RESEARCH ARTICLE

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Uncertainty in geographical estimates of performance and fitness

H. Arthur Woods[1] | Joel G. Kingsolver[2] | Samuel B. Fey[3] | David A. Vasseur[4]

[1]Division of Biological Sciences, University of Montana, Missoula, Montana

[2]Department of Biology, University of North Carolina, Chapel Hill, North Carolina

[3]Biology Department, Reed College, Portland, Oregon

[4]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut

Correspondence
H. Arthur Woods, Division of Biological Sciences, University of Montana, Missoula, MT.
Email: art.woods@mso.umt.edu

Abstract

1. Thermal performance curves (TPCs) have become key tools for predicting geographical distributions of performance by ectotherms. Such TPC-based predictions, however, may be sensitive to errors arising from diverse sources.

2. We analysed potential errors that arise from common choices faced by biologists integrating TPCs with climate data by constructing case studies focusing on experimental sets of TPCs and simulating geographical patterns of mean performance. We first analysed differences in geographical patterns of performance derived from two pairs of commonly used TPCs. Mean performance differed most (up to 30%) in regions with relatively constant mean temperatures similar to those at which the TPCs diverged the most.

3. We also analysed the effects of thermal history by comparing geographical estimates derived from (a) a broad TPC based on short-term measurements of insect larvae (*Manduca sexta*) with a history of exposure to thermal variation versus (b) a narrow TPC based on long-term measurements of larvae held at constant temperatures. Estimated mean performance diverged by up to 40%, and differences were magnified in simulated future climates.

4. Finally, to quantify geographical error arising from statistical error in fitted TPCs, we propose and illustrate a bootstrapping technique for establishing 95% prediction intervals on mean performance at each location (pixel).

5. Collectively, our analyses indicate that error arising from several underappreciated sources can significantly affect the mean performance values derived from TPCs, and we suggest that the magnitudes of these errors should be estimated routinely in future studies.

KEYWORDS
climate, climate change, climatic extremes, model, species distribution, temperature, thermal performance curve

1 | INTRODUCTION

How will new climates alter the performance of organisms and the ranges that they occupy? This question is pressing, as shifting ranges will affect the likelihood that populations will decline or go extinct (Bellard, Bertelsmeier, Leadley, Thuiller, & Courchamp, 2012; Thomas et al., 2004), how local communities and ecosystems will function

(Walther, 2010), whether they will provide new, different, or diminished ecosystem services (Schröter, 2005), and how much change will be required in approaches to agriculture (Nelson et al., 2014). For most species, answers hinge on uncertainty in both biotic and abiotic aspects of the problem (Suzuki, Rivero, Shulaev, Blumwald, & Mittler, 2014). Here we focus on abiotic conditions, which by themselves still pose enormous uncertainties—about future climates (IPCC, 2014),

about how climates are translated into the microclimates where organisms live (Frey et al., 2016; Pincebourde, Murdock, Vickers, & Sears, 2016; Potter, Woods, & Pincebourde, 2013; Woods, Dillon, & Pincebourde, 2015), and about the effects of climates and microclimates on organisms (Deutsch et al., 2008; Dillon, Wang, & Huey, 2010; Kearney & Porter, 2009; Sinclair et al., 2016). Because it is so complex, this last challenge is perhaps the most difficult: how should we translate the abiotic conditions (microclimates) experienced by individuals and populations into metrics of performance and fitness (Childress & Letcher, 2017; Sinclair et al., 2016)?

Thermal performance curves (TPCs) have become key tools in this translation (Angilletta, 2009; Deutsch et al., 2008; Dillon et al., 2010; Huey et al., 2009; Kingsolver, Diamond, & Buckley, 2013; Sinclair et al., 2016; Vasseur et al., 2014). TPCs are appealingly straightforward and easy to model. Although they restrict attention to only a single environmental variable, temperature, it is an important one (Kingsolver, 2009). In addition, recently developed high-resolution datasets (Kearney, Isaac, & Porter, 2014; Levy, Buckley, Keitt, & Angilletta, 2016; Rodell & Beaudoing, 2016) are starting to provide temperature values at scales relevant to small ectotherms nature (Kaspari, Clay, Lucas, Yanoviak, & Kay, 2015; Potter et al., 2013). Indeed, the increasing availability of fine-scale, gridded datasets of temperature provides powerful new approaches to integrating thermal physiology and ecology, which has been a central goal of physiological ecology for the past few decades (Huey & Kingsolver, 1989; Huey & Stevenson, 1979). We anticipate that TPCs will be used with climate data even more broadly in the future (Sinclair et al., 2016).

Despite their simplicity, TPCs can hide significant uncertainty arising from multiple methodological and analytical problems (Childress & Letcher, 2017; Sinclair et al., 2016). Here we consider three such problems, which have been largely unexplored. First, it is often unclear how to choose which underlying model(s) to fit and how to weigh the relative merits of alternatives. Angilletta (2006) analysed this problem by fitting five different equations to data on sprint speed of eastern fence lizards *Sceloporus undulatus*. The overall shape of the data was much like those in other TPCs (Dell, Pawar, & Savage, 2011)—a gradual rise in performance on the left side, an optimum at relatively high temperatures, and then a rapid decline on the right side at temperatures > 37°C. The model with the lowest AIC was a simple Gaussian curve, which had only four parameters and an r^2 of 0.70. However, the problem is worse than Angilletta's analysis indicates, because there are many other parametric models of TPCs that are routinely used (e.g., Krenek, Berendonk, & Petzoldt, 2011) and that could also have been fit. Indeed, it is possible for two (or more) models to have similar AIC scores, yet describe the data quite differently, and this is more likely to happen with increases in the number of models analysed. To further complicate the problem, TPCs can also be estimated nonparametrically, or by combined parametric and nonparametric models (Kingsolver, Diamond, & Gomulkiewicz, 2015; Stinchcombe & Kirkpatrick, 2012).

A related problem is whether to allow performance to go negative (Deutsch et al., 2008; Kingsolver et al., 2013). Many processes

stop at temperatures above some critical thermal maximum (CT_{max})—for example, rate of feeding or sprint speed—and, in these cases, it may be appropriate to use an underlying model that goes to zero. When integrating over time series of temperature, the assumption in such an approach usually is that the organism resumes activity again as the temperature falls below CT_{max}. However, TPCs can use more aggregated metrics of performance such as population growth (Frazier, Huey, & Berrigan, 2006), which can be negative (McKechnie & Wolf, 2009). The consequences of using negative values have largely gone unexamined in broader spatial and temporal contexts. One exception (Kingsolver et al., 2013) found that allowing negative values had a large influence at mid-latitudes, where intra-annual fluctuations in temperature are particularly large.

Second, TPCs can vary depending on the conditions under which they are measured, on life stage, and on the short- and long-term thermal histories of the organisms used in the assays (Dillon et al., 2016; Kingsolver, Higgins, & Augustine, 2015; Kingsolver et al., 2011; Rezende, Castañeda, & Santos, 2014; Sinclair et al., 2016). Historically, TPCs often were measured using organisms that had been held (sometimes for the entire life cycle) in constant temperatures. Extrapolating such TPCs to variable conditions in the wild is of course problematic. For example, Kingsolver, Higgins, et al. (2015) measured TPCs for larval growth rates in the sphingid moth *Manduca sexta* reared at three mean temperatures (20, 25, and 30°C) crossed with three levels of diurnal temperature variation (±0, 5, and 10°C). TPCs from the groups differed in several ways, including that larvae from diurnally fluctuating treatments had higher optimal temperatures and higher maximal growth rates than did larvae from constant-temperature treatments. TPCs can also show pronounced acclimation over seasonal, weekly, and even daily time scales (Dillon et al., 2016; Schulte, Healy, & Fangue, 2011), which can have important consequences for predictions about the effects of climate change (Calosi, Bilton, & Spicer, 2008). For example, how well *Drosophila* survived heat shock depended on the environmental temperature they had experienced during the prior 2 days (Overgaard & Sørensen, 2008). Finally, TPCs can change when organisms are exposed even briefly to temperatures in the zone between T_{opt} and CT_{max}. Exposure to these temperatures can induce significant shifts in the overall biochemistry, of which the best studied is the heat shock response and correlated downregulation of expression of other genes (Lindquist, 1986). These shifts can harden organisms to still higher temperatures, but they can also strongly depress performance above T_{opt} (Kingsolver & Woods, 2016). As a result, the optimal and maximal temperatures for performance can differ for TPCs measured at short (e.g., hourly or daily) versus longer (weekly or monthly) time scales, with important consequences for mean performance in fluctuating temperatures (Kingsolver, Higgins et al., 2015).

Third, researchers often fit models of TPCs to empirical data with the goal of estimating a single "best model" to be used in subsequent analyses without considering the magnitude of the uncertainty associated with that fit, or how that uncertainty translates into geographical uncertainty in patterns of performance. The
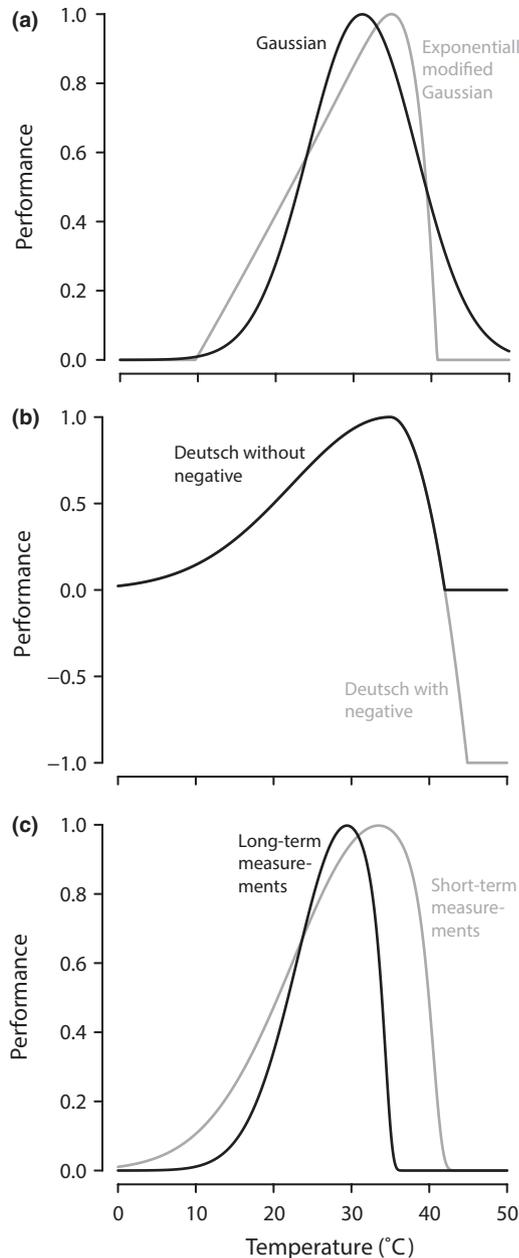
FIGURE 1 Three pairs of performance curves. (a) Gaussian and exponentially modified Gaussian, using parameters derived from fits to the data on lizard sprint speeds analysed by Angilletta (2006). (b) Two versions of the curve used by Deutsch et al. (2008) to model insect fitness as a function of temperature. The curves differ in whether or not they allow performance above CT_{max} to go below zero. (c) Two Thermal performance curves (TPCs) fitted to data on growth rate of larval *Manduca sexta* with different thermal histories (Kingsolver, Higgins, et al., 2015). Short-term measurements were taken over 24 hr in the last larval instar, using larvae that had been reared in diurnally fluctuating temperatures. Long-term measurements were taken from hatching to wandering, using larvae that had been held at constant temperatures. See text for equations and parameter values

underlying data, however, often are quite sparse and variable. Ignoring that variation can provide unrealistically high apparent levels of confidence in geographical patterns of performance. A better

approach would quantify geographical uncertainty in an explicit, systematic way.

Although these limitations have been noted previously, their consequences for ecological predictions have been largely ignored. Indeed, of 20 recent papers published on species distribution and performance curves (see Supporting Information Table S1 for search criteria and results), only one considered alternative forms of TPC models, two the importance of negative performance values, and none the effects of prior thermal history. Instead, authors focused on alternative sources of error, such as variation in curves arising among populations, closely related species, or from using different spatial or temporal scales of environmental data.

Using a set of case studies, we explore the consequences of these three classes of effects for predicting the geography of performance both in recent and future climates. We present a set of simulations in which we construct contrasting pairs of performance curves (see Figure 1)—symmetric versus skewed, non-negative versus negative, and short-term versus long-term—and then evaluate the relative estimates they make about performance of ectotherms in temporally variable environments on a continental scale. In particular, we predict mean performance across Central and North America using 10 years of summertime air temperatures from the GLDAS-2 data product (Rodell & Beaudoing, 2016). We also model what those performance curves predict under a range of future climates, which we generate by altering both mean and variance in temperature from the base environmental data. Finally, we develop a bootstrapping method for establishing 95% prediction intervals on mean performance at each location (pixel). The method establishes a systematic approach to quantifying geographical uncertainty. We illustrate the method using a TPC model fitted to empirical data on growth rates of *M. sexta*.

2 | MATERIALS AND METHODS

2.1 | Climate data and manipulations

We used GLDAS-2 data with a spatial resolution of 0.25° (*c.* 27 km at 30°N latitude) and a temporal resolution of 3 hr (Rodell & Beaudoing, 2016). Data were processed in R using the RASTER package (Hijmans, 2014; R Development Core Team, 2015). We extracted data on near-surface air temperature (band 24) in Central and North America for the months July, August, and September from the years 2001 to 2010. This 3-month "growing season" is shorter than the 6-month and 12-month periods considered by (Deutsch et al., 2008; Kingsolver et al., 2013), and we used the full set of diurnal fluctuations. These data constitute what we call the base environmental data. These data will miss some transient temporal extremes occurring between the 3-hourly time steps, and spatial averaging necessary for producing individual pixels will also reduce the incidence of small-scale spatial extremes. Nevertheless, they provide spatial resolution sufficient for our purposes and good temporal resolution compared to many other analyses that have used mean monthly temperatures or interpolated daily values.

We also produced a set of future climates that have higher mean temperatures (+2 or +4°C) and higher temperature variability (Kingsolver et al., 2013). We generated higher variability by using custom functions (Dillon et al., 2016) written in R to extract the amplitudes of the Fourier transform of the temperature time series at each spatial location. We then added additional variance (either 5 or 10% of the base variance for that location) spread across all frequencies and used an inverse transform to recover the new, manipulated time series. Thus, altogether we had nine sets of environmental data available for evaluating TPCs—3 means × 3 amounts of variability. Results from only two of those datasets are used in the main text (recent data from 2001 to 2010 and future climate with +4°C mean temperature and +10% variability). The full results for all nine climates are shown in Supporting Information Figures S1–S3. The online data do not qualitatively change our interpretations but simply provide greater detail in how shifts in mean and variance of temperatures (and their interactions) affect spatial estimates of performance.

2.2 | Ectotherm performance

Ectotherm body temperature is assumed to be the same as air temperature. This assumption obviously will be inaccurate some of the time, as an ectotherm's body temperature reflects inputs and outputs from the entire heat budget (Gates, 1980). In addition, many ectotherms sustain body temperatures that differ from air temperatures using behavioural thermoregulation to exploit locally available microsites (Huey, Peterson, Arnold, & Porter, 1989; Kaspari et al., 2015; Pincebourde & Casas, 2006; Woods, 2013; Woods et al., 2015). As a first approximation, however, this assumption is reasonable, as our goal is to illustrate the effects of chosen TPCs on estimates of performance and not to model body temperature accurately under all potential sets of environmental factors contributing to heat balance.

Using each TPC described below, we first estimated performance for each time point in each time series (30 months of summertime data) at every geographical location, ignoring the potential effects of thermal history on performance. We then calculated the arithmetic mean performance across that entire 30-month time period.

2.3 | Contrasting thermal performance curves

To evaluate the consequences of choosing one performance curve over another, we constructed three pairs of TPCs that differ in key ways related to choices that researchers commonly confront.

2.3.1 | Gaussian versus exponentially modified Gaussian

These two curves both performed well in the analysis of lizard sprint speeds done by Angilletta (2006). The Gaussian equation (Figure 1a) is

$$P = a \exp\left(-0.5\left(\frac{|T-b|}{c}\right)^2\right) \tag{1}$$

where P is performance and T is temperature, and the parameters a, b, and c are the height, position, and standard deviation (width) of the curve respectively. For both types of TPCs, we obtained parameter values by refitting the original data for *Sceloporus*, using the function *nls* in R, and then scaling the performance values to 1. See Supporting Information Table S2 for all fitted parameter values.

The exponentially modified Gaussian (Figure 1a) is

$$P = \frac{ac\sqrt{2\pi}}{2d} \exp\left(\frac{b-T}{d} + \frac{c^2}{2d^2}\right)\left[\frac{d}{|d|} - \text{erf}\left(\frac{b-T}{\sqrt{2}c} + \frac{c}{\sqrt{2}d}\right)\right] + f \tag{2}$$

where a and b are height and position, $ac\sqrt{2\pi}$ is the area, d is the time constant of the exponential part, and f is the intercept.

2.3.2 | Non-negative versus negative Deutsch

The equation used by Deutsch et al. (2008) is

$$P = \begin{cases} \exp\left\{-\left(\frac{T-T_{opt}}{2\sigma}\right)^2\right\} & \text{for} \quad T \leq T_{opt} \\ 1 - \left(\frac{T-T_{opt}}{T_{opt} - CT_{max}}\right) & \text{for} \quad T > T_{opt} \end{cases} \tag{3}$$

in which T_{opt} is the temperature giving the highest performance, CT_{max} is the critical thermal maximum, and σ defines the width of the Gaussian part on the left side (below T_{opt}).

Without further modification, the equation gives $P < 0$ at temperatures above CT_{max} and can rapidly go to extreme negative values. These values rarely make biological sense for TPCs measuring behavioural rates such as sprint speed, yet this approach allows for a gradient of detrimental impacts to accompany high environmental temperatures. Here, we force performance, at temperatures above CT_{max}, to go either to 0 ("non-negative Deutsch") or not to go below −1 (Figure 1b, "negative Deutsch"). This level gives somewhat more modest negative effects than did the value (−2) chosen by Kingsolver et al. (2013). These levels are arbitrary in the sense that we could have chosen any two values ≤ 0. However, the value of −1 scales the negative effects of very high temperatures realistically to the maximum positive performance value (+1) at T_{opt}. The other parameters were $T_{opt} = 35$, $\sigma = 9$, and $CT_{max} = 42$. Parameter values were chosen to give a curve shape above T_{opt} similar to the one in the short-term curve described below.

2.3.3 | Short- versus long-term measurements of thermal performance

Kingsolver et al. measured TPCs for growth rate of larval *Manduca sexta* reared under different thermal conditions and measured over different periods of time (Kingsolver & Nagle, 2007; Kingsolver, Higgins, et al., 2015). In 24-hr tests during the last instar ("short-term measurements"), larvae reared from hatching under fluctuating conditions (25 ± 10°C) had higher maximal growth rates and thermal optima (35°C) than did larvae reared at a constant temperature (25°C). They also measured hatching-to-wandering growth rates of

larvae held at one of five constant temperatures. In this experiment ("long-term measurements"), TPCs were narrower and shifted to the left: the thermal optimum was 30°C and growth rates were very low at 35°C. To model the TPCs, we fit the natural log of growth rate to a modified Gaussian function (Frazier et al., 2006; Kingsolver & Woods, 2016)

$$P = G_{max} \exp \left\{ - \exp \left(b \left(T - T_{opt} \right) - 6 \right) - a (T - T_{opt})^2 \right\} \qquad (4)$$

where G_{max} controls the height, a controls the rate of rise on the left side, and b the rate of fall on the right side (Figure 1c). The constant 6 controls how rapidly the curve declines to the right of T_{opt}.

2.4 | Bootstrapping the 95% prediction interval

When translating TPCs into geographical estimates of performance, researchers typically use a single curve fitted to empirical data on performance. Such an approach, however, ignores how error in the fitted curve translates into geographical error in performance. We propose to estimate this geographical error by bootstrapping the original data, refitting the TPC model to each bootstrapped dataset, estimating mean performance at each pixel using each model iteraction, and then characterizing the variation in performance values at each pixel. Although computationally expensive, the bootstrapping approach accounts in a straightforward way for variance–covariance relationships among the errors in the 3–6 parameters found in the typical nonlinear TPC model.

We illustrate the process using data on short-term growth rates of larval *M. sexta* across a range of temperatures (Kingsolver, Higgins, et al., 2015). The dataset consisted of 70 measurements of 24-hr growth rates of early fifth-instar larvae, with individuals held at one of seven temperatures (10° to 40°C at 5°C increments, with 10 larvae per temperature). Equation 4 was fitted to the entire dataset using the function nls2() from R package NLS2. The fitted model was then bootstrapped using the function nlsBoot() from the package NLSTOOLS, which uses nonparametric bootstrapping of mean-centred residuals. We used the first 500 bootstrapped sets of parameter values for subsequent geographical analyses. Although more elaborate schemes for choosing the number of bootstrap replicates could be implemented (e.g., (Pattengale, Alipour, Bininda-Emonds, Moret, & Stamatakis, 2009)), 500 exceeds the number recommended by a number of standard texts (Efron & Tibshirani, 1993; Manly, 1991).

For each of the 500 sets of parameter values, we estimated mean performance as described above—except that we used only the first 3 months of air-temperature data (July, August, and September of 2001), to reduce the total computational time. The result was a set of 500 maps of mean performance across North America. These were assembled into a RasterStack (R package RASTER) so that statistical distributions of values at each pixel could be calculated. For each pixel, we calculated the mean performance value and its associated 95% prediction interval (the 2.5% and 97.5% quantiles of the distribution).

3 | RESULTS

The choice of TPC can have substantial effects on the geographical distribution of estimated performance values, but these effects depended both on the location and magnitude of quantitative differences between TPCs, and on how much the location of those differences overlapped with commonly occurring temperatures in each temperature time series.

3.1 | Quantitative effects of climate change on geographical distributions of performance

For all TPCs evaluated (six total), there was a massive northward expansion of the zone of high performance (Figures 2–4) in the future climate scenario. There also was strong depression of performance in the south, especially in the Mojave, Sonoran, and Chihuahuan deserts and into areas of south-central US and northern Mexico. The amount of depression depended on the details of the performance curves. The bottom marginal panels (panels g and h in Figures 2–4) quantify these differences. In northern areas, performance increased by up to 0.4 units (on scale of 0–1), with the strongest increases occurring for the simple Gaussian curve, likely because this curve estimates positive values of performance up to 50°C. The largest local reductions in the southern zone were for the Deutsch curve with negative values (more negative than −0.4). This curve is the only one to allow negative values, and the temperatures in the "red zone" regularly go high enough to produce those negative values. The curve producing the largest total region of depression in the south was the long-term TPC based on measurements of *M. sexta* held at constant temperatures. This result reflects that this TPC is particularly narrow compared to the others, and that the right side is strongly depressed compared to its pair. Future temperatures across the southern US are high enough frequently enough to lie in its right-hand zone of low performance.

3.2 | Effect of choice of TPC from within a pair

The choice within a pair of contrasting TPCs mattered least for Gaussian versus exponentially modified Gaussian (Figure 2c,f). None of the geographical differences in performance was more than +0.2 or less than −0.2 (+20% and −20% respectively), and most values were close to zero. That result also did not depend on climate scenario. In contrast, the choice of TPC mattered most for long-term versus short-term curves (Figure 4c,f), and mattered still more (up to 0.4 units of difference in the southern US) when using the future climate scenario compared to using the recent past. In the Deutsch scenarios, mean performance in most locations did not differ between curves, as temperatures never rose above CT_{max}, the point at which the curves diverge. As a consequence, in the 2001–2010 climate scenario, there were minor differences in mean performance almost everywhere (Figure 3f). In the future scenarios, however, there were pronounced differences in performance in the
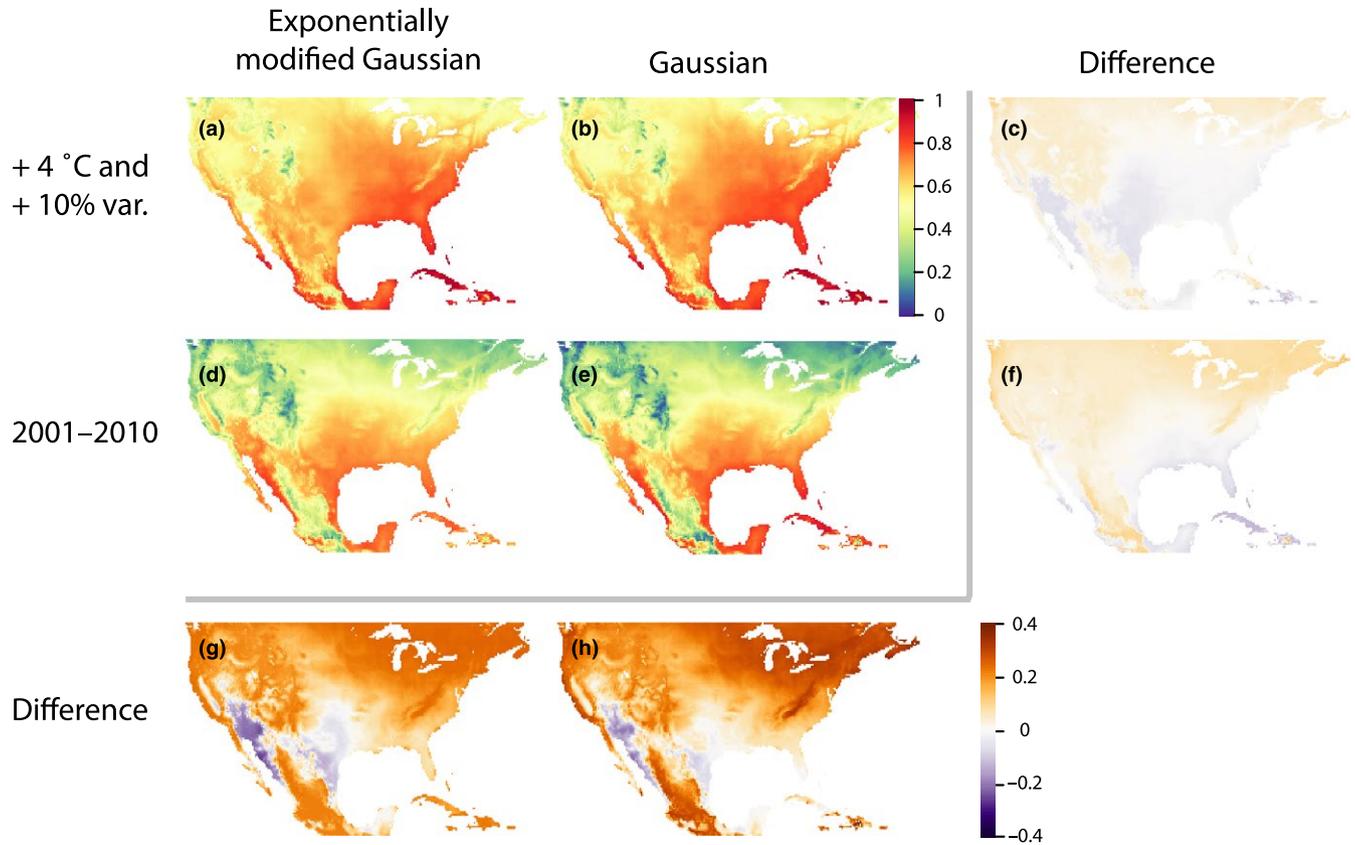
FIGURE 2 Performance maps for thermal performance curves modelled using Gaussian and exponentially modified Gaussian equations fitted to data on temperature-dependent sprint speeds of lizards *Sceloporus undulatas* (see Angilletta, 2006). The four in the upper left (a, b, d, e) show mean performance values (potential range 0–1) across 3 months of summertime temperatures over 10 years, arranged in a 2 × 2 grid. The left-most panels show results for the exponentially modified Gaussian curve and the right two panels show results for the Gaussian curve. The top row shows a simulated future environment in which mean temperature has risen 4°C everywhere, and overall variability has increased by 10%. The bottom row shows the results for the time period 2001–2010. Environmental data are derived from the GLDAS-2 data product (Rodell & Beaudoing, 2016). The maps on the outside of the grey lines show differences among the four primary maps: c is a–b, f is d–e, g is a–d, and h is b–e

southwestern deserts (Figure 3c), where temperatures regularly go above CT_{max}.

3.3 | Correlations between mean performance and mean daily maximum temperature

Mean performance emerges as a complex interaction between local patterns of temperature variation and the shape of the chosen TPC. Nevertheless, mean daily maximum temperature (mean T_{max}) is a useful, simple proxy for local environmental temperature, and how likely it is to reach temperatures at the high ends of the TPCs where performance differences are most pronounced. We calculated mean T_{max} across the entire 2001–2010 dataset (three summertime months only), and then plotted mean performance versus mean T_{max} at 5,000 randomly chosen locations (Figure 5). Within a pair of TPCs, the relative pattern of mean performance reflected the relative heights of the TPCs across temperatures. For example, in the Gaussian pair, the exponentially modified curve estimates higher mean performance at mean T_{max} lower than ~20°C, which is where the two TPCs cross (Figure 1a). Points from the Gaussian

curve were higher at T_{max} of 20–25°C. The pattern again reversed above 25°C, where performance was slightly higher for the exponentially modified curve. This reflects that in this zone of mean T_{max}, the temperature occasionally rose into the range 30–35°C, where the exponentially modified curve is slightly higher. Then at the highest mean T_{max} in the future climate scenario (Figure 5d), the Gaussian curve again gave higher performance, reflecting its relatively long right tail.

In the Deutsch curves, the points overlapped almost entirely in the recent environment, because temperatures everywhere rarely exceeded the CT_{max} of 42°C. In the future environment, however, there was striking divergence at the hottest mean T_{max}, which occurred in the southwestern deserts (see Figure 3). In the short- and long-term curves, the short-term curve estimated higher mean performance at almost all T_{max}, with a few exceptions in the small zone of temperatures near 25°C, which is the only temperature range in which the long-term curve exceeds the short-term curve. Even so, mean performance was higher in this zone only at locations in which temperature fluctuates with very low amplitude around a mean value of about 25°C, as occurs on some Caribbean islands. At the
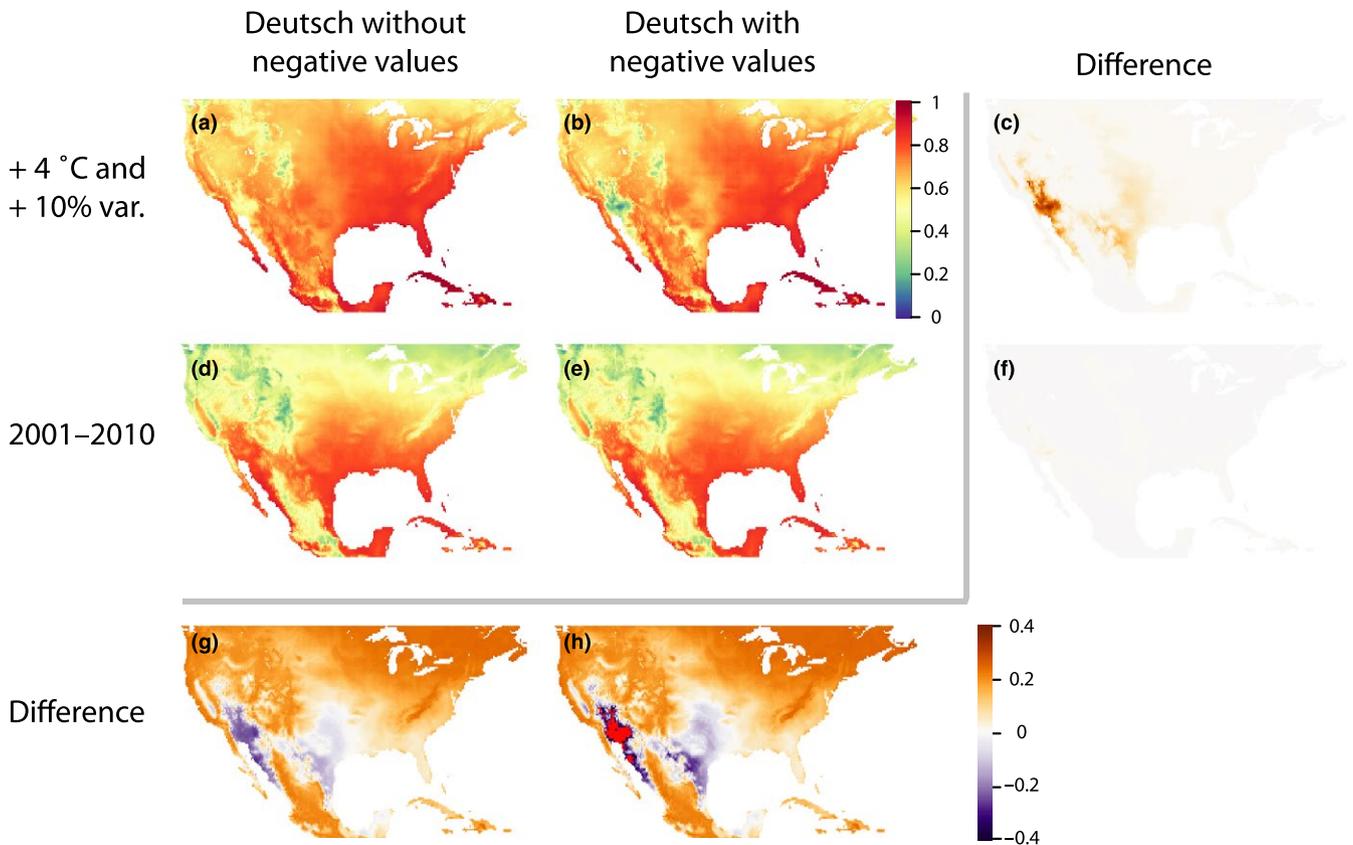
FIGURE 3 Performance maps for thermal performance curves modelled using the Deutsch equation (Deutsch et al., 2008), used to model the temperature dependence of insect fitness. The curves are allowed to go negative or just to zero above CT$_{max}$. Panels are organized identically to those in Figure 2. Bright red pixels in panel H represent values more negative than −0.4

highest mean T_{max}, the short-term curve gave much higher mean performance values than did the long-term curve.

3.4 | Bootstrapping the 95% prediction interval

A bootstrapped set of TPC models provided adequate descriptions of the uncertainty in the data, and the first 100 curves are shown in Figure 6a along with the original raw data. Additional analyses of the 500 bootstrapped performance values at each of 244 randomly chosen locations in North America indicated that distributions of estimated performance values at each pixel were approximately normal (Figure 6b). Normality was confirmed by examining QQ plots for each performance set, all of which were approximately linear (see Supporting Information Figure S4). In addition, a Shapiro–Wilk test of each set detected only about 3% that significantly violated the assumption of normality. These results indicate that the 2.5% and 97.5% quantiles of performance values reasonably approximate the 95% prediction interval.

Mean estimated performance from the bootstrapped dataset (full set of 500 curves) and the width of the 95% prediction interval are shown in Figures 7a,b respectively. The appropriate minimum number of bootstrap replicates necessary for obtaining reasonable estimates of the 95% prediction intervals will depend on the raw

data underlying the TPC; however, it is possible to assess iteratively how additional bootstraps contribute to uncovering additional uncertainty in prediction intervals. For example, the width of the prediction interval based on just the first 100 bootstrapped curves differed by at most 0.02 units, which is approximately 15% of the value of the mean interval width.

The largest 95% prediction intervals occurred in the Caribbean and Gulf of California, maritime areas with relatively small seasonal and diurnal temperature fluctuations (Figure 7a). In contrast, relative interval width (interval size in relation to the mean) showed a different pattern. The 95% prediction interval ranged from approximately 20% of the value of the mean (over much of south and central North America) to over 50% (*c*. 10% percent of all values) in the mountainous regions and in far northern latitudes.

4 | DISCUSSION

Thermal performance curves (TPCs) have become key tools for linking variation in weather and climate to geographical patterns of organismal performance and population fitness. The details of TPCs—their shapes and statistical properties—are important, but the consequences of error are poorly known. Here we show that the
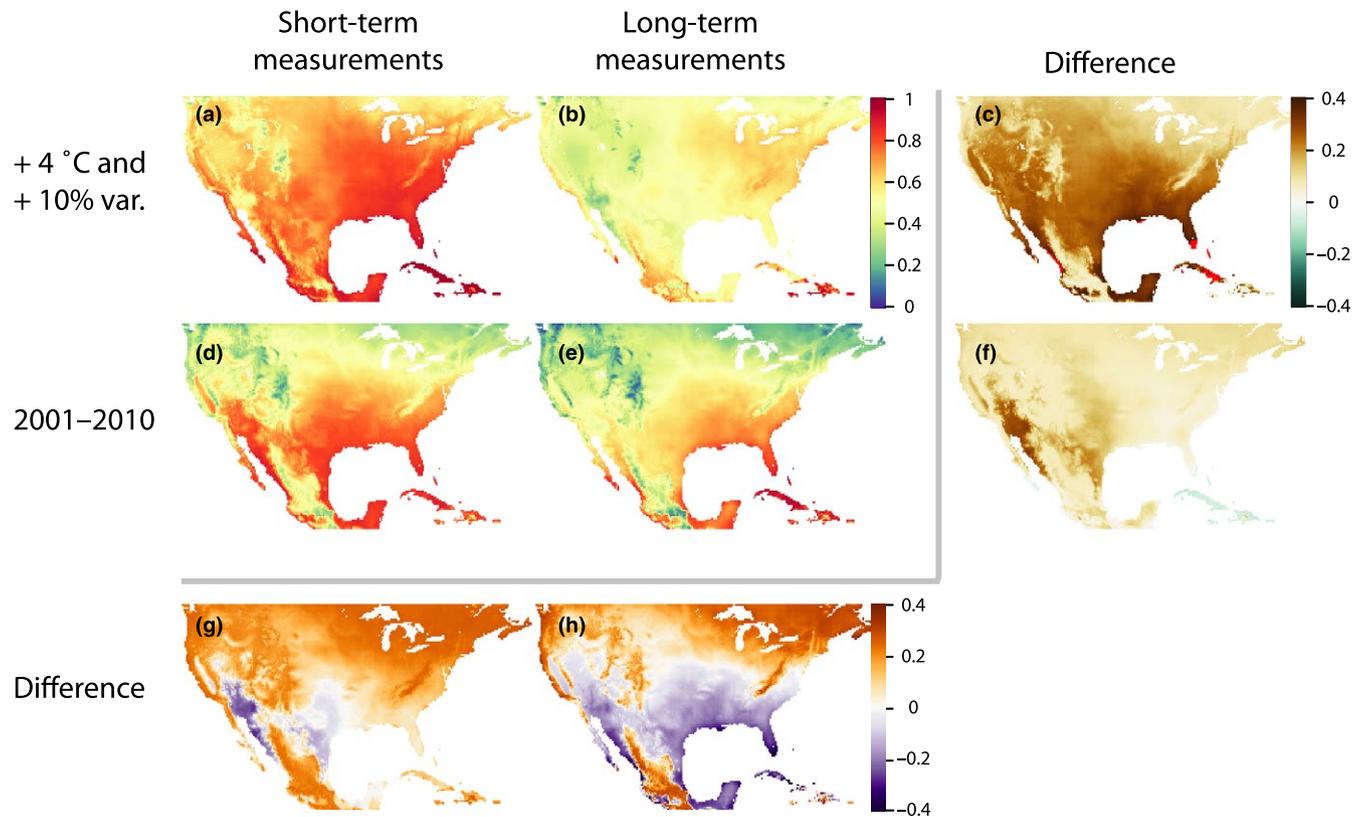
FIGURE 4 Performance maps for thermal performance curves modelled using the equations for short-term and long-term measurements of growth rate of larval *Manduca sexta*. Panels are organized identically to those in Figure 2. Bright red pixels in panel C represent values more positive than 0.4

underlying choice of TPC model, the thermal history of the animals used, and the statistical error inherent in fitting curves to data all can generate large differences in geographical patterns of mean performance. Our conclusions depend in part on a collection of choices made about which TPC models to analyse and which datasets to use to parameterize them. These choices, however, illustrate a phenomenon that we believe is likely to be general.

The consequences of both choice of TPC model and thermal history were greatest in locations with the highest mean temperatures, which is where pairs of curves differed the most. In contrast, higher levels of temperature variability had context-dependent effects (see below). In hot locations, predicted mean performance was disproportionately affected by subtle differences in the location of T_{opt}, by the shape of the decline to zero on the right sides of TPCs, and by whether or not negative performance was allowed. These conclusions reinforce those of other recent work that the biological consequences of high temperatures can be especially important (Buckley & Huey, 2016a, 2016b; Dowd, King, & Denny, 2015; McKechnie & Wolf, 2009; Williams et al., 2016). Our conclusions are likely to be conservative, as our analyses were based on TPCs for two thermophilic organisms (*Sceloporus* and *Manduca*). Differences in predicted patterns of performance are magnified when body temperatures rise frequently to T_{opt} or above, which will occur when using curves tuned to lower temperatures (i.e., for nonthermophilic organisms) or when considering performance in future climates with higher

temperatures. The key limitations of this study are that we did not allow spatial variation in TPCs, which ignores known latitudinal variation in breadth (Deutsch et al., 2008); we did not explicitly incorporate acclimation effects on the position and shape of TPCs (Calosi et al., 2008; Schulte et al., 2011); and we did not allow behavioural thermoregulation, which can strongly affect the range and sequence of body temperatures experienced (Huey et al., 2012; Kearney, Shine, & Porter, 2009; Sunday et al., 2014). These processes could be explored in additional simulations.

The consequences of statistical error in the fitted TPC model varied spatially and were quite large in some locations. The largest absolute 95% prediction intervals occurred in southern localities with relatively stable temperatures (Southern coast, Caribbean, Pacific Northwest coast). In contrast, continental areas with greater variation in temperature had smaller 95% prediction intervals. These patterns reflect that the bootstrapped curves have different shapes, such that greater amounts of temperature variation contributes performance values to the mean that are partially offsetting (see also Figure 8 below).

Quantifying error—both in the data itself and in predictions made by models built from those data—is fundamental to evaluating biological problems. So far, however, no methods have been proposed for propagating error from TPCs to errors in the geography of performance. The proposed bootstrap method does this in a straightforward way, and we suggest that it be used (and reported) routinely whenever TPCs are extrapolated to spatial patterns of
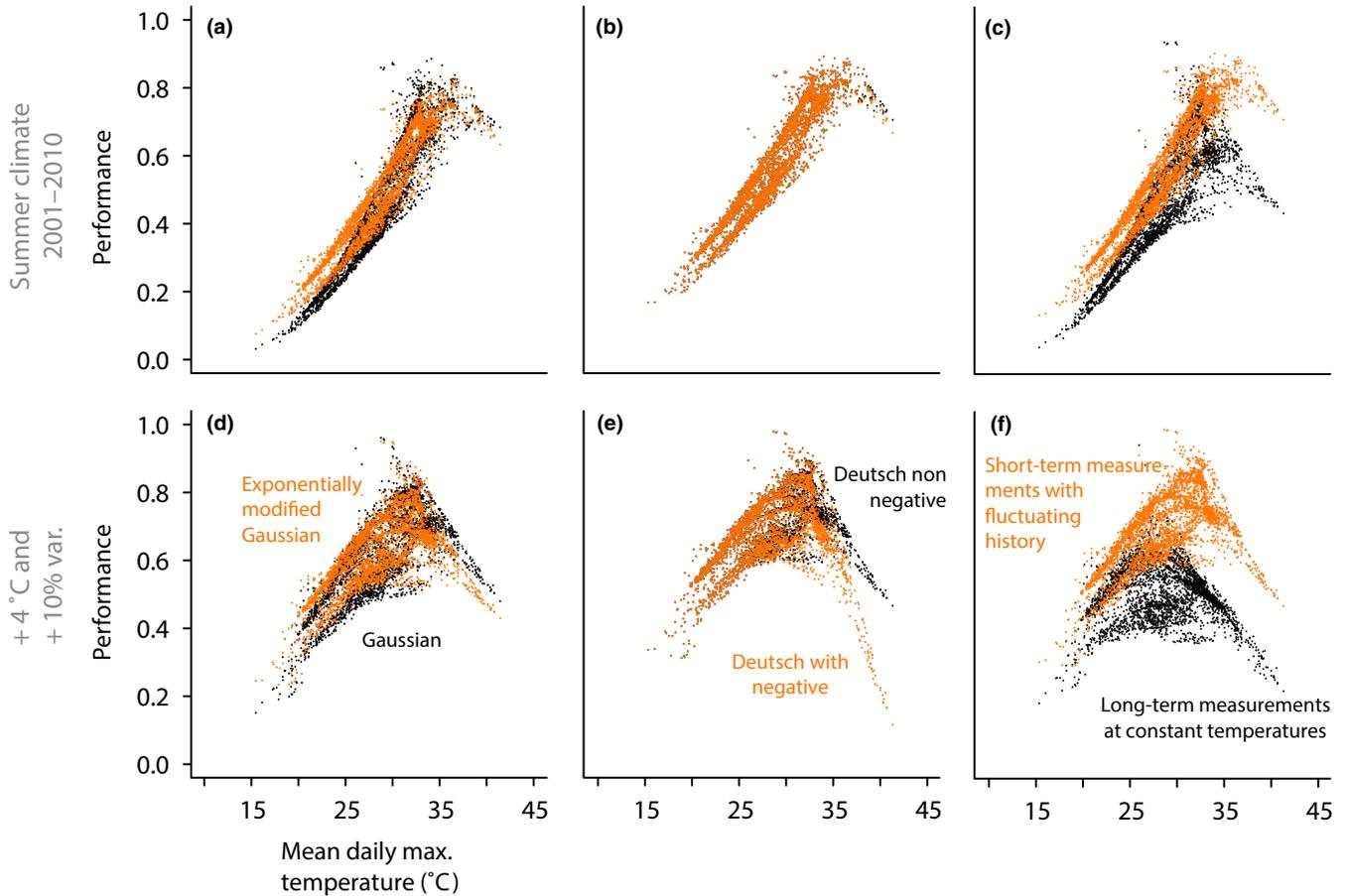
FIGURE 5 Bivariate plots of mean daily maximum temperature over the 30 summertime months used (10 years [2001–2010] × 3 months per year) versus mean performance for all pairs of Thermal performance curves (TPCs) in two sets of climatic conditions. In the entire map, there were 17,471 terrestrial pixels, of which we randomly subsampled 5,000 to make the panels
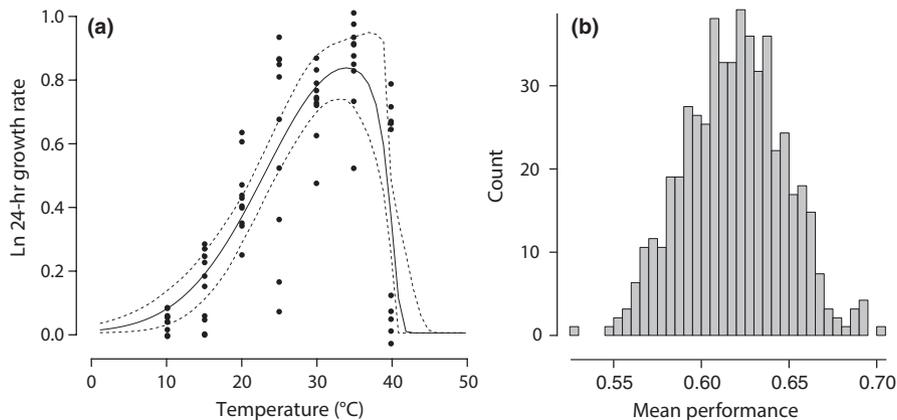

FIGURE 6 (a) Mean fitted Thermal performance curve (TPC) curve (Equation 4) and 95% confidence intervals from the bootstrapped dataset, along with the original dataset on growth rates of *Manduca sexta*. Each of the set of 500 curves was then used to predict mean growth rates at every pixel using the North American dataset on air temperature (GLDAS-2), which generated distributions of performance values for each pixel. (b) One such distribution for a location in central North America (35°N, 100°W) showing an approximately normal distribution of performance values. Additional analyses of distributions from a large set of randomly chosen pixels are provided in the online material (Supporting Information Figure S4)

mean performance or fitness. The bootstrapping approach is general and does not require using the 95% prediction interval. Other reasonable choices would include range, standard deviation, and

skew, all of which are descriptive statistics. Inferential statistics like the 95% CI are less suitable, as they depend on an arbitrarily chosen number of bootstrapped samples—for example, in a very large set
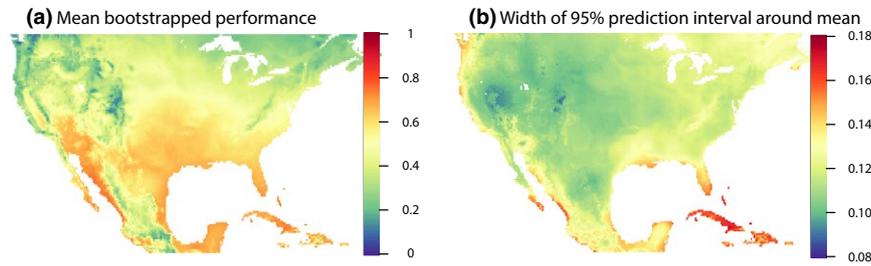
FIGURE 7 (a) The mean performance values estimated using 500 bootstrapped TPC models fitted to data on larval growth rates of *Manduca sexta* (see Figure 6). (b) Width of the 95% prediction interval around the mean, calculated by subtracting the 2.5% quantile from the 97.5% quantile in each pixel. Notice that the scales differ between the two panels
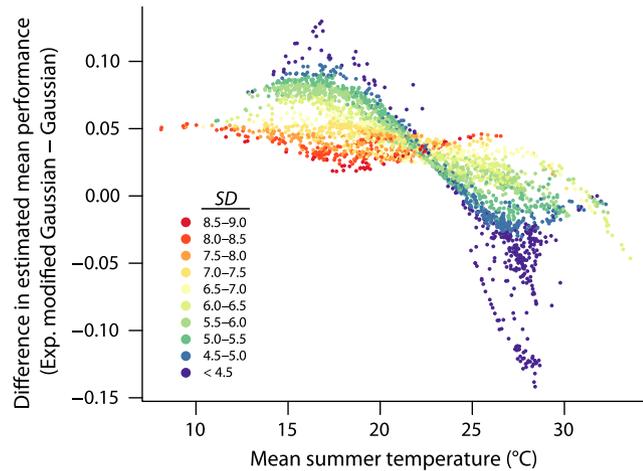

FIGURE 8 Difference between performance estimates of Gaussian (G) and Exponentially Modified Gaussian (EMG) performance curves as functions of mean and standard deviation of temperature at 3,000 randomly sampled locations in North America. Locations with high standard deviations (red points) in temperature had approximately the same differences, regardless of mean temperatures; in contrast, locations with low standard deviations in temperature (purple points) showed differences between curves that depended strongly on mean temperature. The dotted lines indicate the theoretical expectation for *SD* = 4 (heavy line) and *SD* = 8 (light line) assuming normally distributed temperatures, based on the G and EMG performance curves

of bootstrapped samples, the 95% confidence interval will converge onto the mean and thus carries no information.

The problem of error propagation from TPCs is in fact worse than indicated by our analysis—which focuses on TPCs derived from single datasets that are integrated with spatial data on temperature variation. Sinclair et al. (2016) proposed two additional, more integrative approaches to connecting TPCs to fitness. The first is to integrate TPCs describing processes at different levels of biological organization (e.g., (Kingsolver & Woods, 2016). Another is to consider multiple environmental factors at once rather than temperature alone (Harley et al., 2012; Munday, Crawley, & Nilsson, 2009). Both approaches require the integration of two or more curves, each containing its own errors, and how such errors combine to influence overall error in estimated performance remains unknown.

4.1 | Interactions between TPCs and local patterns of temperature variation

In our three paired comparisons, differences in mean performance emerge from how the frequency distribution of environmental temperatures maps onto the thermal performance curves. For the pair of Deutsch curves, the only range in which they differ is above CT_{max}. Thus, as a greater fraction of temperatures surpasses CT_{max}, mean performance exhibits a larger differential (e.g., in the Mojave and Chihuahuan Deserts). For the Gaussian and the short-/long-term response curves, mean performance differences between curves are diminished by temporal variation in temperature. For example, the Gaussian and exponentially modified Gaussian curves intersect four times over the biologically relevant temperature range, and the differences are relatively balanced (areas between the curves are similar between crossover points, with opposite signs); greater temperature variation then tends to have self-cancelling effects on mean performance as it is integrated through time (see Supporting Information Figure S5). In this case, we expect the shape of the curve to matter less in environments with higher temporal variance (Figure 8). However, as temperature distributions shift into portions of the TPCs that have stronger, more persistent differences, more temporal variation will be necessary to obtain this self-cancelling effect. These patterns suggest that differences between TPCs will have larger effects on predicted performance in low-variance thermal environments (i.e., in the tropics) than on high-variance environments (temperate).

Climate change is increasing the frequency and magnitude of extreme climatic events, including extreme high temperatures and heat waves (IPCC, 2014; Smith, 2011; Wang & Dillon, 2014). For several reasons, this has led to renewed focus on the importance of upper thermal limits and the top end of TPCs (Buckley & Huey, 2016a, 2016b; Denny, Hunt, Miller, & Harley, 2009; Dowd et al., 2015; Williams et al., 2016). First, performance changes most rapidly at temperatures above the optimal temperature, so that small changes in mean or variation in temperature can have large effects on mean performance in this temperature range (Ruel & Ayres, 1999). Deutsch curves, for example, gave very small differences in predicted performance in most locations but diverged substantially in those with the highest temperatures (Figures 2–4). The top end is also where alternative choices of the functional forms of TPCs differ the most (Figure 1), adding uncertainty to our predictions about the consequences of

high temperatures. Second, in most study systems, phenotypic and genetic variation in TPCs is greater at higher temperatures, at least for temperatures below upper thermal limits (Chevin & Hoffmann, 2017; Kingsolver, 2009). Greater variation in TPCs increases the expected responses of mean (and variance in) performance, and alters selection and evolutionary responses to high temperatures (Chevin & Hoffmann, 2017; Kingsolver & Buckley, 2017).

A third reason is that stress and acclimation responses may be particularly important at the top ends of TPCs, even at temperatures below the critical thermal limits (Williams et al., 2016). Heat shock protein (HSP) responses to high temperatures can improve performance and heat tolerance in the short run, but continued exposure to such temperatures can cause chronic stress, leading to reduced performance and survival (Feder & Hofmann, 1999). The time-dependent effects of stress and acclimation can generate different TPCs under short-term and long-term conditions, even for the same metric of performance (Kingsolver, Higgins, et al., 2015; Kingsolver & Woods, 2016). Our results show that the choice between short- and long-term TPCs can have dramatic effects on geographical patterns of predicted performance (Figures 4, 5). In this study, we used 3-hourly temperature data, so short-term TPCs are more appropriate. In contrast, some metrics of performance, such as lifetime fitness or population growth rate, can be defined only over longer time scales (generations), so that temperature data over longer time scales may be more appropriate (Kingsolver et al., 2013). Matching the time scales of environmental and biological data remains an important challenge in modelling the ecological and evolutionary responses of organisms to variable weather and climate change (Kingsolver & Buckley, 2017).

4.2 | Choosing TPC models

Our analyses emphasize that links between TPCs and predicted performance depend on how TPCs are measured and which models are used. The sensitivity of predictions to assumptions and methods calls for more work on defining best practices, which we suggest organizing around the following questions. First, what is the shape of TPCs above the thermal optimum? Often TPCs shapes above T_{opt} are inferred from measurements at just one or two temperatures. As a start, relatively more effort should be spent on collecting data at finer increments above T_{opt}; that is, investigators should not simply gather performance data at predetermined and evenly dispersed temperatures in the anticipated range of performance (e.g., 3 or 5°C). Second, when is it reasonable to allow negative performance, and how negative should performance be allowed to go? Biologically, negative performance can be reasonable in some circumstances (e.g., when studying rates of population growth or fluxes of biomass (McKechnie & Wolf, 2009)) but not others (e.g., when analysing sprint speeds, which can go only to zero). If negative performance is possible, then one faces the secondary question of how negative it should be allowed to go in the model, and how to specify the shape of the curve below zero. These are both empirical and theoretical questions, and we encourage research on tractable systems that could provide answers. Third,

how do historical patterns of temperature exposure affect current performance? In our study, the pair of TPCs giving the most divergent geographical patterns of performance was derived from data on *Manduca* caterpillars with different histories of temperature exposure (Figures 1 and 4). This result occurs because the integrated difference in areas between the curves is greatest for this pair. Although it is clear now that such history-dependent shifts are *likely* in many species, there has been little formal exploration of the magnitude of the shifts or how they should be modelled. There is the related question of how to match experimentally imposed histories of temperature exposure to the typical experiences of organisms in the wild.

Thermal performance curves are simple, powerful tools for integrating physiological and climatic data. Despite the central role they have begun to play in climate change biology, TPCs often depend on implicit or unexplored assumptions about how to collect and model data, and they often contain statistical error whose consequences go unexplored. The effects of these assumptions and errors can (and should) be made explicit. We suggest that quantification of geographical error become a routine part of any TPC-based study.

AUTHORS' CONTRIBUTIONS

H.A.W. conceived the overall idea for the paper, and J.G.K., S.B.F., and D.A.V. all contributed major sections; H.A.W. ran the simulations and analysed the outputs; H.A.W. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

Simulation results are available in the Dryad Data Repository at https://doi.org/10.5061/dryad.p2p7j6g (Woods, Kingsolver, Fey, & Vasseur, 2018).

ORCID

H. Arthur Woods (iD) http://orcid.org/0000-0002-3147-516X

Joel G. Kingsolver (iD) http://orcid.org/0000-0002-9839-0208

REFERENCES

Angilletta, M. J. (2006). Estimating and comparing thermal performance curves. *Journal of Thermal Biology*, *31*, 541–545. https://doi.org/10.1016/j.jtherbio.2006.06.002

Angilletta, M. J. (2009). *Thermal adaptation: A theoretical and empirical synthesis*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198570875.001.1

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). Impacts of climate change on the future of biodiversity. *Ecology Letters*, 15, 365–377. https://doi.org/10.1111/j.1461-0248.2011.01736.x

Buckley, L. B., & Huey, R. B. (2016a). How extreme temperatures impact organisms and the evolution of their thermal tolerance. *Integrative and Comparative Biology*, 56, 98–109. https://doi.org/10.1093/icb/icw004

Buckley, L. B., & Huey, R. B. (2016b). Temperature extremes: Geographic patterns, recent changes, and implications for organismal vulnerabilities. *Global Change Biology*, 22, 3829–3842. https://doi.org/10.1111/gcb.13313

Calosi, P., Bilton, D. T., & Spicer, J. I. (2008). Thermal tolerance, acclimatory capacity and vulnerability to global climate change. *Biology Letters*, 4, 99–102. https://doi.org/10.1098/rsbl.2007.0408

Chevin, L.-M., & Hoffmann, A. A. (2017). Evolution of phenotypic plasticity in extreme environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372, pii: 20160138.

Childress, E. S., & Letcher, B. H. (2017). Estimating thermal performance curves from repeated field observations. *Ecology*, 98, 1377–1387. https://doi.org/10.1002/ecy.1801

Dell, A. I., Pawar, S., & Savage, V. M. (2011). Systematic variation in the temperature dependence of physiological and ecological traits. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 10591–10596. https://doi.org/10.1073/pnas.1015178108

Denny, M. W., Hunt, L. J. H., Miller, L. P., & Harley, C. D. G. (2009). On the prediction of extreme ecological events. *Ecological Monographs*, 79, 397–421. https://doi.org/10.1890/08-0579.1

Deutsch, C. A., Tewksbury, J. J., Huey, R. B., Sheldon, K. S., Ghalambor, C. K., Haak, D. C., & Martin, P. R. (2008). Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 6668–6672. https://doi.org/10.1073/pnas.0709472105

Dillon, M. E., Wang, G., & Huey, R. B. (2010). Global metabolic impacts of recent climate warming. *Nature*, 467, 704–706. https://doi.org/10.1038/nature09407

Dillon, M. E., Woods, H. A., Wang, G., Fey, S. B., Vasseur, D. A., Telemeco, R. S., … Pincebourde, S. (2016). Life in the frequency domain: The biological impacts of changes in climate variability at multiple time scales. *Integrative and Comparative Biology*, 56, 14–30. https://doi.org/10.1093/icb/icw024

Dowd, W. W., King, F. A., & Denny, M. W. (2015). Thermal variation, thermal extremes and the physiological performance of individuals. *Journal of Experimental Biology*, 218, 1956–1967. https://doi.org/10.1242/jeb.114926

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Boca Raton, Florida: CRC Press. https://doi.org/10.1007/978-1-4899-4541-9

Feder, M. E., & Hofmann, G. E. (1999). Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology. *Annual Review of Physiology*, 61, 243–282. https://doi.org/10.1146/annurev.physiol.61.1.243

Frazier, M. R., Huey, R. B., & Berrigan, D. (2006). Thermodynamics constrains the evolution of insect population growth rates: 'warmer is better'. *The American Naturalist*, 168, 512–520. https://doi.org/10.1086/506977

Frey, S. J. K., Hadley, A. S., Johnson, S. L., Schulze, M., Jones, J. A., & Betts, M. G. (2016). Spatial models reveal the microclimatic buffering capacity of old-growth forests. *Science Advances*, 2, e1501392. https://doi.org/10.1126/sciadv.1501392

Gates, D. M. (1980). *Biophysical ecology*. New York: Springer-Verlag. https://doi.org/10.1007/978-1-4612-6024-0

Harley, C. D. G., Anderson, K. M., Demes, K. W., Jorve, J. P., Kordas, R. L., Coyle, T. A., & Graham, M. H. (2012). EFfects of climate change on global seaweed communities. *Journal of Phycology*, 48, 1064–1078. https://doi.org/10.1111/j.1529-8817.2012.01224.x

Hijmans, R. J. (2014). Geographic data analysis and modeling. *R Package Version*, 2, 15.

Huey, R. B., Deutsch, C. A., Tewksbury, J. J., Vitt, L. J., Hertz, P. E., Alvarez Pérez, H. J., & Garland, T. (2009). Why tropical forest lizards are vulnerable to climate warming. *Proceedings. Biological Sciences/The Royal Society*, 276, 1939–1948. https://doi.org/10.1098/rspb.2008.1957

Huey, R. B., Kearney, M. R., Krockenberger, A., Holtum, J. A. M., Jess, M., & Williams, S. E. (2012). Predicting organismal vulnerability to climate warming: Roles of behaviour, physiology and adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 1665–1679. https://doi.org/10.1098/rstb.2012.0005

Huey, R. B., & Kingsolver, J. G. (1989). Evolution of thermal sensitivity of ectotherm performance. *Trends in Ecology and Evolution*, 4, 131–135. https://doi.org/10.1016/0169-5347(89)90211-5

Huey, R., Peterson, C., Arnold, S., & Porter, W. (1989). Hot rocks and not-so-hot rocks: Retreat-site selection by garter snakes and its thermal consequences. *Ecology*, 70, 931–944. https://doi.org/10.2307/1941360

Huey, R. B., & Stevenson, R. D. (1979). Intergrating thermal physiology and ecology of ecotherms: A discussion of approaches. *The American Zoologist*, 19, 357–366. https://doi.org/10.1093/icb/19.1.357

IPCC. (2014). *Climate change 2014: Impacts, adaptation, and vulnerability*. Cambridge, UK: Cambridge University Press.

Kaspari, M., Clay, N. A., Lucas, J., Yanoviak, S. P., & Kay, A. (2015). Thermal adaptation generates a diversity of thermal limits in a rainforest ant community. *Global Change Biology*, 21, 1092–1102. https://doi.org/10.1111/gcb.12750

Kearney, M. R., Isaac, A., & Porter, W. P. (2014). microclim: Global estimates of hourly microclimate based on long-term monthly climate averages. *Scientific Data*, 1, 140006.

Kearney, M., & Porter, W. (2009). Mechanistic niche modelling: Combining physiological and spatial data to predict species' ranges. *Ecology Letters*, 12, 334–350. https://doi.org/10.1111/j.1461-0248.2008.01277.x

Kearney, M., Shine, R., & Porter, W. P. (2009). The potential for behavioral thermoregulation to buffer 'cold-blooded' animals against climate warming. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 3835–3840. https://doi.org/10.1073/pnas.0808913106

Kingsolver, J. G. (2009). The well-temperatured biologist. *The American Naturalist*, 174, 755–768.

Kingsolver, J. G., & Buckley, L. B. (2017). Quantifying thermal extremes and biological variation to predict evolutionary responses to changing climate. *Philosophical Transactions of the Royal Society of London Series B (Biological Sciences)*, 372, 20160147.

Kingsolver, J. G., Diamond, S. E., & Buckley, L. B. (2013). Heat stress and the fitness consequences of climate change for terrestrial ectotherms. *Functional Ecology*, 27, 1415–1423. https://doi.org/10.1111/1365-2435.12145

Kingsolver, J. G., Diamond, S. E., & Gomulkiewicz, R. (2015). Curve-thinking: Understanding reaction norms and developmental trajectories as traits. In L. B. Martin, C. K. Ghalambor & H. A. Woods (Eds.), *Integrative organismal biology* (pp. 39–53). Hoboken, NJ: Wiley.

Kingsolver, J. G., Higgins, J. K., & Augustine, K. E. (2015). Fluctuating temperatures and ectotherm growth: Distinguishing non-linear and time-dependent effects. *Journal of Experimental Biology*, 218, 2218–2225. https://doi.org/10.1242/jeb.120733

Kingsolver, J. G., & Nagle, A. (2007). Evolutionary divergence in thermal sensitivity and diapause of field and laboratory populations of *Manduca sexta*. *Physiological and Biochemical Zoology*, 80, 473–479. https://doi.org/10.1086/519962

Kingsolver, J. G., & Woods, H. A. (2016). Beyond thermal performance curves: Modeling time-dependent effects of thermal stress on

ectotherm growth rates. *The American Naturalist*, *187*, 283–294. https://doi.org/10.1086/684786

Kingsolver, J. G., Woods, H. A., Buckley, L. B., Potter, K. A., Maclean, H. J., & Higgins, J. K. (2011). Complex life cycles and the responses of insects to climate change. *Integrative and Comparative Biology*, *51*, 719–732. https://doi.org/10.1093/icb/icr015

Krenek, S., Berendonk, T. U., & Petzoldt, T. (2011). Thermal performance curves of *Paramecium caudatum*: A model selection approach. *European Journal of Protistology*, *47*, 124–137. https://doi.org/10.1016/j.ejop.2010.12.001

Levy, O., Buckley, L. B., Keitt, T. H., & Angilletta, M. J. (2016). A dynamically downscaled projection of past and future microclimates. *Ecology*, *97*, 1888–1888. https://doi.org/10.1002/ecy.1444

Lindquist, S. (1986). The heat-shock response. *Annual Review of Biochemistry*, *55*, 1151–1191. https://doi.org/10.1146/annurev.bi.55.070186.005443

Manly, B. F. J. (1991). *Randomization and monte carlo methods in biology*. Boca Raton, FL: CRC Press. https://doi.org/10.1007/978-1-4899-2995-2

McKechnie, A. E., & Wolf, B. O. (2009). Climate change increases the likelihood of catastrophic avian mortality events during extreme heat waves. *Biology Letters*, *6*, 253–256.

Munday, P. L., Crawley, N. E., & Nilsson, G. E. (2009). Interacting effects of elevated temperature and ocean acidification on the aerobic performance of coral reef fishes. *Marine Ecology Progress Series*, *388*, 235–242. https://doi.org/10.3354/meps08137

Nelson, G. C., van der Mensbrugghe, D., Ahammad, H., Blanc, E., Calvin, K., Hasegawa, T., … Willenbockel, D. (2014). Agriculture and climate change in global scenarios: Why don't the models agree. *Agricultural Economics (United Kingdom)*, *45*, 85–101. https://doi.org/10.1111/agec.12091

Overgaard, J., & Sørensen, J. G. (2008). Rapid thermal adaptation during field temperature variations in Drosophila melanogaster. *Cryobiology*, *56*, 159–162. https://doi.org/10.1016/j.cryobiol.2008.01.001

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., & Stamatakis, A. (2009). How many bootstrap replicates are necessary? Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, pp. 184–200).

Pincebourde, S., & Casas, J. (2006). Multitrophic biophysical budgets: Thermal ecology of an intimate herbivore insect-plant interaction. *Ecological Monographs*, *76*, 175–194. https://doi.org/10.1890/0012-9615(2006)076[0175:MBBTEO]2.0.CO;2

Pincebourde, S., Murdock, C. C., Vickers, M., & Sears, M. W. (2016). Fine-scale microclimatic variation can shape the responses of organisms to global change in both natural and urban environments. *Integrative and Comparative Biology*, *56*, 45–61. https://doi.org/10.1093/icb/icw016

Potter, K. A., Woods, H. A., & Pincebourde, S. (2013). Microclimatic challenges in global change biology. *Global Change Biology*, *19*, 2932–2939. https://doi.org/10.1111/gcb.12257

R Development Core Team. (2015). R: A language and environment for statistical computing.

Rezende, E. L., Castañeda, L. E., & Santos, M. (2014). Tolerance landscapes in thermal ecology. *Functional Ecology*, *28*, 799–809. https://doi.org/10.1111/1365-2435.12268

Rodell, M., & Beaudoing, H. K. (2016). Noah land surface model L4 monthly 0.25 x 0.25 degree version 2.0, version 020. Greenbelt (MD): Goddard Earth Sciences Data and Information Services Center (GES DISC).

Ruel, J. J., & Ayres, M. P. (1999). Jensen's inequality predicts effects of environmental variation. *Trends in Ecology and Evolution*, *14*, 361–366. https://doi.org/10.1016/S0169-5347(99)01664-X

Schröter, D. (2005). Ecosystem service supply and vulnerability to global change in Europe. *Science*, *310*, 1333–1337. https://doi.org/10.1126/science.1115233

Schulte, P. M., Healy, T. M., & Fangue, N. A. (2011). Thermal performance curves, phenotypic plasticity, and the time scales of temperature exposure. *Integrative and Comparative Biology*, *51*, 691–702. https://doi.org/10.1093/icb/icr097

Sinclair, B. J., Marshall, K. E., Sewell, M. A., Levesque, D. L., Willett, C. S., Slotsbo, S., … Huey, R. B. (2016). Can we predict ectotherm responses to climate change using thermal performance curves and body temperatures? *Ecology Letters*, *19*, 1372–1385. https://doi.org/10.1111/ele.12686

Smith, M. D. (2011). An ecological perspective on extreme climatic events: A synthetic definition and framework to guide future research. *Journal of Ecology*, *99*, 656–663. https://doi.org/10.1111/j.1365-2745.2011.01798.x

Stinchcombe, J. R., & Kirkpatrick, M. (2012). Genetics and evolution of function-valued traits: Understanding environmentally responsive phenotypes. *Trends in Ecology and Evolution*, *27*, 637–647. https://doi.org/10.1016/j.tree.2012.07.002

Sunday, J. M., Bates, A. E., Kearney, M. R., Colwell, R. K., Dulvy, N. K., Longino, J. T., & Huey, R. B. (2014). Thermal-safety margins and the necessity of thermoregulatory behavior across latitude and elevation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 5610–5615. https://doi.org/10.1073/pnas.1316145111

Suzuki, N., Rivero, R. M., Shulaev, V., Blumwald, E., & Mittler, R. (2014). Abiotic and biotic stress combinations. *New Phytologist*, *203*, 32–43. https://doi.org/10.1111/nph.12797

Thomas, C. D., Thomas, C. D., Cameron, A., Cameron, A., Green, R. E., Green, R. E., … Williams, S. E. (2004). Extinction risk from climate change. *Nature*, *427*, 145–148. https://doi.org/10.1038/nature02121

Vasseur, D. A., DeLong, J. P., Gilbert, B., Greig, H. S., Harley, C. D. G., McCann, K. S., … O' Connor, M. I. (2014). Increased temperature variation poses a greater risk to species than climate warming. *Proceedings of the Royal Society B*, *281*, 20132612. https://doi.org/10.1098/rspb.2013.2612

Walther, G.-R. (2010). Community and ecosystem responses to recent climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 2019–2024. https://doi.org/10.1098/rstb.2010.0021

Wang, G., & Dillon, M. E. (2014). Recent geographic convergence in diurnal and annual temperature cycling flattens global thermal profiles. *Nature Climate Change*, *4*, 1–5.

Williams, C. M., Buckley, L. B., Sheldon, K. S., Vickers, M., Pörtner, H.-O., Dowd, W.W., … Stillman, J. H. (2016). Biological impacts of thermal extremes: Mechanisms and costs of functional responses matter. *Integrative and Comparative Biology*, *56*, 73–84. icw013. https://doi.org/10.1093/icb/icw013

Woods, H. A. (2013). Ontogenetic changes in the body temperature of an insect herbivore. *Functional Ecology*, *27*, 1322–1331. https://doi.org/10.1111/1365-2435.12124

Woods, H. A., Dillon, M. E., & Pincebourde, S. (2015). The roles of microclimatic diversity and of behavior in mediating the responses of ectotherms to climate change. *Journal of Thermal Biology*, *54*, 86–97. https://doi.org/10.1016/j.jtherbio.2014.10.002

Woods, H. A., Kingsolver, J. G., Fey, S. B. & Vasseur, D. A. (2018). Data from: Uncertainty in geographical estimates of performance and fitness. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.p2p7j6g

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.