Machine-learning model led design to experimentally test species thermal limits: the case of kissing bugs (Triatominae)

Jorge E. Rabinovich[1¶*], Agustín Alvarez Costa[2,3&], Ignacio Muñoz[2,3&], Pablo E. Schilman[2,3&], Nicholas Fountain-Jones[4¶]

[1] Centro de Estudios Parasitológicos y de Vectores (CEPAVE CONICET-CCT La Plata, UNLP), National University of La Plata, La Plata, Argentina.

[2] Laboratorio de Ecofisiología de Insectos, Departamento de Biodiversidad y Biología Experimental, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Buenos Aires, Argentina.

[3] Instituto de Biodiversidad y Biología Experimental y Aplicada (IBBEA). CONICET – Universidad de Buenos Aires, Buenos Aires, Argentina.

[4] School of Natural Sciences, University of Tasmania, Hobart, Australia.

[*] Correspondent author

E-mail: jorge.rabinovich@gmail.com

¶These authors contributed equally to this work.

&These authors also contributed equally to this work

Short title: Machine-learning, laboratory design and Triatominae thermal limits

20

Abstract

Species Distribution Modelling (SDM) determines habitat suitability of a species across geographic areas using macro-climatic variables; however, micro-habitats can buffer or exacerbate the influence of macro-climatic variables, requiring links between physiology and species persistence. Experimental approaches linking species physiology to micro-climate are complex, time consuming and expensive. E.g., what combination of exposure time and temperature is important for a species thermal tolerance is difficult to judge *a priori*. We tackled this problem using an active learning approach that utilized machine learning methods to guide thermal tolerance experimental design for three kissing-bug species (Hemiptera: Reduviidae: Triatominae), vectors of the parasite causing Chagas disease. As with other pathogen vectors, triatomines are well known to utilize micro-habitats and the associated shift in microclimate to enhance survival. Using a limited literature-collected dataset, our approach showed that temperature followed by exposure time were the strongest predictors of mortality; species played a minor role, and life stage was the least important. Further, we identified complex but biologically plausible nonlinear interactions between temperature and exposure time in shaping mortality, together setting the potential thermal limits of triatomines. The results from this data led to the design of new experiments with laboratory results that produced novel insights of the effects of temperature and exposure for the triatomines. These results, in turn, can be used to better model micro-climatic envelope for the species. Here we demonstrate the power of an active learning approach to explore experimental space to design laboratory studies testing species thermal limits. Our analytical pipeline can be easily adapted to other systems and we provide code to allow practitioners to perform similar analyses. Not only does our approach have the potential to save time and money: it can also increase

our understanding of the links between species physiology and climate, a topic of increasing ecological importance.

Author summary

Species Distribution Modelling determines habitat suitability of a species across geographic areas using macro-climatic variables; however, micro-habitats can buffer or exacerbate the influence of macro-climatic variables, requiring links between physiology and species persistence. We tackled the problem of the combination of exposure time and temperature (a combination difficult to judge *a priori*) in determining species thermal tolerance, using an active learning approach that utilized machine learning methods to guide thermal tolerance experimental design for three kissing-bug species, vectors of the parasite causing Chagas disease. These bugs are found in micro-habitats with associated shifts in microclimate to enhance survival. Using a limited literature-collected dataset, we showed that temperature followed by exposure time were the strongest predictors of mortality, that species played a minor role, that life stage was the least important, and a complex nonlinear interaction between temperature and exposure time in shaping mortality of kissing bugs. These results led to the design of new laboratory experiments to assess the effects of temperature and exposure for the triatomines. These results can be used to better model micro-climatic envelope for species. Our active learning approach to explore experimental space to design laboratory studies can also be applied to other environmental conditions or species.

Introduction

The main environmental requirements for any organism to be in thermodynamic equilibrium over a reasonable length of time in order to survive are well known for more than a half century [1]. Of these,

radiation absorbed, wind speed, and air temperature are physiological requirements for a certain body temperature or temperature range, and are referred to as the 'climate space', and constitute the conditions which animals must fulfil in order to survive [2]. Much effort has gone into applying Species Distribution Modelling (SDM) to model climate space across different geographic areas to determine habitat suitability, including species that are disease vectors [3-9]. In general, the SDM methodology utilizes macro-climatic variables to predict the distribution of a species; however, for many disease vector species natural and human-made micro-habitats can buffer or exacerbate the influence of the macro-climatic variables [10,11]. While these models increasingly harness high resolution climatic data down to ~1km^2 (e.g., WorldClim 2 [12]), the links between physiology, disease transmission and vector survival to micro-climate are opaque. Consequently, the use of these climatic variables leads to some caveats on the reliability of the suitability estimates produced by the SDM. However, accurately quantifying micro-climatic relationships and tolerances pose significant problems for most disease vector species.

Mathematical models have helped predict operative temperatures but also show how evaporative cooling and metabolic heating might cause the body temperature of an organism to deviate from the operative temperature [13]. However, despite mathematical models have provided important insights into what factors influence operative temperatures, they are impractical for mapping thermal environments at a sufficient resolution to understand selective pressures on behavior and physiology [14,15]. This is because to compute operative temperatures from a mathematical model many variables must be known (solar radiation, ground reflectivity, air temperature, ground temperature, and wind speed), which represents an overwhelming task for a large number of locations as needed for fine-scale mapping [15]. Operative temperatures have been computed for a limited number of

86 microclimates, such as full sun and full shade, and for a few animal species [16-18], but a general
87 methodology is still needed.

88 Because of their use of a variety of micro-habitats, the trouble of using macroclimatic variables to
89 predict vector species distributions is particularly serious [19]. This is true for the kissing bugs
90 (Hemiptera: Reduviidae: Triatominae), a group of species vectors of *Trypanosoma cruzi*, the parasite
91 causing Chagas disease in Latin America. This disease is endemic in 21 countries, and it is estimated
92 that affects between 6 and 8 million individuals (with 25 million people at risk of infection), resulting in
93 approximately 12,000 deaths per year [20,21]. The triatomines are a subfamily that comprise around
94 150 species grouped in 18 genera and six tribes [22]. A commonly observed behavior of most of these
95 species is to enter domestic and peri-domestic structures ("intrusion"), with some of them trying to
96 colonize the human habitat ("domiciliation"), making the sylvatic species a possible source of infection
97 by *T. cruzi* [23]. Due to their generally nocturnal habits and hidden refuges, they are not only
98 inconspicuous and hard to collect in the field, but also endure extreme values of macro-climatic
99 variables that would not allow them to survive without the use of micro-habitats [24].

100 For triatomines, as with other disease vector species, there is very limited information available to
101 adequately link insect physiology to micro-climatic variables. Much more experimental work is
102 necessary, but those experiments are laborious, costly, and demanding a large number of insects.
103 Additionally, the experimental design is a complex one for, in general, the micro-climatic variables
104 (*e.g.*, temperature) need to be included with at least two factors: the value of the micro-climatic
105 variable itself, and the exposure time (or duration) to each value of the micro-climatic variable. It is
106 difficult to guess beforehand the limits and number of those two factors, or the impact of each of them
107 (and their combination) on the demographic parameters, in order to design a laboratory experiment.

To facilitate the design of such kind of experiments (which, in theory, could involve hundreds or thousands of combinations) we propose a methodology based upon a machine learning pipeline approach [25] in order to predict the survival of triatomines by different combinations of micro-climatic temperature values and exposure times. This approach leverages recent advance in machine learning to construct powerful but interpretable predictive models that can guide what combinations of variables could be important in shaping a species thermal limit and thus can configure feasible experimental designs. Importantly these models can quantify complex non-linear interactions between variables that can be difficult to include *a priori* in a model.

Finding computational solutions to guide experimental design or 'active learning' is not a new idea and has been used to guide experiments to better understand complex gene regulatory networks [26]. Active learning is an iterative process in that a model is formed from preliminary data that guides new experiments that in turn generates novel data that is used to update the original model (Fig 1). For example, machine learning approaches have been successfully used to guide gene knock-out experiments where the number of experiments is quadratic to the number of genes (see [26] for a review on the topic). However, active learning approaches are rarely applied to explore species thermal limits. We show the utility of the approach to better understand the thermal ecology of an important group of disease vector species that could easily be adapted to guide ecological experiments more broadly.

Fig 1. Schematic description of the active learning approach used in this study. We have highlighted the context and purpose of the work, the approach and methods, and an outline of the application of

129 the main results. GBM: Gradient boost model, SVM: Support vector machine, RF: Random forests,
130 BGLM: Bayesian general linear model. *Best model: model with the highest root mean square error
131 (RMSE).

132

133 ## Materials and methods
134

135 **Available laboratory information for triatomines**

136 **Data source.** The only previous data available were the results from thermal shock experiments
137 (usually at 40 °C) applied on different stages and adults of several triatomine species, exposed for 1
138 and 12 hours. In total we were able to find seven bibliographic sources with adequate data for three
139 species of triatomines (*Triatoma infestans*, *Rhodnius prolixus*, and *Panstrongyus megistus*). See Data
140 Cases below.

141 **Data preparation.** As in most of the cases results were presented as survival values (l_x, with values
142 from 1 to ≥0, which express the proportion of the initial number of individuals alive at day x, and where
143 $1 \leq x \leq 30$, 30 being the number of days of the control period of all experimental insects), we
144 converted these survival values to mortality (d_x, or number of individuals dying in day x, where again x
145 is from day 1 to day 30 of the control period; we expressed the result of this conversion as $d_x = (l_x - l_{x-1})$.
146 The mortality values were inserted into a file with a database structure, and were considered as the
147 end points (as percentages), for each combination of temperature, exposure time, and stage (and sex
148 in the case of adults) predictor variables (hereafter 'features' in line with computer science

terminology), making a total of 228 combinations (Supplementary Material 2). Table 1 represents a summary of the different cases analyzed.

Table 1. Source and characteristics of the original information.

Species	Temperature	Exposure	Stage	N	Period	Source
Panstrongylus megistus (D)	28, 40	1, 12	3,4,5,Ad	100	30	[27]
Panstrongylus megistus (D)	5, 28	1, 12	5	53, 100	30	[28]
Panstrongylus megistus (D)	35, 40	1, 12	5	53, 100	30	[28]
Panstrongylus megistus (S)	28, 40	1, 12	3,4,5,Ad	100	30	[27]
Rhodnius prolixus	35-44	1, 24	1	10	1	[29]
Triatoma infestans	30, 40	1, 12	3,4,5,Ad	50	30	[30]
Triatoma infestans	35, 45, 50, 55	0.17	E,1,2,3,4,5,Ad	30	45	[31]
Triatoma infestans	30	1, 12	5	NA	30	[32]

Summary information of the cases with data available on the effect of various temperatures on the survival of triatomines obtained from the bibliography, and fed to the machine learning algorithm. For the species *Panstrongylus megistus* D and S represent domiciliary and sylvatic origin, respectively; Temperature is in °C; Exposure is measured in hours; Stage represents the triatomine developmental stage used in the experiments (E: eggs; 1-5: first to fifth nymphs; Ad: adults); N is the number of individual replicates per temperature and exposure combination; Period: period of observation (days) after experimental exposure to determine survival. The complete dataset can be found in the Supplementary Material 2.

The machine learning pipeline approach

We adapted the framework from [25] using our collated feature set to predict percentage mortality as a regression problem. Initially we checked for correlations between features prior to running our machine learning pipeline. We compared support vector machine, gradient boosting model (GBM), Bayesian generalized linear models as well as linear models fitted with ordinary least squares. In each case we used 10-fold cross validation to compute model performance (root mean square error, RMSE) and to prevent overfitting the model. The model with the lowest RMSE and thus highest predictive performance was further interrogated. We also compared predictive performance by calculating R^2, in this case the correlation between the observed and predicted values [33]. To interpret how each feature was affecting model performance we computed ranked feature importance using model class reliance approach [34] and plotted individual responses using individual conditional expectation curves [35]. Interactions in the model were quantified by calculating Friedman's *H* statistic [36] and visualized plotting multidimensional partial dependency plots. See [25] for more details and https://github.com/nfj1380/ThermalLimits_ActiveLearning for the code used.

Results

The machine learning pipeline results

The GBM model had the highest predictive power compared to the other algorithms we tested, with an R^2 of 0.8 and RMSE of 18.2. In this model, temperature followed by the number of hours exposed were the strongest predictors of mortality based on the preliminary experimental results (Fig 2a). Species played a more minor role in our predictive models with life stage the least important of the variables included in the model (Fig 2a). When we further interrogated our model, we not only found some strong non-linear effects of each variable on mortality (Fig 2b-e) but some striking non-linear

interactions between variables (Fig 3). For example, triatomine mortality was generally stable up until the 39˚C mark when it rapidly increased before plateauing between 40-44˚C, and this response was an essential component of the experimental design for the combination of temperature and exposure time on insect´s survival (see below). However, mortality predictions for some individuals in the dataset were much lower than others which is evidence that interactions are important in this predictive model (indicated by the spread of the lines in Fig 1b, see [37]). The relationship between exposure and mortality was even more variable across individuals showing no clear trend (Fig 2c). Exposure time, however, had the strongest interactions with the other variables (Fig 3a), with the relationship between exposure time and temperature being of the strongest predictive importance. Low mortality was predicted with short exposure times (<7 hours to temperatures between 0-39˚C), yet any exposure to temperatures above 42-44 ˚C was associated with high mortality, although an exact threshold was not evident. Our model predicted that temperatures between 10-39˚C to be optimal for triatomine survival with no effect of exposure time on mortality. Our analysis also revealed that *T. infestans* experienced lower mortality overall (Fig 2c), and this may be linked to temperature as we found lower mortality of this species compared to the others at temperatures >39˚C.

Fig. 2. Results from the best performing machine learning model. Variable importance plot (a) and centered individual expectation plot (cICE) plots (b-e) from the best performing machine learning model (GBM). Y axes represent the marginalized effect of each variable on vector estimated mortality (%) whilst controlling for the effect. The red line represents the average effect of the variable of triatomine survival.

205

206 **Fig. 3**. **Interaction between variables as detected from the machine learning model.** Plots
207 showing the overall interactive strength of each variable (a), the top three interactions between
208 variables based on Friedman's *H* (b), followed by 2D partial dependencies of the most
209 important interactions in our GBM model using Friedman's *H* Index (c-d), where $\hat{y}$ is the
210 estimated mortality (%). Darker red in the heatmap (c) reflect higher predicted mortality. Fig 1
211 of the Supplementary Material 1 shows the interaction between exposure time and species.

212

213 **Machine learning led experimental design**

214 The results of our machine learning model revealed that any exposure to temperatures above 42-44 ˚C
215 was associated with high mortality. Subsequently, we exposed insects to one of three temperatures
216 (40, 42 and 44 °C) for different exposure times (1, 2, 4, 8 and 12 hours). We used fifth-instar nymphs of
217 *T. infestans* provided by the National Chagas Control Service (Córdoba, Argentina). Insects were fed on
218 live chicken two weeks before experiments, and they were acclimated in the laboratory at 25 ± 0.5°C,
219 and 12:12 light/dark photoperiod (light on 08:00 am) for one week, because thermotolerance showed
220 a plastic response in this species [38].

221 The insects were placed in a chamber (PTC-1 Peltier Effect Cabinet; Sable Systems International (SSI),
222 Las Vegas, NV, USA), connected to a temperature controller (Pelt-5 (SSI)) set at 40, 42 or 44°C for each
223 of the exposure times. A group of insects handled in the same way as the experimental groups, but
224 without being exposed to high temperature, and kept at 25°C, were used as the control group (C). After
225 their exposure to the treatment the insects were maintained at the basic laboratory conditions

described above, for 70 days. Ten insects were used for each replicate; the 40°C treatment was replicated twice, while the 42 and 44°C treatments were replicated three times. A total of 480 insects were used. The number of dead and molted insects were recorded each day after the heat treatment. Molted insects were eliminated from the survival analysis. A survival percentage was calculated as the ratio of number of live insects over the total number of treated insects (*i.e.*, the "initial" number) for each replicate and treatment. To determine the effects of heat for different temperatures and exposure periods on survival, a Cox analysis (using the package *survival* of the R language [39]) was performed for each experimental temperature, *i.e.*, 40, 42 and 44 °C. Also, a GLM (Generalized Linear Model) analysis was carried out to determine the importance of the covariates (the treatment temperature and exposure times of the laboratory experiments), as well as their interactions, on mortality; the GLM enables the use of linear models in cases where the response variable has an error distribution that is non-normal. The GLM fit to the data was then used to predict the insects' mortality for various combinations of temperatures and exposure times; those mortality predictions were converted into age specific survival values (l_x) and then used to estimate the life expectancy (days) after the heat treatment (e_0). All these calculations included the 95% confidence intervals and were carried out in R [40](R Core Team, 2020); see https://github.com/nfj1380/ThermalLimits_ActiveLearning for the code used.

The laboratory experiment results

At 40 °C survival did not differ significantly across exposure periods (Log-rank test, $\chi2 = 5.33$, df = 5, p= 0.377) (Table 2). However, survival differed significantly across exposure periods at the other two temperatures, *i.e.*, 42 °C (Log-rank test, $\chi2 = 66.92$, df = 5, $p< 0.001$), and 44 °C (Log-rank test, $\chi2$ = 138.25, df = 5, $p< 0.001$). At 42 °C, the exposure periods of 8 and 12 h showed a lower mean survival

248 time than the other exposure times (Table 2). At 44 °C, the mean survival time of 2 and 4 h treatments
249 were lower than the control and 1 h treatment, and 8 and 12 h showed the lowest mean survival time
250 from all treatments (Table 2).

251 **Table 2. Laboratory experiment results of the effect of temperature on survival of 5th instar T.**
252 **infestans**.

		Exposure periods (h)				
Temp. (°C)	Control	1	2	4	8	12
40	64.1 ± 4.3a	67.0 ± nda	65.6 ± 3.6a	69.5 ± nda	>70 a	64.7 ± 5.1a
42	65.9 ± 2.2a	64.6 ± 3.1a	54.6 ± 5.4a	63.8 ± 4.3a	37.5 ± 5.8b	20.83 ± 5.5b
44	68.8 ± 1.2a	22.2 ± 5.3a	21.97 ± 5.9b	13.9 ± 4.4b	1 ± 0c	1 ± 0c

253 Survival time (mean ± SE) in days of fifth nymphs of *T. infestans* under different exposure periods (h)
254 and temperatures (Temp, °C). Different letters indicate significant differences between the exposure
255 times of each temperature. nd: the parameter could not be estimated.

256

257 The fit of the mortality data using the survival function (with interaction) of the Cox analysis showed
258 that both covariates (temperature and exposure times) were highly significant, as well as their
259 interaction, in determining the mortality level (Concordance= 0.872 (se = 0.015), Likelihood ratio test=
260 295.9 on 3 df, $p \leq$ 2e-16, Wald test = 276 on 3 df, $p \leq$ 2e-16, and Score (logrank) test = 447.6 on 3 df,
261 $p \leq$ 2e-16). Fig 4 shows the result of the survival curve estimation for only one temperature (44 °C) and
262 three exposure times (2, 4, and 6 hours). The survival curves for all combinations of temperature and
263 exposure times are given in Section 2 of the Electronic Supplementary Material 1.

Fig 4. Results of Cox analysis of laboratory experiments. The survival curves (l_x) predicted by the Cox analysis for 44 °C temperature and three exposure times (2, 4, and 6 hours).

The GLM analysis confirmed this kind of results, except that it showed a much weaker effect of temperature (NS at the 5% level: p= 0.0552) and a highly significant effect of exposure time and the interaction of temperature with exposure time (p= 0.00408, and p= 0.00285, respectively); the multiple adjusted R^2 was 0.822, with a p value of 4.274e-06. Fig 5 shows the proportion of insects alive after 70 days of the heat treatment for various simulated combinations of temperatures and exposure times.

Fig. 5. Results of GLM analysis of laboratory experiments. Proportion of insects alive (l_x) after 70 days of the heat treatment for various simulated combinations of temperatures and exposure times.

Another useful parameter to represent the effects of the heat treatments is the expectation of life (e_x) after the day of treatment (x); this parameter is defined as the average number of days of life remaining to an individual alive at age or time x, and is calculated as $0.5 + \frac{l_{x+1} + l_{x+2} + \ldots + l_w}{l_x}$, where l_x is the probability of being alive at age or time x, and l_w is the same where w refers to the time of death [41]. Fig 6 shows the number of days expected to be lived by an average insect after each heat treatment (e_0). These e_0 estimates were calculated from both the experimental data and the survival curves predicted by the Cox analysis. It is clearly seen that the variability of the insects' mortality response becomes more variable as the temperature increases.

Fig. 6. Comparison of observed and expected life expectancies. Life expectancy (e_0, in days) after the day of the heat treatment from both the experimental data and the survival curves predicted by the Cox analysis. The dashed lines are the 95% confidence intervals.

Discussion

We demonstrate the power of an active learning approach to explore experimental space to design studies that can provide greater understanding of species thermal limits. From a preliminary and very limited dataset, we were able to identify complex but biologically plausible nonlinear interactions between temperature and exposure times shaping mortality, and setting potential thermal limits of triatomine species for SDM (species distribution modelling). This led to successful new experiments that generated novel insights into this important vector group that can be utilized by species distribution models based upon micro-climatic information. This approach can not only guide experimental efforts for the specific study of disease vectors but can also be extended to reduce the experimental envelope in any systems with multiple interacting variables.

The harmful effect on insects´ survival of exposing them to extreme temperatures (close to the upper lethal temperature) for extended periods of time is not linear, and thus difficult to predict [42]. Without our active learning approach, researchers would have had to do a much larger number of experimental combinations to find these non-linear thresholds between exposure and temperature for each species; not only impractical but unaffordable. Our experimental temperatures of between 40-44 °C and exposure times of 1 to 6 hours at high temperatures are found on *T. infestans´* natural habitats during summer [24]; we extended the recorded exposure times to 8 and 12 hours to provide some margin for potential global climate change. Similar to [30], our results showed that survival of fifth nymph of *T. infestans* was not affected by any of the exposure times tested at 40 °C. However, it could

affect other physiological traits that were not measured like the germinal cells [43]. In addition, a survival reduction was observed after 8 or more hours of exposure at 42°C or 4 or more hours of exposure at 44°C. Similarly, in nymphs of the related species, *Pastrongylus megystus*, a brief exposure of 1 h to 40°C did not compromise its survival, but a long period of exposure, *i.e.*, 12 h, showed an important drop on survival [27,28]. The combined effect of temperature and exposure time on survival is quite variable, possibly due to thermal tolerance differences across triatomine species, as shown by [44] for seven species of triatomines. Those species were chosen because of their epidemiological relevance, and it was observed that thermo-tolerance range increases with increasing latitude mainly due to better cold tolerances, and limiting their southern distribution [44].

A sub-lethal temperature acquires a considerable ecological relevance, when an extended time of exposures to such a temperature turns into lethal [42]. In our experiments this was the case for 42 and 44 °C. These temperatures did not reduce insects´ survival when exposures times were small, *e.g.*, 1 h, while longer exposure times, *i.e.*, 2 or 8 h at 44 or 42°C respectively, highly decreases survival. There is little information about the effects of exposure times and temperatures in kissing bugs. There is one study where a brief exposure (0.17 h) at a high temperature (55 °C) produced 100 % mortality in nymphs and adults of *T. infestans*, while lower temperatures presented various effects: at 40 °C there was an increase of insect´s activity, while 50 °C produced severe damage, like knock-down, proboscis extension and leg paralysis [31]. Although the exposure times of our experiments were longer, we observed similar damage, especially knock-down, for almost all exposure times at 44°C, also in *T. infestans* (A.A.C. and I.M. personal observations).

Lastly, our methodology opens up further opportunities to model population dynamics based upon the responses to micro-climatic environments (as [45] have done for in a tephritid fly *Bactrocera dorsalis*).

The effects of temperature and exposure on other demographic parameters like development time and fecundity in order to estimate the thermal effects on the population growth rate are reasonably understood. Under laboratory conditions [46] found in the triatomine *Rhodnius prolixus* that daily temperature fluctuations (DTF) did not affect development time and fertility. However, fecundity was lower in females reared at DTF than at constant temperature, and males had higher body mass reduction rate and lower survival in the DTF regime, suggesting higher costs associated to fluctuating thermal environments [46]. Also, the humidity factor has to be considered when performing SDM, which interacts with temperature (expressed as the vapor pressure saturation deficit [47]). Using an active learning approach could further help to understand this interaction and thus, the dryness dimension of the fundamental niche of small ectotherms such as insects, because of their high surface area-to-volume ratios, are usually at risk of dehydration in arid environments. Undoubtedly the lower thermal tolerance has also to be taken into account for a complete bioclimatic analysis of ectotherm population dynamics. The ultimate goal of all these physiological and SDM modelling efforts is to predict population density in different areas of the triatomine's geographical range; population density of pathogens' vectors is an important component of transmission risk, so this kind of micro-SDM modelling is of important epidemiological significance. We hope that our successful results will prompt experiments with other species in that direction.

Incorporating computational advances and active learning into ecological experimental design decision making is rare, but this study provides a case study on how this approach can be used in the context of species thermal ecology. As the world climate shifts even more swiftly, being able to search rapidly through an experimental space to design better experiments to identify species tolerance thresholds is increasingly important. Moreover, as ecological systems are inherently complex and ecological

experiments often costly, an active learning approach has the potential to be of more general use in various ecological sub-disciplines.

Acknowledgements

We are grateful to Yuedong Wang, University of California, Santa Barbara, who made valuable suggestions for the implementation of the GLM analysis to our laboratory results. Raúl Stariolo, of the National Chagas Control Service (Córdoba, Argentina), kindly provided the live specimens of *T. infestans* used to carry out the experiments.

Authors' contributions

JR and NFJ conceived the ideas and designed the methodology; PES, AAC and IM designed and carried out the laboratory experiments; all five authors analyzed the data; JR and NFJ led the writing of the manuscript. All five authors contributed critically to the drafts and gave final approval for publication.

Data availability: The original data and the scripts for running the analyses are hosted at a github site: https://github.com/nfj1380/ThermalLimits_ActiveLearning

References

1. Porter WP, Gates DM. Thermodynamic Equilibria of Animals with Environment. Ecological Monographs. 1969; 39(3):227-244.
2. Grant BW, Porter WP. Modeling Global Macroclimatic Constraints on Ectotherm Energy Budgets. American Zoologist. 1992; 32(2):154-178. URL: http://www.jstor.org/stable/3883756

3. Townsend Peterson A. Ecologic Niche Modeling and Spatial Patterns of Disease Transmission. Emerging Infectious Diseases. 2006; 12(12):1822-1826.

4. Nieto P, Malone JB, Bavia ME. Ecological niche modeling for visceral leishmaniasis in the state of Bahia, Brazil, using genetic algorithm for rule-set prediction and growing degree day-water budget analysis. Geospatial Health. 2006; 1:115-126.

5. Fuller DO, Ahumada ML, Quiñones ML, Herrera S, Beier JC. Near-present and future distribution of Anopheles albimanus in Mesoamerica and the Caribbean Basin modeled with climate and topographic data. International Journal of Health Geographics. 2012; 11(1):1-12.

6. Gurgel-Goncalves R, Galvao C, Costa J, Townsend Peterson, A. Geographic Distribution of Chagas Disease Vectors in Brazil Based on Ecological Niche Modeling. Journal of Tropical Medicine. 2012; Article ID 705326, 15 pages. doi:10.1155/2012/705326.

7. Moriguchi S, Onuma M, Goka K. Potential risk map for avian influenza A virus invading Japan. Diversity and Distributions. 2013; 19:78-85.

8. Escobar LE, Romero-Alvarez D, Leon R, Lepe-Lopez MA, Craft ME, Borbor-Cordova MJ, Svenning J-C. Declining Prevalence of Disease Vectors Under Climate Change. Scientific Reports. 2016; 6(39150). DOI: 10.1038/srep39150.

9. Bender A, Python A, Lindsay SW, Golding N, Moyes CL. Modelling geospatial distributions of the triatomine vectors of *Trypanosoma cruzi* in Latin America. bioRxiv. 2019; DOI: http://dx.doi.org/10.1101/738310.

10. Jatta E, Jawara M, Bradley J, Jeffries D, Kandeh B, Knudsen JB, Wilson AL, Pinder M, Alessandro UD, Lindsay SW. How house design affects malaria mosquito density, temperature, and relative

390 humidity: an experimental study in rural Gambia. The Lancet Planetary Health. 2018;
391 2(11):e498-e508. DOI: https://doi.org/10.1016/s2542-5196(18)30234-1.
392 **11.** Ribeiro AC, Sarquis O, Lima MM, Abad-Franch F. Enduring extreme climate: Effects of severe
393 drought on *Triatoma brasiliensis* populations in wild and man-made habitats of the Caatinga.
394 PLoS Neglected Tropical Diseases. 2019; 13(10):e0007766.
395 https://doi.org/10.1371/journal.pntd.0007766. 3.2-3.
396 **12.** Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land
397 areas. International Journal of Climatology. 2017; 37(12):4302-4315.
398 **13.** Casey TM. Biophysical Ecology and Heat Exchange in Insects. American Zoologist. 1992;
399 32(2):225–237. doi:10.1093/icb/32.2.225.
400 **14.** O'Connor MP, Spotila JR. Consider a spherical lizard: animals, models, and approximations.
401 American Zoologist. 1992; 32:179-193.
402 **15.** Angilletta Jr. MJ. Thermal Adaptation: A Theoretical and Empirical Synthesis. Oxford: Oxford
403 University Press. 2009. ISBN 10:0198570872.
404 **16.** Christian KA, Bedford GS. Seasonal changes in thermoregulation by the frillneck lizard,
405 *Chlamydosaurus kingii*, in tropical Australia Ecology. 1995; 76:124–132.
406 **17.** Seebacher F, Grigg GC. Patterns of body temperature in wild freshwater crocodiles, Crocodylus
407 johnstoni: thermoregulation versus thermoconformity, seasonal acclimatization, and the effect
408 of social interactions. Copeia. 1997; 1997(3):549–557. DOI: 10.2307/1447558
409 **18.** Van Damme R, Bauwens D, Castilla AM, Verheyen RF. Altitudinal variation of the thermal
410 biology and running performance in the lizard Podarcis tiliguerta. Oecologia. 1989; 80:516–524.

411 **19.** Parlin AF, Schaeffer PJ, Jezkova T. Modelling the effect of environmental temperatures,
412 microhabitat and behavioural thermoregulation on predicted activity patterns in a desert lizard
413 across its thermally diverse distribution. Journal of Biogeography. 2020 (In press), DOI:
414 10.1111/jbi.13936.
415 **20.** Lee BY, Bacon KM, Bottazzi ME, Hotez PJ. Global economic burden of Chagas disease: a
416 computational simulation model. The Lancet Infectious Diseases. 2013; 13(4):342-348.
417 **21.** Cucunubá ZM, Okuwoga O, Basáñez M-G, Nouvellet P. Increased mortality attributed to Chagas
418 disease: a systematic review and meta-analysis. Parasites & Vectors. 2016; 9:42.
419 https://doi.org/10.1186/s13071-016-1315-x.
420 **22.** Alevi KCC, Moreira FFF, Jurberg J, Azeredo-Oliveira MTV. Description of the diploid
421 chromosome set of *Triatoma pintodiasi* (Hemiptera, Triatominae). Genetics and Molecular
422 Research. 2016; 15(2):1-10. DOI: 10.4238/gmr.15026343.
423 **23.** Noireau F, Dujardin J-P. Biology of Triatominae. In: Telleria J, Tibayrenc M, editors. American
424 Trypanosomiasis, Chagas Disease One Hundred Years of Research. Elsevier, London, UK. 2017.
425 Chapter 7.
426 **24.** Balsalobre A. ¿Qué especies de vinchucas modificarán su distribución geográfica en la
427 Argentina? Un análisis de los microhábitats y microclimas de los triatominos vectores de la
428 enfermedad de Chagas. Doctoral Thesis. The National University of La Plata. La Plata, Buenos
429 Aires, Argentina. 2016. Available from:
430 http://naturalis.fcnym.unlp.edu.ar/repositorio/_documentos/tesis/tesis_1425.pdf

25. Fountain-Jones NM, Machado G, Carver S, Packer C, Recamonde-Mendoza M, Craft ME. How to make more from exposure data? An integrated machine learning pipeline to predict pathogen exposure. Journal of Animal Ecology. 2019; 88(10):1447-1461. DOI: 10.1111/1365-2656.13076.

26. Sverchkov Y, Craven M. A review of active learning approaches to experimental design for uncovering biological networks. PLoS computational biology. 2017; 13(6):e1005466. https://doi.org/10.1371/journal.pcbi.1005466

27. García SL, Rodrígues VLCC, García NL, Ferraz Filho AN, Mello MLS. Survival and Molting Incidence after Heat and Cold Shocks in *Panstrongylus megistus* Burmeister. Memórias do Instituto Oswaldo Cruz. 1999; 94(1):131-137.

28. García SL, Mello MLS, García NL, Rodrígues VLCC. Experimentally Induced Heat-Shock Tolerance in *Panstrongylus megistus* (Hemiptera: Reduviidae). Journal of Medical Entomology. 2001; 38(4):510-513.

29. Buxton PA. The Thermal Death-Point of *Rhodnius* (Rhynchota, Heteroptera) Under Controlled Conditions of Humidity. Journal of Experimental Biology. 1931; 8:275-278.

30. Rodrígues VLCC, Mello MLS, Ferraz Filho AN, Dantas MM. Sobrevivência e ocorrência de muda em *Triatoma infestans* Klug (Hemiptera, Reduviidae) após choque de temperatura. Revista de Saúde Pública. 1991; 26(5):461-467.

31. Gentile AG, Sartini JL Campos MC, Sánchez JF. La aerotermia como alternativa para el control de *Triatoma infestans* (Hemiptera, Reduviidae) resistentes a deltametrina. Cadernos de Saúde Pública/ 2004; 20(4):1014-1019.

451 **32.** Campos SGP, Rodrígues VLCC, Wada CY, Mello MLS. Effect of Sequential Cold Shocks on Survival
452 and Molting Rate in *Triatoma infestans* Klug. Memórias do Instituto Oswaldo Cruz. 2002;
453 97(4):579-582.
454 **33.** Kvalseth TO. Cautionary note about R2. The American Statistician. 1985; 39(4):279-285.
455 **34.** Fisher A, Rudin C, Dominici F. Model class reliance: Variable importance measures for any
456 machine learning model class, from the "Rashomon" Perspective. 2018. Retrieved from
457 http://arxiv.org/abs/1801.01489
458 **35.** Goldstein A, Kapelner A, Bleich J, Pitkin E. Peeking inside the black box: Visualizing statistical
459 learning with plots of individual conditional expectation. Journal of Computational and
460 Graphical Statistics. 2015; 24(1):44–65. https://doi.org/10.1080/10618600.2014.907095
461 **36.** Friedman JH., Popescu BE. Predictive learning via rule ensembles. The Annals of Applied
462 Statistics. 2008; 2(3):916–954. https://doi.org/10.1214/07-AOAS148
463 **37.** Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.
464 2019. URL https://christophm. github. io/interpretable-ml-book.
465 **38.** Belliard SA, de la Vega GJ, Schilman PE. Thermal tolerance plasticity in Chagas disease´s vectors
466 *Rhodnius prolixus* (Hemiptera: Reduviidae) and *Triatoma infestans*. Journal of Medical
467 Entomology. 2019; 56(4):997-1003. doi: 10.1093/jme/tjz022.
468 **39.** Therneau TM. survival: A Package for Survival Analysis in R. R package version. 2020. Available
469 from https://CRAN.R-project.org/package=survival.
470 **40.** R Core Team. R: A language and environment for statistical computing. R Foundation for
471 Statistical Computing, Vienna, Austria. 2020. URL https://www.R-project.org/.

472 **41.** Carey JR. Longevity. The Biology and Demography of Life Span. Princeton and Oxford: Princeton
473 University Press. 2003.
474 **42.** Chown SL, Nicolson SW. Insect Physiological Ecology: mechanisms and patterns. Oxford: Oxford
475 University Press. ISBN: 0 19 851548 0. 2004
476 **43.** Mello MLS, Maria SS, Tavares MCH. Heat shock-induced apoptosis in germ line cells of *Triatoma*
477 *infestans* Klug. Genetics and Molecular Biology. 2000; 23(2):301-304.
478 http://dx.doi.org/10.1590/S1415-47572000000200011.
479 **44.** de La Vega GJ, Schilman PE. Ecological and physiological thermal niches to understand
480 distribution of Chagas disease vectors in Latin America. Medical and Veterinary Entomology.
481 2018; 32(1):1-13 DOI: 10.1111/mve.12262.
482 **45.** Mutamiswa R, Tarusikirwa V, Nyamukondiwa C, Chidawanyika F. Fluctuating environments
483 impact thermal tolerance in an invasive insect species *Bactrocera dorsalis* (Diptera:
484 Tephritidae). Journal of Applied Entomology. 2020; In press. https://doi.org/10.1111/jen.12795
485 **46.** Rolandi C, Schilman PE. The costs of living in a thermal fluctuating environment for the tropical
486 haematophagous bug, *Rhodnius prolixus*. Journal of Thermal Biology. 2018; 74:92-99.
487 https://doi.org/10.1016/j.jtherbio.2018.03.022.
488 **47.** de la Vega GJ, Schilman PE. Using eco-physiological traits to understand the realized niche: the
489 role of desiccation tolerance in Chagas disease vectors. Oecologia. 2017; 185(4):607–618.
490 https://doi.org/10.1007/s00442-017-3986-1.
491
492

Active learning

Tolerance predictions

Heating and cooling

Micro-habitat survival

Prelim temp shock experiments

% survival

Machine learning pipeline

GBM SVM RF BGLM

*Best model**

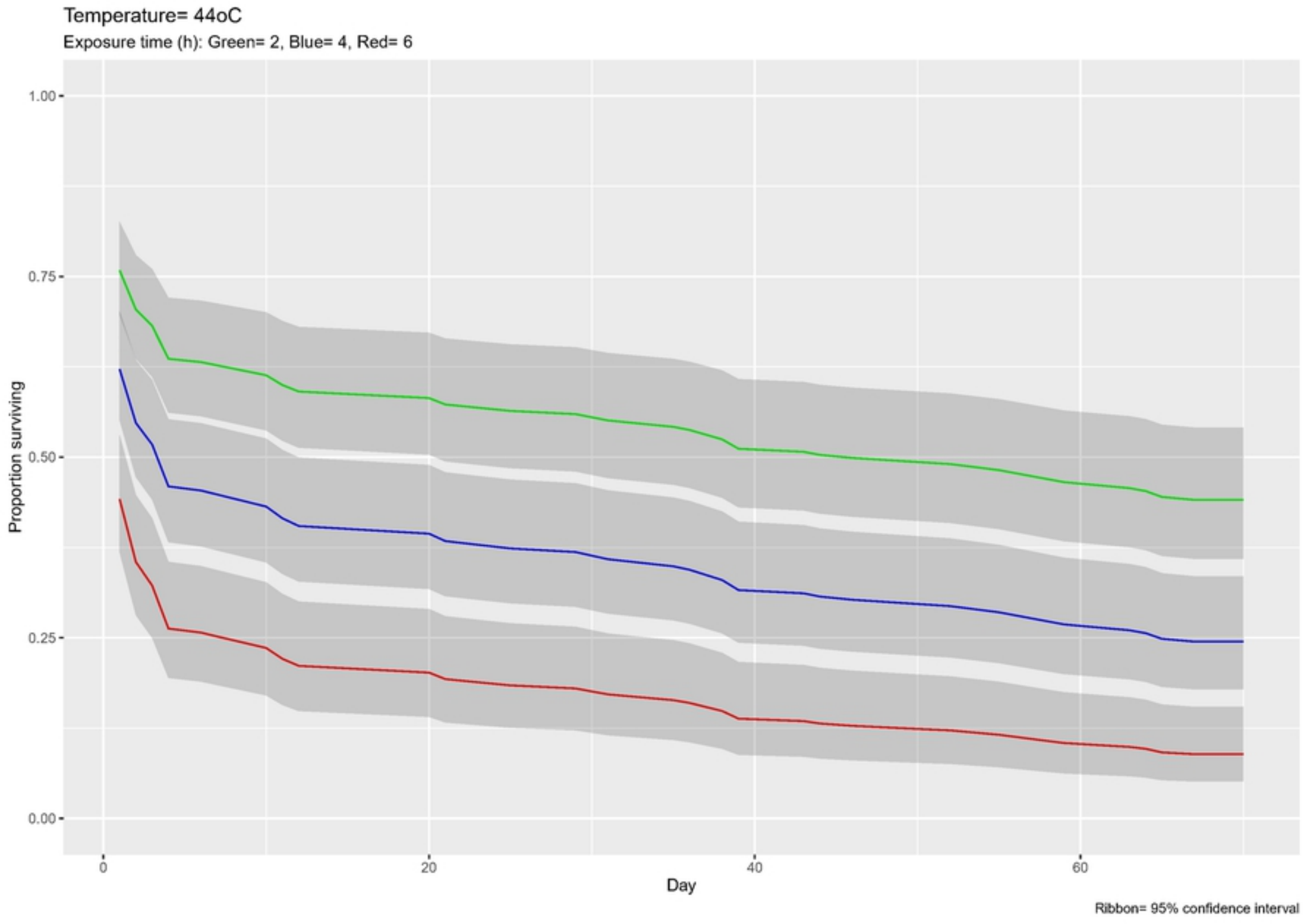
Variable importance

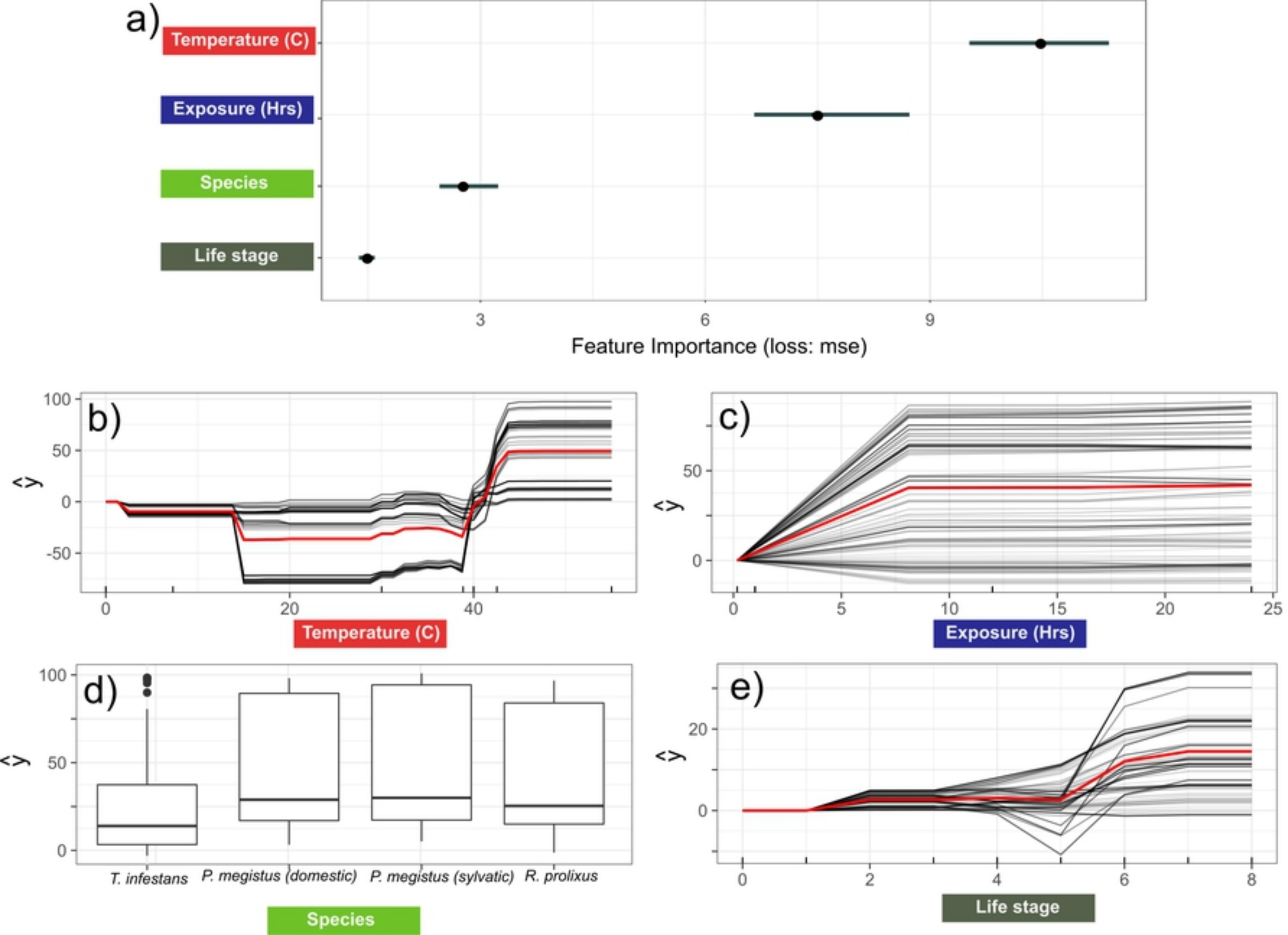
Interaction strength

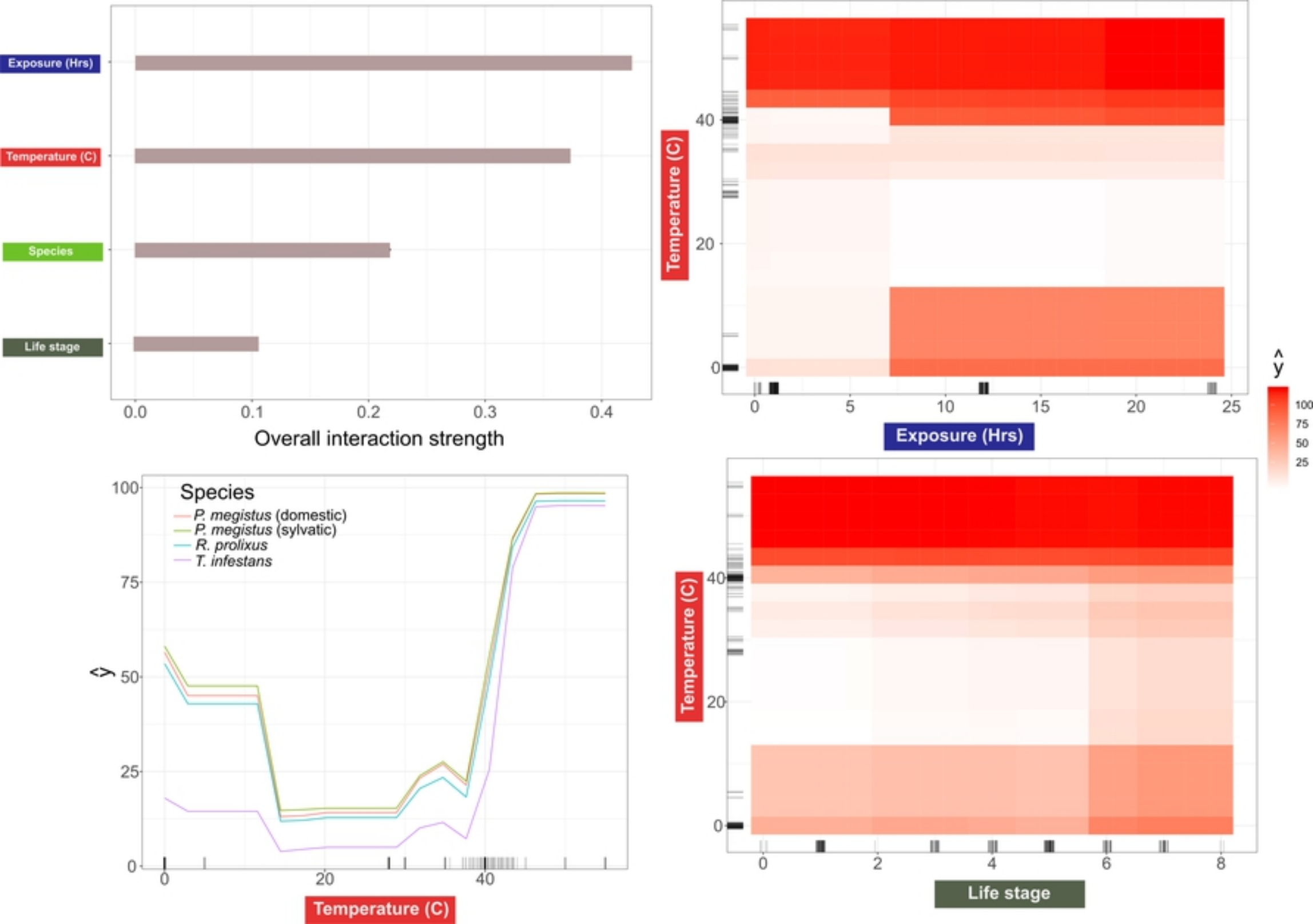
Experimental design (laboratory)

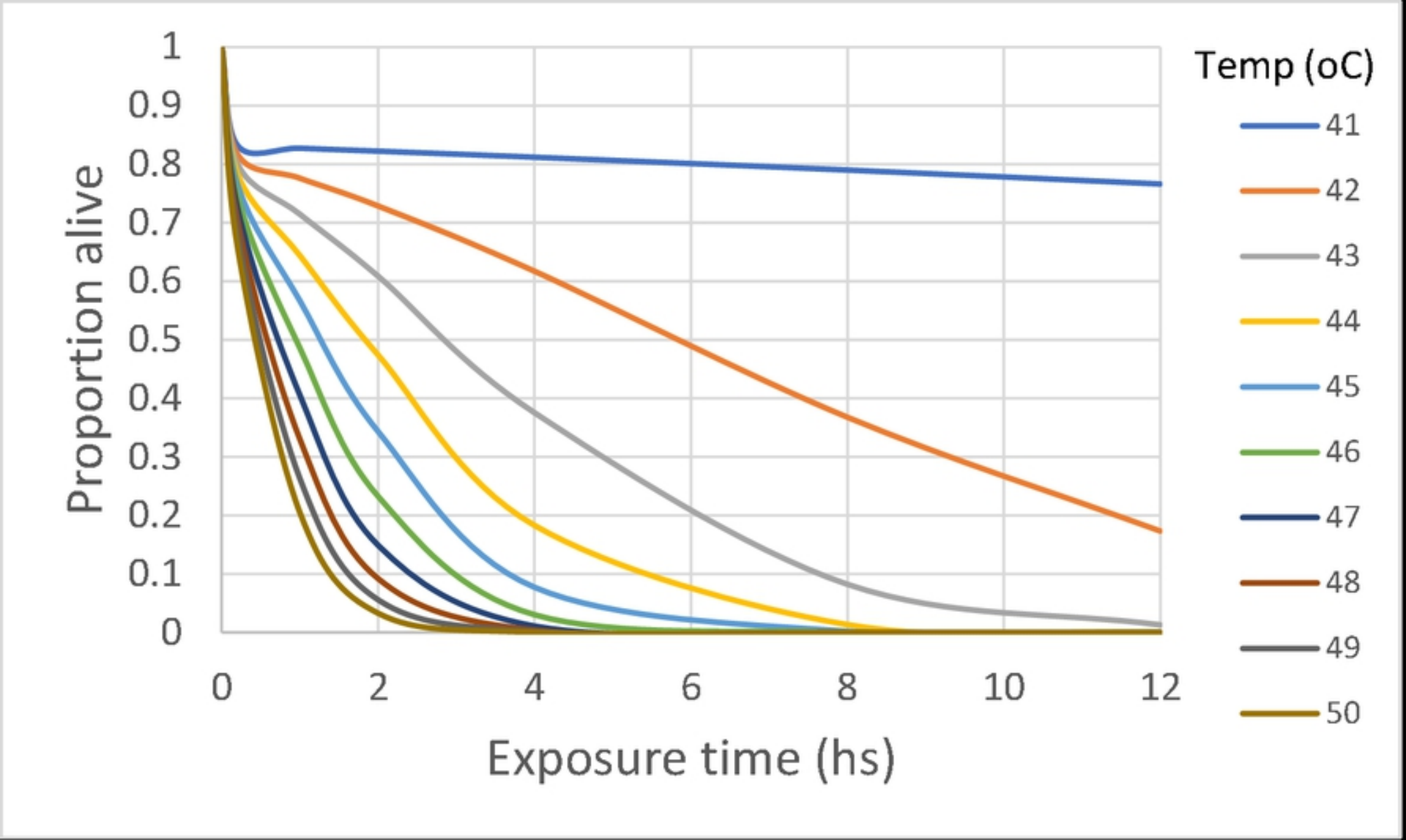
Updated predictions

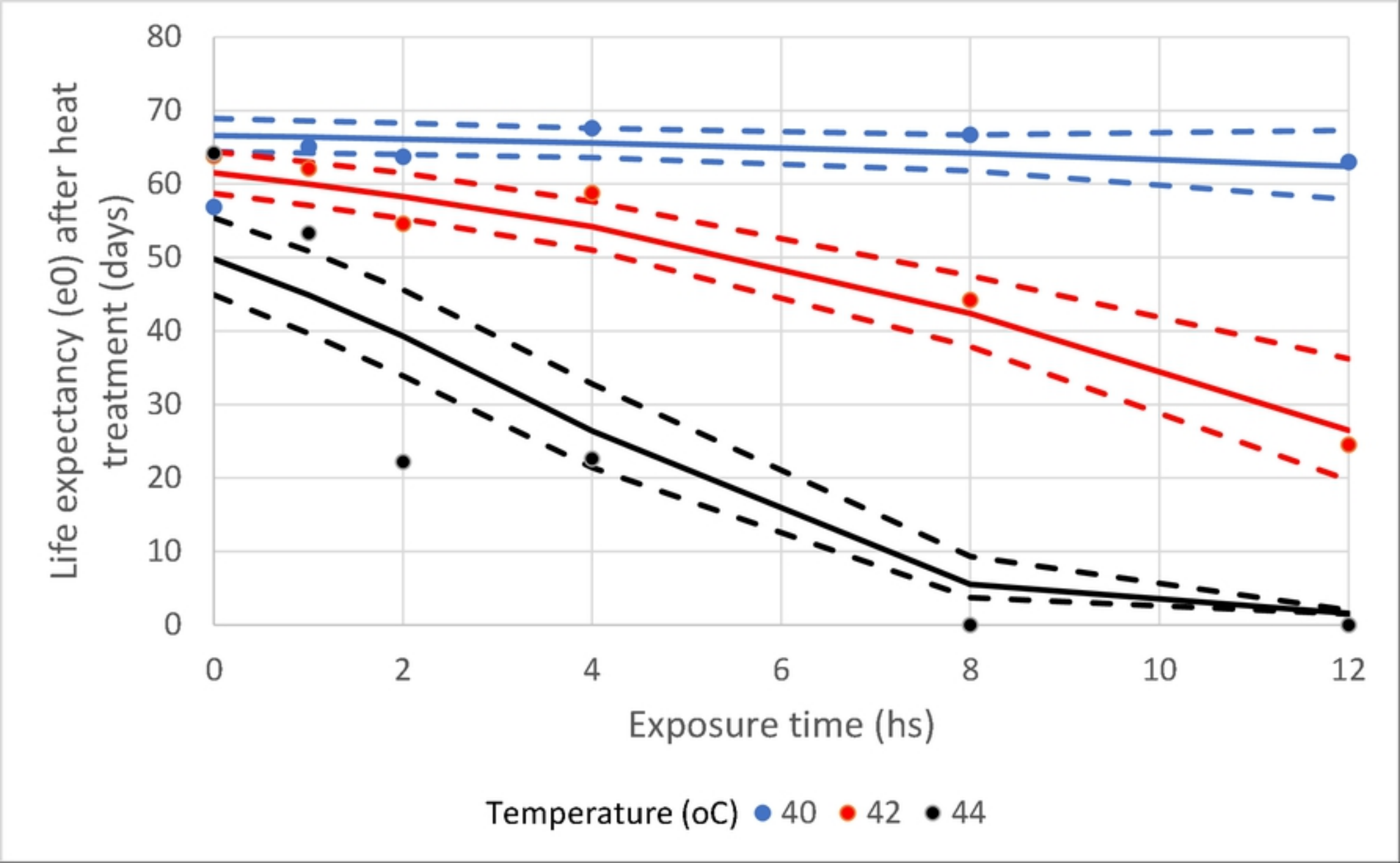
Macro SDM


Temperature= 44oC
Exposure time (h): Green= 2, Blue= 4, Red= 6
1.00
0.75
0.50
0.25
0.00
Proportion surviving
0
20
40
60
Day
Ribbon= 95% confidence interval


a)
Temperature (C)
Exposure (Hrs)
Species
Life stage
3
6
9
Feature Importance (loss: mse)
b)
100
50
0
-50
ŷ
0
20
40
Temperature (C)
c)
50
0
ŷ
0
5
10
15
20
25
Exposure (Hrs)
d)
100
50
0
ŷ
T. infestans
P. megistus (domestic)
P. megistus (sylvatic)
R. prolixus
Species
e)
20
0
ŷ
0
2
4
6
8
Life stage


Exposure (Hrs)
Temperature (C)
Species
Life stage
0.0
0.1
0.2
0.3
0.4
Overall interaction strength
Temperature (C)
0
20
40
0
5
10
15
20
25
Exposure (Hrs)
ŷ
100
75
50
25
Species
P. megistus (domestic)
P. megistus (sylvatic)
R. prolixus
T. infestans
<y
100
75
50
25
0
0
20
40
Temperature (C)
Temperature (C)
0
20
40
0
2
4
6
8
Life stage


Proportion alive
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
Exposure time (hs)
Temp (oC)
41
42
43
44
45
46
47
48
49
50


Life expectancy (e0) after heat treatment (days)
80
70
60
50
40
30
20
10
0
0
2
4
6
8
10
12
Exposure time (hs)
Temperature (oC)
40
42
44