

RESEARCH ARTICLE

Leveraging Natural History Collections to Understand the Impacts of Global Change

Applying computer vision to digitised natural history collections for climate change research: Temperature-size responses in British butterflies

Rebecca J. Wilson^{1,2}  | Alexandre Fioravante de Siqueira³  | Stephen J. Brooks² | Benjamin W. Price²  | Lea M. Simon¹ | Stéfan J. van der Walt³  | Phillip B. Fenberg^{1,2} 

¹School of Ocean and Earth Sciences, University of Southampton, Southampton, UK

²Department of Life Sciences, Natural History Museum, London, UK

³Berkeley Institute for Data Science, University of California, Berkeley, CA, USA

Correspondence

Phillip B. Fenberg

Email: p.b.fenberg@soton.ac.uk

Funding information

Alfred P. Sloan Foundation, Grant/Award Number: 2013-10-27; Gordon and Betty Moore Foundation, Grant/Award Number: GBMF3834; Natural Environmental Research Council, Grant/Award Number: NE/L002531/1

Handling Editor: Natalie Cooper

Abstract

1. Natural history collections are invaluable resources for understanding biotic response to global change. Museums around the world are currently imaging specimens, capturing specimen data and making them freely available online. In parallel to the digitisation effort, there have been great advancements in computer vision: the computer trained automated recognition/detection, and measurement of features in digital images. Applying computer vision to digitised natural history collections has the potential to greatly accelerate the use of these collections for biotic response to global change research. In this paper, we apply computer vision to a very large, digitised collection to test hypotheses in an established area of biotic response to climate change research: temperature-size responses.
2. We develop a computer vision pipeline (Mothra) and apply it to the NHM collection of British butterflies (>180,000 imaged specimens). Mothra automatically detects the specimen and other objects in the image, sets the scale, measures wing features (e.g. forewing length), determines the orientation of the specimen (pinned ventrally or dorsally) and identifies the sex. We pair these measurements and specimen collection data with temperature records for 17,726 specimens across a subset of 24 species to test how adult size varies with temperature during the immature stages of species. We also assess patterns of sexual size dimorphism across species and families for 32 species trained for automated sex ID.
3. Mothra accurately measures the forewing lengths of butterfly specimens compared to manual measurements and accurately determines the sex of specimens, with females as the larger sex in most species. An increase in adult body size with warmer monthly temperatures during the late larval stages is the most common temperature-size response. These results confirm suspected patterns

Rebecca J. Wilson and Alexandre Fioravante de Siqueira contributed equally to the paper and share joint first authorship.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

and support hypotheses based on recent studies using a smaller dataset of manually measured specimens.

4. We show that computer vision can be a powerful tool to efficiently and accurately extract phenotypic data from a very large collection of digital natural history collections. In the future, computer vision will become widely applied to digital collections to advance ecological and evolutionary research and to accelerate their use to investigate biotic response to global change.

KEYWORDS

butterfly, climate change, computer vision, deep learning, digitisation, lepidoptera, Mothra, natural history collections

1 | INTRODUCTION

The world's natural history collections contain at least 2 billion specimens (Ariño, 2010). Tens of millions of these specimens (and counting) are making their way out of the halls and cabinets of natural history museums and into the virtual world as digital images and specimen data, either through data portals (<https://data.nhm.ac.uk/>) or aggregators (e.g. <https://www.gbif.org/>) (Nelson & Ellis, 2019). The purpose of this vast effort is twofold: to provide a digital copy of these priceless collections and to advance the core research of natural history museums for understanding the living world. But as the Anthropocene progresses, digitised natural history collections can also be leveraged for understanding the biological impacts of global change (Johnson et al., 2011; Meineke et al., 2019). Not only will the widespread availability of specimen images and data increase the rate at which scientists can perform this essential research, but also the sheer taxonomic, spatial and temporal scope of these digitised collections will help provide a more holistic understanding of how the biosphere has and will respond to global change.

Digitised natural history collections have been used to investigate multiple aspects of biotic response to global change, including documenting changes in geographic range and biodiversity (Ewers-Saucedo et al., 2021; Kharouba et al., 2019), phenology (Brooks et al., 2017) and body size of species (Wilson et al., 2019; Wonglersak et al., 2020). While such studies are incredibly important, the number of specimens used are often limited due to the time required to physically measure and record each specimen. For example, until recently, studies examining change in body size using images must first open images in software, set the scale and manually measure body size or its proxies (Fenberg et al., 2016). Thus, despite their availability, specimen images still require time-consuming manipulation and manual measurement, limiting the amount of data available for individual research projects.

In parallel to mass digitisation efforts by museums, major advancements have been made in computer vision technologies. Computer vision is a rapidly developing field in which computers are trained to recognise, extract and measure information from digital images or video. While practical applications of computer vision

have been made in several fields, such as object recognition/detection for medical purposes (e.g. tumour detection; Svoboda, 2020) and ecologists have recently used computer vision for biodiversity analyses in the field (Bjerger et al., 2021), computer vision is only starting to be used for ecology and evolution research.

Given the rapid advancements in computer vision technology and its many applications, it is thought that computer vision will become an essential tool for ecology and evolutionary biologists (Lürig et al., 2021). For example, computer vision can be used along with molecular data to help identify cryptic species and other eco-evolutionary questions (Høye et al., 2021). Currently, however, there are very few studies showcasing the powerful utility of pairing computer vision with natural history collections for the purposes of climate change research (Hsiang et al., 2019; McAllister et al., 2019). In this paper, we apply computer vision to a very large, digitised collection to test hypotheses in an established area of biotic response to climate change research: temperature-size responses (Sheridan & Bickford, 2011).

1.1 | Temperature-size responses

Body size is one of the most important traits of an organism due to its correlation with many aspects of the life history, ecology and evolution of species. However, climate warming is thought to be causing widespread reduction in body size and is even suggested to be a 'universal' response to warming (Sheridan & Bickford, 2011). When reared at warmer temperatures, ectotherms often mature at smaller sizes and reach smaller adult body sizes, compared to when reared at cooler temperatures; this is known as the 'temperature-size rule' (Atkinson, 1994). Such a reduction in size appears to be most pronounced among aquatic species, which is likely due to oxygen limitation in warmer waters (Forster et al., 2012). On the other hand, many terrestrial species exhibit variable temperature-size responses (Horne et al., 2015; Na et al., 2021; Tseng et al., 2018; Wonglersak et al., 2020). This is especially true for insects, which, due to their complex and diverse life cycles, can lead to a variety of temperature-size responses. Each life stage of holometabolous

insects can experience different environmental conditions, which may cause each stage to respond in a different way to temperature (Kingsolver et al., 2011; Wilson et al., 2019). In addition, each sex may have different temperature-size responses, which may affect the magnitude of sexual size dimorphism (Fenberg et al., 2016). Thus, it is important that life stages, sex and the environmental conditions experienced by them, are considered when investigating temperature-size responses.

Lepidoptera are useful study taxa for examining temperature-size responses as their life stages are clearly defined, the sexes of many species can be easily identified, and they have relatively short generation times. If adult body size measurements are paired with temperature records across multiple generations, years, per sex and for each immature life stage (e.g. early to late larval and pupal stages), then it is possible to determine the direction and strength of adult body size responses to temperature and which factors are most predictive of observed responses (Bowden et al., 2015; Davies, 2019; Wilson et al., 2019).

Natural history collections paired with temperature records can provide a useful resource for studying temperature-size responses because specimens often span many decades, over which a large range of inter- and intra-annual (i.e. seasonal) temperature records may be available. In recent years, the use of natural history collections to study temperature-size responses in insects has become common, but responses often vary among taxa (Baar et al., 2018; Tseng et al., 2018). For example, the body sizes of Zygoptera (damselflies) are more sensitive to temperature than Anisoptera (dragonflies) (Wonglersak et al., 2020). This suggests that, at least in some insect groups, phylogenetic relationships are also an important predictor of the direction and magnitude of temperature-size responses. Butterflies often increase in adult size with increasing temperature (MacLean et al., 2016; Na et al., 2021; but see Bowden et al., 2015) and analysis of four UK butterfly species found that the strongest prediction of adult size was temperature during the late larval stage (Fenberg et al., 2016; Wilson et al., 2019). But in order to determine if these are general responses, more species and specimens need to be analysed.

Here, we use a newly developed computer vision pipeline to automatically measure body size attributes (e.g. wing lengths) and orientation (pinned ventrally or dorsally) of British butterfly specimens housed at the NHM ($n = 184,533$, 94 species). We also trained the pipeline to identify the sexes for species that can reliably be identified by eye from images (32 species). We test the accuracy of the pipeline measurements by comparing them to manual measurements of 30 butterfly species. We also test if there are patterns of sexual size dimorphism (SSD) across 32 species, testing the hypothesis that females are larger than males (Teder, 2014).

For temperature-size responses, we pair wing length measurements with monthly temperature records experienced by the immature stages of 24 species across four families to determine the direction and strength of responses per species and to look for general patterns across species. We hypothesise that the adult sizes of univoltine species (one generation per year), and first generation

of bivoltine species (two generations per year) will increase with increasing temperatures during the late larval stages, and that males and females will respond differently, based on previous work (Fenberg et al., 2016; Wilson et al., 2019). It has been suggested that the body sizes of second generation individuals of bivoltine species may not be as responsive to temperature (compared to first generation individuals) due to the constrained amount of time available for growth and emergence before the flight period ends and/or the likelihood that more food will be available to second generation larvae, given that they appear later in the growing season when temperatures are warmer (Wilson et al., 2019). These same studies, however, also show that increasing temperatures during the early larval stage causes some species to become smaller as adults and that response to temperature during the pupal stages varies. We therefore hypothesise that (a) warmer temperatures during the late larval stages will be correlated with larger adults, (b) warmer temperatures during early larval and pupal stages will result in variable responses across species and (c) sex and family will be important factors.

1.2 | Study system

The British butterfly specimens housed at the Natural History Museum (London) were among the first very large scientific collections to be mass digitised. A total of 184,533 specimens, comprising 94 species of butterflies collected from 1803 to 2006 have been digitised during the iCollections project (Paterson et al., 2016). Each pinned specimen is imaged with a ruler (scale bar) and associated data labels. All specimen data have been extracted and databased for specimens with sufficient information, which include the location, date of collection and collector. See Paterson et al. (2016) for a detailed description of the geographic and temporal coverage of the iCollections dataset. We use these data and life-history information paired with historical temperature records in order to test our temperature-size hypotheses.

2 | MATERIALS AND METHODS

2.1 | Mothra development

Mothra is a Python package for analysing images of Lepidoptera specimens, inferring sex and measuring body size attributes using a combination of deep learning and image processing techniques. It is built on NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), matplotlib (Hunter, 2007), scikit-image (van der Walt et al., 2014), PyTorch (Paszke et al., 2019) and fastai (Howard & Guggler, 2020). Mothra processes images that include: the pinned specimen, a ruler (scale bar) and several printed or hand-written labels (Figure 1a). Mothra identifies these image elements, finds key points on the specimen, makes measurements, and translates pixel distances to millimetres after interpreting the scale bar. Mothra can be applied to any images of pinned Lepidoptera specimens if a millimetre scale bar is present (Figure 1a) and can be trained to

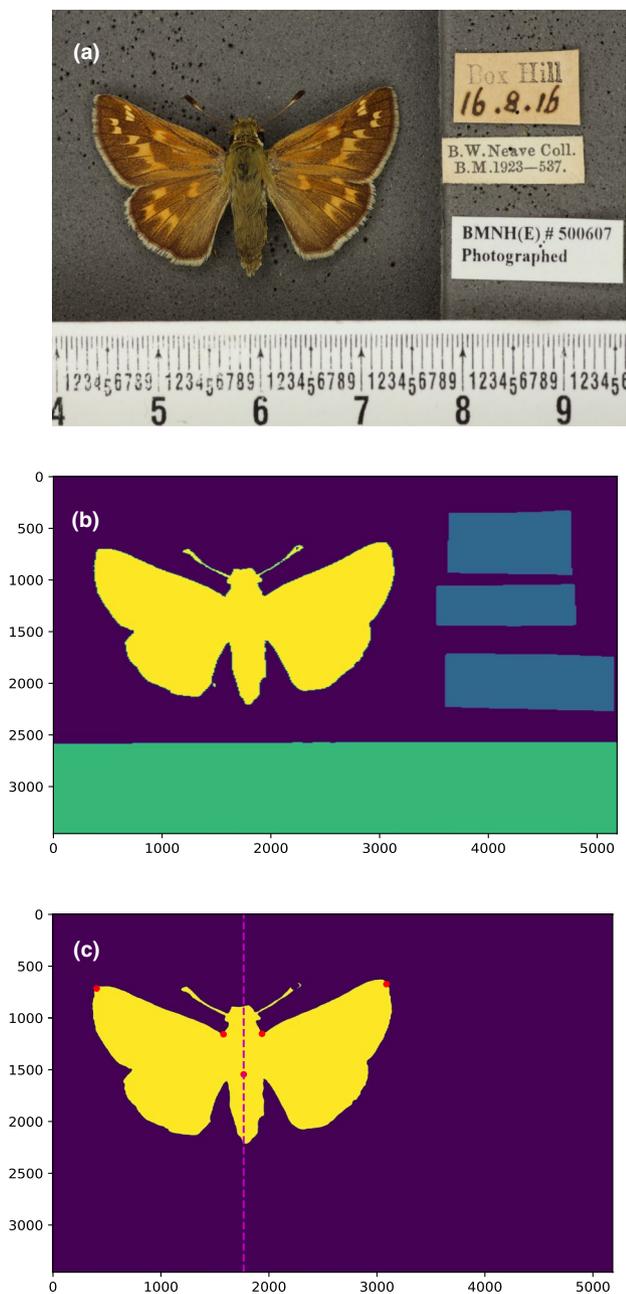


FIGURE 1 (a) Example input image (female *Hesperia comma*) containing the pinned specimen, a scale bar and data labels. (b) Image returned by Mothra, containing predictions to the specimen (yellow), scale bar (green), labels (blue) and background (purple). (c) Wing tips, shoulders and centre (red dots) of the specimen (yellow). These points are used for the measurements of forewing lengths (shoulders to wingtips), wingspan (wingtip to wingtip), centre to wingtips and shoulder width (shoulder to shoulder). Axis values in (b) and (c) are pixel numbers

identify other scale bars as needed. While Mothra also works on many moth species, we focus on butterflies for this paper as they were used to train the current segmentation algorithm.

To recognise image elements (specimen, scale bar and data labels), we use a U-Net convolutional neural network (Ronneberger

et al., 2015) with ResNet-34 (He et al., 2016) in the analysis path (Howard & Gugger, 2020; Paszke et al., 2019; Zhang et al., 2018). The ResNet-34 implementation from PyTorch is pre-trained on the ImageNet image database (Deng et al., 2009). The U-Net is trained using 150 manually segmented images of different Lepidoptera species. Labels correspond to the three elements (specimen, scale bar and data labels) as well as the background. Each iteration of training uses a batch of four images, and training completes after 26 epochs (i.e. after all data have been seen 26 times).

The network is trained using the one cycle policy (Smith, 2018), whereby learning rates start low, increase, then drop back to below the initial value. The first epoch only trains the last U-Net layer (bottom of the “U”) with a learning rate of 2×10^{-3} while the rest of the network is frozen. In subsequent epochs, the entire network is unfrozen. We use a discriminative learning rate (i.e. a different learning rate for each layer; Smith, 2018) of 10^{-5} for the first layer, 10^{-3} for the last layer and logarithmically interpolated values for the middle layers. Training continues for a further 25 epochs.

The input dataset is augmented by changes in orientation, scale, exposure and warp. Input and labelled images were resized from their original size, $5,184 \times 3,456$ pixels, to 448×448 pixels. After classification, Mothra returns an image with labels corresponding to four classes: specimen, ruler (scale bar), data labels and background (Figure 1b). The central axis of the specimen is taken as the horizontal centre of gravity. The image is then split into left and right sides. Wing tips and shoulders are located (Figure 1c) for each side: the wing tip is defined as the most distant pixel from the centre of the specimen, while the shoulder is where the upper-central part of the body dips lowest in the vertical direction.

To convert between pixel distances and millimetres, the scale bar is analysed. Its image coordinates are returned by the classification step, after which the scale bar image is extracted and turned into a binary image using an automated Otsu threshold (Otsu, 1979). Numbers are removed by filtering objects on their area and eccentricity, and the image is then summed vertically to produce a one-dimensional vector of values. Summing across the scale bar increases robustness against noise. That summation is, in turn, thresholded, since we are only interested in transition periods, not in amplitudes. A fast Fourier transform is then performed to determine the most dominant frequency. This frequency is given in pixels per cycle and corresponds to the minor ticks on the scale bar: using it, we can convert the measurements from pixels to millimetres.

Next, we want to predict sex and orientation: either the specimen is pinned dorsally (with the upper surface of the wings shown), or ventrally (with the underside of the wings shown). For that purpose, we trained a ResNet-50 network using 2,986 images separated into three classes: 1,549 pinned ventrally (where we did not classify sex), 722 male, and 715 female (both latter classes being pinned dorsally). Training images were resampled to 256×256 pixels, and data augmentation was performed using the Albumentations library (Buslaev et al., 2020) which adds random changes of hue, saturation and value in the interval $(-0.2, 0.2)$, as well as coarse dropout of rectangular regions in the image (DeVries & Taylor, 2017). Each

augmentation was applied with a probability of 0.5 per generated augmented sample.

Mothra, the collection of algorithms and functions implemented for this study, is permissively licensed under the BSD-3 clause licence and is available on Zenodo (Feng et al., 2021). Mothra automatically downloads the latest pre-trained version of the neural network. The training data accompanying this study, including the networks trained and images used in training are available on Zenodo (de Siqueira, 2021). The images we used are part of the iCollections project, released under the CC0-1.0 licence (Paterson et al., 2016).

For each analysis Mothra takes an input folder of images or a text file listing the location of the input images, and then outputs the following data as a CSV file: length (mm) of each forewing, distance from each wing tip to the centre of the specimen (mm), wingspan from wing tip to wing tip (mm), shoulder width between shoulders (mm), pinned orientation and sex. For each image an output image can be provided with the measurements overlaid (Figure 2).

2.2 | Mothra testing: Manual versus automated measurements and sex ID

We manually measured the forewing lengths of 3,145 specimens of 30 species across four families using ImageJ software. Measurements of four species are from previously published research by the co-authors (Fenberg et al., 2016; Wilson et al., 2019). We then measured the same specimens using Mothra. For each specimen, we calculated the average between the left and right forewing lengths for both the manual and Mothra measurements. We then compared the correlation between measurements across all specimens; testing if the slope is equal to 1 (i.e. a one-to-one correlation). We also performed *t* tests of measurements grouped per family and species to test if the manual versus automated measurements are statistically different. We categorised specimens by sex for species in which the sexes are reliably detectable by eye from images (2,807 specimens from 20 species). A further 5,127 specimens were identified to sex by Wilson (2021). We then compared the sex IDs for all specimens combined ($n = 7,934$ specimens) to the Mothra outputs to determine the accuracy of the automated sex identifications.

2.3 | Mothra measurements of the iCollections

Once we determined the accuracy of the automated wing length measurements and sex identification (see below), we ran Mothra on all butterfly specimens within the iCollections dataset (Paterson et al., 2016) using the NHM HPC cluster. This dataset constitutes 184,533 specimens. For analysis purposes, we only focus on the four main families that constitute 99% of the collections (Hesperiidae, Lycaenidae, Nymphalidae and Pieridae) and removed species ($n = 32$) which have either very few specimens (<100) or are not native to Britain (e.g. rare occurrences). Ventrally pinned specimens ($n = 51,646$) were removed to keep forewing length measurements and sex identification consistent. Measurements of forewing length for 130,173 specimens across 60 species and four families were analysed. For each species, we removed any specimens in which the absolute value difference between the right and left forewing lengths were larger than 2 mm in order to remove any specimens with wing damage. We also removed specimens for which the Mothra measurements were clearly incorrect (e.g. measurements that were too large or small given the size of the species) by examining the output images for the biggest outliers. We also checked the remaining output images for the largest and smallest individuals per species to determine if they were incorrect measurements. In total, only 1.8% of specimens were removed as clear outliers/incorrect measurements ($n = 2,360$), leaving 127,813 specimens for analysis (Table S2). For species which we trained Mothra for sex identification ($n = 32$), we tested the hypothesis that females are, on average, larger than males and looked for patterns across families.

2.4 | Temperature-size responses: Individual species analyses

We analysed a subset of the Mothra measurements for temperature-size responses (24 species). These species were chosen as they have good specimen data, are representative of each family, and have varying life histories and habitat requirements. We only included specimens if there was a known year, location and month of collection, and collected on the island of Great Britain. Where applicable, we separated specimens into generations (see Wilson et al., 2019).

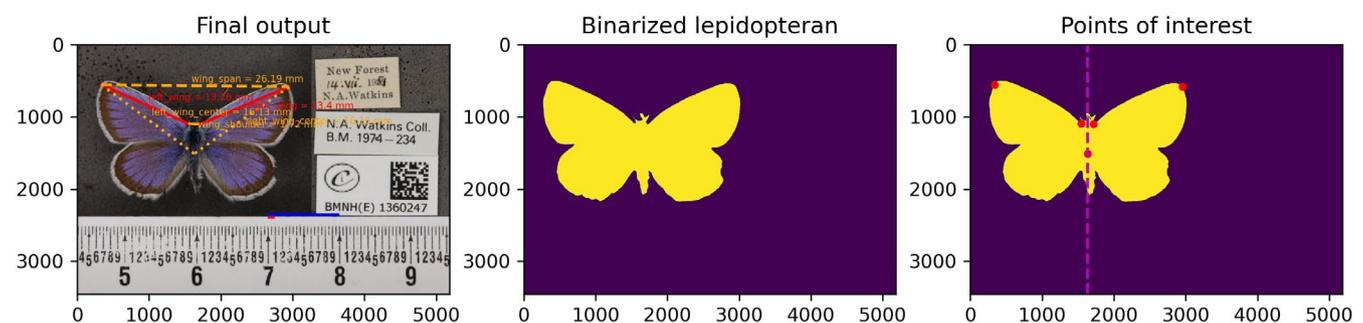


FIGURE 2 Example Mothra output image (male *Plebejus argus*) with the final output showing the measurements overlaid on the image, the binarized specimen and the points of interest

If a species had a partial second generation, or a variable number of generations from year to year, specimens were removed to keep the number of generations per year consistent. For example, *Aglais urticae* has one generation in Scotland and two generations a year in other parts of the United Kingdom, so Scottish specimens were removed. Additionally, we removed specimens with collection dates outside the expected range of adult flight season for that species, based on Thomas and Lewington (2014). We did not include specimens of a species if there were fewer than three specimens available per year (and sex where applicable). We used information about the life cycles of each species given in Thomas and Lewington (2014) to determine which monthly temperatures were appropriate for analyses. We used temperatures from months when species were in the early larval, late larval and pupal stages; winter months were not used as growth would be limited. We used mean monthly temperature data from the Central England Temperature (CET) Record for all analyses (<https://www.metoffice.gov.uk/hadobs/hadcet/>). The CET dataset has been shown to be highly representative of the wider United Kingdom (Croxton et al., 2006) and encompasses the temporal range of specimens in our dataset.

Following Fenberg et al. (2016) and Wilson et al. (2019), we compared average forewing length to average monthly temperatures using multiple linear regression analyses to determine if temperatures experienced during the immature stages affect adult size. We used R statistical packages MASS and MuMIn to run stepwise regression in both directions to select variables for the final model and information theoretic (IT) model selection with model averaging based on Akaike information criterion (AIC). Where applicable, we ran separate models for each sex and generation.

A total of 17,726 specimens and 24 species are in the final analysis, collected between 1830 and 2000, with most collections from ~1900 to 1950 (Figure S1). The geographic spread of specimens is reflective of the overall iCollections dataset (i.e. most specimens are from southern England; Paterson et al., 2016). In 15 species, males and females could be identified, and three species had two generations that could be analysed separately, giving a total of 44 models. For each species with a significant model, we calculated the percentage change in adult size per °C for the most significant month in early larval, late larval and pupal stages. Where there was not a significant variable for a particular life stage, the most important non-significant variable was used. We calculated percentage changes from slopes of the natural log of average forewing length versus temperature: $((\exp[\text{slope}] - 1) \times 100)$.

2.5 | Temperature-size responses: Multi-species analyses

We compiled data to look for general patterns of temperature-size responses across species. First, we compiled the percentage change in adult size per °C of the three immature stages for each species and, where applicable, each sex and generation. Second, we compiled the natural log of average forewing lengths for all specimens

used in the individual species analyses. Natural logs were used to allow for species of different sizes to be compared without the effects of scaling. We used temperature data from the most important month for predicting adult size during each immature stage for the multi-species analyses. We also included four other variables (family, habitat, size category and overwintering stage) in the form of multi-level factors (Table S1) to determine which may affect the strength and direction of temperature-size responses. We selected these four factor variables a priori as likely having an influence on temperature-size response based on previous research (Davies, 2019; Fenberg et al., 2016; Tseng et al., 2018; Wilson et al., 2019).

We compared percentage change in adult size per °C increase in temperature during each immature stage between the four factor variables (Table S1). We performed linear mixed effects models using the natural log of average forewing lengths of specimens from all 24 species, with temperature during the early larval, late larval and pupal stages as fixed effects and the random effects of family, overwintering stage, habitat and size category in each model. ANOVAs and AIC values were used to determine which model gave the best fit. We repeated analyses for species where sex could be determined, with sex included as a fixed effect. The datasets used for all above analyses can be found in Wilson et al. (2022).

3 | RESULTS

3.1 | Automated versus manual measurements and sex ID

The Mothra measurements are nearly identical to the manual measurements (Figure 3). The correlation between average forewing length of the Mothra versus manual measurements is 0.98 and the slope is 1.0. After six clear outliers were removed, the correlation is 0.99 with a slope of 1.03. These results indicate that there is a nearly perfect one-to-one relationship between the Mothra and manual measurements. For all specimens combined, there is no difference between measurements (*t* test, $p = 0.33$). When grouped by family, manual versus Mothra measurements are not statistically different except for Hesperidae, where there is a slight difference ($p < 0.001$) in mean forewing length between manual (13.34 mm) and Mothra measurements (13.12 mm). These differences are driven by *Hesperia comma*, due to a consistent difference in where the wingtip was manually located by Fenberg et al. (2016). When grouped by species, only four species (out of 30) had significant differences between the Mothra versus manual measurements of forewing lengths (*Hesperia comma*, *Apatura iris*, *Aglais urticae* and *Favonius quercus*). However, all differences are small (<1 mm on average) and driven by a consistent difference in the placing of the wingtip between Mothra and manual measurements.

The Mothra sex identifications were highly accurate. Out of 7,934 specimens, the sex identifications differed by only 2.9% ($n = 149$ specimens) between the manual versus Mothra identifications. After inspection of a subset of specimens that have a discrepancy

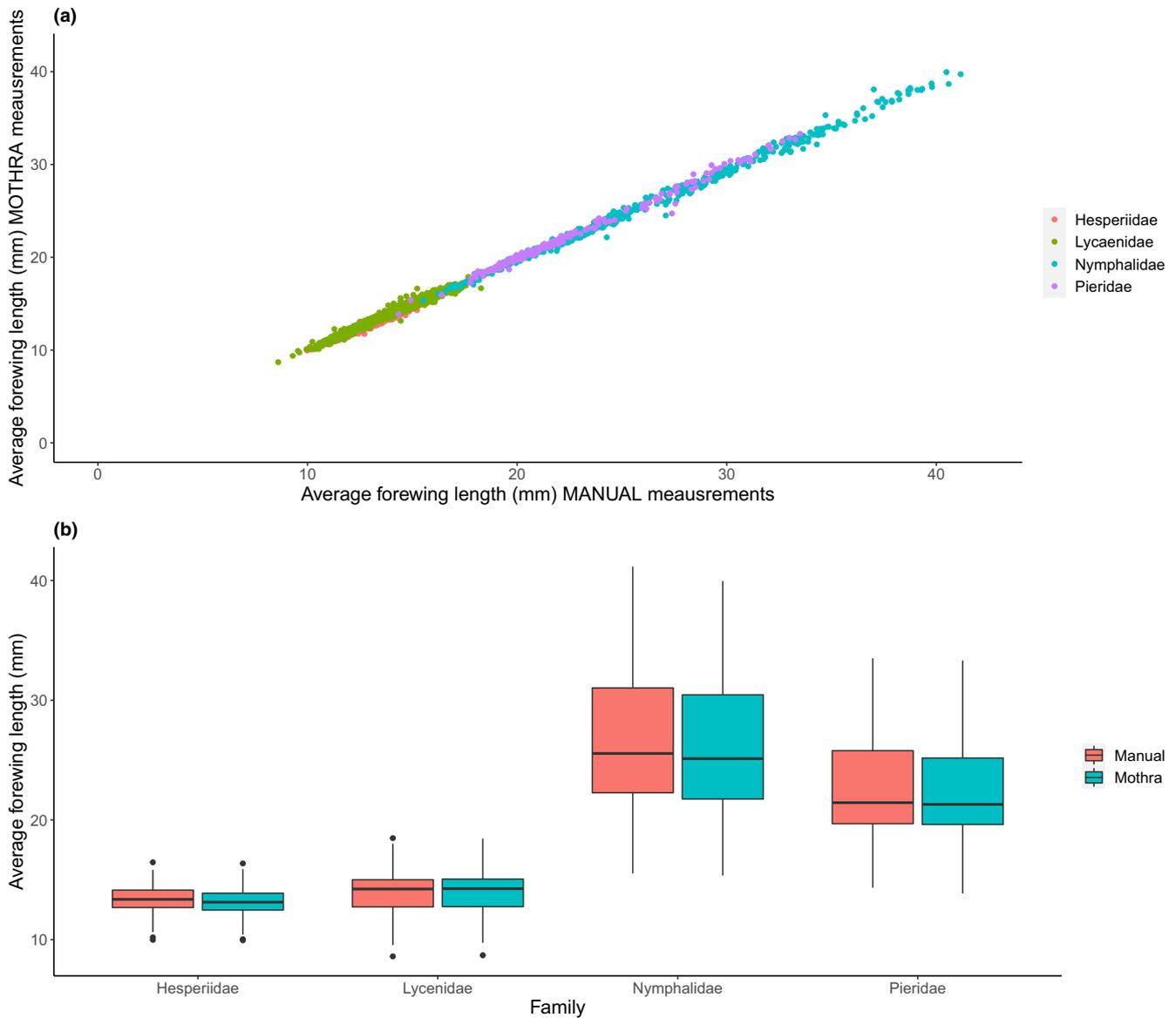


FIGURE 3 (a) Correlation between Mothra versus manual measurements for 3,145 specimens across 30 species from four families. Pearson's correlation $R = 0.99$ and the slope = 1.03, revealing a nearly one to one relationship between manual and Mothra measurements. (b) Boxplots comparing the manual versus Mothra measurements grouped by family. Except for Hesperidae (see main text), there are no significant differences between the two measurements. Six outliers were removed from these figures due to incorrect Mothra measurements (see main text)

in sex ID ($n = 41$ specimens), it was noted that 17 specimens were mis-identified by eye and nine were misidentified by Mothra, the remaining 24 specimens were discoloured or gynandromorphs where sex ID is not possible.

3.2 | Size distribution and patterns of sexual size dimorphism

Given the accuracy of the wing length measurements and the sex identifications, we felt confident to run Mothra on all specimens in the iCollections dataset (all results available here: Price & Fenberg, 2021). The number of inaccurate measurements (either

damaged specimens or incorrect Mothra measurements) removed from the dataset was very small (1.8% of specimens, see above), with the resulting size distributions per species seen in Figure 4. As an initial test of the utility of this massive dataset, we tested the hypothesis that females are larger than males per species (as is the case for many insect species, largely due to their longer developmental times; Teder, 2014). Our results show that this is broadly true for British butterflies (Figure 5). Out of 32 species, 30 have significant SSD, but males are the larger sex in only seven species (five are in Lycaenidae and two in Pieridae; Table S2). Four of the Lycaenidae are in the subfamily Polyommatae (i.e. the blues). For the remaining species ($n = 23$), the females are the larger sex, including all species in Hesperidae and Nymphalidae.

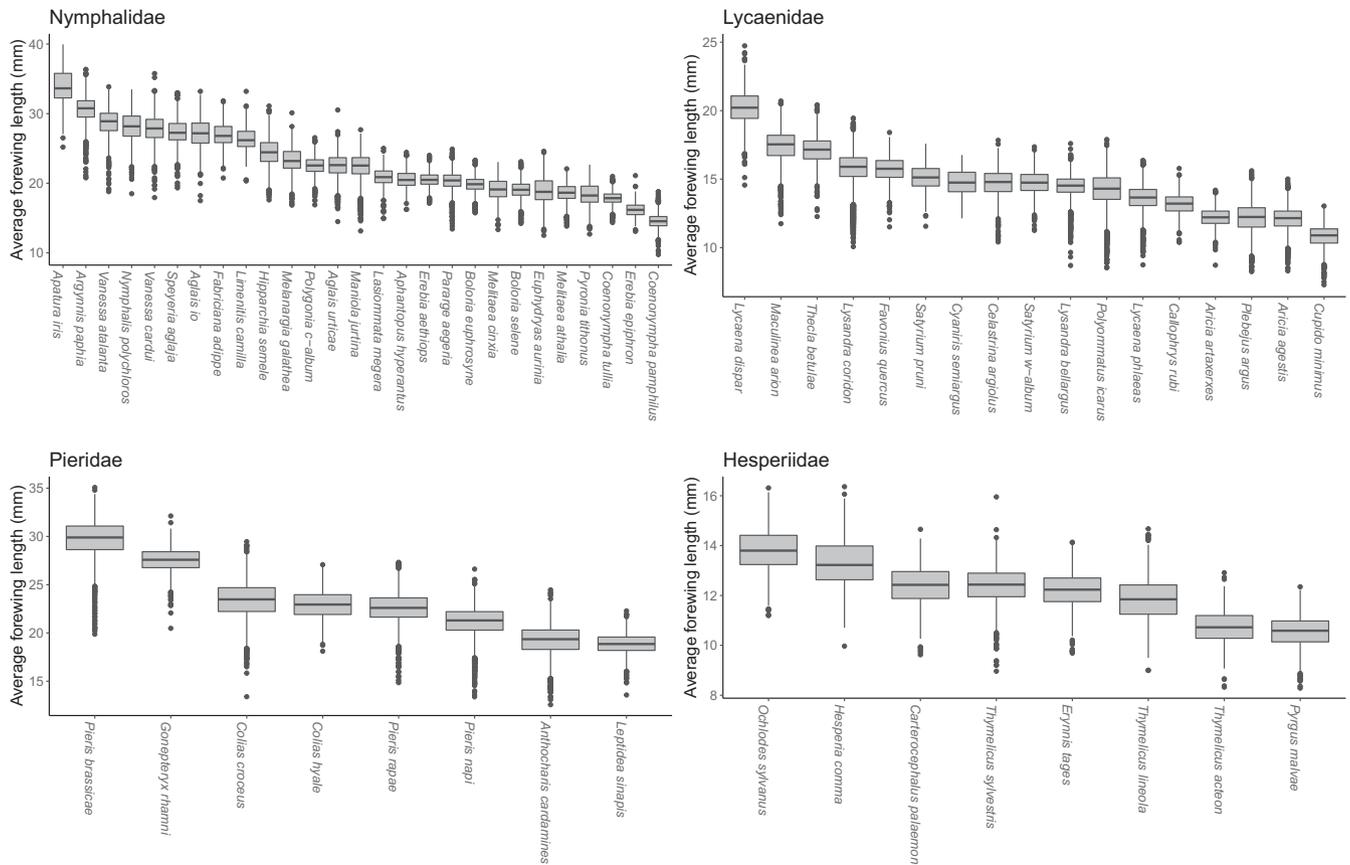


FIGURE 4 Size distributions of the Mothra measurements of the dorsally pinned specimens for each species in the iCollections native to the island of Great Britain from the four main families (60 species). This figure represents measurements from a total of 127,813 specimens, excluding incorrect measurements and damaged specimens ($n = 2,360$ specimens)

3.3 | Temperature-size responses: Individual species analyses

When average forewing lengths were compared to monthly temperatures using multiple linear regression models, 20 of the 44 models were significant. This accounted for 17 of the 24 species analysed. In all but four of the significant models, an increase in adult size with increasing temperature during the late larval stage was significant. The responses of adult size to temperatures experienced during the early larval and pupal stages were less consistent. Only eight of the 20 models had a significant change in adult size in relation to changes in temperature during the early larval stage and eight models had significant changes in adult size in the pupal stage, with both having a mix of increases and decreases in size with increasing temperatures. The percentage changes in adult size per °C increase during each immature stage are given in Table S5, and detailed individual model results are in the supplementary information (Tables S3 and S4).

3.4 | Temperature-size responses: Multi-species analyses

The influence of temperature during the immature stages on adult size for each species was compared in two ways: using percentage

change in size from all species and using only those with significant individual models (Table S5). There was little difference in the results between the two methods and, therefore, the results presented here are for species with significant models only. There was no significant difference in the mean percentage change in adult size between species in different size categories (early larval: $F = 0.011$, $df = 2$ and 17 , $p = 0.989$; late larval: $F = 0.633$, $df = 2$ and 17 , $p = 0.543$; pupal: $F = 1.087$, $df = 2$ and 15 , $p = 0.362$) or between species that overwintered in different life-history stages (early larval: $F = 0.439$, $df = 3$ and 16 , $p = 0.728$; late larval: $F = 0.985$, $df = 3$ and 16 , $p = 0.425$; pupal: $F = 1.114$, $df = 3$ and 14 , $p = 0.377$). There were no significant differences in mean percentage change in adult size between species occurring in different habitat types for the early and late larval stages (early larval: $F = 0.377$, $df = 3$ and 16 , $p = 0.771$; late larval: $F = 0.601$, $df = 3$ and 16 , $p = 0.624$), but there was for the pupal stage ($F = 3.649$, $df = 3$ and 14 , $p = 0.0392$), with the largest difference being between those in a woodland habitat and those in grassland habitats ($p = 0.0777$, $CI = -0.0935, 2.135$).

There was a significant difference in percentage change in adult size according to the three developmental stages for species with significant individual models ($F = 12.21$, $df = 2$ and 55 , $p < 0.001$), with differences between percentage changes per °C in the late larval stage and both the early larval ($p < 0.001$, $CI = 0.488, 1.472$) and pupal stages ($p = 0.00388$, $CI = -1.213, -0.202$; Figure 6). On

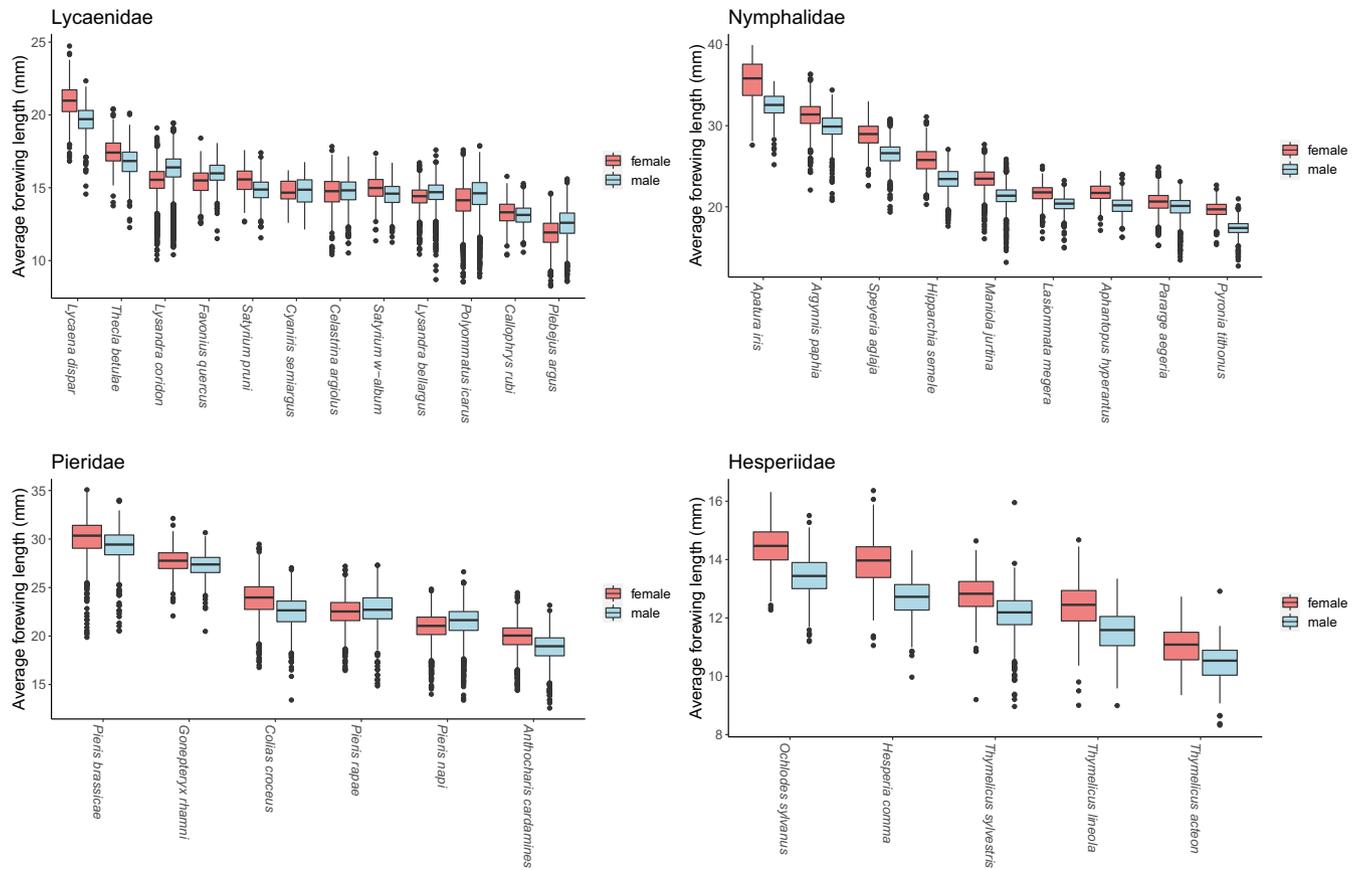


FIGURE 5 Size distributions of the Mothra measurements for each species trained for sex identification ($n = 32$ species). Most species ($n = 23$) have female biased sexual size dimorphism (including all species in Hesperidae and Nymphalidae). Seven species have male biased sexual size dimorphism (five species in Lycaenidae and two species in Pieridae) and two species do not have sexual size dimorphism

average, forewing length increased by 0.69% per °C increase in temperature during the late larval stage ($SE = 0.13$), decreased by 0.29% per °C in the early larval stage ($SE = 0.14$) and decreased by 0.02% per °C in the pupal stage ($SE = 0.17$). Additionally, there were significant differences in percentage change in adult size per °C temperature increase in the early larval stage between species from different families ($F = 15.74$, $df = 3$ and 16 , $p < 0.001$), with differences between the Hesperidae and all other families (Lycaenidae: $p < 0.001$, $CI = -1.383, -1.982$; Nymphalidae: $p = 0.0129$, $CI = -0.661, -1.195$; Pieridae: $p = 0.0142$, $CI = -1.305, -2.373$), and the Lycaenidae and the Nymphalidae ($p = 0.0101$, $CI = 0.722, 0.158$). On average, there was a 0.39% increase in average forewing length per °C temperature increase in the early larval stage for the Hesperidae ($SE = 0.09$), a 0.99% decrease in size per °C for the Lycaenidae ($SE = 0.13$), and a 0.28% decrease in size per °C in the Nymphalidae ($SE = 0.15$; Figure 6). A two-way ANOVA to test for differences in percentage changes in adult size between stages and families also found a significant interaction between life stage and family ($F = 3.004$, $df = 6$ and 46 , $p = 0.0146$). There is a large difference in responses to temperatures between larval stages for the Lycaenidae (Figure 6); on average, there is a large increase in adult size per °C temperature increase in the late larval stage (0.99%; $SE = 0.27$), but a large decrease in adult size per °C in the early larval stage (-0.99%; $SE = 0.13$).

For the linear mixed effects models, various models were tested using combinations of factors for the random effects. The simplest models tested included only one random effect (e.g. family or habitat type), then more factors were added, with the most complicated model including family, habitat type, overwintering stage and size category as random effects. Family was given priority in selection of factors as the above analyses showed there were significant differences in responses between families. The model with the lowest AIC value was Average forewing length (Ln) ~ Early larval temperature + Late larval temperature + Pupal temperature + (1|Family) (AIC = -20,579). For the species where sex could be determined, the model which was most significant was Average forewing length (Ln) ~ Early larval temperature + Late larval temperature + Pupal temperature + Sex + (1|Family) (AIC = -15,330). In both models (for all species and those which can be sexed), family explained the highest proportion of variance in the results (Table S6).

4 | DISCUSSION

The huge effort currently underway to digitise natural history collections will make museum specimens and their associated collecting data accessible to scientists all over the globe. A major reason for

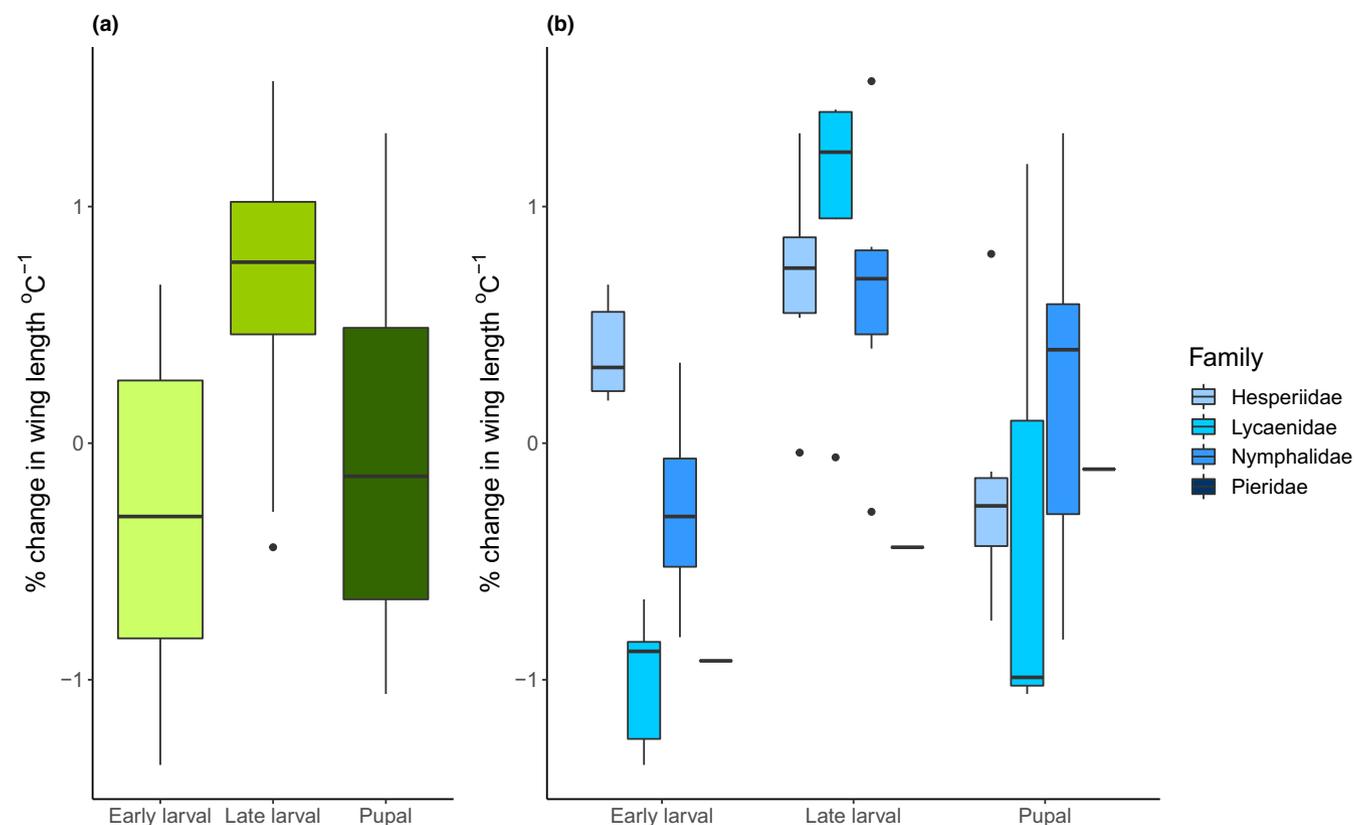


FIGURE 6 Boxplots of percentage change in adult size per $^{\circ}\text{C}^{-1}$ during (a) the early larval, late larval and pupal stages and (b) for species grouped by family within each stage. NB: There is only one species of Pieridae with a significant response (*Pieris napi*)

this mass digitisation effort is to accelerate their usage for global change research (Hedrick et al., 2020). There is now an increasing need for ecologists and museum scientists to collaborate with computer vision scientists in order to help make sense of these massive datasets. Our study is among the first to show that computer vision can accurately measure phenotypic data from very large digitised NHC datasets in order to test biotic response to climate change hypotheses.

We show that Mothra accurately measures multiple phenotypic aspects of butterfly specimens (Figures 1–3). It can also tell whether a specimen is pinned ventrally or dorsally, and its sex (for species where sexes are detectable by eye). While each of these attributes can be measured manually from images, the time involved would be immense: manual measurements of all imaged butterfly specimens ($n = 184,533$) by a single person would take $>3,000$ hr (or ~ 2 years, assuming regular working hours, and only forewing length measurements). Using Mothra, we were able to run all specimens in under 1 week by performing 10 analyses in parallel on a computer cluster, and could have reduced the time further by running more analyses in parallel (e.g. 50 analyses in parallel would have reduced the time to a mere 30 hr and remained within the capacity of the NHM cluster: 96 CPUs, 2 TB RAM, Centos 7 OS at the time of analysis).

Computer vision applied to digitised natural history collections will become a common tool in ecology and evolution research (Lürig et al., 2021). It will help scientists uncover unknown aspects of the biology and morphology of species, but also to confirm/

test hypotheses or suspected patterns based on previous research using manual measurements. For example, we test hypotheses that were formulated based on recent studies on temperature-size responses using manual measurements (Fenberg et al., 2016; Wilson et al., 2019). For most species with a significant temperature-size response (14/17), adult size increases with increasing temperature during the late larval stage (Figure 6), which is consistent with these studies. While some species did not show this response, there were no species, sexes or generations that showed the reverse response. This pattern, while suspected, is now clearer thanks to the application of computer vision to many more specimens and species. We suggest that this is because a higher volume and/or quality of food is available during years with warmer temperatures during the late larval stages. Therefore, late larval stage individuals may increase their feeding during years with warmer or longer growing seasons (Higgins et al., 2014), when food quality and quantity should be more plentiful. However, the optimum temperature for growth and the highest rate of growth will vary between species, sexes and generations (Gotthard, 2008).

As expected, different generations did not respond in the same way to temperature (Wilson et al., 2019). For the three bivoltine species, each responded in the first generation but not in the second (*P. bellargus*, *A. urticae* and *P. napi*). The different responses between generations were expected as the larvae of each generation experience different environmental conditions, which can affect adult body size (Horne et al., 2017). In addition, different temperature-size

responses between the sexes can also occur (Fenberg et al., 2016; Wilson et al., 2019). Of the 15 species in which the sexes were analysed separately, males had a significant temperature-size response in eight species and females responded in five species (three species had a significant response in both sexes), and there was no response from either sex for five species. In all but one of the species with significant results, the responses to temperature differed between males and females (i.e. the significance or direction of the temperature response was different in at least one developmental stage).

In the multi-species analyses, family explained the highest proportion of variance. Although significant responses to temperature in the late larval stage were always positive, the magnitude was greatest for Lycaenidae (Figure 6). The response to early larval stage temperatures shows the clearest differences between families: all Hesperidae species with significant models showed an increase in adult size with increasing temperature and Lycaenidae species showing a decrease in adult size with increasing temperature. Meanwhile, the species analysed within the Pieridae showed very little response; the only significant response was a decrease in adult size of generation one male *P. napi* with increasing temperature in the early larval stage. Overall, the Lycaenidae show the largest variation in responses to temperature between the immature stages, with a relatively large increase in adult size (0.99% per °C on average) with increasing temperatures in the late larval stage and a decrease in adult size (−0.99% per °C on average) with increasing temperature in the early larval stage. In the pupal stage, there was a range of positive and negative responses within each family. There are also some differences in response between species from different habitat types, particularly to temperature during the pupal stage, which may be due to differences in the microclimates within the habitats experienced by each stage (Figure S2).

Although the temperature-size responses across species are relatively modest (Figure 6), our results are on par with similar studies of temperature-size responses in insects. For example, a meta-analysis of laboratory studies shows that univoltine species (which constitute most of the species in our dataset) increase in size by 1.03% per °C on average (Horne et al., 2015). In addition, Wonglersak et al. (2020) find a similar per cent change (−1.10%) using museum collections paired with historic temperature records of British damselflies. Thus, whether temperature values are controlled in the laboratory or based on historic monthly averages (e.g. our study), the magnitude of size change appears to be broadly similar across some insect groups.

We can also now confirm that females are the larger sex for most species of British butterflies. While this is not particularly surprising given that female biased sexual size dimorphism (SSD) is commonly reported across insect species (Teder, 2014), our study represents the largest test of this phenomenon in terms of sample sizes. All species of Hesperidae and Nymphalidae have female biased SSD, but at least five species of Lycaenidae and two species of Pieridae have male biased SSD (Figure 5). Interestingly, four of the Lycaenidae species with male biased SSD are in the subfamily Polyommatae. In these species, there is also a strong colour dimorphism between the sexes. While the reason some species of this subfamily have

male biased SSD requires more research, we can make some inferences based on their natural history. In most species of insects, the males emerge earlier than females, termed protandry (Teder et al., 2021). In Polyommatae, males actively compete and swarm upon freshly emerged females to mate (e.g. in *P. bellargus*; Thomas & Lewington, 2014). Larger males may therefore be at a competitive advantage and promote male biased SSD. While the causes of SSD in insects is an ongoing debate and are likely to vary among taxa, our research and that from other invertebrate groups (e.g. Høye et al., 2009) shows that the direction and strength of temperature-size responses often varies by sex. Thus, the magnitude of SSD may increase, decrease or stay the same with increasing temperature.

Clearly, temperature-size responses in insects are a complex interaction between many different ecological, geographic, environmental, life history, evolutionary and historical variables. While the use of natural history collections can give us valuable clues to how temperature affects size, and computer vision can greatly accelerate data collection and analysis, there will always be a need to conduct field, laboratory and long-term monitoring studies to better understand the complexities of how insects will respond to climate change.

ACKNOWLEDGEMENTS

We thank the iCollections team (NHM) for capturing the images and data, Paul Ward (NHM) for providing server access to the images, Robert Foster (NHM) for access to and training on use of the HPC cluster, and James Durrant for developing an early wing measurement prototype. We thank Gary Fisher, Graham Wilson and Hannah O'Sullivan for their help with the image analysis, and Dennis Feng, Sera Yang, Teddy Tran and Théo Bodrito for their work on preliminary versions of Mothra. This work was supported by the Natural Environmental Research Council (grant number NE/L002531/1), and in part by the Gordon and Betty Moore Foundation (Grant GBMF3834) and by the Alfred P. Sloan Foundation (Grant October 27, 2013) to the University of California, Berkeley.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

P.B.F., R.J.W., B.W.P. and S.J.B. conceived of the ideas for the paper; A.F.d.S. and S.J.v.d.W. developed Mothra and wrote the accompanying methods section; B.W.P. and P.B.F. analysed the images and Mothra outputs; L.M.S. and R.J.W. performed manual measurements; R.J.W. performed all temperature-size analyses; R.J.W. and P.B.F. wrote most of the paper with helpful comments and edits from all co-authors.

DATA AVAILABILITY STATEMENT

All data and Mothra code are archived on Zenodo and iCollections images are accessible via the NHM data portal. Mothra code are available in Feng et al. (2021): <https://doi.org/10.5281/zenodo.5551110>. Mothra training data, including the networks trained

and images used in training are available in de Siqueira (2021): <https://doi.org/10.5281/zenodo.6066153>. The results of Mothra measurements of all iCollections butterflies are available in Price and Fenberg (2021): <https://doi.org/10.5281/zenodo.5759758>. All data used for Figures 3–6, Figures S1–S2 and the supplementary information tables are available in Wilson et al. (2022): <https://doi.org/10.5281/zenodo.6066070>. All images used are part of the iCollections project (Paterson et al., 2016) and are freely available on the NHM data portal: <https://doi.org/10.5519/0038559>

ORCID

Rebecca J. Wilson  <https://orcid.org/0000-0002-5705-6078>

Alexandre Fioravante de Siqueira  <https://orcid.org/0000-0003-1320-4347>

Benjamin W. Price  <https://orcid.org/0000-0001-5497-4087>

Stéfan J. van der Walt  <https://orcid.org/0000-0001-9276-1891>

Phillip B. Fenberg  <https://orcid.org/0000-0003-4474-176X>

REFERENCES

- Ariño, A. H. (2010). Approaches to estimating the universe of natural history collections data. *Biodiversity Informatics*, 7, 81–92.
- Atkinson, D. (1994). Temperature and organism size—A biological law for ectotherms. *Advances in Ecological Research*, 25, 1–58.
- Baar, Y., Friedman, A. L. L., Meiri, S., & Scharf, I. (2018). Little effect of climate change on body size of herbivorous beetles. *Insect Science*, 25, 309–316.
- Bjerger, K., Nielsen, J. B., Sepstrup, M. V., Helsing-Nielsen, F., & Høye, T. T. (2021). An automated light trap to monitor moths (Lepidoptera) using computer vision-based tracking and deep learning. *Sensors*, 21, 343.
- Bowden, J. J., Eskildsen, A., Hansen, R. R., Olsen, K., Kurle, C. M., & Høye, T. T. (2015). High-Arctic butterflies become smaller with rising temperatures. *Biology Letters*, 11, 20150574.
- Brooks, S. J., Self, A., Powney, G. D., Pearce, W. D., Penn, M., & Paterson, G. L. (2017). The influence of life history traits on the phenological response of British butterflies to climate variability since the late-19th century. *Ecography*, 40, 1152–1165.
- Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information*, 11, 125.
- Croton, P. J., Huber, K., Collinson, N., & Sparks, T. H. (2006). How well do the Central England temperature and the England and Wales precipitation series represent the climate of the UK? *International Journal of Climatology*, 26, 2287–2292.
- Davies, W. J. (2019). Multiple temperature effects on phenology and body size in wild butterflies predict a complex response to climate change. *Ecology*, 100, e02612.
- de Siqueira, A. F. (2021). Mothra training data [Data set]. *Zenodo*. <https://doi.org/10.5281/zenodo.6066153>
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255.
- DeVries, T., & Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*.
- Ewers-Saucedo, C., Allspach, A., Barilaro, C., Bick, A., Brandt, A., Fiege, D., Fütting, S., Hausdorf, B., Hayer, S., & Husemann, M. (2021). Natural history collections recapitulate 200 years of faunal change. *Royal Society Open Science*, 8, 201983.
- Fenberg, P. B., Self, A., Stewart, J. R., Wilson, R. J., & Brooks, S. J. (2016). Exploring the universal ecological responses to climate change in a univoltine butterfly. *Journal of Animal Ecology*, 85, 739–748.
- Feng, D. C., de Siqueira, A., Yang, S., Tran, T., Bodrito, T., & van der Walt, S. J. (2021). machine-shop/mothra: v1.0-rc.2 [Data set]. *Zenodo*. <https://doi.org/10.5281/zenodo.5551110>
- Forster, J., Hirst, A. G., & Atkinson, D. (2012). Warming-induced reductions in body size are greater in aquatic than terrestrial species. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 19310–19314.
- Gotthard, K. (2008). Adaptive growth decisions in butterflies. *Bioscience*, 58, 222–230.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., & Smith, N. J. (2020). Array programming with NumPy. *Nature*, 585, 357–362.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
- Hedrick, B. P., Heberling, J. M., Meineke, E. K., Turner, K. G., Grassa, C. J., Park, D. S., Kennedy, J., Clarke, J. A., Cook, J. A., & Blackburn, D. C. (2020). Digitization and the future of natural history collections. *Bioscience*, 70, 243–251.
- Higgins, J. K., Maclean, H. J., Buckley, L. B., & Kingsolver, J. G. (2014). Geographic differences and microevolutionary changes in thermal sensitivity of butterfly larvae in response to climate. *Functional Ecology*, 28, 982–989.
- Horne, C. R., Hirst, A. G., & Atkinson, D. (2015). Temperature-size responses match latitudinal-size clines in arthropods, revealing critical differences between aquatic and terrestrial species. *Ecology Letters*, 18, 327–335.
- Horne, C. R., Hirst, A. G., & Atkinson, D. (2017). Seasonal body size reductions with warming covary with major body size gradients in arthropod species. *Proceedings of the Royal Society B: Biological Sciences*, 284, 20170238.
- Howard, J., & Gugger, S. (2020). Fastai: A layered API for deep learning. *Information*, 11, 108.
- Høye, T. T., Årje, J., Bjerger, K., Hansen, O. L., Iosifidis, A., Leese, F., Mann, H. M., Meissner, K., Melvad, C., & Raitoharju, J. (2021). Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences of the United States of America*, 118, e2002545117.
- Høye, T. T., Hammel, J. U., Fuchs, T., & Toft, S. (2009). Climate change and sexual size dimorphism in an Arctic spider. *Biology Letters*, 5, 543–544.
- Hsiang, A. Y., Brombacher, A., Rillo, M. C., Mlencek-Vautraviers, M. J., Conn, S., Lordsmith, S., Jentzen, A., Henehan, M. J., Metcalfe, B., & Fenton, I. S. (2019). Endless Forams: >34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks. *Paleoceanography and Paleoclimatology*, 34, 1157–1177.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9, 90–95.
- Johnson, K. G., Brooks, S. J., Fenberg, P. B., Glover, A. G., James, K. E., Lister, A. M., Michel, E., Spencer, M., Todd, J. A., & Valsami-Jones, E. (2011). Climate change and biosphere response: Unlocking the collections vault. *Bioscience*, 61, 147–153.
- Kharouba, H. M., Lewthwaite, J. M., Guralnick, R., Kerr, J. T., & Vellend, M. (2019). Using insect natural history collections to study global change impacts: Challenges and opportunities. *Philosophical Transactions of the Royal Society B*, 374, 20170405.
- Kingsolver, J. G., Arthur Woods, H., Buckley, L. B., Potter, K. A., MacLean, H. J., & Higgins, J. K. (2011). *Complex life cycles and the responses of insects to climate change*. Oxford University Press.
- Lüriig, M. D., Donoughe, S., Svensson, E. I., Porto, A., & Tsuboi, M. (2021). Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. *Frontiers in Ecology and Evolution*, 9, 148.
- MacLean, H. J., Kingsolver, J. G., & Buckley, L. B. (2016). Historical changes in thermoregulatory traits of alpine butterflies reveal

- complex ecological and evolutionary responses to recent climate change. *Climate Change Responses*, 3, 1–10.
- McAllister, C. A., McKain, M. R., Li, M., Bookout, B., & Kellogg, E. A. (2019). Specimen-based analysis of morphology and the environment in ecologically dominant grasses: The power of the herbarium. *Philosophical Transactions of the Royal Society B*, 374, 20170403.
- Meineke, E. K., Davies, T. J., Daru, B. H., & Davis, C. C. (2019). Biological collections for understanding biodiversity in the Anthropocene. *Philosophical Transactions of the Royal Society B*, 374, 20170386.
- Na, S., Lee, H., Choi, S., & Yi, H. (2021). The relationship of mean temperature and 9 collected butterfly species' wingspan as the response of global warming. *Journal of Ecology and Environment*, 4, 1–8.
- Nelson, G., & Ellis, S. (2019). The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B*, 374, 20170391.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9, 62–66.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., & Antiga, L. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32, 8026–8037.
- Paterson, G., Albuquerque, S., Blagoderov, V., Brooks, S., Cafferty, S., Cane, E., Carter, V., Chainey, J., Crowther, R., & Douglas, L. (2016). iCollections—Digitising the British and Irish butterflies in the Natural History Museum, London. *Biodiversity Data Journal*, 4, e9559.
- Price, B. W., & Fenberg, P. B. (2021). Results of Mothra analysis of all iCollections butterflies [data set]. *Zenodo*. <https://doi.org/10.5281/zenodo.5759758>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241). Springer.
- Sheridan, J. A., & Bickford, D. (2011). Shrinking body size as an ecological response to climate change. *Nature Climate Change*, 1, 401–406.
- Smith, L. N. (2018). A disciplined approach to neural network hyperparameters: Part 1—Learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*.
- Svoboda, E. (2020). Artificial intelligence is improving the detection of lung cancer. *Nature*, 587, S20–S22.
- Teder, T. (2014). Sexual size dimorphism requires a corresponding sex difference in development time: A meta-analysis in insects. *Functional Ecology*, 28, 479–486.
- Teder, T., Kaasik, A., Taits, K., & Tammaru, T. (2021). Why do males emerge before females? Sexual size dimorphism drives sexual bimaturism in insects. *Biological Reviews*, 96, 2461–2475.
- Thomas, J., & Lewington, R. (2014). *The butterflies of Britain and Ireland*. Bloomsbury Publishing.
- Tseng, M., Kaur, K. M., Soleimani Pari, S., Sarai, K., Chan, D., Yao, C. H., Porto, P., Toor, A., Toor, H. S., & Fograscher, K. (2018). Decreases in beetle body size linked to climate change and warming temperatures. *Journal of Animal Ecology*, 87, 647–659.
- van der Walt, S. J., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). Scikit-image: Image processing in python. *PeerJ*, 2, e453.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17, 261–272.
- Wilson, R. J. (2021). *Disentangling the effects of ecology and life history on ectothermic temperature-size responses*. University of Southampton.
- Wilson, R. J., Brooks, S. J., & Fenberg, P. B. (2019). The influence of ecological and life history factors on ectothermic temperature-size responses: Analysis of three Lycaenidae butterflies (Lepidoptera). *Ecology and Evolution*, 9, 10305–10316.
- Wilson, R. J., Fenberg, P. B., Price, B. W., & Simon, L. (2022). Data from Wilson et al.: Applying computer vision to digitised natural history collections for climate change research: Temperature-size responses of British butterflies [Dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.6066070>
- Wonglersak, R., Fenberg, P. B., Langdon, P. G., Brooks, S. J., & Price, B. W. (2020). Temperature-body size responses in insects: A case study of British Odonata. *Ecological Entomology*, 45, 795–805.
- Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15, 749–753.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Wilson, R. J., de Siqueira, A. F., Brooks, S. J., Price, B. W., Simon, L. M., van der Walt, S. J. & Fenberg, P. B. (2022). Applying computer vision to digitised natural history collections for climate change research: Temperature-size responses in British butterflies. *Methods in Ecology and Evolution*, 00, 1–13. <https://doi.org/10.1111/2041-210X.13844>