the United States increase their engagement in conservation. Biol Conserv 208:106–112.

Lobo JM, Martín-Piera F (2002) Searching for a predictive model for Iberian dung beetle species richness (Col., Scarabaeinae) using spatial and environmental variables. Conserv Biol 16:158-173.

Lobo JM, Hortal J, Yela JL, Millán A, Sánchez-Fernández D, García-Roselló E, González-Dacosta J, Heine J, González-Vilas L, Guisande C (2018) KnowBR: an application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. Ecol Ind 91:241–248.

Lomolino MV, Heaney LR (2004) Frontiers of biogeography: New directions in the geography of nature. Sinauer Associates, Inc. Publishers Sunderland, Massachusetts.

MacGregor CJ, Thomas CD, Roy DB et al (2019) Climate-induced phenology shifts linked to range expansions in species with multiple reproductive cycles per year. Nature Comm 10:4455.

MacKenzie DI, Nichols JD, Royle JA, Pollack KH, Bailey LL, Hines JE (2006) Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence. Academic Press, New York.

Mallet J (1986) Dispersal and gene flow in a butterfly with home range behavior: *Heliconius erato* (Lepidoptera: Nymphalidae). Oecologia 68:210-217.

Menéndez R, González Megías A, Hill JK, Braschler B, Willis SG, Collingham Y, Fox R, Roy DB, Thomas CD (2006) Species richness changes lag behind climate change. Proc. R. Soc. B. 273: 1465–1470.

Montgomery GA, Dunn RR, Fox R, Jongejans E, Leather SR, Saunders ME, Shortall CR, Tingley MW, Wagner DL (2020) Is the insect apocalypse upon us? How to find out. Biol Conserv 241:108327.

Meyer C, Kreft H, Guralnick R, Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. Nat Commun 6:8221

Norberg A, Abrego N, Blanchet FG et al (2019) A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. Ecol Monogr 89:e01370.

Outhwaite CL, Chandler RE, Powney GD, Collen B, Gregory RD, Isaac NJ (2018) Prior specification in Bayesian occupancy modelling improves analysis of species occurrence data. Ecol Ind 93:333-343.

Ovaskainen O, Roy DB, Fox R, Anderson BJ (2016) Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. Methods Ecol Evol 7:428-436.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37-42.

Pereira HM, Ferrier S, Walters M et al (2013) Essential biodiversity variables. Science 339:277–278.

Peterson TA, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, Araújo MB (2011) Ecological Niches and Geographic Distributions. Princeton University Press, Princeton, NJ.

Pocock MJO, Roy HE, Preston CD, Roy DB (2015) The Biological Records Centre: a pioneer of citizen science. Biol J Linn Soc 115:475–493.

Powney GD, Isaac NJB (2015) Beyond maps: a review of the applications of biological records. Biol J Linn Soc 115:532–542.

Pyke GH, Ehrlich PR (2010) Biological collections and ecological/environmental research: a review, some observations and a look to the future. Biol Rev 85:247–266.

Ripple WJ, Wolf C, Newsome TM, Barnard P, Moomaw WR (2020) World scientists' warning of a climate emergency. BioScience 70:8–12.

Romo H, Garcia-Barros E, Lobo JM (2006) Identifying recorder-induced geographic bias in an Iberian butterfly database. Ecography 29:873– 885.

Sánchez-Fernández D, Lobo JM, Abellán P, Ribera I, Millán A. (2008) Bias in freshwater biodiversity sampling: the case of Iberian water beetles. Divers Distrib 14:754-762.

Santos AMC, Jones OR, Quicke DLJ, Hortal J (2010) Assessing the reliability of biodiversity databases: identifying evenly inventoried island parasitoid faunas (Hymenoptera: Ichneumonoidea) worldwide. Insect Conserv Diver 3:72–82.

Saunders ME, Janes JK, O'Hanlon JC (2020) Moving On from the insect apocalypse narrative: engaging with evidence-based insect conservation. BioScience 70:80–89.

Scudder GGE (2017) The Importance of Insects. In: R.G. Foottit & P.H. Adler (eds.), Insect Biodiversity: Science and Society. 2nd ed. Chichester, UK: John Wiley & Sons, Ltd, pp. 9-43.

Soberón J, Jiménez R, Golubov J, Koleff P (2007) Assessing completeness of biodiversity databases at different spatial scales. Ecography 30:152–160.

Stork NE (2018) How many species of insects and other terrestrial arthropods are there on Earth? Ann Rev Entomol 63:31–45.

Thompson GG, Withers PC, Pianka ER, Thompson SA (2003) Assessing biodiversity with species accumulation curves; inventories of small reptiles by pit-trapping in Western Australia. Austral Ecol 28:361-383.

Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F (2017) Taxonomic bias in biodiversity data and societal preferences. Sci Rep 7:9132.

Turrini T, Dörler D, Richter A, Heigl F, Bonn A (2018) The threefold potential of environmental citizen science - Generating knowledge, creating learning opportunities and enabling civic participation. Biol Conserv 225:176–186.

Varela S, Anderson RP, García-Valdés R, Fernández-González F (2014) Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography 37:1084–1091.

Wagner DL, Fox R, Salcido DM, Dyer LA (2021) Moth declines: What can they tell us? Proc Natl Acad Sci USA DOI: 10.1073/pnas.2002549118.

Warren MS, Hill JK, Thomas JA, Asher J, Fox R, Huntley B, Roy DB, Telfer MG, Jeffcoate S, Harding P, Jeffcoate G, Willis SG, Greatorex-Davies JN, Moss D, Thomas CD (2001) Rapid responses of British butterflies to opposing forces of climate and habitat change. Nature 414: 65-69.

Whittaker RJ, Araújo MB, Jepson P, Ladle RJ, Watson JEM, Willis KJ (2005) Conservation Biogeography: assessment and prospect. Diver Distrib 11:3-23.

Yang W, Ma K, Kreft H (2013) Geographical sampling bias in a large distributional database and its effects on species richness–environment models. J Biogeogr 40:1415–1426.

Table 1.- Total species richness (S) and number of data base records (DR) for each family of butterflies recorded in Great Britain.

Family	S (%)	DR (%)
Hesperidae	8 (13.79)	641,627 (6.39)
Lycaenidae	14 (24.14)	1,185,705 (11.80)
Nymphalidae	27 (46.55)	5,437,602 (54.13)
Papilionidae	1 (1.72)	2,723 (0.03)
Pieridae	7 (12.07)	2,767,276 (27.55)
Riodinidae	1 (1.72)	11,432 (0.11)
Total	58	10,046,365

Table 2.- Spatial resolutions in minutes at which accumulation curves have been produced, their approximate resolution in kilometres, total number of cells (N), number of well surveyed cells (WSC) and their corresponding percentages. S_{OBS} and S_{PRE} are the mean number of observed and predicted species according to a rational function adjustment of the accumulation curve.

resolution	**N**	**WSC**	**S_{OBS}**	**S_{PRE}**
60’ (≈120 x 120 km)	70	64 (91.4 %)	34.2	34.9
30’ (≈60 x 60 km)	208	183 (88.0 %)	32.5	33.4
20’ (≈40 x 40 km)	420	357 (85 %)	31.4	32.4
10’ (≈20 x 20 km	1456	1025 (70.4 %)	27.8	29.1
5’ (≈10 x 10 km)	2775	1721 (62.0 %)	26.1	27.8

Figure 1. Number of database records for each species.

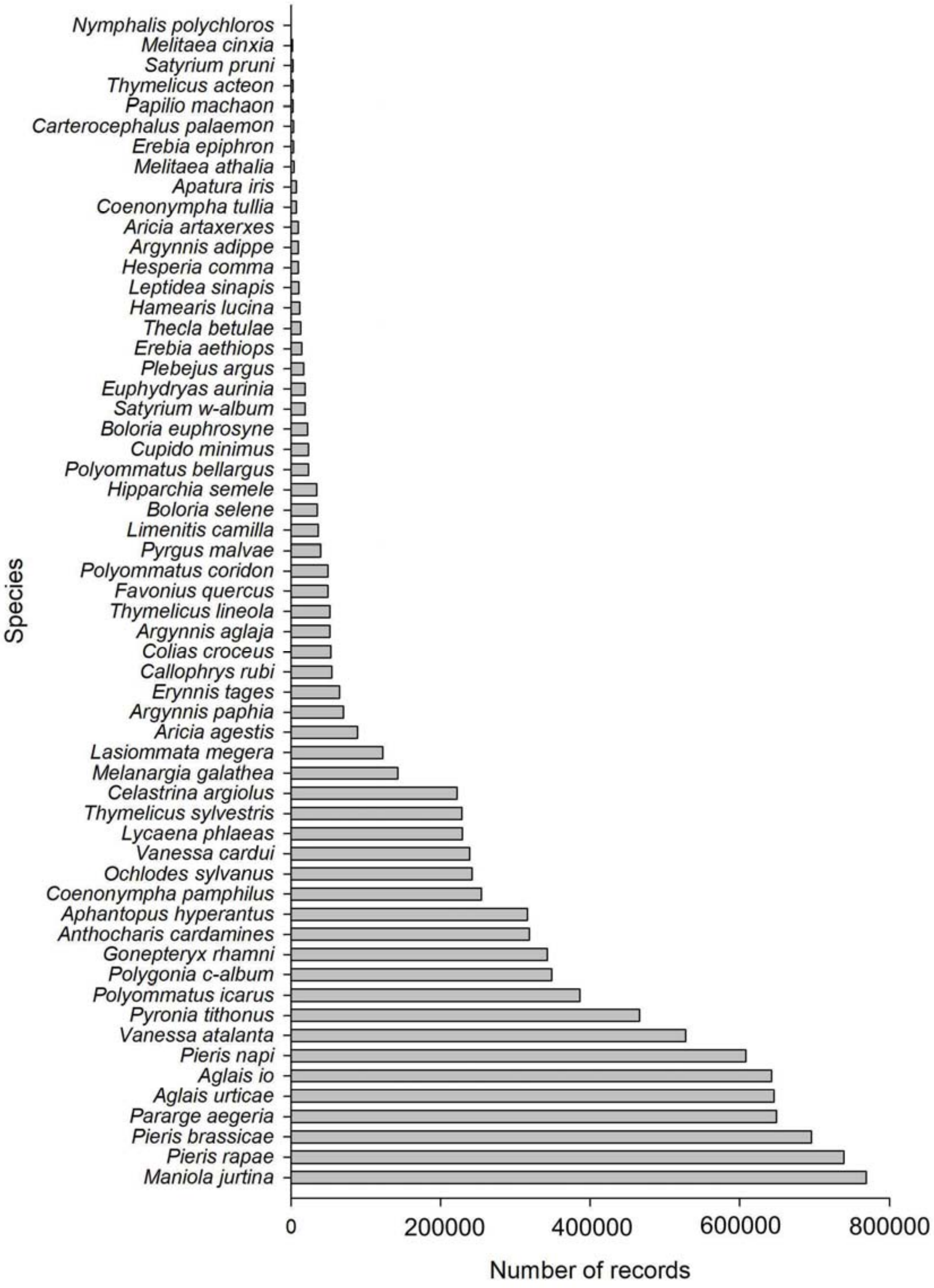

Figure 2.- Spatial variation (5’ cell resolution) in the number of database records (A), completeness percentages (B), number of observed butterfly species (C), and number of predicted species (D) after adjusting accumulation curves to a rational function (see Lobo et al., 2018 for details).

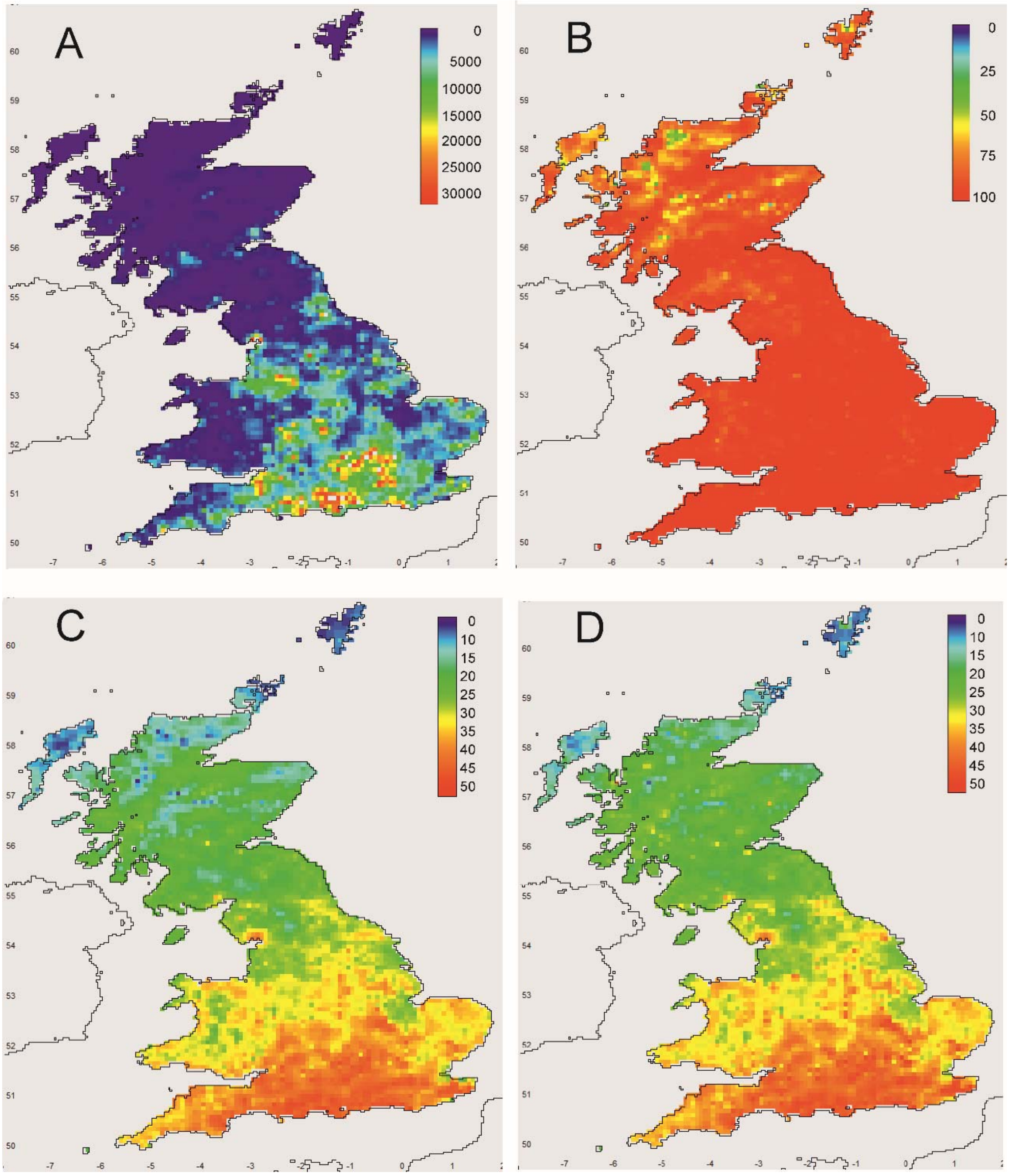

Figure 3.- Well surveyed cells at the five considered resolutions. Well surveyed cells are those have a final slope in the accumulation curve ≤0.01, a completeness value ≥ 90%, and at least 20 times more database records than species

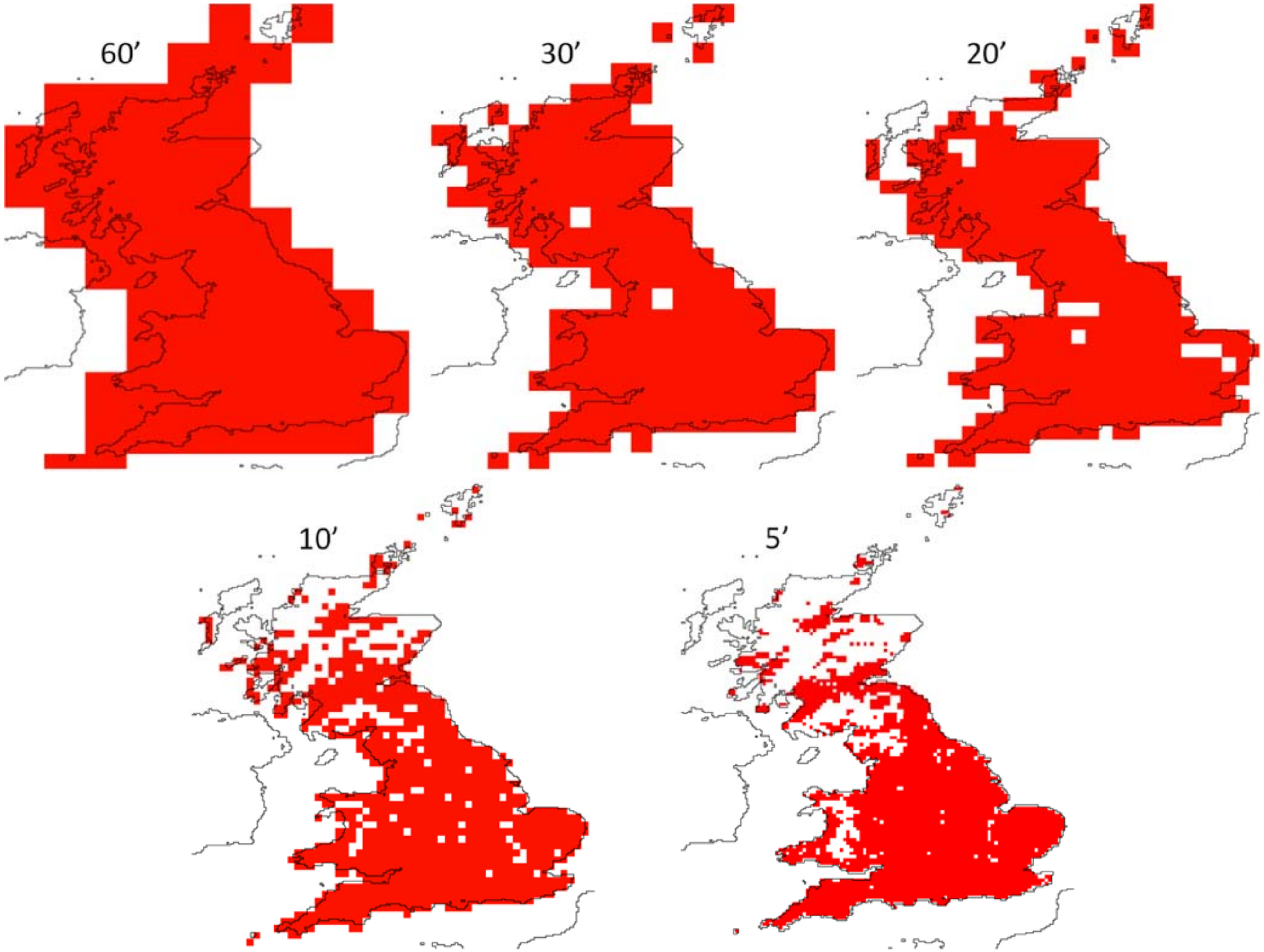

Figure 4.- Variation in the percentage of well surveyed cells (%WSC) according to five different (grey dots) spatial resolutions of the cells (in minutes) and a Morgan-Mercer-Flodin adjusted curve on these data (*r*= 0.999; broken line). The grey area represents the variation in the observed mean number of species in well surveyed cells, while the continuous black line represents the predicted mean number of species in well surveyed cells following the adjustment of the accumulation curves to a rational asymptotic function.

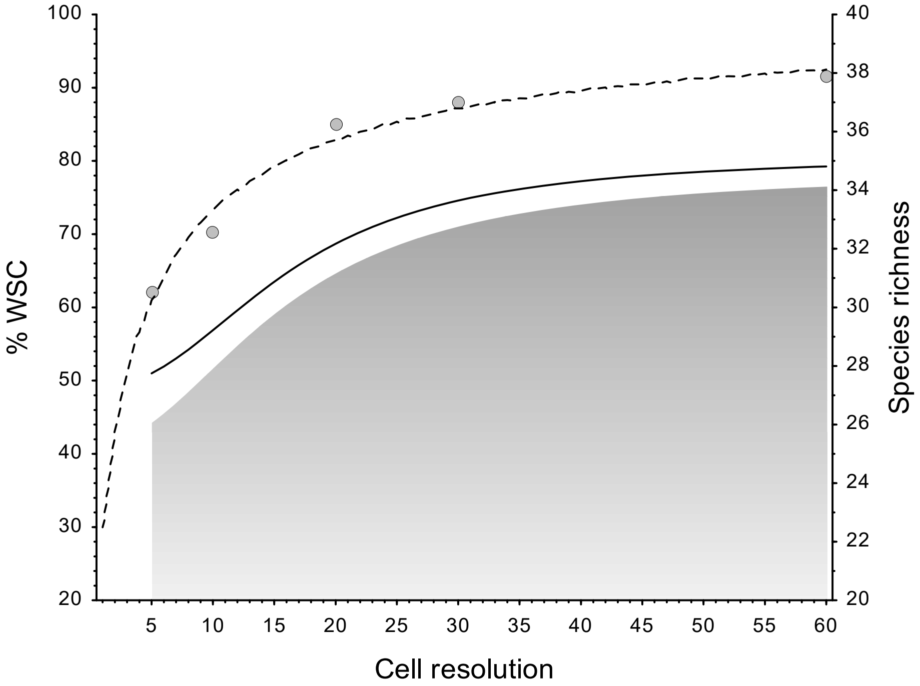

Figure 5.- Relationship between the completeness and the proportion of locally uncommon species in all the 5’ cells of Great Britain. Completeness was calculated as the proportion of observed species against those predicted by adjusting the accumulation curve of each cell to a rational function. Locally uncommon species are those with five or less database records in a cell. The size of the dots is proportional to the frequency of similar values. The broken line represents a cubic polynomial fit of the data.